

Chapter 17

Historical Data Analysis through Data Mining from an Outsourcing Perspective: The Three-Phases Model

Arjen Vleugel

Utrecht University, The Netherlands

Marco Spruit

Utrecht University, The Netherlands

Anton van Daal

In Summa, The Netherlands

ABSTRACT

The process of historical data analysis through data mining has proven valuable for the industrial environment. There are many models available that describe the in-house process of data mining. However, many companies either do not have in-house skills or do not wish to invest in performing in-house data mining. This paper investigates the applicability of two well-established data mining process models in an outsourcing context. The authors observe that both models cannot properly accommodate several key aspects in this context; therefore, this paper proposes the Three-phases method, which consists of data retrieval, data mining and results implementation within an organization. Each element is presented as a visual method fragment, and the model is validated through expert interviews and an extensive case study at a large Dutch staffing company. Both validation techniques substantiate the authors' claim that the Three-phases model accurately describes the data mining process from an outsourcing perspective.

INTRODUCTION: ON THE NEED FOR A NEW DATA MINING METHOD

A miner who has to work with only very current information can never detect trends and long-term patterns of behavior. Historical information is crucial to understanding the seasonality of business and the larger cycles of business to which every corporation is subject (Inmon, 1996).

The crucial element of this quote is ‘the patterns of behavior’. The main technique used to retrieve those patterns is called data mining. There are several definitions which describe data mining. We use a definition from Shaw et al. (2001): “*Data mining is the process of searching and analyzing data in order to find implicit, but potentially useful, information. It involves selecting, exploring and modeling large amounts of data to uncover previously unknown patterns, and ultimately comprehensible information, from large databases*”. In the early nineties, data mining was often described as “*a blend of statistics, AI, and data base research*” and was not considered to be a field of interest for statisticians, where some of them described it as “*a dirty word in Statistics*” (Pregibon, 1997). Nevertheless, the research area of data mining has increasingly become an important field of interest to both academics and practitioners.

Data mining can be positioned as a corollary from business intelligence (Kudyba et al., 2001; Shmueli et al., 2006). This claim is also supported by business intelligence tool providers such as Microsoft and Oracle, who both position their data mining tool as an integral part of their overall business intelligence solution (Microsoft, 2008; Oracle, 2007). Business Intelligence (BI) can be defined as the process of turning data into information and then into knowledge (Golfarelli et al., 2004). It was first introduced in the early nineties, “*to satisfy the managers’ request for efficiently and effectively analyzing the enterprise data in order to better understand the situation of their business and improving the decision process.*”

(Golfarelli et al., 2004). Data mining supports this by providing companies the unique ability to review historical data to help improve the managers’ decision-making processes (Golfarelli et al., 2004).

Most research performed in the area of data mining is aimed at adjusting existing data mining techniques to solve a specific problem, thus creating a new data mining technique (e.g., Hui et al., 1999; Rygielski et al., 2002). This research, on the other hand, has a different goal, which is the creation of a method concerning the whole process of data mining. Two methods (one emerged from the field of statistics, one emerged from business needs) have become the standards with regard to the description of the process. The first method was suggested by Fayyad et al. (1996) and involves five different stages. Its input is data, which eventually leads to knowledge (see Figure 1). The method embraces the description of the process, but does not include the use of specific tools or include a section of how to implement data mining results. Furthermore, the method does not include business needs. The business environment needs a practical model to apply data mining, one which also includes the business aspects of specific organizations.

Because of this, the Cross Industry Standard Process for Data Mining (CRISP-DM) was developed by three major industrial players in the field of data mining: DaimlerChrysler, NCR and SPSS (CRISP-DM, 2000). Unlike the Fayyad et al. (1996) approach, the trigger was the need for a model which describes the process of implementing data mining results from a business point of view. Nowadays, the CRISP-DM method has become the standard approach concerning data mining applied in the field of business (KDNuggets, 2007).

It differs from Fayyad in various ways. As said above, its key focus is on organizational implementation. This translates directly into the way the CRISP model starts with creating business understanding (see Figure 2). The KDD

23 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/historical-data-analysis-through-data/63978

Related Content

Transportation Network Optimization

A. Ogunbanwo, A. Williamson, M. Veluscek, R. Izsak, T. Kalganova and P. Broomhead (2014).

Encyclopedia of Business Analytics and Optimization (pp. 2570-2583).

www.irma-international.org/chapter/transportation-network-optimization/107437

Probabilistic Preferences Composition in the Classification of Apparel Retail Stores

Rodrigo Otávio de Araújo Ribeiro, Lidia Angulo Meza and Annibal Parracho Sant'Anna (2015). *International Journal of Business Analytics* (pp. 64-78).

www.irma-international.org/article/probabilistic-preferences-composition-in-the-classification-of-apparel-retail-stores/132802

Performance Measurement Systems and Firms' Characteristics: Empirical Evidences from Nigerian Banks

Oyewo Babajide Michael (2015). *International Journal of Business Analytics* (pp. 67-83).

www.irma-international.org/article/performance-measurement-systems-and-firms-characteristics/126834

Rethinking Knowledge Sharing Barriers: A Content Analysis of 103 Studies

Simon Cleveland and Timothy J. Ellis (2016). *Business Intelligence: Concepts, Methodologies, Tools, and Applications* (pp. 1857-1880).

www.irma-international.org/chapter/rethinking-knowledge-sharing-barriers/142706

Decision Making and Behavior: Use of Micro-Influencers as a Creative Strategy in the SMEs in México

Jovanna Nathalie Cervantes Guzmán (2020). *Handbook of Research on IT Applications for Strategic Competitive Advantage and Decision Making* (pp. 390-404).

www.irma-international.org/chapter/decision-making-and-behavior/262487