PREFACE

The new generation of computing techniques collectively called data mining methods are now applied to stock market analysis, predictions, and other financial applications. In this book we discuss the relative merits of these methods for financial modeling and present a comprehensive survey of current capabilities of these methods in financial analysis.

The focus is on the specific and highly topical issue of adaptive linear and non-linear "mining" of financial data. Topics are progressively developed. First, we examine the distinction between the use of such methods as ARIMA, neural networks, decision trees, Markov chains, hybrid know-ledge-based neural networks, and hybrid relational methods. Later, we focus on examining financial time series, and, finally, modeling and forecasting these financial time series using data mining methods.

Our main purpose is to provide much needed guidance for applying new predictive and decision-enhancing hybrid methods to financial tasks such as capital-market investments, trading, banking services, and many others.

The very complex and challenging problem of forecasting financial time series requires specific methods of data mining. We discuss these requirements and show the relations between problem requirements and the capabilities of different methods. Relational data mining as a hybrid learning method combines the strength of inductive logic programming (ILP) and probabilistic inference to meet this challenge. A special feature of the book is the large number of worked examples illustrating the theoretical concepts discussed.

The book begins with problem definitions, modern methodologies of general data mining and financial knowledge discovery, relations between data mining and database management, current practice, and intellectual challenges in data mining.

Chapter 2 is devoted to numerical data mining learning models and their financial applications. We consider ARIMA models, Markov chains, instance-based learning, neural networks, methods of learning from experts ("expert" mining"), and new methods for testing the results of data mining.

Chapter 3 presents rule-based and hybrid data mining methods such as learning propositional rules (decision trees and DNF), extracting rules from learned neural networks, learning probabilistic rules, and knowledge-based stochastic modeling (Markov chains and hidden Markov models) in finance.

Chapter 4 describes a new area of data mining and financial applications -- relational data mining (RDM) methods. From our viewpoint, this approach will play a key role in future advances in data mining methodology and practice. Topics covered

in this chapter include the relational data mining paradigm and current challenges, theory, and algorithms (FOIL, FOCL and MMDR).

Numerical relational data mining methods are especially important for financial analysis where data commonly are numerical financial time series. This subject is developed in chapters 4, 5 and 6 using complex data types and representative measurement theory. The RDM paradigm is based on highly expressive first-order logic language and inductive logic programming. Chapters 5 and 6 cover knowledge representation and financial applications of RDM. Chapter 6 also discusses key performance issues of the selected methods in forecasting financial time series. Chapter 7 presents fuzzy logic methods combined with probabilistic methods, comparison of fuzzy logic and probabilistic methods, and their financial applications. Well-known and commonly used data mining methods in finance are attribute-based learning methods such as neural networks, the nearest neighbours method, and decision trees. These are relatively simple, efficient, and can handle noisy data. However, these methods have two serious drawbacks: a limited ability to represent background knowledge and the lack of complex relations. The purpose of relational data mining is to overcome these limitations. On the other hand, as Bratko and Muggleton noted [1995], current relational methods (ILP methods) are relatively inefficient and have rather limited facilities for handling numerical data. Biology, pharmacology, and medicine have already benefited significantly from relational data mining. We believe that now is the time for applying these methods to financial analyses. This book is addressed to researchers, consultants, and students interested in the application of mathematics to investment, economics, and management. We also maintain a related website http://www.cwu.edu/~borisk/finance.