

PART I

**KNOWLEDGE DISCOVERY AND
DATA MINING IN SOCIAL
SCIENCE RESEARCH**

Chapter 1

INTRODUCTION

A DVANCES IN TECHNOLOGY—the internet, mobile devices, computers, digital sensors, and recording equipment—have led to exponential growth in the amount and complexity of data available for analysis. It has become difficult or even impossible to capture, manage, process, and analyze these data in a reasonable amount of time. We are at the threshold of an era in which digital data play an increasingly important role in the research process. In the traditional approach, hypotheses derived from theories are the driving forces behind model building. However, with the rise of big data and the enormous wealth of information and knowledge buried in this data mine, using data mining technologies to discover interesting, meaningful, and robust patterns has become increasingly important. This alternative method of research affects all fields, including the social sciences. The availability of huge amounts of data provides unprecedented opportunities for new discoveries, as well as challenges.

Today we are confronted with a data tsunami. We are accumulating data at an unprecedented scale in many areas of industry,

government, and civil society. Analysis and knowledge based on big data now drive nearly every aspect of society, including retail, financial services, insurance, wireless mobile services, business management, urban planning, science and technology, social sciences, and humanities. Google Books has so far digitalized 4 percent of all the books ever printed in the world, and the process is ongoing. The Google Books corpus contains more than 500 billion words in English, French, Spanish, German, Chinese, Russian, and Hebrew that would take a person eighty years to read continuously at a pace of 200 words per minute. This entire corpus is available for downloading (<http://storage.googleapis.com/books/ngrams/books/datasetsv2.html>), and Google also hosts another site to graph word usage over time, from 1800 to 2008 (<https://books.google.com/ngrams>). The Internet Archive, a digital library of internet sites and other cultural artifacts in digital form, provides free access to 279 billion web pages, 11 million books and texts, 4 million audio recordings, 3 million videos, 1 million images, and 100,000 software programs (<https://archive.org/about/>). Facebook generates 4 new petabytes of data and runs 600,000 queries and one million map-reduce jobs per day. Facebook's data warehouse Hive stores 300 petabytes of data in 800,000 tables, as reported in 2014 (<https://research.fb.com/facebook-s-top-open-data-problems/>). The GDELT database monitors global cyberspace in real time and analyzes and extracts news events from portals, print media, TV broadcasts, online media, and online forums in all countries of the world and extracts key information such as people, places, organizations, and event types related to news events. The GDELT Event Database records over 300 categories of physical activities around the world, from riots and protests to peaceful appeals and diplomatic exchanges, georeferenced to the city or mountaintop, across the entire world from January 1, 1979 and updated every 15 minutes. Since February 2015, GDELT has brought together 940 million messages from global cyberspace in a volume of 9.4TB (<https://www.gdeltproject.org/>). A report by McKinsey (Manyika et al. 2011) estimated that corporations, institutions, and users stored more than 13 exabytes of new data, which is over 50,000 times larger than the amount of data in the Library of Congress. The value of global personal location data is estimated to be \$700 billion, and these data can reduce costs as much as 50 percent in product development and assembly.

Both industry and academic demands for data analytical skills have soared rapidly and continue to do so. IBM projects that by 2020 the number of jobs requiring data analytical skills in the United States will increase by 15 percent, to more than 2.7 million, and job openings requiring advanced data science analytical skills will reach more than 60,000 (Miller and Hughes 2017). Global firms are focusing on data-intensive sectors such as finance, insurance, and medicine. The topic of big data has been covered in popular news media such as the *Economist* (2017), the *New York Times* (Lohr 2012), and National Public Radio (Harris 2016), and data mining has also been featured in *Forbes* (2015; Brown 2018), the *Atlantic* (Furnis 2012), and *Time* (Stein 2011), to name a few.

The growth of big data has also revolutionized scientific research. Computational social sciences emerged as a new methodology, and it is growing in popularity as a

result of dramatic increases in available data on human and organizational behaviors (Lazer et al. 2009). Astronomy has also been revolutionized by using a huge database of space pictures, the Sloan Digital Sky Survey, to identify interesting objects and phenomena (<https://www.sdss.org/>). Bioinformatics has emerged from biological science to focus on databases of genome sequencing, allowing millions or billions of DNA strands to be sequenced rapidly in parallel.

In the field of artificial intelligence (AI), scientists have developed AlphaGo, which was trained to model expert players from recorded historical games of a database of 30 million game moves and was later trained to learn new strategies for itself (<https://deepmind.com/research/alphago/>). AlphaGo has defeated Go world champions many times and is regarded as the strongest Go player in the game's history. This is a major advancement over the old AI technology. When IBM's DeepMind beat chess champion Gary Kasparov in the late 1990s, it used brute-force AI that searched chess moves in a space that was just a small fraction of the search space for Go.

The Google Books corpus has made it possible to expand quantitative analysis into a wider array of topics in the social sciences and the humanities (Michel et al. 2011). By analyzing this corpus, social scientists and humanists have been able to provide insights into cultural trends that include English-language lexicography, the evolution of grammar, collective memory, adoption of technology, pursuits of fame, censorship, and historical epidemiology.

In response to this fast-growing demand, universities and colleges have developed data science or data studies majors. These fields have grown from the confluence of statistics, machine learning, AI, and computer science. They are products of a structural transformation in the nature of research in disciplines that include communication, psychology, sociology, political science, economics, business and commerce, environmental science, linguistics, and the humanities. Data mining projects not only require that users possess in-depth knowledge about data processing, database technology, and statistical and computational algorithms; they also require domain-specific knowledge (from experts such as psychologists, economists, sociologists, political scientists, and linguists) to combine with available data mining tools to discover valid and meaningful knowledge. On many university campuses, social sciences programs have joined forces to consolidate course offerings across disciplines to teach introductory, intermediate, and advanced courses on data description, visualization, mining, and modeling to students in the social sciences and humanities.

This chapter examines the major concepts of big data, knowledge discovery in databases, data mining, and computational social science. It analyzes the characteristics of these terms, their central features, components, and research methods.

WHAT IS BIG DATA?

The concept of big data was conceived in 2001 when the META analyst D. Laney (2001) proposed the famous “3V's Model” to cope with the management of increasingly

large amounts of data. Laney described the data as of large *volume*, growing at a high *velocity*, and having great *variety*. The concept of big data became popular in 2008 when *Nature* featured a special issue on the utility, approaches, and challenges of big data analysis. Big data has since become a widely discussed new topic in all areas of scientific research. *Science* featured a special forum on big data in 2011, further highlighting the enormous potential and great challenge of big data research. In the same year, McKinsey's report "Big Data: The Next Frontier for Innovation, Competition, and Productivity" (2011) announced that the tsunami of data will bring enormous productivity and profits, adding enthusiasm to this already exciting development. Mayer-Schönberger and Cukier (2012) focused on the dramatic impacts that big data will have on the economy, science, and society and the revolutionary changes it will bring about in society at large.

A variety of definitions of big data all agree on one central feature of this concept: data enormity and complexity. Some treat data that are too large for traditional database technologies to store, access, manage, and analyze (Manyika et al. 2011). Others define big data based on its characteristic four big V's: (1) big *volume*, measured at terabytes or petabytes; (2) big *velocity*, which grows rapidly and continuously; (3) big *variety*, which includes structured numerical data and unstructured data such as text, pictures, video, and sound; and (4) big *value*, which can be translated into enormous economic profits, academic knowledge, and policy insights. Analysis of big data uses computational algorithms, cloud storage, and AI to instantaneously and continuously mine and analyze data (Dumbill 2013).

There are just as many scholars who think big data is a multifaceted and complex concept that cannot be viewed simply from a data or technology perspective (Mauro, Greco, and Grimaldi 2016). A word cloud analysis from the literature shows that big data can be viewed from at least four different angles. First, big data contains information. The foundation of big data is the production and utilization of information from text, online records, GPS locations, online forums, and so on. This enormous amount of information is digitized, compiled, and stored on computers (Seife 2015). Second, big data includes technology. The enormous size and complexity of the data pose difficulties for computer storage, data processing, and data mining technologies. The technology component of big data includes distributed data storage, cloud computing, data mining, and artificial intelligence. Third, big data encompasses methods. Big data requires a series of processing and analytical methods that are beyond the traditional statistical approaches, such as association, classification, cluster analysis, natural language processing, neural networks, network analysis, pattern recognition, predictive modeling, spatial analysis, statistics, supervised and unsupervised learning, and simulation (Manyika et al. 2011). And fourth, big data has impacts. Big data has affected many dimensions of our society. It has revolutionized how we conduct business, research, design, and production. It has brought and will continue to bring changes in laws, guidelines, and policies on the utility and management of personal information.

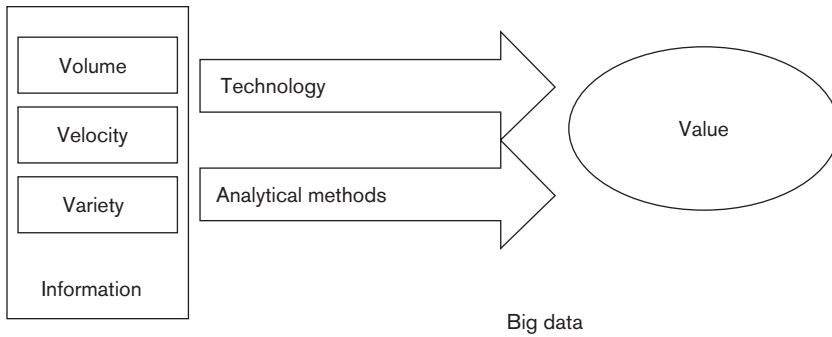


FIGURE 1.1
What Is Big Data?

To summarize, the essence of big data is big volume, high velocity, and big variety of information. As shown in figure 1.1, it also comprises *technology* and *analytical methods*, to transform the information into insights that are worth economic *value*, thus having an impact on society.

WHAT IS KNOWLEDGE DISCOVERY IN DATABASE?

Knowledge discovery in database (KDD) is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data (Fayyad, Piatetsky-Shapiro, and Smyth 1996, 84). It consists of nine steps that begin with the development and understanding of the application domain and ends with actions on the knowledge discovered, as illustrated in figure 1.2.

KDD is a nine-step process that involves understanding domain knowledge, selecting a data set, data processing, data reduction, choice of data mining method, data mining, interpreting patterns, and consolidating discovered knowledge. This process is *not* a one-way flow. Rather, during each step, researchers can backtrack to any of the previous steps and start again. For example, while considering a variety of data mining methods, researchers may go back to the literature and study existing work on the topic to decide which data mining strategy is the most effective one to address the research question.

KDD has already been applied to a variety of fields, including astronomy, investment, marketing, manufacturing, public policy, sports, and telecommunications. An example of KDD system is the Sky Image Cataloging and Analysis Tool (SKICAT), which can automatically analyze, classify, and catalog sky objects as stars or galaxies using machine learning, machine-assisted discovery, and other AI technologies (http://www.ifa.hawaii.edu/~rgal/science/dposs/dposs_frames_skicat.html). Another KDD application is Advanced Scout, which is used by NBA coaching staffs to discover interesting patterns in basketball game data and allows users to relate these patterns to videos (<https://www.nbastuffer.com/analytics101/advanced-scout/>).

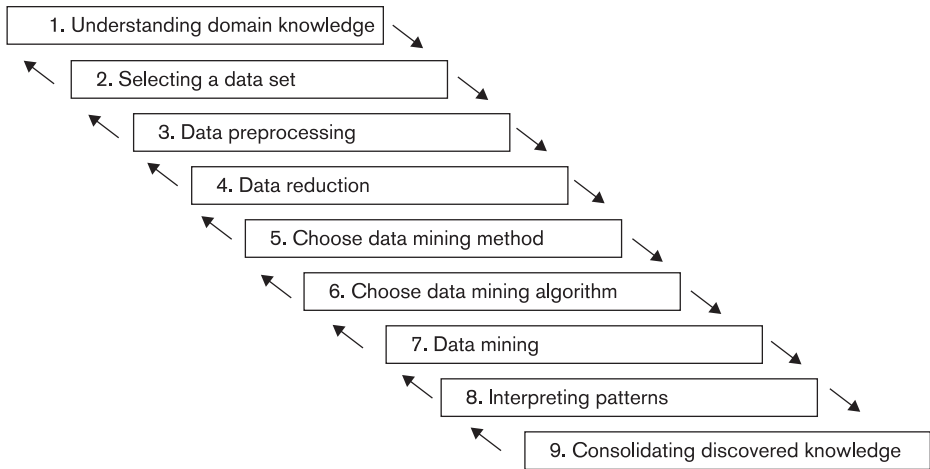


FIGURE 1.2
Components of Knowledge Discovery in a Database.

WHAT IS DATA MINING?

Data mining has two definitions. The narrow definition is that it is a step in the KDD process of applying data analysis and discovery algorithms to produce certain patterns or models on the data. As shown in figure 1.2, data mining is step 7 in the nine steps of the KDD model. It is usually the case that data mining operates in a pattern space that is infinite, and data mining searches that space to find patterns.

Based on this narrow definition of data mining, its techniques and terminology come from three sources. Statistics are the basic source of data mining and bring well-defined techniques to identify systematic relationships between variables. Data visualization, such as histograms and various plots, presents information in visual forms that provide attractive and powerful methods of data exploration. Computational methods include descriptive statistics, correlation, frequency tables, multivariate exploratory techniques, and advanced and generalized linear models. Figure 1.3 shows the three foundations of data mining according to this narrow definition.

AI, another foundation of data mining techniques, contributes to data mining development with information processing techniques based on a human reasoning model that is heuristic. Machine learning represents an important approach in data mining that “trains” computers to recognize patterns in data. An artificial neural network (ANN) consists of structures of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs are modeled after human brains in how they process information, and they learn by example by adjusting to the synaptic connections that exist between the neurons.

The last foundation of data mining is database systems that provide the support platform for information processing and mining. Increases in computing power and advancements in computer science have made it possible to store, access, and retrieve

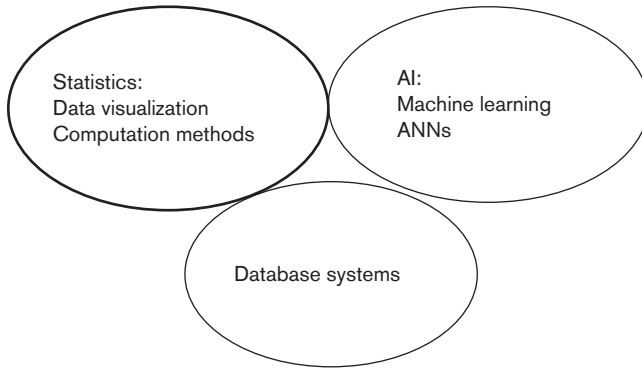


FIGURE 1.3
A Narrow Definition of Data Mining: Three Foundations.

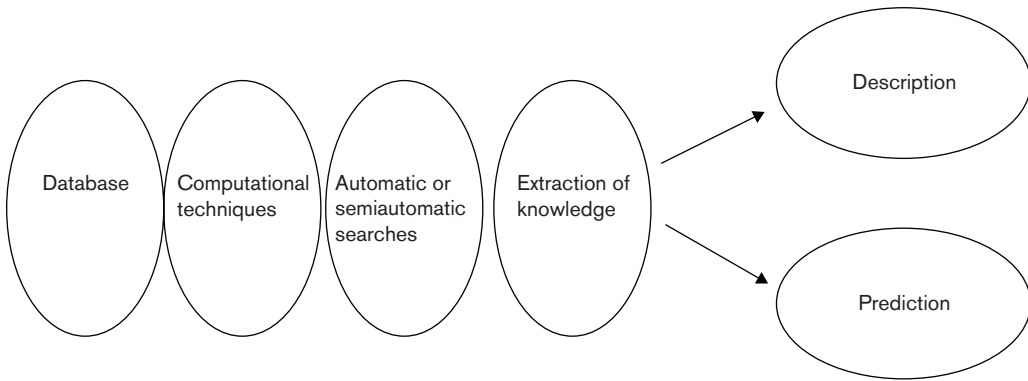


FIGURE 1.4
A Broad Definition of Data Mining.

a huge amount of data, enabling the use of new methods for revealing hidden patterns. These advancements made discoveries of new ideas and theories possible.

The second, and broader, definition of data mining conceptualizes it in a way similar to KDD (Gorunescu 2011). According to this definition, data mining has several components: (1) use of a huge database; (2) computational techniques; (3) automatic or semiautomatic search; and (4) extraction of implicit, previously unknown and potentially useful patterns and relationships hidden in the data. The information expected to be extracted by data scientists are of two types: descriptive and predictive (Larose and Larose 2016). Descriptive objectives are achieved by identifying relations among variables that describe data, and these patterns can be easily understood. Predictive objectives are achieved by using some of the variables to predict one or more of the other outcome variables, thus making it possible to accurately estimate future

outcomes based on existing data (Larose and Larose 2016). Figure 1.4 shows the broad definition of data mining.

This book uses the terms *knowledge discovery* and *data mining* interchangeably, according to the broadest conceptualization of data mining. Knowledge discovery and data mining in the social sciences constitute a research process that is guided by social science theories. Social scientists with deep domain knowledge work alongside data miners to select appropriate data, process the data, and choose suitable data mining technologies to conduct visualization, analysis, and mining of data to discover valid, novel, potentially useful, and ultimately understandable patterns. These new patterns are then consolidated with existing theories to develop new knowledge. Knowledge discovery and data mining in the social sciences are also important components of computational social science.

WHAT IS COMPUTATIONAL SOCIAL SCIENCE IN THE ERA OF BIG DATA?

Computational social science (CSS) is a new interdisciplinary area of research at the confluence of information technology, big data, social computing, and the social sciences. The concept of computational social science first gained recognition in 2009 when Lazer and colleagues (2009) published “Computational Social Science” in the journal *Science*. Email, mobile devices, credit cards, online invoices, medical records, and social media have recorded an enormous amount of long-term, interactive, and large-scale data on human interactions. CSS is based on the collection and analysis of big data and the use of digitalization tools and methods such as social computing, social modeling, social simulation, network analysis, online experiments, artificial intelligence to research human behaviors, collective interactions, and complex organizations (Watts 2013). Only computational social science can provide us with the unprecedented ability to analyze the breadth and depth of vast amounts of data, thus affording us a new approach to understanding individual behaviors, group interactions, social structures, and societal transformations.

Scholars have formulated a variety of conceptualizations of CSS. One version argues that there are two fundamental components in CSS: substantive and instrumental (Cioff-Revilla 2010). The substantive, or theoretical, dimension entails complex systems and theories of computer programming. The instrumental dimension includes tools for data processing, mining, and analysis, such as automatic information retrieval, social network analysis, socio-GIS, complex modeling, and computation simulation.

Another conceptualization postulates that CSS has four important characteristics. First, it uses data from “natural” samples that document actual human behaviors (unlike the more artificial data collected from experiments and surveys). Second, the data are big and complex. Third, patterns of individual behavior and social structure

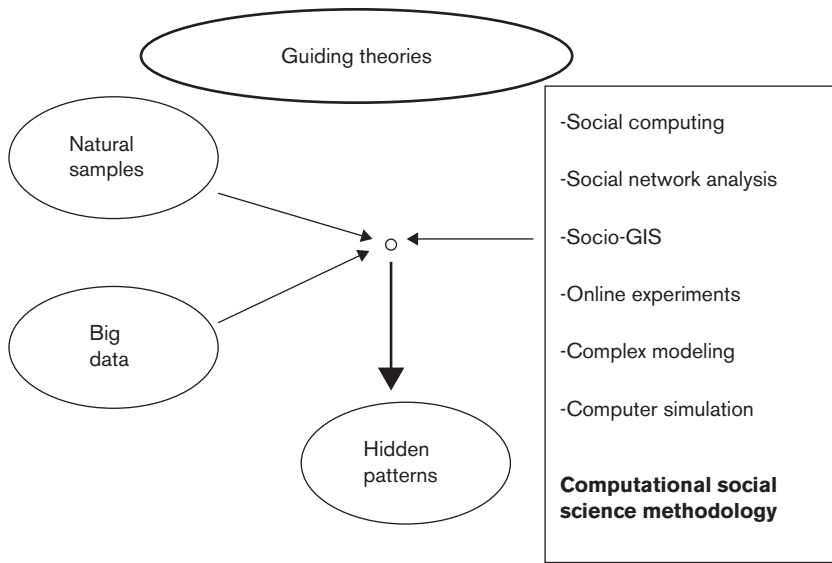


FIGURE 1.5
Computational Social Science.

are extracted using complex computations based on cloud computing with big databases and data mining approaches. And fourth, scientists use theoretical ideas to guide data mining of big data (Shah et al. 2015).

Others believe that CSS should be an interdisciplinary area at the confluence of domain knowledge, data management, data analysis, and transdisciplinary collaboration and coordination among scholars from different disciplinary training (Mason, Vaughan, and Wallach 2014). Social scientists provide insights on research background and questions, deciding on data sources and methods of collection, while statisticians and computer scientists develop appropriate mathematical models and data mining methods, as well as the necessary computational knowledge and skills to maintain smooth project progress.

Methods of computational social science consist primarily of social computing, online experiments, and computer simulations (Conte 2016). Social computing uses information processing technology and computational methods to conduct data mining and analysis on big data to reveal hidden patterns of collective and individual behaviors. Online experiments as a new research method use the internet as a laboratory to break free of the confines of conventional experimental approaches and use the online world as a natural setting for experiments that transcend time and space (Bond et al. 2012; Kramer, Guillory, and Hancock 2014). Computer simulations use mathematical modeling and simulation software to set and adjust program parameters to simulate social phenomena and detect patterns of social behaviors (Bankes 2002;

Gilbert et al. 2005; Epstein 2006). Both online experiments and computer simulations emphasize theory testing and development.

As figure 1.5 shows, in CSS researchers operate under the guidance of social scientific theories, apply computational social science methodology to data (usually big data) from natural samples, detect hidden patterns to enrich social science empirical evidence, and contribute to theory discovery.

OUTLINE OF THE BOOK

The book has six parts. Part I, comprising this chapter and chapter 2, explains the concepts and development of data mining and knowledge and the role it plays in social science research. Chapter 2 provides information on the process of scientific research as theory-driven confirmatory hypothesis testing. It also explains the impact of the new data mining and knowledge discovery approaches on this process.

Part II deals with data preprocessing. Chapter 3 elaborates issues such as privacy, security, data collection, data cleaning, missing data, and data transformation. Chapter 4 provides information on data visualization that includes graphic summaries of single, bivariate, and complex data.

Part III focuses on model assessment. Chapter 5 explains important methods and measures of model selection and model assessment, such as cross-validation and bootstrapping. It provides justifications as well as ways to use these methods to evaluate models. This chapter is more challenging than the previous chapters. Because the content is difficult for the average undergraduate student, I recommend that instructors selectively introduce sections of this chapter to their students. Later chapters on specific approaches also introduce some of these model assessment approaches. It may be most effective to introduce these specific methods of model assessment after students acquire knowledge of these data mining techniques.

Part IV is devoted to the methods of unsupervised learning: clustering and association. Chapter 6 explains the different types of cluster analysis, similarity measures, hierarchical clustering, and cluster validity. Chapter 7 concentrates on the topic of associations, including association rules, the usefulness of association rules, and the application of association rules in social research.

Part V continues with the topic of machine learning: supervised learning that includes generalized regression, classification and decision trees, and neural networks. Chapter 8 focuses on models of parameter learning that include linear regression and logistic regression. Chapter 9 covers inductive machine learning, decision trees, and types of algorithms in classification and decision trees. Chapter 10 focuses on neural networks, including the structure of neural networks, learning rules, and the advantages and disadvantages of neural networks.

Part VI ends the book with mining of text data and network data. Chapter 11 focuses on text mining and elaborates the topics of web mining, web crawling and scraping;

web mining of content, structure, and usage; and text mining topics including bag of words, N-grams, and topic models. Chapter 12 elaborates the network/link analysis that uses social network analysis, with various measures of network cohesion, and modes of centrality and prestige in social networks.

Chapters 10 through 12 tackle cutting-edge topics that are emerging as important techniques drawn from a variety of fields to address novel questions. As a result, the materials in these chapters are more cursory than those in others. These approaches—neural networks, network analysis, and text mining—are very broad and involve more advanced methods, so a fuller treatment cannot be provided here. These chapters thus serve as an introductory text for interested students and scholars. I provide a recommended reading list for those who are interested in pursuing these approaches further.

TARGET AUDIENCE AND GUIDE FOR INSTRUCTIONAL USE

The target audience of this book includes a wide range of people with varying backgrounds and readiness: faculty wishing to update or sharpen their skills, graduate students, select undergraduates, and applied data analysts in the nonprofit and corporate sectors. Each group can tailor the content to its individual needs and requirements.

This book is suitable for a variety of courses. It can serve as a textbook for an upper division social science methods course, as an introductory-level textbook for graduate students, as a companion to a traditional statistics textbook in data management or regression modeling, as an ancillary to a more in-depth statistics textbook that would overlap with material presented in chapters 4 through 8, and as a primer for faculty and researchers interested in advanced statistical methods.

Given these various purposes and audiences, below is a guide recommending different teaching plans targeted at different audiences in different courses.

1. An upper division undergraduate course on data mining: chapters 1–8.
2. A graduate course on knowledge discovery: chapters 1–12.
3. A course on machine learning: chapters 5–9.
4. A course on data management: chapters 1–4.
5. Supplementary reading for an undergraduate research methods course: chapters 1–4.
6. Supplementary reading for a graduate statistics course on new and modern methods in data science: chapters 10–12.

Instructors can also tailor the contents to the specific campus or training demands using some combination of chapters, as follows:

- Overview: chapters 1 and 2;
- Data and visualization: chapters 3 and 4;

- Model assessment: chapter 5;
- Unsupervised learning: chapters 6 and 7;
- Supervised learning: chapters 8–10; and
- Advanced approaches in neural networks, social networks, and text mining: chapters 10–12.

SUMMARY

The rapid growth of big data has provided us with new opportunities and challenges in knowledge discovery. To take advantages of these opportunities and resolve these challenges, data analytical skills are in high demand.

Big data has also revolutionized scientific research with the emergence of many inter- and multidisciplinary fields. Computational social science, bioinformatics, the Sloan Digital Sky Survey in astronomy, AlphaGo in artificial intelligence, and the Google Books corpus in language and culture studies have all brought structural transformations in these scientific research communities.

Data mining projects require both knowledge in handling data (i.e., data processing, database technology, statistical and computational algorithms) and domain expertise to detect and discover valid and meaningful knowledge and theory. It is thus important to train students and scholars who possess expertise in fields such as psychology, economics, sociology, political science, and linguistics to acquire data mining skills to work alongside computer scientists and statisticians.

Big data is characterized by big volume, high velocity, and big variety of information. It also entails technologies and analytical methods to transform information into insights that possess economic value and have societal impacts.

Knowledge discovery in database (KDD) is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. It consists of nine steps that begin with the development and understanding of the application domain and end with action on the knowledge discovered. The KDD process searches for patterns of interest in a particular representational form or a set of these representations.

Data mining has both a narrow definition and a broad definition. Data mining can be narrowly defined as one step in the KDD process that applies data analysis and discovery algorithms to produce patterns or models on the data. Data mining techniques come from statistics, artificial intelligence, and database systems.

A broader definition of data mining conceptualizes it as similar to the entire process of KDD, which involves use of huge databases, computational techniques, automatic or semiautomatic searches, and extraction of knowledge. The extracted information from data are of two types: descriptions of relationships and patterns in data and predictive estimates of future outcomes based on existing data.

Computational social science is a new interdisciplinary area of research at the confluence of information technology, big data, social computing, and social sciences. It is guided by social science theories, applies computational social science methodology to

data that is usually big and from natural samples, detects hidden patterns to enrich social science empirical evidence, and contributes to theory discovery and development.

PRACTICE PROBLEMS

1. Discuss what new opportunities have emerged in your field of research resulting from big data. What types of new research are made possible by this transformation? Can you identify potentially uncharted topics of research in your field using big data?
2. In addition to being big, what are the important features of data in the concept of big data? Does the concept of big data only entail data? If not, what else is included?
3. What is knowledge discovery in database? What are the steps involved in KDD?
4. What is data mining when it is narrowly defined? What are the three foundations of a narrow definition of data mining?
5. What is data mining broadly defined? What are the components in data mining? How do you characterize information extracted from data?
6. What is computational social science? Read Hanna Wallach's 2018 article, "Computational Social Science \neq Computer Science + Social Data" (<https://cacm.acm.org/magazines/2018/3/225484-computational-social-science-computer-science-social-data/fulltext>), and discuss why CSS is more than the sum of computer science and social data. What do social scientists bring to this new interdisciplinary area?

REFERENCES

- Banks, Steven C. 2002. "Agent-Based Modeling: Revolution?" *Proceedings of the National Academy of Science of the United States of America* 99 (S3): 7199–7200.
- Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature* 489: 295–98.
- Brown, Meta S. 2018. "Data Mining Concepts That Business People Should Know." *Forbes*, July 31. <https://www.forbes.com/sites/metabrown/2018/07/31/basics-of-data-mining-that-business-people-should-know/#74504e89515f>.
- Cioff-Rivilla, Claudio. 2010. "Computational Social Science." *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (3): 259–71.
- Conte, Rosaria. 2016. "Computational Social and Behavioral Science." In *New Frontiers in the Study of Social Phenomena*, ed F. Cecconi. New York: Springer.
- Dumbill, Edd. 2013. "A Revolution That Will Transform How We Live, Work, and Think: An Interview with the Author of Big Data." *Big Data* 1 (2): 73–77. <https://doi.org/10.1089/big.2013.0016>.
- Economist. 2017. "The World's Most Valuable Resource Is No Longer Oil, but Data." *Economist*, May 6. <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>.
- Epstein, Joshua M. 2006. "Remarks on the Foundations of Agent-Based Generative Social Science." In *Handbook of Computational Economics*, vol. 2, 1585–1604. New York: Elsevier. <https://www.sciencedirect.com/handbook/handbook-of-computational-economics/vol/2/suppl/C>.