Chapter 4
The Use of Data Mining for Assessing Performance of Administrative Services

Zdravko Pečar
University of Ljubljana, Slovenia

Ivan Bratko
University of Ljubljana, Slovenia

ABSTRACT

The aim of this research was to study the performance of 58 Slovenian administrative districts (state government offices at local level), to identify the factors that affect the performance, and how these effects interact. The main idea was to analyze the available statistical data relevant to the performance of the administrative districts with machine learning tools for data mining, and to extract from available data clear relations between various parameters of administrative districts and their performance. The authors introduced the concept of basic unit of administrative service, which enables the measurement of an administrative district's performance. The main data mining tool used in this study was the method of regression tree induction. This method can handle numeric and discrete data, and has the benefit of providing clear insight into the relations between the parameters in the system, thereby facilitating the interpretation of the results of data mining. The authors investigated various relations between the parameters in their domain, for example, how the performance of an administrative district depends on the trends in the number of applications, employees' level of professional qualification, etc. In the chapter, they report on a variety of (occasionally surprising) findings extracted from the data, and discuss how these findings can be used to improve decisions in managing administrative districts.

INTRODUCTION

The aim of this research was to assess the performance of 58 Slovenian administrative districts (state government offices at local level), which provide administrative services for eight state ministries. These administrative tasks are commonly organised in four departments of each administrative district. The research covered only one of them – the departments for environment and spatial planning, whose task is to issue various permits (planning permits, building permits and others) upon applications,

DOI: 10.4018/978-1-60566-906-9.ch004

under the laws and supervision of Ministry for environment and spatial planning. The following three hypotheses were set at the beginning of the research:

- The administrative districts have very different productivity.
- The level of employee education has the major influence on productivity.
- The increased number of applications results in longer times for processing.

The analysis showed several findings of interest. Among them it was found that the organizational productivity among administrative districts varied enormously, up to the ratio of 10: 1. Also, the number of new applications plays a major role in predicting the future trends in productivity. Level of education of employees, and to a lesser degree their age and gender, also influence the productivity.

In our experience, machine learning methods proved to be a very efficient tool for quick, automatic and holistic analysis of large sets of different data. It was especially effective at exposing most characteristic patterns of behavior. According to our experience in this study, the analyses with classical statistical methods is much more rigid and more costly in that it requires more time for recognizing various hidden patterns of behavior such as ones generated by machine learning methods. In this sense, machine learning is particularly good at data exploration stage when hypotheses are formulated. Of course, when we get to the question of proving statistical significance of hypotheses, then we face essentially the same problems as in classical statistics.

BACKGROUND

Among practitioners, there is unfavourable and prevailing general opinion (shared also by professionals and politicians) that the work performance at administrative services cannot be measured. The authors of this paper and an emerging group of innovative public managers which participated with data gathering and discussions during the present research have organized "committee for quality". Members of this committee do not share this opinion and believe that the performance of these services can gradually be more systematically measured and managed, very much like all those in the private sector (Asbjorn, 1995). This committee was a facilitator of new ideas in this respect.

The main idea was to analyse the available statistical data (Annual reports of Administrative statistics 1996–1999) relevant to the performance of the administrative districts with tools of machine learning, to obtain clear relations between various parameters of administrative districts and the performance. These machine learning tools include those that are usually employed in data mining. Additional objective was to set the requirements for better performance measurement system and suggest the need for new ways of decision making by public managers in the fields of strategic planning and performance management (including performance based budgeting and performance based pay).

The main data analysis tool used in this study was the method of regression trees, one of rather common machine learning techniques (Witten and Frank 2005). We will describe this technique in more detail in Section 3.

Developing and Organizing Data for Analysis

All 58 administrative districts, which employ more than 3000 administrative workers, provide administration services at local level for eight state ministries. The performance of these districts is not properly measured, monitored and thus not well managed.

Data sets that were used for the analysis were gathered for the period of four years: 1996–99,