Smart Data Analysis for Financial Services

Mathias Bauer[1]

Abstract. This talk addresses opportunities for the application of intelligent data analysis techniques at various stages of the value added chain for financial services. After introducing some basic notions and explaining the fundamental steps of data mining, we will have a closer look at various recent and ongoing projects and discuss issues of practical relevance such as data quality and expert knowledge. The talk concludes with some remarks on the potential impact of new developments, e. g. in the context of Big Data.

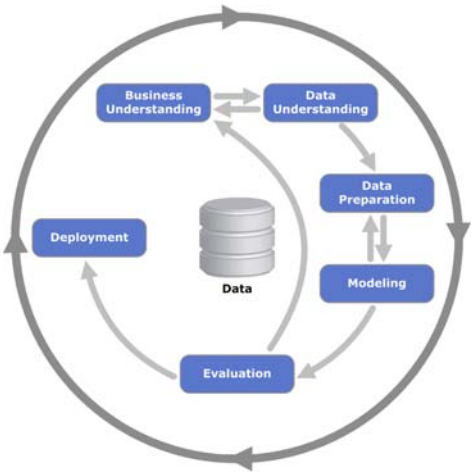

Figure 1: The CRISP-DM process model for data mining.

1 DATA MINING

Data mining – this notion will be used as a synonym for all kinds of smart data analysis – is a complex process that aims at turning raw data into actionable knowledge (see Figure 1 which depicts a standard process model). We will introduce the basic notions, discuss the various steps and in particular have a closer look at the choices to be made and a few pitfalls to be avoided.

In particular, we will address the crucial aspects of how to choose an appropriate modeling approach and how to assess the quality of a solution found by a data analyst.

We show that in many cases it is not a good idea to simply apply the data analyst's favorite modeling technique. Instead we describe the various dimensions of such a choice and encourage the end users of a data analysis to clearly state their requirements.

2 SAMPLE APPLICATIONS

Data Analysis can (and should) play a central role at various stages of the value added chain in the financial industry. In the following we will have a closer look at some relevant activities in this context.

2.1 Appraisal of real economic goods

Scoring and rating processes are at the heart of financial industry. Here we will demonstrate an approach to appraise vessels as typical representatives of real economic goods which form an important class of investments.

2.2 Fraud detection

In B2B scenarios a company's annual accounts form the basis for their credit rating and all further negotiations. Usually, the numbers reported are accepted as a correct representation of last year's business activities. But what if they are manipulated? We describe an approach that identifies abnormalities in annual accounts, thus facilitating the detection of intentional manipulations.

2.3 Identifying interesting customers

There are numerous aspects that can make a customer particularly interesting to a company – his/her interest in certain products, credit-worthiness and default risk, churn rate etc. We describe an integrated approach to identify these individuals that reduces the marketing effort required while simultaneously improving the company's insight into their customer base and the quality of customer contact.

In particular, we will see how the modeling technique applied affects the usefulness of the analytical findings.

2.4 Stock selection

From an abstract point of view, selecting a relevant set of stocks is similar to the previous task as it mainly involves segmentation and classification efforts. However, we will see that data preprocessing in this case is significantly more complex and requires some advanced expert knowledge.

3 Perspective

Big data is more than a buzzword – even if it's not the silver bullet for all problems ahead. We will discuss various techniques and attempts to commercially make use of huge, largely unstructured data sets and briefly discuss potential future applications.

[1] mineway GmbH, Saarbrücken, Germany, email: mbauer@mineway.de