



ON APPROACH FOR THE IMPLEMENTATION OF DATA MINING TO BUSINESS PROCESS OPTIMISATION IN COMMERCIAL COMPANIES

Aleksander PIVK^a, Olegas VASILECAS^b, Diana KALIBATIENE^c, Rok RUPNIK^d

^a*Department of Intelligent Systems, Jozef Stefan Institute, Jamova 39,
SI-1000 Ljubljana, Slovenia*

^{b,c}*Department of Information Systems, Vilnius Gediminas Technical University,
Sauletekio al. 11, LT-10223 Vilnius, Lithuania*

^d*Faculty of Computer and Information Science, University of Ljubljana,
Trzaska cesta 25, SI-1000 Ljubljana, Slovenia*

Received 14 March 2013; accepted 28 March 2013

Abstract. Nowadays, organisations aim to automate their business processes to improve operational efficiency, reduce costs, improve the quality of customer service and reduce the probability of human error. Business process intelligence aims to apply data warehousing, data analysis and data mining techniques to process execution data, thus enabling the analysis, interpretation, and optimisation of business processes. Data mining approaches are especially effective in helping us to extract insights into customer behaviour, habits, potential needs and desires, credit associated risks, fraudulent transactions and etc. However, the integration of data mining into business processes still requires a lot of coordination and manual adjustment. This paper aims at reducing this effort by reusing successful data mining solutions. We propose an approach for implementation of data mining into a business process. The confirmation of the suggested approach is based on the results achieved in eight commercial companies, covering different industries, such as telecommunications, banking and retail.

Keywords: business process, data mining, CRISP-DM, ontology, SOA.

Reference to this paper should be made as follows: Pivk, A.; Vasilecas, O.; Kalibatiene, D.; Rupnik, R. 2013. On approach for the implementation of data mining to business process optimisation in commercial companies, *Technological and Economic Development of Economy* 19(2): 237–256.

JEL classification: C8.

Introduction

The goal of a commercial organisation is to perform business activities in alignment with the vision and mission statement, which should clearly define what the organisation intends to become and achieve at some point in the future, stated in competitive terms. A business activity may encompass many different objectives, such as increasing the company's market share, gaining the competitive edge with offered products and services, providing excellent customer service, or promoting profitable and sustainable actions that meet customer needs. These activities involve cooperation of several business units that make the underlying processes highly connected, inter-dependent and extensive.

Therefore, contemporary organisations aim to automate their business processes to improve operational efficiency, reduce costs, improve the quality of customer service and reduce the probability of human error. Various tools and techniques are available for designing business processes. They fall into two categories: graphical (such as flow charts, data flow diagrams (e.g. Yourdon's technique), role activity diagrams (RAD), role interaction diagrams (RID) etc.) and mathematical (such as Petri nets and etc.) (Aguilar-Savén 2004; Recker *et al.* 2009). Besides, some simulation tools can be used for detecting performance bottlenecks. However, current business process management systems lack analytical tools to quantify performance of business processes on business metrics.

Business process intelligence aims to apply data warehousing, data analysis, and data mining techniques to process execution data, thus enabling the analysis, interpretation, and optimisation of business processes. In this paper, we focus on the use of data mining in a business process.

Data mining approaches are most effective in helping us extract the insights into customer behaviour, habits, potential needs and desires, credit associated risks, fraudulent transactions etc. Data mining results should be used at an appropriate time and for the right purpose, e.g. in line with established strategies or for better, faster and more accurate decision-making on customer service strategies to be used on a case-by-case basis. These efforts result in greater value for customers as well as greater revenue and improved productivity for the company.

However, such problems demand for complex solutions that would consider two perspectives, namely: (1) intense cooperation and inter-connectivity among business units, underlying activities and processes; as well as (2) comprehensive knowledge of customer behaviours, needs and demands. Traditionally, such business problems have not been addressed at all or at best, experience-based approaches have been used. Lately, with the adoption of data mining technology and proper business process renovation approaches, they can be addressed in a controlled and systematic manner. In terms of domain knowledge capturing, the authors used ontologies, which allowed reaching a higher maturity level.

The article describes the approach, which focuses on (a) faster and improved decision-making, (b) retention of most valuable customers, (c) assurance of satisfied and loyal customers, and (d) identification of potential loan defaulters. Provided both perspectives are appropriately considered, a company can expect additional business benefits, such as sustainable revenue growth. The aforementioned approach was substantiated using results of eight commercial companies operating in different industries, namely telecommunications, banking and retail.

The paper is structured as follows: Section 1 presents the related background information; Section 2 reviews relevant works; Section 3 describes the approach for implementation of data mining into a business processes; Section 4 presents a case study in eight companies on the use of data mining to supporting direct marketing with the help of the presented approach; Section 5 reviews the future prospects; and the last section concludes and summarises the paper.

1. Background

This section provides a brief introduction to areas that are important for the paper.

Data mining is the process of data analysis that results in discovery of implicit, but potentially useful information as well as previously unknown patterns and relationships, which are hidden in data (Witten, Frank 2005). In the last decade, the digital revolution has provided relatively inexpensive and accessible means for data collection and storage. The increase in data volumes causes greater difficulties related to extraction of useful information that supports decision-making. Traditional manual data analysis has become insufficient, which makes methods for efficient computer-based analysis indispensable. This need gave rise to a new interdisciplinary field of data mining. Consequently, statistical methods, pattern recognition and machine learning tools are used to support data analysis and discovery of principles that lie within data.

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 industry leaders in collaboration with 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). Special Interest Group (CRISP-DM SIG) was formed in order to further develop and refine CRISP-DM process model to ensure the appropriate level of service for the data mining community. CRISP-DM version 1.0 was presented in 2000 and it is now 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, each of which includes a variety of tasks (Shearer 2000; Rupnik *et al.* 2007):

- *Business understanding*: focuses on understanding the project objectives from the business perspective and transforming them 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 aimed at getting familiar with data, discovering first insights into the data and identifying data quality problems.
- *Data preparation*: covers all activities pertaining to construction of the final data set starting with the *initial* raw data and including selection, cleaning, construction, integration and formatting of data.
- *Modelling*: covers the creation of various data mining models. The phase starts with the selection of data mining methods followed by creation of data mining models and finishes with their assessment. As some data mining methods have specific requirements on the form of data, a return to the data preparation phase is often necessary.

- *Evaluation*: evaluates the data mining models created during the modelling phase. The model evaluation aims to confirm that models are of high quality to achieve business objectives.
- *Deployment*: covers activities aimed at organising knowledge gained through data mining models and presenting it in a way that would capacitate users exploiting it for decision-making.

Even though CRISP-DM recognises that a project does not generally finish with creation of a model and introduces four tasks of the deployment phase, they lack a more detailed description on deployment of data mining results into a business process, which requires implementation of a repeatable data mining process (Chung, Gray 1999).

Throughout the last twenty years, the business community started giving more attention to **business processes**. There are several definitions of a business process. One of the most commonly used is by Davenport and Short (1990) describing a business process as a set of logically related tasks performed to achieve a defined business outcome. However, it should be considered that contemporary business processes, such as order processing, account replenishment, etc. are typically integrated and automated (often partially). Therefore, they have a long duration, involve coordination across many manual and automated tasks, require access to several different databases and several application systems (e.g. ERP systems), and are governed by complex business rules (Dayal et al. 2001).

Information is vital to all business processes that comprise operations and management of an organisation (Chaffey, Wood 2005). Therefore, it should not come as a surprise, that business process renovation research demonstrated the critical role of information technologies in business process renovation (Broadbent et al. 1999). Nowadays, IT offers very good solutions for business process renovation.

In addition to IT, business process renovation requires consideration of organisational and managerial issues, such as cross-functional integration, stakeholder involvement, leadership qualities, and employee motivation. According to (Chaffey, Wood 2005), high failure rates of information system projects are often a consequence of managerial neglect of user reactions towards new working methods. This issue is very important for data mining implementation into business processes.

2. The use of data mining in business processes

Although the use of data mining in business processes has become more frequent, it has not reached the level that would produce potential benefits. The literature review reveals that it is mainly used to support decision-making (Rupnik et al. 2007). There are only few examples on introduction of the daily use of data mining in business processes (Kohavi, Provost 2001; Feelders et al. 2000; Chung, Gray 1999). It can be stated that data mining is predominantly used to support decision-making on a tactical level in decision processes.

Chung and Gray (1999) argue that there is a lot of research made in the area of data mining model creation; however, insufficient attention has been given to the use of data mining models in business processes.

Kohavi and Provost (2001) argue that it is important to enable the use of data mining in business processes through automated solutions. They discuss the importance of the ease of integration of data mining in business processes. In their paper, the authors regard the integration of data mining in business processes as the consequence of the need to incorporate background knowledge in business processes. They state that deploying automated solutions to previously manual processes can be rife with pitfalls. Authors also argue that one must deal with social issues when deploying automated solutions to previously manual processes.

Feelders *et al.* (2000) discuss the use of data mining in business processes with the emphasis on the integration of data mining models and solutions into existing application systems within information systems. Authors argue that it is essential that the results of data mining are used to support such business processes as direct mail for a group of potential customers.

Gray (2005) discusses data mining as an option of knowledge sharing within the enterprise. He states that through the integration of data mining in business processes, knowledge sharing is not only present in business processes on the tactical level, but also on the operational level.

Ciflikli and Ozjirmidokuz (2010) discuss the use of data mining in manufacturing processes and state that the knowledge acquired through data mining improves processes.

Kurgan and Musilek (2006) state that the use of knowledge acquired through data mining can improve work of business users through semi-automating or automating activities. He argues that semi-automating is more realistic than automating and declares this as important new trend in knowledge discovery.

Wegener and Ruping (2011) propose a pattern-based approach for integrating data mining in a business process. Their approach is based on CRISP and includes the definition of data mining patterns and a hierarchy of tasks to guide the specialisation of abstract patterns to concrete processes, and a meta-process for applying patterns to business processes. These data mining patterns allow representing the reusable parts of a data mining process at different levels of generalisation and provide a simple formal description for the reuse and integration of data mining. They evaluated the suggested approach in a fraud detection case study in the healthcare domain. Definition, architecture and implementation of the underlying services are yet to be detailed in the future. In addition, the following question would have to be addressed: How to model data within a process?

As the review suggests, only Wegener and Ruping (2011) propose a practicable approach on integration of data mining into a business process. However, a lot of questions are yet to be answered.

The research presented in this chapter was also motivated by prior work delivered by the authors of this article. They developed a data mining process model and an appropriate decision support system based on data mining for a telecommunications company (Rupnik *et al.* 2007). The aforementioned research explored the use of data mining application systems. In terms of this approach, data mining is not used for ad-hoc projects, but through data mining based decision support systems. The authors aimed defining the ways business users could use data mining models to facilitate decision-making on the tactical level where data mining experts create and describe data mining models while business users use them. The feedback received from the telecommunications company indicated that in terms of business processes, the use of data mining models resulted in an added value.

3. An approach for the implementation of data mining into business processes

We propose a methodology for the implementation of data mining into business processes, which is based on the following contributions: 1) approaches to BI implementations (e.g. Moss, Atre 2003; Williams, S., Williams, N. 2007; Shearer 2000); 2) a methodological framework to business process renovation and IS development (Kovačič, Bosilj-Vukšič 2005; Shearer 2000); 3) CRISP-DM framework, and 4) our experience with the implementation of data mining in analytical business process.

The main elements of the proposed methodology are:

- Assessment of organisational processes and their structure;
- Assessment of resources and potential improvements;
- Business process renovation;
- Extended CRISP-DM methodology;
- Operationalization of data mining results, i.e. the use of data mining models in business processes.

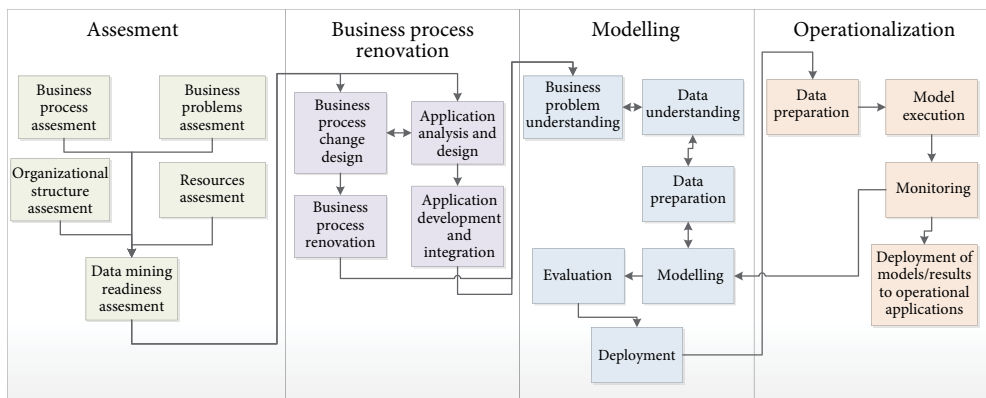


Fig. 1. The approach for successful implementation of data mining into business processes

Fig. 1 depicts the proposed approach, its major phases, related activities and the flow. The following subsections provided a detail description of each phase, its key activities and related issues.

3.1. Assessment

The purpose of the first phase is to analyse and assess readiness of an organisation to incorporate data mining operations into its business processes. The assessment of each activity must consider the following evaluation elements:

- *Risks*: the assessment of risk includes a careful examination of diverse factors that can bring about different types of risks, ranging from strategic, compliance, financial, and operational;

- *Challenges*: the assessment must consider all types of challenges the organisation is expected to face when initiating new business improvement processes. Types of challenges include strategic, legal and compliance, financial, operational, and portfolio;
- *Opportunities*: the assessment considers potential improvements on all levels of the organisation;
- *Potential value*: it identifies the potential impact on the organisation's performance in terms of money or market position.

The assessment phase covers the following activities:

- *Assessment of business processes*: business processes to be affected need to be identified and potential improvement scenarios assessed for all evaluation elements;
- *Assessment of business problems*: business problems to be solved using data mining approach need to be identified and evaluated in terms of all 4 perspectives. Business problems should be identified during the assessment of business processes;
- *Assessment of the organisational structure*: the impact of the new or renovated business processes on the existing organisational structure must be assessed;
- *Assessment of resources*: this activity comprises two main areas for analysis:
 - availability and skills of business analysts, business users, and support teams;
 - availability, sophistication, and interoperability of the data mining platform to be used;
- *Assessment of data mining readiness*: this activity comprises of the assessment results from previous activities to determine the level of readiness for incorporation of data mining operations into business processes. Data mining ontology is useful at this stage. It is used to match and combine data mining activities into business process activities.

3.2. Business process renovation

To ensure successful implementation of data mining, two key activities must be executed more or less simultaneously: *process change design* and *application analysis and design*. The aim of the former is to define and design the changes required for a business process to ensure data mining brings added value. The aim of the latter is to make analysis and design of the potential application that would use data mining. It not only relates to new applications that would have to be developed but also to existing operational and analytical applications that would need to be adapted to the use of data mining models and support changes in business process. Both activities are very dependent on each other. For example, a particular application must provide the functionalities suitable for changes designed in terms of the activity "*process change design*". Once the activity "*application analysis and design*" is finalised, the development of an application is initiated through the activity "*application development and integration*". The aim of this activity is not only to develop an application but also to integrate it into information system of the company, i.e. integrate it with other applications.

When the activity "*business process change design*" is finished, the activity "*business process renovation*" is initiated. The aim of this activity is to renovate the business process according to changes defined and designed during the previous activity.

3.3. Modelling

The purpose of the modelling phase is to create a stable data mining model that needs to tackle the identified business problem. The core activities of this phase are defined by CRISP-DM process model, which breaks down the life cycle of a data mining project into the following six activities that all cover a variety of tasks (Shearer 2000):

- *Understanding of a business problem*: focuses on understanding a business problem and transforming it into a data mining problem, i.e. domain definition;
- *Data understanding*: starts with an initial data collection, which is followed by activities required to get familiar with data, to discover first insights and identify data quality problems;
- *Data preparation*: covers all activities to construct the final data set from the initial raw data including data integration, selection, cleaning, transformation (i.e. log transformations and power transformations), imputation (i.e. creating new variables on the basis of existing ones, such as averages, sums, proportions, ratios);
- *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 their assessment. Some data mining methods have specific requirements on the form of data, thus a step back to data preparation phase is often necessary;
- *Evaluation*: evaluates the data mining models created during the modelling phase. The aim of the model evaluation is to confirm that the models are well performing and stable thus able to achieve required business objectives;
- *Deployment*: covers the activities related to arrangement of knowledge gained through data mining models in a way users could exploit it within decision-making in business processes.

It is important to note, that this phase demands an intense cooperation among several teams within an organisation, namely business analyst, IT experts, business users and data mining experts. A special attention must be paid to the assessment phase in order to avoid risk factors associated with collaboration of several teams, in particular in the *Organisational structure* and *Assessment of Resources* activities.

3.4. Operationalization

In the operationalization phase, the daily use of data mining models in business processes is achieved through the following activities:

- The *data preparation* activity is implemented through the automated procedures defined and tested during the previous phase;
- Data mining models being developed, evaluated, and deployed during the previous phase, are executed during this phase as part of the “*model execution*” activity. The execution could be run as a batch process (automated execution at scheduled time-frames, e.g. daily, weekly, or monthly) or on demand when the updated information is required (i.e. dynamic scoring);

- Data mining models are monitored during the “*monitoring*” activity. The aim is to track the quality, stability, and performance of deployed models through time and hence measure their ability to consistently support resolution of business problems. If the performance parameters of a particular model fall below some threshold, the model execution for that model is aborted until the model is refreshed;
- During the activity “*deployment of models to various applications*” data mining results (scores) are deployed to various applications, which are adapted to use them. Additionally, the results must be written to a data warehouse in order to keep and track the history.

The methodology is structured into phases in order to be able to track and control the entire procedure of incorporating data mining operations into business processes thus providing the organisation with the ability to learn, enhance, and improve.

4. The case study on the use of data mining in eight companies to support direct marketing

This section presents a case study on the impact of data mining within direct marketing business processes in eight organisations: four banks, three telecommunications companies and one retail company. All eight data mining implementation cases were performed within the last year using the approach introduced in the previous section.

The following part of the section introduces typical process models according to those particular to eight companies. One could say that typical process models were introduced regardless of the industry. Common elements of a process model are demonstrated while unimportant or insignificant elements are ignored. First, the direct marketing process model entitled “BEFORE data mining implementation” is introduced. It summarises the model of direct marketing business process for all eight organisations. Next, direct marketing process model “AFTER data mining implementation” is described. As in some companies, product managers have the main role in process model discussions, in other – the main role is given to segment managers. Consequently, process model descriptions use a common term *business user*.

4.1. Process model “BEFORE data mining implementation”

The summarised process model “*BEFORE data mining implementation*” of eight companies includes the following activities (Fig. 2):

- *Defining a business problem*: a business user decides to devise a marketing campaign aimed at potential customers with a particular business goal or problem;
- *Creating the list of export-based rules*: a list of export-based rules, typically expressed in natural language, is created by a business user and constitute the initial target groups that have to be contacted;
- *Creating the list of initial leads*: an IT specialist resolves human-language rules and builds queries to create an initial list of leads; then, returns the list to the business user together with requested data fields;

- *Reviewing the list of leads and creating additional refinement rules*: a business user reviews the list of initial leads and further explores the target groups using provided data fields. As part of the review, the business user defines additional refinement rules;
- *Creating the refined list of leads*: the IT specialist creates a new (refined) list of leads based on additional refinement rules defined by the business user. The last two steps may be repeated several times making this a very time-consuming process, in particular due to misunderstandings as well as time and resource availability constraints;
- *Printing out offers*: once the lists of target groups are finalised and approved, the marketing offer is prepared and the lists with corresponding offers are sent to the printing office. Offers are printed and mailed to customers;
- *Response tracking*: as customers can respond to offers using different channels (e.g. physically arriving to a bank or a shop, over a telephone or e-channel), tracking of responses is typically very limited mostly because data appears in a number of systems. Some companies that participated in the study do not track responses at all.

4.2. Process model “AFTER data mining implementation”

The summarised process model “*AFTER data mining implementation*” (Fig. 3) of eight companies includes the following activities:

- *Creating the list of leads*: a business user has access to an operational application (e.g. campaign management), which uses the dedicated customer focused datamart. Access to this data enables the business user to analyse the customer database, select target groups based on expertise, test several scenarios and perform required refinements. Note that IT experts are no longer required to create target lists;
- *Optimisation*: the dedicated customer focused datamart also includes results (scores) of data mining models for particular business problems, where these results are typically populated on the regular basis. Additionally, the operational application allows users to run models on demand providing the most up-to-date information on customers. Business users can exploit this data to further refine and optimise final target groups, for example, to calculate the expected value or profit in case the offer is accepted, or determine the tendency to respond to the offer;
- *Communicating with potential customers*: once the lists of target groups are finalised and approved and the offers are created, communications is executed using appropriate channels (e.g. direct mail, e-mail, SMS, retail branch, shop);
- *Response and performance tracking*: all responses from associated systems are gathered into a dedicated customer-focused datamart. The automation of this process enables automated response tracking and performance measurement.

Organisations provide different response mechanisms, such as coupons, call centres, shops or subsidiaries, e-mails, personal URLs, or activation codes. The purpose of response tracking is to gather information from all mechanisms to better understand customer reactions to the offer, the numbers that accepted it (converted), the numbers that showed some interest but were not converted, and the amount of non-responders. Based on response data,

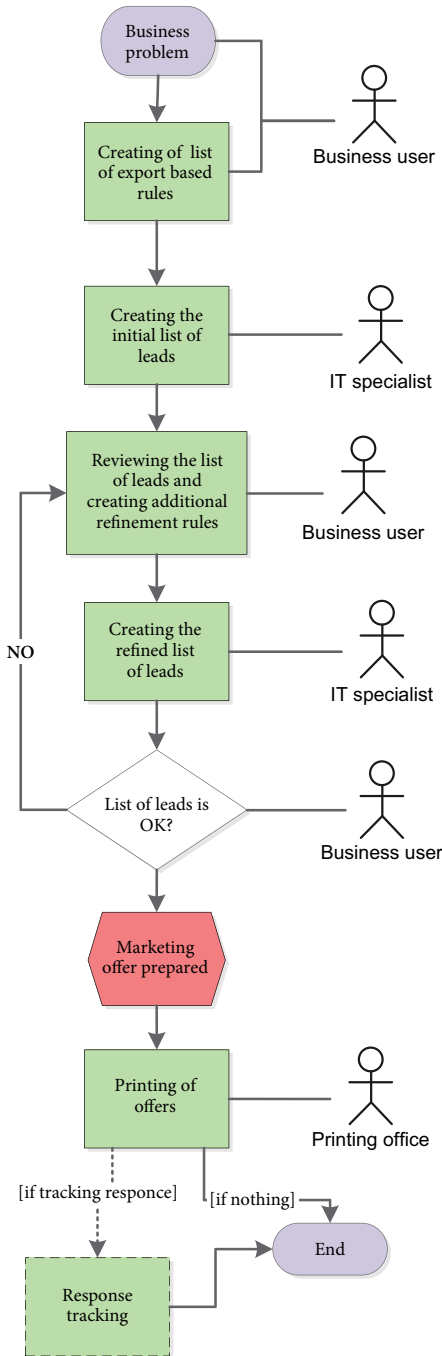


Fig. 2. Process model “BEFORE data mining implementation”

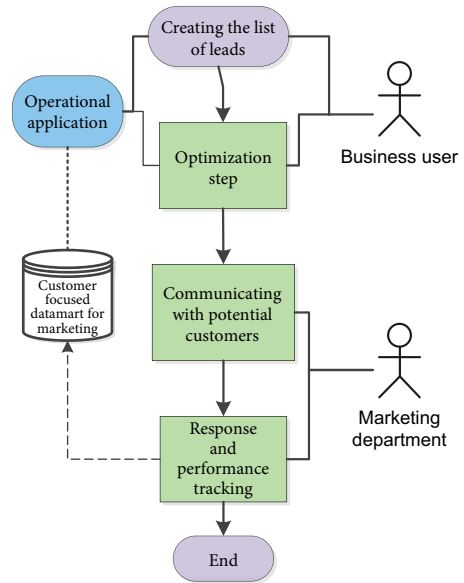


Fig. 3. Process model “AFTER data mining implementation”

a performance analysis is undertaken. The primary goal of such analysis is to determine the return on (marketing) investment. It comprises such information as cost and profit per acquisition, cost and expected profit per sale, time to convert a prospect into a customer and marketing response mechanisms that pull a better response rate for contacted leads.

4.3. Discussion

Data mining has been implemented in eight companies. The progress of implementation projects from their start to the end has revealed that companies with a data warehouse (DWH) had a significant advantage. Those companies implemented data mining at a higher level of sophistication. Values pertaining to *data mining level of sophistication* and *data level of sophistication* are presented in tables below (Tables 1 and 2).

Table 1. The level of data mining sophistication

Data mining level – numerical value	Data mining level – descriptive value	Description
1	Low	No data mining
2	Mid	Some data mining modelling in place with manual work to operationalise
3	High	Data mining modelling regularly undertaken; operational applications adapted for the use of data mining results; automated monitoring processes; etc.

Table 2. Data level of sophistication

Data level – numerical value	Data level – descriptive value	Description
1	Low	No DWH
2	Mid	DWH in place with some dedicated datamarts
3	High	DWH in place; dedicated datamarts; automated data flows;

Project monitoring was used to create the snapshot of the status of projects in autumn of 2010, which is presented in Table 3.

Table 3. The status of data sophistication and data mining sophistication in autumn of 2010

Company	DM Level	Data level	DM Level	Data Level
Bank A	2	2	Mid	Mid
Bank B	2	3	Mid	High
Bank C	1.5	2	Mid	Mid
Bank D	3	3	High	High
Telco A	2	3	Mid	High
Telco B	2	1	Mid	Low
Telco C	1	1	Low	Low
Retail	1	2	Low	Mid

In order to demonstrate the relationship between the *data level* and *data mining level* of sophistication more clearly, the mosaic plot (Fig. 4) is presented below: a higher data sophistication level implies a higher value of data mining sophistication level.

As the mosaic plot (Fig. 4) and the table (Table 3) suggest, only high data level of sophistication facilitates high data mining level of sophistication. The presence of DWH indicates a higher level of data integration and, thus, a much better basis for data mining.

Several advantages of the use of data mining in direct marketing business process were observed, for example: business users are more independent from IT users; besides, direct marketing business process is better controlled and more efficient.

The analysis of data mining implementation in eight companies revealed the following benefits of the process model “AFTER data mining implementation”:

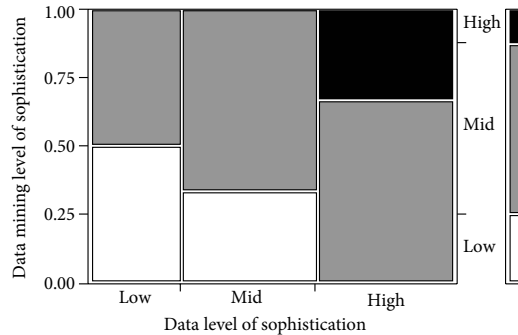


Fig. 4. Relationship between data level of sophistication and data mining level of sophistication

- Significantly less involvement of the IT department: once business processes were implemented and in production, the IT involvement was reduced by 60–90%;
- Independence of business users;
- The re-defined and optimised business processes led to the following improvements observed on a monthly basis:
 - The number of executed campaigns increased by 3–12 times;
 - Marketing cost reduced by 50–75%;
 - The number of contacted leads reduced by 30–87.5%;
 - The return on (marketing) investment grew by 50–400%;
 - Time to market dropped by 45–85%;
 - The conversion rate improved by 20–100%;
 - Effectiveness of business users improved by 20–30%.
- Improved control of the process;
- Consistent tracking and reporting that provides the ability to learn and improve.

5. The potential role of ontologies for the use of data mining in business processes

This section presents advices on the ways to deal with some challenges related to the use of data mining presented in the paper. Nowadays, *Ontologies* coupled with *Service-Oriented Architecture* (SOA) can be used to make new concepts for various approaches. Therefore, this section presents the main definitions of these concepts and a vision for using them in future research.

Motivation of the authors for research on the use of ontologies and SOA for data mining was based on the aspiration to move the use of data mining in business processes to a higher level of maturity. In the opinion of the article authors, in terms of the use of data mining, the following levels of maturity can be named:

- The first level of maturity: data mining used on demand or ad-hoc basis;
- The second level of maturity: data mining is used on a regular basis within business processes or decision processes. This is the level, which basically reflects the methodology

for the implementation of data mining into business processes, which is introduced in this paper (process model “AFTER data mining implementation”).

Observing those two levels and their characteristics, the authors saw a promising option for a higher added value of the use of data mining in the following areas:

- The possibility to define and model characteristics of a business process, within which data mining will be used;
- The possibility to define and model the business domain and business rules within the domain;
- The possibility to define and model the ways data mining will be used in those business processes;
- The possibility to store those definitions and models in an independent location and enable the use of models to every authorised user and application system.

In the opinion of the article authors, the third level of maturity of the use of data mining in business processes is based on ontologies. The use of ontologies for the use of data mining is beneficial in the following situations: the importance of prior knowledge, the knowledge and understanding of the problem, the choice of the proper data mining method to undertake knowledge discovery and the expertise in similar situations or problems. The remaining part of the paper, discusses ideas for future work.

5.1. Ontologies

There are many definitions of an ontology, which mostly depends on the task and reasons of its use. In computer science, Gruber (1995) defines ontology as a specification of a conceptualisation. Based on the definition by Guarino (1998), ontology consists of concepts, their definitions and relationships between concepts and constrains expressed as axioms. However, the modern definition of an ontology is extended in terms of instances (in Protégé¹ ontologies) or individuals (in Ontology Web Language (OWL) (OMG 2005) based ontologies), which represent objects in the domain of interest², e.g. in the context of data mining. According to the components, ontology defines the basic concepts, their definitions and relationships comprising the vocabulary of a domain and the axioms for constraining interpretation of concepts and expressing complex relationships between concepts (Vasilecas *et al.* 2009; Kalibatiene, Vasilecas 2012). Some authors, for example (Falbo *et al.* 1998), make a distinction between properties and concepts. According to (Li *et al.* 2010), the core ontology can be decomposed into various sub-ontologies according to different kinds of knowledge.

The creation of *data mining ontologies* is motivated by the need for unification of the data mining domain and the demand for formalised representation of outcomes of the data mining process. Some authors (Panov *et al.* 2010) propose heavy-weight ontology, named *OntoDM*, based on a general framework for data mining. It represents entities such as data, data mining tasks and algorithms, and generalisations. Other authors (Gong *et al.* 2009) use a data mining ontology for selecting data mining algorithms and semantic searching. E.g. ontology is used

¹ <http://protege.stanford.edu>

² Also known as the *domain of discourse*.

to match a user query with a resource description and to select the appropriate service. As stated by the authors, this method is especially helpful for data mining beginners that have no extensive knowledge on the kind of algorithm, which could perform their tasks.

As stated in (Ankolekar *et al.* 2002), the *Process Ontology* describes a service in terms of its inputs, outputs, preconditions, effects, and, where appropriate, its component sub-processes. The primary kind of entity in the Process Ontology is a process. The basic Process class has several associated properties. A process can have any number of *inputs*, representing the information that is, under some conditions, required for the execution of the process. It can have any number of *outputs*, the information that the process provides, conditionally, after its execution. Another important type of parameter specifies the *participants* in a process. There can also be any number of preconditions, which must all be considered in order for the process to be invoked. Finally, the process can have any number of *effects*, which are the side effects that result from execution of the program in real life. Outputs and effects can have conditions associated with them. *Conditions* are the implementation of business rules, under which business process is performed. For more details see Ankolekar *et al.* (2002).

This research took some interest in the context of data mining and process. In the context of process ontology, *domain ontology* presents information, which is used as input or output of a process. Axioms of a domain ontology present constraints of a domain, which have to be satisfied for the input or output of the process.

In the context of data mining ontology, the domain ontology presents a structure of data in terms of concepts, relationships among concepts, definitions, and axioms, restricting concepts and their relationships, and data in terms of instances, e.g. semantic descriptors to the data and background knowledge (Lavrač *et al.* 2011; Pinto, Santos 2009; Arantes 2011). Therefore, the domain ontology is helpful for data analysis and finding different data patterns. Moreover, the domain ontology is useful for the integration of data from different sources.

The domain ontologies can also incorporate specifics of different industries, such as telecommunications, banking, retail, insurance and etc. For example, in telecommunications industry there are clients with different relationships, using various networks, products and services.

5.2. Service oriented architecture (SOA)

Another promising solution for just-in-time, distributed and privacy-protected data mining process is the use of Service Oriented Architecture (SOA), which is presented in Cheung *et al.* (2006).

SOA provides a seamless integration of self-contained computational services that can communicate and coordinate with each other to perform goal-directed activities. The concept has originated in the past few years due to created Web-service related standards and technologies, such as WSDL, Universal Description, Discovery, and Integration (UDDI), SOAP and etc. Web-service-enabled SOAs are now widely accepted for on-demand computing as well as for developing more interoperable intra- or inter-organisational systems.

According to Cheung *et al.* (2006), adopting SOA for distributed data mining (DDM) has at least three advantages. They are as follows: 1) It allows focusing on implementation of

data mining services without having to deal with interfacing details such as the messaging protocol. 2) It allows extending and modifying DDM applications by creating or discovering new services and then reconfiguring the service-flow declaration specified at the mediators. 3) It makes on-demand DDM possible so that users can think about their business or science problems without having to worry about data mining implementations.

Authors of Guedes *et al.* (2006) propose *Anteater, service-oriented architecture for data mining*, that relies on Web services to achieve extensibility and interoperability, offers simple abstractions for users, and supports computationally intensive processing on large amounts of data through massive parallelism. For more details see Guedes *et al.* (2006).

In the future, the authors of the article plan to use SOA for the implementation of the proposed approach.

5.3. Ontology and SOA-based data mining

This chapter presents the view of the article authors on the possible future use of ontologies for data mining analysis. The future piece of work could be entitled *Ontology-based use of data mining through the cloud* to support data mining analysis.

As presented in Section 3, the approach can be extended by the following contribution: ontologies for defining and integrating business processes, and data mining processes and domain specifics. Ontologies overlay all four phases. Only the first ontology out of the set of three is mandatory. The ontologies are as follow:

- *Business process ontology*, which describes business process properties, such as inputs, outputs, participants, effects, preconditions, constraints and components (activities), have to be identified and created;
- *Domain ontology*, which describes the domain in terms of main concepts, their properties, relationships among concepts and axioms; and
- *Data mining ontology*, which describes data mining process and is used for selecting data mining algorithms and semantic searching.

The following part of this Section reveals the ways ontologies can be used in other activities of the methodology.

Assessment using ontology:

- *Business process assessment*: business process ontology is used to assess a business process and to generate potential improvement scenarios;
- *Business problems assessment*: business process ontology and domain ontology are used to identify business problems to be solved using the data mining approach;
- *Organizational structure assessment*: organisational structure is part of domain ontology. Therefore, after generating potential improvement scenarios, potential organisational structures can be generated from domain and business process ontologies;
- *Resources assessment*: domain and business process ontologies are useful here for assessment of availability and skills of business analysts, business users, and support teams;
- *Data mining readiness assessment*: Data mining ontology is used to match and combine data mining activities with business process activities.

Business process renovation using ontologies: previously generated scenarios and three ontologies are used for process change design.

Modelling using ontologies:

- *Business problem understanding*: domain definition using domain ontology;
- *Data understanding*: using domain ontology;
- *Modelling*: covers the creation of various data mining models. Therefore, data mining ontology is useful here for selecting various data mining algorithms and models;
- *Evaluation*: simulation of created models using an ontology management and creation tool, such as Protégé.

During the Operationalisation phase, ontologies are useful for documentation of processes, domain and data mining for users.

Summary, conclusions and future work

The area of integrating data mining into business processes is very complex and requires a lot of working hours. Nowadays, there are a number of approaches for integrating data mining into business processes, since it is not trivial to implement them in real situations, such as e-commerce, fraud detection and etc. The proposed approach on integration of data mining into business processes is based on the CRISP-DM model. It was extended by adding the data mining process into assessment and re-engineering of the business process.

The undertaken experiment on the use of data mining to support marketing demonstrates that companies with a data warehouse had a significant advantage. This allows eliminating unnecessary operations and optimising business process. Moreover, the presence of a data warehouse indicates a higher level of data integration and, thus, a much better basis for data mining.

The authors of the article observed several advantages of the use of data mining in the business process of direct marketing once data mining was introduced, namely: 1) business users were more independent from IT users, 2) marketing process was better controlled and more efficient.

On the basis of the results, the authors of the article propose using ontologies and SOA for data mining, this way moving the use of data mining in business processes to a higher level of maturity. It is suggested to use three ontologies – domain ontology, business process ontology and data mining ontology – to allow for a possibility of defining and modelling the characteristics of a business process, business domain and business rules within the domain, the ways data mining would be used in those business processes, and store those definitions and models in an independent location as well as ensure the use of models to every authorised user and application system. SOA is going to be used in data mining to ensure data exchange, distribution and protection during the data mining process.

References

- Aguilar-Savén, R. S. 2004. Business process modelling: review and framework, *International Journal of Production Economics* 90(2): 129–149.
- Ankolekar, A.; Burstein, M.; Hobbs, J. R.; Lassila, O.; McDermott, D.; Martin, D.; McIlraith, S. A.; Narayanan, S.; Paolucci, M.; Payne, T.; Sycara, K. 2002. DAML-S: web service description for the semantic web, in Horrocks, I.; Hendler, J. (Eds.). *Proc. of the 1st International Semantic Web Conference*, June 9–12, 2002, Sardinia, Italy, 2342: 348–363.
- Arantes, E. A. Y. F. L. 2011. Meta-DM: an ontology for the data mining domain, *Revista de Sistemas de Informacao da FSMA* 8(2011): 36–45.
- Broadbent, M.; Weill, P.; St. Clair, D. 1999. The implications of information technology infrastructure for business process redesign, *MIS Quarterly* 23(2): 159–182.
- Chaffey, D.; Wood, S. 2005. *Business Information management: improving performance using information systems*. Harlow: Prentice Hall.
- Cheung, W. K.; Zhang, X.-F.; Wong, H.-F.; Liu, J.; Luo, Z.-W.; Tong, F. C. H. 2006. Service-oriented distributed data mining, *IEEE Internet Computing* 10(4): 44–54.
- Chung, H. M.; Gray, P. 1999. Special Section: data mining, *Journal of Management Information Systems* 16(1): 11–16.
- Ciflikli, C.; Ozjirmidokuz, E. K. 2010. Implementing a data mining solution for enhancing carpet manufacturing productivity, *Knowledge-Based Systems* 23(8): 783–788.
- Davenport, T. H.; Short, J. E. 1990. The new industrial engineering: information technology and business process redesign, *Sloan Management Review* 31(4): 11–27.
- Dayal, U.; Hsu, M.; Ladin, R. 2001. Business process coordination: state of the art, trends, and open issues, in *Proc. of the 27th VLDB Conference*, 2001, Roma, Italy, 3–13.
- Falbo, R. A.; Menezes, C. S.; Rocha, A. R. C. 1998. A systematic approach for building ontologies, in Coelho, H. (Ed.). *Proc. of the 6th Ibero-American Conference on AI (IBERAMIA'98)*, 1998, Lisbon, Portugal, 1484: 349–360.
- Feelders, A.; Daniels, H.; Holsheimer, M. 2000. Methodological and practical aspects of data mining, *Information & Management* 37(5): 271–281.
- Gong, X.; Zhang, T.; Zhao, F.; Dong, L.; Yu, H. 2009. On service discovery for online data mining trails, in *the Second International Workshop on Computer Science and Engineering*, *IEEE Computer Science*, Qingdao, China, 478–482.
- Gray, P. 2005. New thinking about the enterprise, *Information Systems Management* 11(1): 91–95.
- Gruber, T. R. 1995. Toward principles for the design of ontologies for knowledge sharing, *International Journal of Human and Computer Studies* 43(4–5): 907–928.
- Guarino, N. 1998. Formal ontology and information systems, in *Proc. of FOIS'98*, 1998, Trento, Italy, 3–15.
- Guedes, D.; Meira, Jr. W.; Ferreira, R. 2006. Anteater: a service-oriented architecture for high-performance data mining, *IEEE Internet Computing* 10(4): 36–43.
- Kalibatiene, D.; Vasilecas, O. 2012. Application of the ontology axioms for the development of OCL constraints from PAL constraints, *Informatica* 23(3): 369–390.
- Kohavi, R.; Provost, F. 2001. Applications of data mining to electronic commerce, *Data Mining and Knowledge Discovery* 5(1–2): 5–10.
- Kovačič, A.; Bosilj-Vukšić, V. 2005. *Business process management*. GV založba (in Slovene).
- Kurgan, L. A.; Musilek, P. 2006. A survey of knowledge discovery and data mining process models, *The Knowledge Engineering Review* 21(1): 1–24.

- Lavrač, N.; Vavpetič, A.; Soldatova, L.; Trajkovski, I.; Kralj Novak, P. 2011. Using ontologies in semantic data mining with SEGS and g-SEGS, in Elomaa, T.; Hollmen, J.; Mannila, H. (Eds.). *Discovery Science* October 5–7, 2011, Espoo, Finland, 165–178.
- Li, L. S.; Dang, Y. Z.; Sun, J.; Ping, J. Y. 2010. A multi-dimensional ontology model for product lifecycle knowledge management, in Wang, H. (Ed.). *International Conference on E-Product E-Service and E-Entertainment*, 2010, Henan, China, IEEE, 1–4.
- Moss, L. T.; Atre, S. 2003. *Business intelligence roadmap: the complete project lifecycle for decision-support applications*. Indiana: Addison-Wesley Professional.
- OMG. 2005. *Ontology definition metamodel* [online], [cited 8 March 2013]. Available from Internet: <http://www.omg.org/spec/ODM/1.0/PDF/>.
- Panov, P.; Džeroski, S.; Soldatova, L. N. 2010. representing entities in the OntoDM data mining ontology, in Džeroski, S.; Goethals, B.; Panov, P. (Eds.). *Inductive databases and constraint-based data mining*, Part 1, Springer, 27–58.
- Pinto, F. M.; Santos, M. F. 2009. Considering application domain ontologies for data mining, *WSEAS Transactions on Information Science and Applications* 6(9): 1478–1492.
- Recker, J.; Rosemann, M.; Indulska, M.; Green, P. 2009. Business process modeling – a comparative analysis, *Journal of the Association for Information Systems* 10(4): 333–363.
- Rupnik, R.; Kukar, M.; Krisper, M. 2007. Integrating data mining and decision support though data mining based decision support system, *Journal of Computer Information Systems* 47(3): 89–104.
- Shearer, C. 2000. The CRISP-DM model: the new blueprint for data mining, *Journal for Data Warehousing* 5(4): 13–22.
- Vasilecas, O.; Kalibatiene, D.; Guizzard, G. 2009. Towards a formal method for transforming ontology axioms to application domain rules, *Information Technology and Control* 38(4): 271–282.
- Wegener, D.; Ruping, S. 2011. Integration and reuse of data mining in business processes? A pattern-based approach, *International Journal of Business Process Integration and Management* 5(3): 218–228.
- Williams, S.; Williams, N. 2007. *The profit impact of business intelligence*. San Francisco: Morgan-Kaufmann.
- Witten, I. H.; Frank, E. 2005. *Data mining: practical machine learning tools and techniques*. San Francisco: Morgan-Kaufmann.

Aleksander PIVK. Dr Aleksander Pivk is a part-time researcher at Jozef Stefan Institute, Department of Intelligent Systems, and a full-time practice leader of analytical department at SAS Adriatic Region. At SAS, he is responsible for implementation of advanced data and text mining solutions to improve business results in various industries, such as telecommunication, banking, and retail. He has actively collaborated in European research projects ALVIS (Superpeer Semantic Search Engine) and SEKT (Semantically Enabled Knowledge Technologies), and still participates at several national research projects. He is the author of more than 20 scientific papers and 1 book chapter in the field of knowledge acquisition and modelling. He is a member of SLAIS (Slovenian AI Society). His research interests include intelligent agents, information extraction/retrieval, and semantic technologies, in particular ontology learning from heterogeneous web and text resources.

Olegas VASILECAS. Professor, Dr Habil Olegas Vasilecas is a full time Professor at the Department of Information Systems and a principal researcher and the Head of Information Systems Research Laboratory of Vilnius Gediminas Technical University. He is the author and co-author of more than 250 research papers and 5 books in the field of information systems development. He delivered lectures in 7 European universities including London, Barcelona, Athens and Ljubljana. Prof. Vasilecas is constantly invited to give training sessions at universities of Germany, Holland, China, Latvia and Slovenia. He supervised 10 successfully defended doctoral theses and currently is supervising 4 additional doctoral students. He was the leader of a number of international and local projects. The latest project under his management was

entitled “Business Rules Solutions for Information Systems Development (VeTIS)”, which was carried out under the High Technology Development Programme of Lithuania. His research interests include knowledge, represented by business rule and ontology, information systems development.

Diana KALIBATIENE. Dr Diana Kalibatiene is a full time Associate Professor at the Department of Information Systems and a researcher at the Information Systems Research Laboratory of Vilnius Gediminas Technical University. She participated in the project “Business Rules Solutions for Information Systems Development (VeTIS)” of the High Technology Development Programme. She is the member of the European Committee and Lithuanian Government supported SOCRATES/ERASMUS Thematic Network projects “Doctoral Education in Computing” (ETN DEC) and “Teaching, Research, Innovation in Computing Education” (ETN TRICE). She is the author and co-author of more than 30 papers and 1 book in the field of information systems development. Research interests: development of information systems based on business rules and ontology and conceptual modelling.

Rok RUPNIK. Dr Rok Rupnik is Associate Professor of Information Systems at the University of Ljubljana, Faculty of Computer and Information Science. He received his Masters in Information Systems Engineering from the University of Ljubljana, Slovenia, in 1998, and his PhD in Mobile Applications from the University of Ljubljana, Slovenia, in 2002. His articles have appeared in proceedings of many international conferences and journals. He had an important role in various information systems development and other types of projects. He has strong practical and teaching interests in project management, information systems strategic planning, developing data mining applications and other areas. He is a senior member of PMI (Project Management Institute) at the Slovenian Chapter. In 2009, he received PMP (Project Management Professional) certificate by the PMI (Project Management Institute). His research interests include IT governance, information systems development methodologies, mobile applications, data mining, and project management.