DECISION SUPPORT SYSTEM TO SUPPORT THE SOLVING OF CLASSIFICATION PROBLEMS IN TELECOMMUNICATIONS

Rok Rupnik

University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenija

Key words: data mining; decision support; decision support system; knowledge discovery; classification; telecommunications.

Abstract: Traditional techniques of data analysis do not enable the solution of all kind of problems and for that reason they have become insufficient. This caused a new interdisciplinary field of data mining to arise, encompassing both classical statistical, and modern machine learning techniques to support the data analysis and knowledge discovery from data. Data mining methods are powerful in dealing with large quantities of data, but on the other hand they are difficult to master by business users to facilitate decision support. In this paper we introduce our approach to integration of decision support system with data mining method called classification. We discuss the role of data mining to facilitate decision support, the use of classification method in decision support system, discuss applied approaches and introduce a data mining decision support system called DMDSS (Data Mining Decision Support System). We also present some obtained results and plans for future development.

Sistem za podporo odločanju za reševanje klasifikacijskih problemov na področju telekomunikacij

Kjučne besede: odkrivanje zakonitosti v podatkih; podpor odločanju; sistem za podporo odločanju; odkrivanje znanja; klasifikacija; telekomunikacije.

Izvleček: Tradicionalne tehnike za analizo podatkov ne omogočajo reševanja vseh vrst problemov. Nekatere vrzeli na tem področju zapolnjuje multidisciplinarno področje odkrivanja zakonitosti v podatkih, ki omogoča analizo podatkov na podlagi klasičnih statističnih tehnikah in tehnikah strojnega učenja. Metode odkrivanja zakonitosti v podatkih so učinkovite tudi na večjih količinah podatkov, vendar so praviloma prezahtevne, da bi jih poslovni uporabniki sami obvladovali za potrebe podpore odločanju. V okviru članka je predstavljen naš pristop uporabe metode klasifikacije za potrebe podpore odločanju. V uvodu je predstavljena vloga odkrivanja zakonitosti v podatkih za potrebe podpore odločanju s poudarkom na uporabi metode klasifikacije. V nadaljevanju pa je predstavljen DMDSS (Data Mining Decision Support System), sistem za podpor odločanju, ki temelji na uporabi metode klasifikacije. V zaključku so predstavljeni rezultati uporabe sistema DMDSS ter plani za njegov nadaljnji razvoj.

1. Introduction

Companies use several types of decision support systems to facilitate decision making. Traditionally, OLAP tools are used for an advanced data analysis and decision support in the business area. OLAP tools follow what is in essence a deductive approach (JSR-73 Expert Group, 2004). The drawback of this approach is that it depends on coincidence or even luck of choosing the right dimensions at drilling-down to acquire the most valuable information, trends and patterns. It lacks algorithmic approach and depends 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 accurate and valuable information, trends and patterns (Goebel and Gruenwald, 1999). It seems that with the increase in data volume, traditional data analysis has become insufficient, and new methods for data analysis are needed.

To satisfy this need, a new interdisciplinary field of data mining appeared. Data mining encompasses statistical, pattern recognition, and machine learning tools to support the discovery of patterns, trends and rules that lie within data given (Heinrichs and Lim, 2003). 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 (JSR-73 Expert Group, 2004). Data mining reduces or even eliminates the above mentioned drawback. 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. Data mining methods also extend the possibilities of discovering information, trends and patterns by using richer model representations (e.g. decision rules, decision trees, ...) than the usual statistical methods, and are therefore well-suited for making the results more comprehensible to the non-technically oriented business users. OLAP tools mostly enable the answers to the questions like: "What has been going on?" On the other hand, data mining enables the answers to different kind of questions, e.g.: "What are characteristics of our best customers?" Indeed, the use of data mining advances the whole field of data analysis including its role in the decision-making process to a higher level (Nemati and Barko, 2002).

Data mining can be used through two approaches. The first approach is called *data mining software tool approach*. In this approach data mining is used through ad hoc data mining projects by the use of data mining software tools (Goebel and Gruenwald, 1999; Kohavi and

Sahami, 2000; Holsheimer, 1999). Data mining software tools require a significant expertise in data mining methods, databases and statistics. They are rather complex, because they offer a variety of methods and parameters that the user must understand in order to use them effectively (Kohavi and Sahami, 2000). Data mining software tool approach has a disadvantage in a number of various experts needed to collaborate in a project, the complexity of software tools and in transferability of results and models (Srivastava et al., 2000; Hirji, 2001). The disadvantages mentioned call for different approach, which in this paper we call *data mining application system approach*. Data mining application systems approach signifies the possibility to develop decision support systems which use data mining methods and do not demand expertise in data mining for business users. It is an approach which focuses on business users and other decision makers, enabling them to view data mining models which are presented in a user-understandable manner through a user friendly and intuitive GUI using standard and graphical presentation techniques (Aggarwal, 2002). Through the use of data mining application systems approach, data mining becomes better integrated in business environments and their decision processes (Goebel and Gruenwald, 1999; Holsheimer, 1999; Kohavi and Sahami, 2000; Bayardo and Gehrke, 2001). We hope to demonstrate the latter by introducing the DMDSS (Data Mining Decision Support System), which we developed.

The paper is structured as follows. In the second section we are making a brief introduction of data mining and decision support systems. In the third section we are introducing the motivation for the use of data mining to facilitate decision support. We are presenting two approaches of the use of data mining and data mining standards. In the fourth section we are going to introduce DMDSS, decision support system based on data mining method called classification which we developed for telecommunication service provider. We are presenting the process model for the use of DMDSS and the platform of DMDSS. We are also introducing the functionalities of DMDSS for classification data mining method supported by DMDSS. In fifth section we are representing the results and the experiences of the use of DMDSS after five months of production. We are also representing the brief list of enhancements planed for new version of DMDSS. And finally, we are introducing summary and concluding remarks.

2. Decision support systems and data mining: a brief introduction

Decision support systems (DSS) are defined as interactive application systems which are intended to help decision makers utilize data and models in order to identify problems, solve problems and make decisions. They incorporate both data and models and they are designed to assist decision makers in decision making processes. They provide support for decision making, they do not replace it (Steblovnik et.al., 2005). The mission of decision support systems is to improve effectiveness, rather than the efficiency of decisions (Mladenić et al., 2003b). A decision support system can take many different forms and in general we can say that every decision support system is developed for a specific objective and bases on a particular decision process and set of methods, techniques and approaches. The DSS can be developed for the purpose of simulation (Chen, 2004; Kuan, 2004), analysis (Bohanec, 2001; Heinrichs and Lim, 2003; Ward, 2000), forecasting (Zhong et al., 2005; Patelis et al. 2005) and optimization (Heinrichs and Lim, 2003). The design of DSS is very dependant on decision-making process and decision problems which the DSS is going to support (Heinrichs and Lim, 2003). In the context of our paper especially important are repetitive decision problems, which must be supported daily in a non ad-hoc manner.

The objective of data mining is to discover relationships, patterns and knowledge hidden in data (Sherry and Xiaoui, 2005; Kukar, 2006). Data mining is the process of analyzing data in order to discover implicit, but potentially useful information and uncover previously unknown patterns and relationships hidden in data. Data mining is an interdisciplinary field which encompasses statistical, pattern recognition, and machine learning tools to support the analysis of data and discovery of principles that lie within the data. The data mining learning problems that we consider can be roughly categorized as either *supervised* or *unsupervised* (Witten and Frank, 2000). In supervised learning, the goal is to predict the value of an outcome based on a number of input measures. In unsupervised learning, there is no outcome measure, and the goal is to describe associations and patterns among a set of input measures.

3. Integrating data mining and decision support

Companies use several types of decision support systems to facilitate decision support. For the purposes of analysis and decision support in the business area traditionally OLAP based decision support systems are used. OLAP systems represent a tool enabling decision support on a tactical level. They 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 and Sugumaran, 1999). OLAP systems support analysis processes and decision processes, where the analysts are supposed to look for information, trends and patterns. They do it by viewing OLAP forms swapping dimensions and drilling-down through them (Goebel and Gruenwald, 1999). Performing analysis through OLAP follows a deductive approach of analyzing data (JSR-73 Expert Group, 2004). The disadvantage of such an approach is that it depends on coincidence or even luck of choosing the right dimensions at drilling-down to acquire the most valuable information, trends and patterns. We could say that OLAP systems provide analytical tools enabling userled analysis of the data, where the user has to start the right query in order to get the appropriate answer (Mladenić et al., 2003a). Such an approach enables mostly the answers to the questions like: "What has been going on?" What about the answers to the questions like: "What are characteristics of our best customers?" Those answers can not be provided by OLAP systems, but can be provided by the use of data mining, which follows an inductive approach of analyzing data (JSR-73 Expert Group, 2004; Kukar, 2003).

When discussing the relation between data mining and OLAP it is not the question of which one of them is better or worse. Data mining enables the answers to different questions than OLAP, i.e. it enables the solution of different problems and to acquire different information. Decision processes in general, depending on problem, need both, OLAP and data mining, to get the appropriate level of support of decision processes (Forgionne and Rubenstein-Montano, 1999).

Several authors discuss the use of data mining to facilitate decision support and they all confirm the value of it. Chen and Liu (2004) argue that the use of data mining helps institutions make critical decisions faster and with a greater degree of confidence. They believe that the use of data mining lowers the uncertainty in decision process. Nemati and Barko (2002) state that the use of data mining offers companies an indispensable decision-enhancing process to exploit new opportunities by transforming data into valuable knowledge and a potential competitive advantage. Authors also introduce their survey which indicates that the use of data mining can improve the quality and accuracy of decisions. Lee and Park (2003) state that the knowledge gained from data sources by the use of data mining methods can be crucial for the decision making processes. Mladenic et al. claim that the integration of data mining and decision support can lead to the improved performance of decision support systems and can enable the tackling of new types of problems that have not been addressed before. They also argue that the integration of data mining and decision support can significantly improve current approaches and create new approaches to problem solving, by enabling the fusion of knowledge from experts and knowledge extracted from data (Mladenić et. al., 2003c). Chen and Liu (2005) state that data mining is a very useful technology which opens new opportunities for data analysis. Tseng and Lin (2007) have used data mining for efficient discovery of temporal movement patterns or objects in sensor networks.

3.1. Data mining software tool approach

Data mining can be used through two different approaches. The first approach is called *data mining software tool approach* where the use of data mining is typically initiated through ad hoc data mining projects (Goebel and Gruenwald, 1999; Kohavi and Sahami, 2000; Holsheimer, 1999; Radivojevic et. al., 2003). Ad hoc data mining projects are initiated by a particular objective on a chosen area which represents a basis for the defining of the domain. They are performed using data mining software tools which require a significant expertise in data mining methods, databases and/or statistics (Kohavi and 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, post-processing and data transformation (Holsheimer, 1999; Goebel and 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. Data mining software tool approach represents the first generation of data mining (Holsheimer, 1999).

3.2. Data mining application system approach

The data mining software tool approach has revealed some disadvantages. The most important of them is the fact that due to the complexity of data mining software tools, they can not be directly used by business users. Data mining models are produced for business users. For that reason we need applications which will enable them to view and exploit data mining models effectively to facilitate decision support (Kohavi and Sahami, 2000; Goebel and Gruenwald, 1999). This implies to the new approach of the use of data mining which we call *data mining application system approach*. It is an approach which focuses on business users and other decision makers, enabling them to view and exploit data mining models. Models are presented in a user-understandable manner through a user friendly and intuitive GUI using standard and graphical presentation techniques (Aggarwal, 2002). Decision makers can focus on specific business problems covered by areas of analysis 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 and their decision processes (Goebel and Gruenwald, 1999; Holsheimer, 1999; Kohavi and Sahami, 2000; Bayardo and Gehrke, 2001).

Data mining standards undoubtedly represent an important issue for data mining application systems approach (Holsheimer, 1999). Employing common standards simplifies the development of data mining application systems and business application systems utilizing data mining models. 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 where application interface (API) is probably one of the most important of them (Grossman, 2003). The standardized data mining API represents the key issue for data mining application systems approach. Its main advantage in the possibility to leverage data mining functionality using standard API shared by all application systems within information system. JDM (Java Data Mining) is a Java based API specification which has reached final release status in 2004 (JSR-73 Expert Group, 2004). Another important standard for the area of data mining is PMML. PMML is an XML-based standard and language which in theory provides a way for applications to define statistical and data mining models and to share models between PMML compliant applications (Grossman, 2003). In practice, this is limited to sharing the models between different systems that use the same platform (e.g., ODM). All models in DMDSS are stored in the PMML format.

4. Data mining based decision support system to facilitate classification problems

We have developed a data mining based decision support system for a telecommunication service provider. In the following part of the paper it will be called simply a company. We first did a survey about 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 important statement of the survey was that company intolerably needs data mining for performing analysis in the area of analytical CRM. It was stated that the application system approach is more suitable for the introduction of data mining. The main reason for choosing the application system approach was the fact that the area of analytical CRM in the company represents a rather dynamic environment with continual need for repeated analyses. Right after the survey, the development project for decision support system was initiated. The decision support system is called DMDSS and will be introduced in the following part of the paper.

4.1. Related Work

Some decision support systems that use data mining have already been developed and introduced in the literature. Lee and Park (2003) presented *Customized Sampling Decision Support System* (CSDSS) which uses data mining. CSDSS is a web-based system that enables the user to select a process sampling method that is most suitable according to his needs at purchasing semiconductor products. The system enables the autonomous generation of available customized sampling methods and provides the performance information for those methods. CSDSS uses clustering data mining method within the generation of sampling methods. The system is not designed to support other domains; it only supports the domain mentioned.

Bose and Sugumaran (1999) introduced *Intelligent Data Miner* (IDM) decision support system. 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 and multidimensional analysis through enabling OLAP views on multidimensional data. In the data mining part of IDM it supports the creation of models, manipulation of models and presentation of models in various presentation techniques of, among others, the following data mining methods: association rules, clustering and classifiers (classification). The system also performs data cleansing and data preparation and provides necessary parameters for data mining algorithms. Interesting characteristic of IDM is that it makes a connection to external data mining software tool which performs data mining model creation. The system enables predefined and ad-hoc data mining model creation. The authors state that the disadvantage of IDM is the fact that non-technical users (business users) need to have a fair amount of understanding of data mining and that the use of data mining and the creation of data mining models still needs to be clearly directed by the user, especially with ad-hoc model creation.

Polese et. al. (2002) introduced decision support system based on data mining. The system was designed to support tactical decisions of a basketball coach during a basketball match through suggesting tactical solutions based on the data of the past games. The decision support system only supports association rules data mining method and uses association rule algorithm called *Apriori* algorithm combined with *Decision query* algorithm. The decision support system enables the coach to submit data about his tactical strategies and data about the game and the rival team. After that the system provides the coach with opinion about the chosen strategies and with suggestions. The system is not designed to support other domains; it only supports the basketball domain.

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".

Sasa et.al. (2008) presented interesting results in the area of improving and automating decision making through ontologies. Their research did not base on data mining, but the results are important in context of improving and facilitating decision making by business users.

4.2. The Process Model TO SUPPORT decision Making

To determine the process model for DMDSS was one of the key issues within the design of DMDSS. We needed the model which would be appropriate to enable decision support for the purposes of analytical CRM in the company. The process model shows how decision processes are supported by the decision support system. The design of process model was based on CRISP-DM (CRoss-Industry Standard Process for Data Mining) process model. CRISP-DM is a data mining process model which was developed by the industry leaders and the collaboration of experienced data mining users and data mining software tool providers (Shearer, 2000; Clifton and Thuraisingham, 2001; Grossman, 2003). 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 and Gruenwald, 1999; Li et al., 2002). The analysis of data mining process models confirmed CRISP-DM as the most appropriate process model for DMDSS. CRISP-DM process model breaks down the data mining activities into the following six phases which all include a variety of tasks (Shearer, 2000; Clifton and Thuraisingham, 2001): business understanding, data understanding, data preparation, modelling, evaluation and deployment.

CRISP-DM process model was adapted to the needs of DMDSS as a two stage model: the *preparation stage* and the *production stage* (Figure 1). The division into two stages is based on the following two demands, which are the consequence of data mining application system approach used. First, DMDSS should enable repeated creation of data mining models based on up-to-date data set for every area of analysis. Second, business users should only use it within the deployment phase with only basic level of understanding of data mining concepts.

The preparation stage represents the process model for the use of DMDSS for the purposes of preparation of the area of analysis for the production use. During the preparation stage the CRISP-DM phases are performed in multiple iterations with the emphasize on the first five phases: from business understanding to evaluation. The aim of executing multiple iterations of all CRISP-DM phases for every area of analysis is to achieve step-by-step improvements in all of the phases. In the business understanding phase the slight redefinitions of the objectives can be made, if necessary, according to the results of other phases, especially the results of evaluation phase. In the data preparation phase the improvements in the procedures which execute the recreation of data set can be achieved. Data set must be recreated automatically every night based on

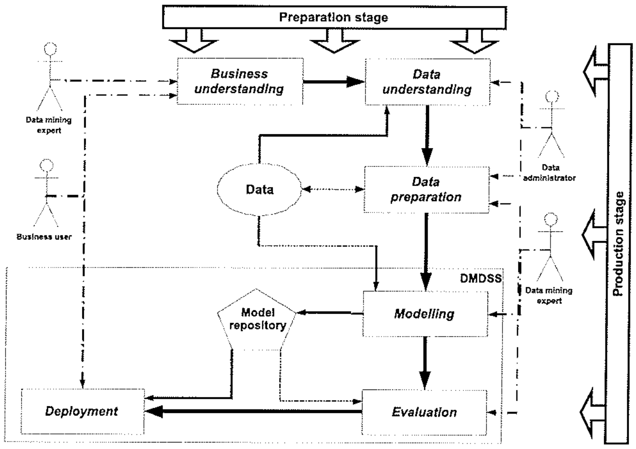

Fig. 1: The process model of DMDSS

the current state of data warehouse and transactional databases. The problems detected in the data preparation phase can also demand changes in the data understanding phase. In the data preparation phase the attribute names of data sources are transformed to aliases which are user-friendly and enable the user to easily understand the meaning of the attribute. In cases when the attribute has the limited set of discrete codes, the codes are decoded to values which are more appropriate to the user than the codes.

In the modelling phase and evaluation phase the model is created and evaluated for several times with the aim to do the fine-tuning of data mining algorithms through finding the proper values of the parameters for the algorithms. It is essential to do enough iterations in order to monitor the level of changes in data sets and data mining models acquired. Through multiple iterations the stability of data preparation phase is reached and optimal values of parameter values for data mining algorithms are identified. According to the results gained in the evaluation phase through multiple iterations, the preparation stage can either reject the area of analysis due to the insufficient quality of models; either approves it with or without minor redefinitions of objectives in the business understanding phase and consequently the changes in other phases. The mission of the preparation stage is to confirm the fulfilling of the objectives of area of analysis for decision support and to assure the stability of data preparation.

The production stage represents the production use of DMDSS for an area of analysis. In the production stage the emphasis is on the phases of modelling, evaluation and deployment, what does not mean that other phases are not encompassed in the production stage. Data preparation, for example, is executed automatically based on procedures developed in the preparation stage. The modelling and evaluation are performed by data mining expert, while the deployment phase is performed by business users. Figure 1 introduces the schema of the process model of DMDSS showing a preparation stage and production stage in a joint view. The schema reflects the phases of both stages of the process model. Beside that schema also shows the roles of data administrator, data mining expert and business user and the phases where they take part in: either actively as the executor of the phase, either only collaborating. The schema also reveals that DMDSS only supports the phases of modelling, evaluation and deployment. The functionalities of DMDSS will be introduced later on in this section.

4.3. The Platform

DMDSS uses Oracle database and was developed on J2EE platform through several iterations (Bajec et.al., 2007). The selection of Oracle was the consequence of the fact that Oracle 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. The second component is Java-based data mining application interface (ODM API), which enables access to services provided by DME (JSR-73 Expert Group, 2004). Another factor that influenced the platform selection was the fact that practically all of the data needed for initial set of areas of analysis was available in Oracle databases in the company. Before finally accepting Oracle 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 gave positive results.

4.4. The functionalities of dmdss to support the solving of classification problems

The design of functional demands for DMDSS and the design of data mining process model were done simultaneously. Both activities are very interrelated, because the process model implies the functionality of a decision support system to a great extent. The functionalities of DMDSS will be introduced based on the production stage of the process model proposed and the functionalities offered by forms to a user. DMDSS supports two roles: *data mining expert* and *business user.* Each of the roles has the access to the forms and their functionalities according to the production stage of the process model (Figure 1).

DMDSS enables the data mining expert to create classification models by using model creation form (Figure 2). When creating the model he inputs 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 set of discrete values which was defined as appropriate in the preparation stage of process model for a particular area of analysis. At the bottom of the form there are recommended values for parameters to acquire a model with fewer or more rules: recommended settings for fewer rules in a model, and settings for more rules in a model.

The examples of forms in the figures shown in the following part of the section are for the area of analysis called "Customers classification". Model testing is performed automatically as the last step of the model creation.

Within the evaluation phase data mining expert can view and evaluate the model using data mining expert model viewing form. This form is very similar to the form for model viewing which is used by business user in the deployment phase and will be introduced later on. While evaluating, the data mining expert can input comments for the model. The role of comments is to help the business users to understand and interpret the models better. In case of classification model evaluating signifies the evaluation of the quality of the model. Model evaluation is performed based on results of the model testing step within the model creation.

As the final step of evaluation phase data mining expert can change the published status of a model to a value *true* if the relative model quality reaches a certain level, and if the model is different from the previously created model of particular area of analysis. It is the duty of data mining expert to evaluate the practical quality of the model, both qualitative and quantitative, and the level of difference of newly created model in relation to previously created model and decide about it. Business users have access only the models with published status set to *true*.

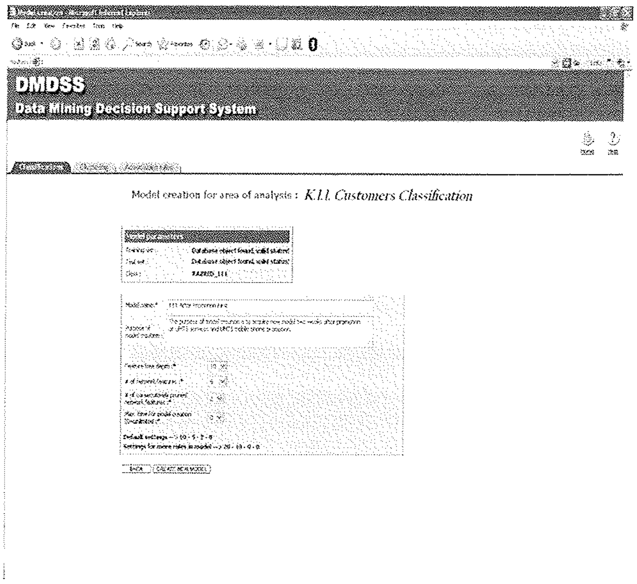

Fig. 2: A model creation form for data mining expert

Within the deployment phase business users have access to the form for model viewing for a business user (Figure 3). There are two visualization techniques available to support model viewing. The first technique is a table where classification rules are presented in IF-THEN form. The second technique is decision trees, where classification rules are converted into decision trees showing equivalent information as classification rules. The decision trees technique is a graphical technique, which enables visual presentation of rules and for that reason it is very appropriate for business users. It is appropriate for business users because it enables them obtain information in an easier way and it can be especially useful in case when model has a lot of rules (over 10, for example). The form shows general information about the model: model name, date of model creation, number of rules, number of comments and number of classification classes. The form enables the filtering of classification rules according to the class. The user can choose either to view the rules for all classes, either only the rules for a chosen class. The form also shows the last comment of a model written by the data mining expert and enables the user to view all other comments through the opening of the form for comments.

In order to present the information about the quality of the model the form also presents testing parameters relative model quality and classification accuracy. For a business user the parameter relative model quality is presented in qualitative form. It is more appropriate than the quantitative form, because this way business user doesn't have to remember exact values. For example: *middle* is when the value is greater than 0 and smaller than 0.5, that is, the relative improvement of model quality is between 0% and 50% with respect to the prediction of the most frequent class (*mode*).

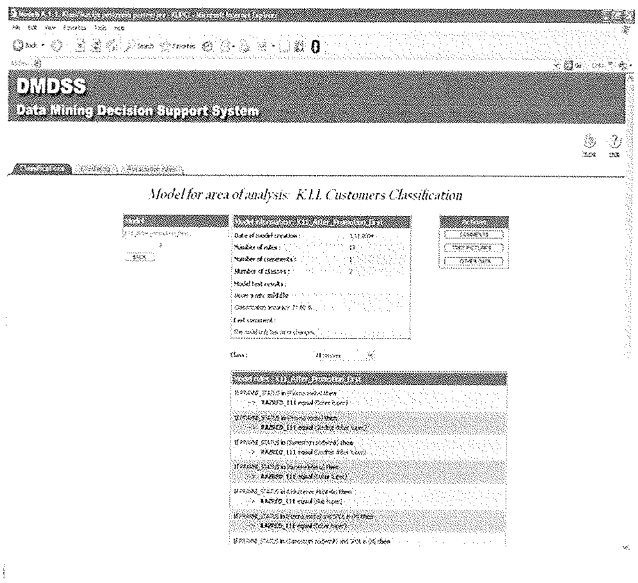

Fig. 3: A model viewing form for business users

The example of the area of analysis which uses classification method in DMDSS is called *Customers classification*. For the purpose of area of analysis customers are ranked into three categories, i.e. classes: a good customer, a normal customer and a bad customer. The aim of the area of analysis is to acquire the model for each customer category. The model acquired enables business users to monitor characteristics of a particular customer category and plan better customer category focused marketing campaigns for acquiring new customers. DMDSS supports additional areas of analysis for the purposes of mobile phone sales analysis, other areas of customer analysis and vendor analysis.

4.5. Example of the Scenario of Use

The area of analysis called *Loyalty model* enables the company to acquire the classification model for customers which are possible defectors, i.e. customers who potentially could switch to another telecommunication service provider. This area of analysis is very important for the company due to intense campaigns for acquiring new customers launched by the competition. Data mining expert creates the new model every day in order to detect changes as soon as possible. If the new model is different comparing to previous one, the data mining expert changes the published status to the value *true*. Analyst in marketing department first checks the new model to evaluate changes. After that the model is applied on the whole set of existing customers. Every customer who is according to the model acquired a possible defector and who hasn't been sent a special offer (a new mobile phone for a very low price) in the last 30 days, gets a special offer to increase the loyalty of the customer and to lower the possibility for him to switch to another company. The results of this area of analysis will be introduced later on in the paper.

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 and Krisper, 2004). 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 4). 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.

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 (Furlan and Bajec, 2008).

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 paper.

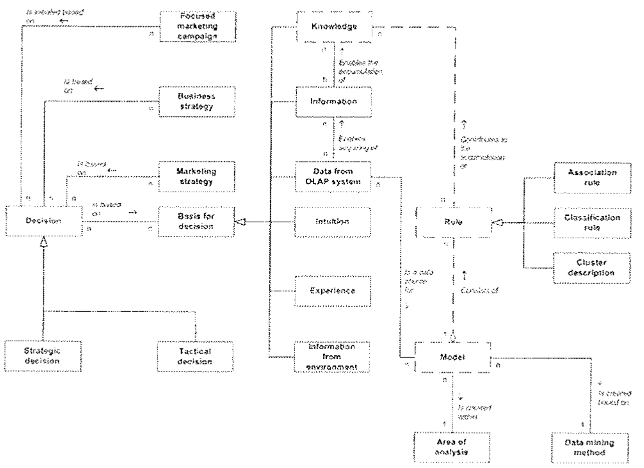

Fig. 4: A meta-model presenting the semantic contribution of the use of DMDSS

5. The experience of use

DMDSS has now been in production use in the company for five months. Before the start of production first the tutorial for data mining for business users was organised. The aim of tutorial was to educate business users with the basic concepts of data mining with the emphasis on teach then how to interpret data mining models correctly. After the tutorial the training was organised where business users learned how to use and truly exploit DMDSS to facilitate decision support through the use of data mining. For one employee from the company, who was planed to take over the role of data mining expert, there was no special training organised. He was a member of the development team and he all time actively collaborated in the preparation stage for every area 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 to facilitate decision support. The most important achievement after five months of use is the fact that business users have really started to understand the potentials of data mining. Suddenly they have got many new ideas for 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 is regularly updated, and after the first year of production it will be discussed and evaluated. Selected areas of analysis will then be implemented and introduced to DMDSS according to the process model introduced in the paper, which also determines the methodology of introducing new areas of analysis to DMDSS.

The use of DMDSS to facilitate decision support has already shown some measurable and non-measurable results:

- Based on the before mentioned area of analysis called *Loyalty model* the company's managers have defined a strategy for continual focused campaigns to increase the loyalty of possible defectors. The first three-month campaign has already been launched and two months after the campaign there were 9% less defectors according to two months before the campaign. Business results from previous years do not show any significant seasonal changes between the two compared time frames. The achievement was gained despite of intensive campaigns launched by the competition.
- The area of analysis called *Customer classification* enables the possibility to acquire the classification model for good, normal and bad customers. The company's managers are defining a strategy for treating new customers based on that model, because it enables the forecasting of customer category for a new customer. Especially important for the company are of course customers who are potentially good. Beside that they are defining a strategy for continual focused marketing campaigns for acquiring good customers.
- The area of analysis called *Customers and mobile phone categories* enables the company to acquire the classification model for customers and three mobile phones categories they are purchasing: high-price, mid-price and low-price mobile phones. For each mobile phone category the model describes the properties of the customers who purchased mobile phones of that category. The company's managers are defining a strategy for continual focused campaigns based on the model to increase the sales of mobile phones. The start of implementation of the strategy is planned in two or three months.

It is important to stress that the strategies mentioned are being defined based on the fact that DMDSS enables the daily creation of models and consequently repeated campaigns with the frequency which is determined by decision maker. It in general mainly depends on the policy covering the area of analysis and also on the frequency of changes in data mining models within the area of analysis.

5.1. The plan for future development

There are some new functionalities planned besides introducing new areas of analysis. The experience of the use of DMDSS has 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 gained by viewing and analysing the rules. We also plan to enhance DMDSS with other data mining methods.

6. Summary and conclusion

DMDSS is decision support system which supports decision processes through classification method. It is a passive DSS, because it supports decision processes through new knowledge acquired without producing explicit decision suggestions or solutions. The mission of DMDSS is to offer an easy-to-use tool which enables business users to exploit classification data mining models with only a basic level of understanding of the classification concepts, which enables them to interpret the models correctly. The process model of DMDSS defines the roles of business user and data mining expert, where phases that demand expertise in data mining are performed by data mining expert and are hidden from business user. Business users only exploit data mining models, which are created by data mining expert. The experience shows that through such a process model we have achieved a rather high level of integration of data mining into daily decision processes through DMDSS. DMDSS was developed for a telecommunications service provider, but its process model and architecture enable its use in any business environment.

Comparing DMDSS to decision support systems introduced in the related work section, we would like to put out the following remarks. First, DMDSS uses data mining services of data mining engine in Oracle database. It does not use external data mining software like IDM, it also does not contain self programmed data mining algorithms like CSDSS (Lee and Park, 2003) and system introduced by Polese et al. (2002). The solution and platform used by DMDSS enables the deployment of data mining models to other J2EE based application systems, what we plan to utilize in the future for the purposes of prediction. The dilemma between data mining software tools and database built-in data mining engines is justified, because data mining software tools have reached a very high level of maturity. But, on the other hand we intensively follow the development of the platform of choice, Oracle Data Mining and accompanying tools, which have already gained a certain level of maturity.

Second, DMDSS only enables predefined areas of analysis like system introduced by Polese et al. (2002) and CSDSS (Lee and Park, 2003) do, whereas IDM enables also ad-hoc data mining analysis. The architecture and the design of DMDSS enable to include any area of analysis as long as data set is available in Oracle database. We believe that in order decision support system to be more appropriate for business users, it should not enable ad-hoc data mining model creation. The process of creation of ad-hoc models should be controlled rigorously in this case and this demands a rather high level of knowledge of data mining. We see that the problem of predefined and ad-hoc areas of analysis is interrelated to the problem the knowledge of data mining and the division of user's roles. We hold the opinion that the users of data mining based decision support system should be divided into two groups: data mining experts who create and evaluate models and business users who exploit models to support decision processes. We believe that with the concepts of predefined areas of analysis and the use of two user roles mentioned, we managed to make DMDSS a useful tool. It is accepted by business users as a tool which enables decision support through data mining models with only a basic level of knowledge of data mining, which is needed to interpret the models.

Although DMDSS is a rather new application system, there exists a plan for further 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. We intend to provide our users with more data mining methods 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 to facilitate decision support.

7. References

Aggarwal, C.C. (2002) 'Towards-Effective and Interpretable Data Mining by Visual Interaction', *SIGKDD Explorations*, Vol. 3, No. 2, pp. 11-22.

Agrawal, R. and Srikant, R. (1994) 'Fast Algorithms for Mining Association Rules in Large Databases', *Proceedings of the 20th International Conference on Very Large Data Bases VLDB'94*, pp. 487-499.

Bajec, M., Vavpotic, D. and Krisper, M. (2007) 'Practice-Driven Approach for Creating Project Specific Software Development Methods', *Information and Software Technology,* Vol. 49, No. 4, pp. 345-365.

Bajec, M. and Krisper, M. (2004) 'A Methodology and Tool Support for Managing Business Rules in Organizations', *Information Systems,* Vol. 30, No. 6, pp. 423-443.

Bayardo, R. and Gehrke, J.E. (2001) 'Report on the Workshop on Research Issues in Data Mining and Knowledge Discovery Workshop (DMKD 2001)', *SIGKDD Explorations,* Vol. 3, No. 1, pp. 43-44.

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

Bose, R. and Sugumaran, V. (1999) 'Application of Intelligent Agent Technology for Managerial Data Analysis and Mining', *The DATABASE for Advances in Information Systems*, Vol. 30, No. 1, pp. 77-94.

Chen, K.C. (2004) 'Decision Support System for Tourism Development: System Dynamic Approach' *Journal of Computer Information Systems*, Vol. 45, No. 1, pp. 104-112.

Chen, S.Y. and Liu, X. (2004) 'The Contribution of Data Mining to Information Science', *Journal of Information Science*, Vol. 30, No. 6, pp. 550-558.

Chen, S.Y. and Liu, X. (2005) 'Data Mining From 1994 to 2004: an Application Oriented Review', *International Journal of Business Intelligence and Data Mining*, Vol. 1, No. 1, pp. 4-21.

Clifton, C. and Thuraisingham. M. (2001) 'Emerging Standards for Data Mining', *Computer Standards & Interfaces*, Vol. 23, No. 3, pp. 187-193.

Forgionne, G.A. and Rudenstein-Montano, B. (1999) 'Post Data Mining Analysis for Decision Support Through Econometrics', *Information, Knowledge, Systems Management*, Vol. 1, No. 2, pp. 145-157.

Furlan, S. and Bajec, M. (2008) 'Holistic Approach to Fraud Management in Health Insurance', *Journal Of Information and Organizational Sciences*, Vol. 32, No. 2, pp. 99-114.

Goebel, M. and Gruenwald, L. (1999) 'A Survey of Data Mining Knowledge Discovery Software Tools', *SIGKDD Explorations*, Vol. 1, No. 1, pp. 20-33.

Grossman, R. (2003) 'KDD-2003 Workshop on Data Mining Standards, Services and Platforms (DM-SSP 03)', *SIGKDD Explorations*, Vol. 5, No. 2, pp. 197.

Heinrichs, J.H. and Lim, J.S. (2003) 'Integrating Web-based Data Mining Tools with Business Models for Knowledge Management', *Decision Support Systems*, Vol. 35, No. 1, pp. 103-112.

Hirji, K.K. (2001) 'Exploring Data Mining Implementation', *Communications of ACM*, Vol. 44, No. 7, pp. 87-93.

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

JSR-73 Expert Group (2004) 'Java™ Specification Request 73: Java™ Data Mining (JDM)', Java Community Process.

Kohavi, R. and Sahami, M. (2000) 'KDD-99 Panel Report: Data Mining into Vertical Solutions', *SIGKDD Explorations*, Vol. 1, No. 2, pp. 55-58.

Kuan, C.C. (2004) 'Decision Support System for Tourism Development: System Dynamics Approach', *Journal of Computer Information Systems*, Vol. 45, No. 1, pp. 104-112.

Lee, H.L. and Park, C. (2003) 'Agent and Data Mining Based Decision Support System and its Adaptation to a New Customer Centric Electronic Commerce', *Expert Systems with Applications*, Vol. 25, No. 4, pp. 619-635.

Li, T., Li, Q., Zhu., S. and Ogihara, M. (2002) 'A Survey on Wavelet Applications in Data Mining', *SIGKDD Explorations*, Vol. 4, No. 2, pp. 49-68.

Kukar, M. (2003) 'Transductive reliability estimation for medical diagnosis', *Artificial Intelligence in Medicine,* Vol. 29, pp. 81-106.

Kukar, M. (2006) 'Quality assessment of individual classifications in machine learning and data mining', *Knowledge and Information Systems,* Vol. 9, No. 3, pp. 364-384.

Mladenić, D., Lavrač, N., Bohanec, M. and Moyle, S. (2003) *'Data Mining and Decision Support: Integration and collaboration (Chapter 1: Data Mining: Authors; Lavrac, N. and Grobelnik, M.)'*, Kluwer Academic Publishers, Dordrecht, The Nederlands.

Mladenić, D., Lavrač, N., Bohanec, M. and Moyle, S. (2003) *'Data Mining and Decision Support: Integration and collaboration (Chapter 3: Decision Support, Author: Bohanec, M.)'*, Kluwer Academic Publishers, Dordrecht, The Nederlands.

Mladenić, D., Lavrač, N., Bohanec, M. and Moyle, S. (2003) *'Data Mining and Decision Support: Integration and collaboration (Chapter 4: Integration of Data Mining and Decision Support, Authors: Lavrac, N. and Bohanec, M.)'*, Kluwer Academic Publishers, Dordrecht, The Nederlands.

Nemati, H. and Barko, C.D. (2002) 'Enhancing Enterprise Decisions Through Organizational Data Mining', *Journal of Computer Information Systems*, Vol. 42, No. 4, pp. 21-28.

Patelis, A., Metaxiotis, K., Nikolopoulos, K. and Assimakopoulos, V. (2003) 'ForTV: Decision Support System for Forecasting Television Viewership', *Journal of Computer Information Systems,* Vol. 46, No. 1, pp. 25-34.

Polese, G., Troiano, M. and Tortora, G. (2002) 'A data Mining Based System Supporting Tactical Decisions', *Proceedings of the 14th international conference on Software engineering and knowledge engineering*, pp. 681-684.

Radivojevic, Z., Cvetanovic, M., Milutinovic, V. and Sievert, J. (2003) 'Data Mining: A Brief Overview and Recent IPSI Research', *Annals of Mathematics, Computing & Teleinformatics,* Vol. 1, No. 1, pp. 84-90.

Sasa, A., Juric, M.B. and Krisper, M. (2008) 'Service-Oriented Framework for Human Task Support and Automation', *IEEE Transactions on Industrial Informatics*, Vol. 4, No. 4, pp. 292-302.

Shearer, C. (2000) 'The CRISP-DM Model: The New Blueprint for Data Mining', *Journal of Data Warehousing*, Vol. 5, No. 4, pp. 13-22.

Srivastava, J., Cooley, R., Deshpande, M., and Tan, P.N. (2000) 'Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data', *SIGKDD Explorations*, Vol. 1, No. 2, pp. 12-23.

Steblovnik, K., Tasič, J., and Zazula D. (2005) 'Intelligent Home Assistant', *Informacije MIDEM*, Vol. 35, No. 4, pp. 240-252.

Tseng, V.S. and Lin, K.W. (2007) 'Energy Efficient Strategies for Object Tracking in Sensor Networks: A Data Mining Approach', *Journal of Systems and Software*, Vol. 80, No. 10, pp. 1678-1698.

Ward, S.G. (2000) 'Decision Support for What-If Analysis and the Confirmation Bias', *Journal of Computer Information Systems*, Vol. 40, No. 4, pp. 84-92.

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

Zhong, M., Klein, G., Pick, R.A. and Jiang, J.J. (2005) 'Vector Error-Correction Models in a Consumer Packaged Goods Category Forecasting Decision Support System', *Journal of Computer Information Systems*, Vol. 46, No. 1, pp. 25-34

dr. Rok Rupnik
University of Ljubljana, Faculty of Computer and Information Science
Tržaška 25, 1000 Ljubljana, Slovenija
Email: rok.rupnik@fri.uni-lj.si
Fax: +386-1-300-7879
Tel: +386-1-4768-814

Prispelo (Arrived): 18.03.2009 *Sprejeto (Accepted): 09.09.2009*