Chapter XI
Data Mining in Aviation Safety
Data Analysis

Reima Suomi
Turku School of Economics, Finland

Olli Sjöblom
Turku School of Economics, Finland

ABSTRACT
This chapter introduces aviation safety data analysis as an important application area for data mining. In the beginning of the chapter, the reader is introduced to the basic concepts of data mining. After that, the field of aviation safety management is discussed, and in that connection data mining is identified as a key technology to study through flight incidents reports. Afterwards the test runs for four data mining products, for possible use in the Finnish civil aviation authority, are described in detail. However, before the testing of tools the preparation of the test data for the tools is described in detail. The chapter ends with conclusions that tell that even sophisticated data mining tools are just tools: they do not provide any automatic tools, but skilled users can use them for searching clues in the data.

INTRODUCTION
Since the very beginning of aviation history there has been some kind of investigation after accidents as well as other, less fatal events and incidents. However, the incredible rapid development in huge volume of air travel, mainly because of jet airliners, created the need of systematic research for aviation safety. Different interest groups in aviation, like airlines and aviation authorities, began to collect data about air traffic. The consequences of this development have been a significant reduction in incidents and occurrences as well as aviation accidents, especially during the sixties. It has now been widely recognised among aviation safety experts that the global rate of accidents is stabilising. The situation can now be regarded as satisfactory, but because of the growth in air traffic, the absolute number of fatal accidents per year might increase, if the
flight safety is not improved. The collection of data and reporting systems have reached their top level. The focal point in this context is analysis. The source of aviation safety data are the incident reports that include both structured and narrative fields. The structured data can be analysed easily using queries from databases and running their results through graphic tools. Among narrative data the situation is totally different. There were no tools to analyse textual data until data mining tools have been developed. Still their use, at least among aviation, is at a moderate level. Flight safety can be significantly improved through the development of data analysis. Narrative text mining is demanding also because of the multiplicity of languages spoken in the world. Especially languages with small user groups have to wait for efficient tools longer than big languages such as English.

AVIATION SAFETY DATA MANAGEMENT

It is widely recognised that air transport is among the safest modes of transport. The global rate of accidents is stabilising. If nothing is done to improve it, the growth in air traffic will lead to an increase in the absolute number of fatal accidents per year. This kind of development would be, clearly, totally unacceptable and therefore new ways of improving air safety need to be explored (European Commission, 2000). Fundamental to every Safety Management System is the principle of collecting and analysing operational data (GAIN Working Group B, 2004). Aviation industry has placed significant investments into collecting aviation safety information from multiple sources. As a result of this process large databases exist, and at the same time, enormous challenges have appeared in analysing the information (Megaputer Intelligence, 2004c).

To improve air safety, the significant allocation of the newest research resources and data processing techniques are unavoidable. Data has been collected, more or less, in different forms during the whole history of aviation. The data collected can roughly be divided into two types: structured data and unstructured, narrative data. The exactness of the incident and accident reports using structured data depends on the number of details and alternatives of the system classification. So, with structured data the explanation of the case usually tells the truth till a certain rate, but completed with narrative data it can reach the level of 100 per cent, at least theoretically. During the execution of various operational and support processes among operations of aircraft fleets large volumes of data are collected. Analysis and review of data is typically complicated requiring human involvement significantly. Usually the data accumulates faster than it can be processed (Wang, Huang, Cao, Shi, & Shu, 2007). The need for automated means to process the data is also increasing rapidly, because the amount of generated and stored unstructured data is increasing fast (Delen & Crossland, 2008).

Till the end of the 1990s, there have not been, at least in a wide range, other tools than report generators using queries with data stored into a relational database. Narrative data were practically not been processed with computing. Finding trends from structured data has never been a problem, but from narrative data it has required huge amounts of work made by man. Successful analysis work for aviation safety has been done during decades: In Finland, since beginning of the 1960s till today, through systematic safety work the accident rate has been reduced to a very small percentage from what it was. This trend seems to be global, at least among developed aviation countries: almost every statistics about this theme shows the same development. If one looks at different commercial airlines’ statistics or even military aviation operation/accident rates, the view is very similar.

However, airlines and other operators use at large extent “normal” analytical tools, Excel and