

PREFACE

The explosion of information technology, which continues to expand data-driven markets and business, has made data mining an even more relevant topic of study. Books on data mining tend to be either broad and introductory or focus on some very specific technical aspect of the field.

Lecture Notes in Data Mining is a series of seventeen “written lectures” that explores in depth the core of data mining (classification, clustering and association rules) by offering overviews that include both analysis and insight. Written by graduate students from the University of Tennessee as part of an Advancements in Data Mining course, *Lecture Notes* is an ideal companion to either an introductory data mining textbook or a technical data mining book.

The Advancements in Data Mining Course was developed not only to allow graduate students an opportunity to investigate certain topics in data mining, but it was also planned as a vehicle for graduate students to sharpen their writing and presentation skills. After selecting one of the assigned topics, researching, and writing his or her paper, each student was required to make a lengthy presentation to the class. Both the presentation and the paper was extensively peer reviewed and each student was required to make revisions before submitting a final paper. The discussion following the presentations was spirited but respectful, as each student was aware that at some point in the semester — he or she would be facing their peers as well.

The initial chapters of *Lecture Notes* lay a framework of data mining techniques by explaining some of the basics such as applications of Bayes Theorem, similarity measures and decision trees. Before focusing on the pillars of classification, clustering and association rules, the book also provides information about alternative candidates such as point estimation and genetic algorithms.

The book’s discussion of classification includes an introduction to decision tree algorithms, rule-based algorithms (a popular alternative to decision trees) and distance-based algorithms. Five of the lecture-chapters are

devoted to the concept of clustering or unsupervised classification. The functionality of clustering's hierarchical and partitional algorithms is also covered as well as the efficient and scalable cluster algorithms used in large databases. The concept of association rules in terms of basic algorithms, parallel and distributive algorithms and advanced measures that help determine the value of association rules are included in this book too. The final chapter covers spatial mining algorithms.

In addition to the technical descriptions of the algorithms and methods, the students were encouraged to provide simple examples and to provide commentary based on references that could assess both strengths and weaknesses of the respective techniques. This type of analysis gives the student an understanding of what it is to go behind just reporting the particulars of an algorithm and the value of providing additional insights. With this in mind, each student was required to include a complete bibliography of journal articles, books, conference papers, technical reports, electronic journals and web sites for each report. These references have been compiled and appear at the end of the book.

Lecture Notes can be used as a supplement that would accompany a data mining textbook for graduate level computer students. With its exploration of data mining's core competencies, *Lecture Notes* also appeals to the database professional who is looking for a concise background and evaluation of the various tools that are now available to data miners.

One of the more common mistakes data miners make is becoming too focused on one area of technique. It's not unusual to become an expert in one area and forget that there are other approaches that merit consideration based on the problem at hand. *Lecture Notes* offers a good sampling of what other techniques are available presented with a fresh set of eyes. Hopefully, these hardworking students will open your senses to the burgeoning world of data mining.

All proceeds from the sales of this book will be donated to the student travel fund in the Department of Computer Science, University of Tennessee, Knoxville.

Michael W. Berry, Murray Browne
Knoxville, TN
December 2005