Foreword

There has been much interest developed in the data mining field both in the academia and the industry over the past 10-15 years. The number of researchers and practitioners working in the field and the number of scientific papers published in various data mining outlets increased drastically over this period. Major commercial vendors incorporated various data mining tools into their products, and numerous applications in many areas, including life sciences, finance, CRM, and Web-based applications, have been developed and successfully deployed.

Moreover, this interest is no longer limited to the researchers working in the traditional fields of statistics, machine learning and databases, but has recently expanded to other fields, including operations research/management science (OR/MS) and mathematics, as evidenced from various data mining tracks organized at different INFORMS meetings, special issues of OR/MS journals and the recent conference on Mathematical Foundations of Learning Theory organized by mathematicians.

As the *Encyclopedia of Data Warehousing and Mining* amply demonstrates, all these diverse interests from different groups of researchers and practitioners helped to shape data mining as a broad and multi-faceted discipline spanning a large class of problems in such diverse areas as life sciences, marketing (including CRM and e-commerce), finance, telecommunications, astronomy, and many other fields (the so called "data mining and X" phenomenon, where X constitutes a broad range of fields where data mining is used for analyzing the data). This also resulted in a process of cross-fertilization of ideas generated by these diverse groups of researchers interacting across the traditional boundaries of their disciplines.

Despite all this progress, data mining still faces several challenges that make the field ripe with future research opportunities. First, despite the cross-fertilization of ideas spanning various disciplines, the convergence among different disciplines proceeds gradually, and more work is required to arrive at a unified view of data mining widely accepted by different groups of researchers. Second, despite a considerable progress, still more work is required on the theoretical foundations of data mining, as was recently stated by the participants of the Dagstuhl workshop "Data Mining: The Next Generation" organized by R. Agrawal, J.-C. Freytag and R. Ramakrishnan and also expressed by various other data mining researchers. Third, the data mining community must address the privacy and security problems for data mining to be accepted by the privacy advocates and the Congress. Fourth, as the field advances, so is the scope of data mining applications. The challenge to the field is to develop more advanced data mining methods that would work in these increasingly demanding applications. Fifth, despite a considerable progress in developing more user-friendly data mining tools, more work is required in this area with the goal of making these tools accessible to a large audience of "naïve" data mining users. In particular, one of the challenges is to devise methods that would smoothly embed data mining tools into corresponding applications on the front-end and would integrate these tools with databases on the back-end. Achieving such capabilities is very important since this would allow data mining to "cross the chasm" (using Geoffrey Moore's terminology) and become a mainstream technology utilized by millions of users. Finally, more work is required on actionability and on the development of better methods for discovering actionable patterns in the data. Currently, discovering actionable patterns in data constitutes a laborious and challenging process. It is important to streamline and simplify this process and make it more efficient.

Given significant and rapid advancements in data mining and data warehousing, it is important to take periodic "snapshots" of the field every few years. The data mining community addressed this issue by producing publications covering the "state of the art" of the field every few years starting with the first volume "Advances in Knowledge Discovery and Data Mining" (edited by U. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy) published by AAAI/MIT Press in 1996. This encyclopedia provides the latest "snapshot" of the field and surveys a broad array of topics ranging from the basic theories to the recent advancements in the field and covers a diverse range of problems

from the analysis of microarray data to the analysis of multimedia and Web data. It also identifies future directions and trends in data mining and data warehousing. Therefore, this volume should become an excellent guide to researchers and practitioners.

Alexander Tuzhilin
New York University, USA