

Chapter 1

Introduction

The last ten to fifteen years have seen a decided focus on business intelligence, the ability of an organization to understand and use information relevant to its gainful operations. The goal of business intelligence is to enable business leaders to make better decisions that will translate into increased profitability. While query and report, statistical analysis, and multidimensional tools have become the staples of Business Intelligence systems, increasing numbers of enterprises are turning to data mining to productively uncover information hidden in their data.

Data mining helps to identify and extract high-value business intelligence from enterprise data assets, including high-volume transaction data. It makes it possible to identify patterns, detect deviations, explain past performance and predict future outcomes. It is, therefore, one of the most effective ways of enabling an organization to gain a competitive advantage, by improving customer information, improving sales and customer retention, and improving decision making based on that customer information.

The purpose of this project is to use various predictive and descriptive data mining techniques (statistical methods, neural nets, decision trees) to identify and/or rank customers who are likely to purchase a certain hardware or software product, and to compare and contrast the performance of these mining techniques. Also of interest would be an analysis of which factors are differentiators of product preference. These factors could then be used to determine customer segments using clustering algorithms. Customer factors of interest would include geographical data, the business or principal industrial activity of the customer, the

size of the company, its number of employees, the number of its remote sites, the estimated revenue amount, the number of mainframes, midrange systems or workstations installed at the customer site, whether a network is installed, which CPU-platform does the customer have, past purchasing history of the customer. The resulting market segmentation would allow IBM to develop distinct market strategies for each segment thereby improving IBM's competitive position in the marketplace.

The topics addressed in chapter 2 are introductory considerations on data mining, while the subsequent chapters deal with the knowledge discovery process. Chapter 3 focuses on the business problem and the mining objectives. Also the analysis goes through all the steps that precede the data mining process itself from data collection to data exploration and preparation. Chapter 4 covers the process of data mining using the different statistical, neural nets and decision tree models. The last chapter of this work compares those models and evaluates their predictive performance.