Chapter 1
Data Mining and Business Intelligence:
A Comparative, Historical Perspective

Ana Azevedo
Algoritmi R&D Center/University of Minho, Portugal & Polytechnic Institute of Porto/ISCAP, Portugal

ABSTRACT

Data Mining (DM) is being applied with success in Business Intelligence (BI) environments, and several examples of applications can be found. BI and DM have different roots and, as a consequence, have significantly different characteristics. DM came up from scientific environments; thus, it is not business oriented. DM tools still demand heavy work in order to obtain the intended results. On the contrary, BI is rooted in industry and business. As a result, BI tools are user-friendly. This chapter reflects on this difference from a historical perspective. Starting with a separated historical perspective of each one, BI and DM, the author then discusses how they converged into the current situation, when DM is used, and integrated, in BI environments with success.

INTRODUCTION

BI is one area of the Decision Support Systems (DSS) discipline and refers to information systems aimed at integrating structured and unstructured data in order to convert it into useful information and knowledge, upon which business managers can make more informed and consequently better decisions. The term Business Intelligence (BI) was made popular by Gartner in 1989 (Power, 2007) (Zeller, 2007), but the first reference was made by Luhn in 1958 (Lunh, 1958), not necessarly with the same meaning. Being rooted in the DSS discipline, BI has suffered a considerable evolution over the last years and is, nowadays, an area of DSS that attracts a great deal of interest from both the industry and researchers (Azevedo & Santos, 2012).

The term Knowledge Discovery in Databases (KDD) was coined in 1989 (Wixon & Watson, 2010) to refer to the broad process of finding knowledge in data, and to emphasize the "high-level" application of particular data mining (DM) methods (Fayyad, Piatetski-Shapiro, & Smyth,

DOI: 10.4018/978-1-4666-6477-7.ch001

1996). The DM phase concerns, mainly, to the means by which patterns are extracted and enumerated from data. In recent years, DM has been applied with success to several diversified fields, such as bioinformatics, ecology and sustainability, finance, industry, marketing, scientific research, telecommunications, and several other applications, including BI.

DM is being applied with success in BI and several examples of applications can be found (Linoff, 2008) (Vercellis, 2009) (Hu & Cercone, 2004) (Cheung & Li, 2012) (Phan & Vogel, 2010). BI and DM have different roots and, as a consequence, have significantly different characteristics. DM came up from scientific environments, thus, in its roots, it is not business oriented (Kriegel, Borgwardt, Kröger, Pryakhin, Schubert, & Zimek, 2007) (Piatetsky-Shapiro, 2007). DM tools still demand heavy work in order to obtain the intended results, hence needing the knowledge of DM specialists to explore its full potential value (Azevedo A., 2012). The main focus for DM researchers still is the improvement of algorithms and/or finding new algorithms that behaves better than others in some particular application, as can be recognized by a search in the main conferences and journal in the area of DM. On the contrary, BI is rooted in industry and business (Yermish, Miori, Yi, Malhotra, & Klimberg, 2010), thus it is business oriented. As a result, BI tools are user-friendly and can easily be accessed and manipulated by business users. The main focus for BI researchers is how to better use BI in organizations in order to improve decision making (Wixon & Watson, 2010).

Nevertheless, during the last years DM and BI tend to converge. The gap existing between BI and DM is being filled (Azevedo & Santos, 2012) (Wang & Wang, 2008) (Hang & Fong, 2009). Researchers efforts are shifting towards the integration of DM in BI systems, and focusing on how DM can be used to improve decision making, and creating the possibility of DM tools being accessed and manipulated by business users at the same level as the other BI tools, thus making DM relevant to business. The book in which this chapter is included, is an example of the growing interest of researchers for this topic of research.

This papers presents DM and BI from a comparative historical perspective. As far as our knowledge, there is no similar approach in the literature. The main contribution of this chapter is to provide a comparison between the two different and yet convergent areas of DM and BI, from an historical perspective. We consider that it is not possible to understand the present without knowing the past, thus this is an important issue to analyse.

The structure for the rest of the chapter is the following. Firstly Business Intelligence is introduced. Next, Data Mining is also presented. Finally, Data Mining and Business Inteligence are compared. The chapter ends with sections Future Research Directions and Conclusion.

BUSINESS INTELLIGENCE

The first reference to Business Intelligence was made by Luhn in 1958 (Lunh, 1958). In that article, Luhn describes a system whose main goal is to ultimately deliver the rigth documents, to the rigth "persons" (which are named as "points of action"), at the right time. Since then, some articles can be found dealing with the topic but, like in the case of Lunh's article, not necessarly with the same meaning that is used nowadays. It is widely accepted that the current notion of Business Intelligence (BI) was coined by Howard Dresner, from Gartner, in 1989 (Power, 2007) (Zeller, 2007) (Wixon & Watson, 2010). From then on, BI gained popularity, firstly in business environments and latter in academia.

Figure 1 presents the evolution of the scientific publications with the topic Business Intelligence that can be found in "Web of Science - Cross Search (ISI Ver.3.0)"[1] between 1958 and 2013 (none can be found before that). As can be seen through the analysis of the graph, the number of scientific publications was insignificant until