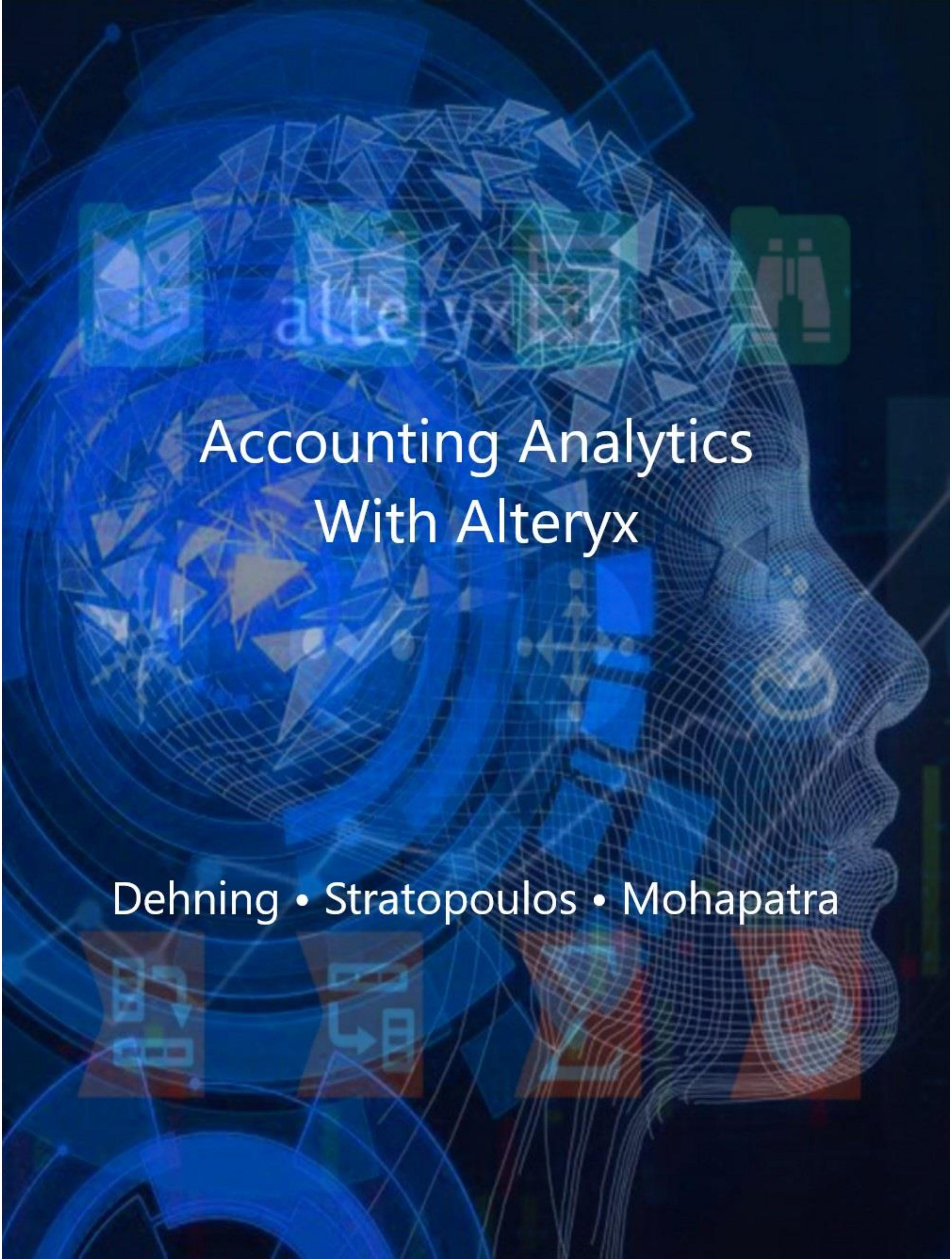

Accounting Analytics With Alteryx

Dehning • Stratopoulos • Mohapatra

PART I - Foundation

"Ithaka
As you set out for Ithaka
hope your road is a long one,
full of adventure, full of discovery."

CHAPTER 1: Textbook Orientation

1.1 Learning Objectives

The theme of this Chapter: To help you prepare for the journey (Constantinos Cavafis, N.A.). Understand the main learning objectives and structure of this textbook. By the end of this chapter, students should be able to:

1. Define and explain data analytics in their own words.
2. To explain why data analytics matters to accounting students.

3. Understand the textbook structure, teaching methods, assignments, and assessment.

1.2 Students: Advance Preparation

This textbook will use Alteryx, a relatively new data science and analytics software program. Alteryx was designed with ease of use at the forefront, making advanced analytics available to everyone. To help you become more familiar with Alteryx and maximize your benefits from this textbook, we have prepared a list of videos to watch and tasks to complete before the beginning of the second week of the term. Use the following hyperlinks to access Appendix C: Alteryx Checklist and Appendix D: Student Preparation Worksheet.

For individuals that feel lost or behind, perhaps because they are in a class with other students with much greater familiarity with Alteryx, we recommend the Datacamp Introduction to Alteryx course: https://www.datacamp.com/courses/introduction-to-alteryx

There is also a link on the textbook website. Before students can go beyond chapter 1, their professor must create a Datacamp classroom through the Datacamp for Universities program: https://www.datacamp.com/universities/

1.3 Background and an Analytics Map

The term *business intelligence* has been around since 1865 (Bogost 2018). According to Bogost (2018), "*The term business intelligence was first coined way back in 1865, in Richard Miller Devens's book Cyclopaedia of Commercial and Business Anecdotes …*" It refers to the ability of firms to leverage information on subject matters such as war, competition, and weather.

The introduction of relational databases by Edgar Codd in the late 1960s allowed firms and analysts to query large datasets. However, it was not until the nineties, when enterprise systems (i.e., manifestation of integrated relational databases) became popular, that firms like Harrah's Entertainment developed and implemented data-enabled strategies. While several other firms pioneered the use of data analytics during the late nineties and the beginning of the new millennium, data analytics did not reach mainstream adoption for another decade.

Interest in analytics exploded in the early 2000s. The publication of the best-selling book "Competing with Analytics" in 2007 (Davenport and Harris 2007) and the release of the movie *Moneyball* fueled this renewed interest (Stratopoulos 2018).

NOTE: Textbook Topics

The arrangement of topics in this textbook follows an approach that resembles the above historical presentation. We will start working with datasets contained in a single file (e.g., an Excel file with just one worksheet) in Part I. In Part II, we will learn what it means to link data in organized and meaningful ways (i.e., form a relational database). These data may relate, for example, to sales, purchases, inventory, and payroll. In Part III, we will discuss extracting, preparing, and analyzing data and leveraging data analytics in forecasting (e.g., forecast sales).

Data Mining

Data Mining (DM) became popular in the late 1990s and early 2000s. It refers to discovering patterns or knowledge from large datasets and applying statistical and logical methods to categorize data or make predictions based on available data.

For example, suppose that a bank has demographic and financial data about its clients, including, among other things, age, education level, and annual income. The bank also has its clients' loan repayment history, which includes information on whether borrowers have repaid loans on time. Based on this data, the bank could create a *classification* model that *labels* potential borrowers with various credit risk categories. The organization could then use the model to support or automate the decision to approve or pre-approve additional loans to its clients.

As another example, an auditor could use historical data on a company's revenue, production, and other factors, such as weather, to create a *regression* model. The current year's data on production and weather could be input into the model to predict the revenue for the year. A difference between the predicted and actual revenue could indicate a material misstatement of actual revenue and warrant additional investigation by the auditor.

Data Science

Data science has emerged in the 21st century as the evolution of data mining with increased access to data and computing power. *Data science* combines data manipulation, visualization, and statistical methods with other machine learning techniques such as neural nets, deep learning, boosting, random forests, and support vector machines (Efron and Hastie 2016).

The driving force of these developments was the commercial viability of prediction algorithms. *Data products* emerged as the commercial implementation of prediction algorithms: Amazon makes personalized recommendations and shows a catalog of its products. Similarly, Netflix uses its recommendation engine to suggest movies to viewers based on their past viewing habits and preferences.

Classification of Data Analytics

The four most commonly used classifications of analytics are descriptive, diagnostic, predictive, and prescriptive analytics.

1. **Descriptive analytics** focuses on preparing and analyzing historical data to identify patterns and report trends (i.e., what has happened?).

2. **Diagnostic analytics** focuses on patterns that have been discovered and tries to dig deeper to find causes (i.e., why did it happen?).
3. **Predictive analytics** focuses on predicting future probabilities and trends (i.e., what is likely to happen?). Predictive analytics aims to uncover patterns and trends that may not be apparent based on descriptive analysis or diagnostic analytics.
4. **Prescriptive analytics** is associated with the highest level of maturity in business analytics. While predictive analytics aims to tell us what will happen, prescriptive analytics focuses on what to do with this information (i.e., what should we do about it?). This often means identifying the best option given internal and external constraints.

Data Analytics for Decision-Making

An alternative approach to viewing these different types of analytics is their usefulness in reactive/proactive decision-making and the level of analytical sophistication (See Figure 1.1). By definition, descriptive and diagnostic analytics are primarily used to support reactive decision-making and tend to rely on relatively simpler analytical methods. For example, one can use simple descriptive statistics to identify patterns in a dataset (e.g., outliers, exceptions) and then leverage customized queries and pivot tables to dig deeper and discover why these outliers occur. Hence, diagnostic and descriptive analytics are relatively low in terms of supporting proactive decision-making and analytics sophistication.

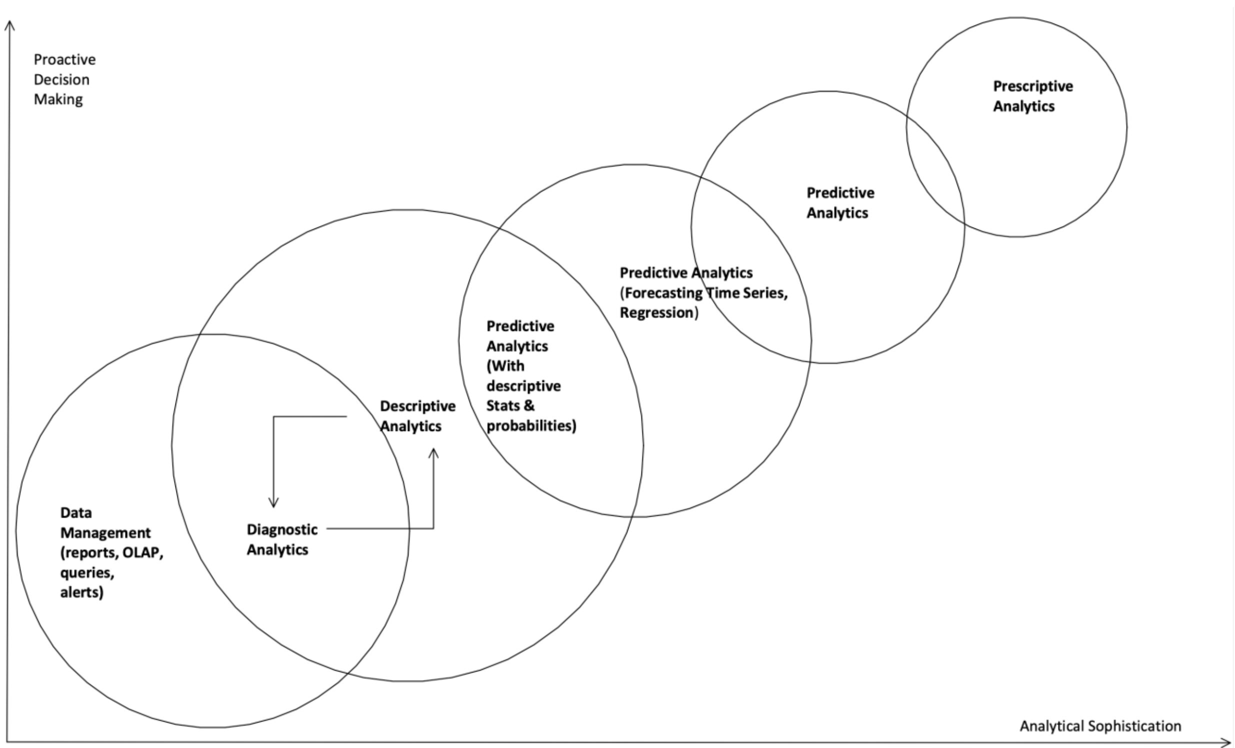

Figure 1.1: Data Analytics for Decision-Making

On the other hand, a combination of predictive analytics (e.g., neural networks) and prescriptive analytics (e.g., optimization algorithms) will enable managers to generate scenarios that reflect different assumptions related to internal resource constraints and market conditions and based on them, make predictions and plan appropriate actions given these different scenarios. Hence, predictive and prescriptive analytics are relatively high in terms of supporting proactive decision-making and analytics sophistication.

Predictions and Forecasting

While the terms "forecast" and "predict" tend to be used interchangeably, the former is typically associated with time series (e.g., forecast sales for the next four quarters), while the latter tends to be used with cross-sectional data (e.g., probability of churn, default on a loan, or response to a new marketing campaign). Forecasting is one of the most popular business analytics applications. Considering the ripple effects of forecasted sales to practically any area within the firm, one can easily understand the popularity of forecasting applications. In its simpler form, this may refer to identifying the basic trend behind a metric (e.g., sales, spending) over time. More advanced techniques will rely on a combination of regression and time series analysis.

Predictions based on cross-sectional data refer to the use of either traditional statistical analysis or data mining to predict the behavior of a target audience. There are countless opportunities for decision-makers to leverage predictive analytics. For example, decision-makers may use decision trees, logistic regression, or neural networks to predict which potential borrowers are more likely to default, credit card holders that will likely carry a balance, or insured motorists more likely to have a claim. Marketers may want to perform cluster analysis to segment customers into groups more likely to respond to a marketing campaign or market basket analysis to find customers purchasing patterns. This would be helpful if the marketer wants to bundle products likely to be purchased together.

1.4 Why Data Analytics?

Since the early 1990s, and primarily due to technological changes, the business landscape has become very competitive. It has become *hyper-competitive* (D'Aveni, Dagnino, and Smith 2010; D'Aveni 1994). The most successful companies embrace and leverage this constant state of change (i.e., the *red queen effect*). This means they need well-trained employees prepared to keep up with these technological changes. Hence, the question:

> … how can we best prepare students for the opportunities and challenges that lie ahead—and ready them for careers in professional services? At PwC, we believe data analytics should be integrated into accounting coursework. (PwC 2015)

In simple terms, the idea behind data analytics, business analytics, or accounting analytics - is that decision-makers should look at all available data to support their decision-making. Decision-makers need to understand what happened in the past and why it happened, what is happening now and why it is happening, and what is likely to happen in the future.

However, accounting students are not here to learn about data analytics; they are here to learn data analytics to make better accounting and business decisions. Hence, this textbook intends to provide an intuitive and practical introduction to data analytics tools/concepts using problems/applications in financial and managerial accounting, auditing, taxation, and accounting information systems settings.

Applications of data analytics in accounting include topics such as:

- Financial Accounting: Compare competing strategies (product differentiation and cost leadership) through ROA decomposition and establish a company's relative position (competitive advantage, parity, disadvantage) versus its peers.
- Managerial Accounting: Understand how we translate data into the information needed to monitor the performance of a business. For example, work with retail businesses to analyze their sales and develop an interactive business dashboard.
- Auditing: Audit client records to identify fraud and/or inventory valuation. Generate a model to predict sales and compare predicted versus actual sales. Any material discrepancies might be indicators of fraud.
- Taxation: Analyze client data for compliance with IRS rules. Forecast sales to calculate tax prepayments.
- Accounting Information Systems: Evaluate payoffs from technology investments. Understand emerging technologies (e.g., AI, blockchain) and consider their implications in using analytics in the profession.

1.5 Expectations & Learning Objectives

CRISP-DM stands for the Cross Industry Standard Process for Data Mining, and it is a process that data analysts/scientists use to approach data analytics problems. According to CRISP-DM (Figure 1.2), each data analytics problem goes through the following steps: 1) business understanding, 2) data understanding, 3) data preparation, 4) modeling, 5) evaluation and communication, and 6) deployment.

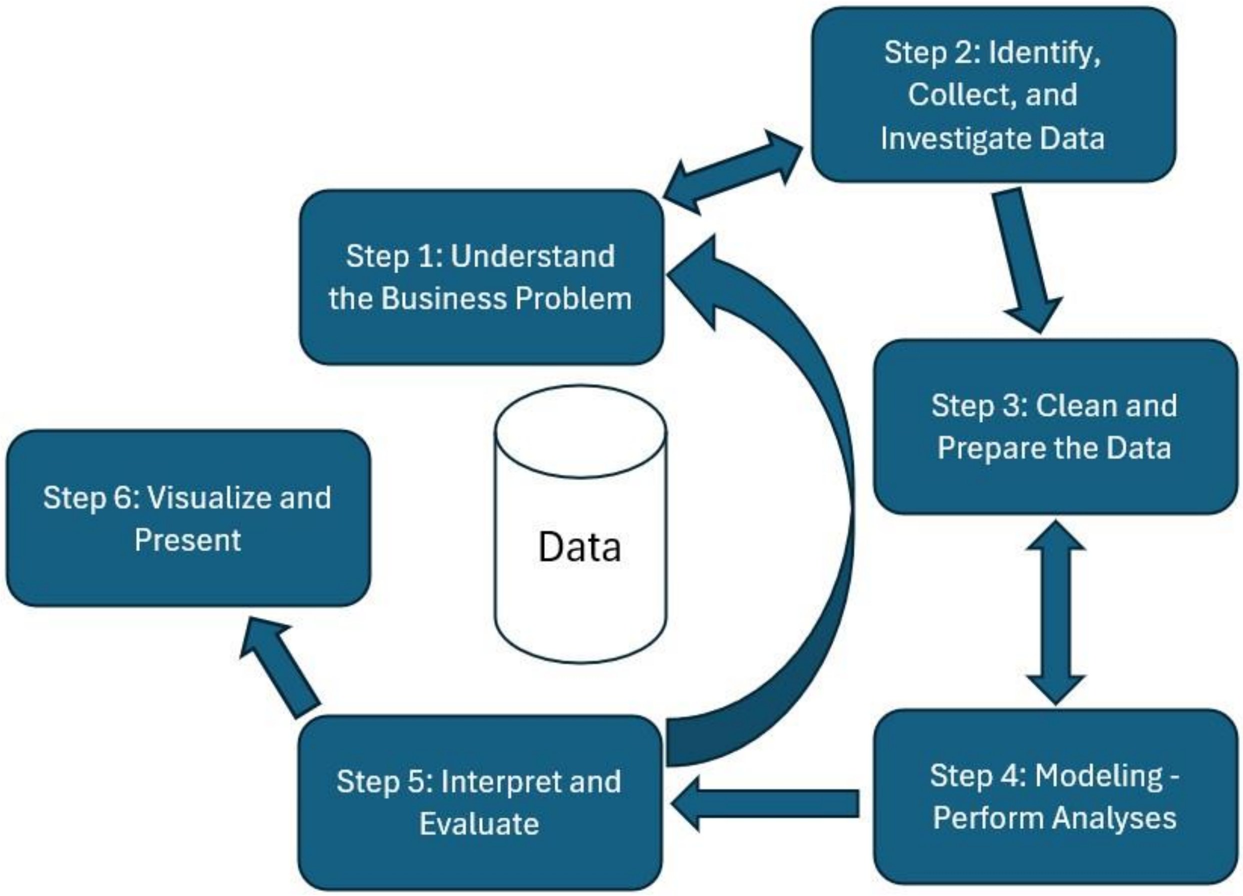

Figure 1.2: CRISP-DM Process Diagram

The steps in CRISP-DM reflect the expectations and learning objectives of this textbook. More specifically, the textbook is designed to provide accounting students with the competencies, professionalism, and practical experience they need to excel in their chosen careers. By the end of the textbook, students should be able to achieve the following objectives:

1. Business Understanding: Identify business and accounting applications where we can use data analytics concepts and tools to answer questions and solve problems.
2. Data Understanding: Identify data sources, collect and extract data, familiarize with data structure.
3. Data Preparation: Identify quality issues and clean and transform data for analysis.
4. Model (tool ambidexterity): Select the appropriate tool from within Alteryx to build a data analytics model to solve the business/accounting problem.
5. Evaluation: Leverage mathematical (i.e., test statistics) and domain-specific knowledge (e.g., financial accounting, managerial accounting, audit, tax, and business strategy) to evaluate how valuable a model is, what it has found, and what you may want to do with the results.
6. Deployment: Communicate your results to stakeholders and use the new insight to answer new questions and solve new problems. Leverage appropriate visualization

tools to communicate your message.

Alternatively, the goal is to help students develop an analytics mindset, i.e., the ability to:

- Ask the right questions.
- Extract, transform, and load relevant data.
- Apply appropriate data analytics techniques.
- Interpret and share the results with stakeholders.

1.6 Textbook Structure and Resources

The textbook is based on data analytics and technology-related teaching material developed explicitly for accounting students. The focus in each chapter will be as follows:

1. Introduce concepts/tools and case-based problems (business understanding).
2. Apply concepts using appropriate tools and available case data.
3. Communicate the key findings from the data analysis and debrief.

For data analysis, we will use Alteryx. A free limited license for Alteryx Designer is free for students under the Alteryx SparkED program (see Appendix C: Alteryx Checklist for details).

1.7 Textbook Assignments

The textbook contains in-class and out-of-class assignments that instructors can assign as individual or team assignments. Assignments are designed to help students develop an analytics mindset. Active participation leads to higher retention and understanding. Students should complete the "Advance Preparation" part of every chapter and review assigned material before they come to class. Active participation means asking and answering questions, commenting, helping other students understand the material, and working individually and in teams on class assignments and presentations.

1.8 Preview for Next Chapter

Before beginning chapter 2, students should review Appendix C: Alteryx Checklist for instructions on getting an Alteryx SparkED student license and installing Alteryx Designer. It is also recommended that students sign up for an Alteryx Community account.

In addition, go to Appendix D: Student Preparation Worksheet and watch at least the first five videos under chapter 1, which takes approximately one hour if you watch them on 1x speed. Many of the recommended videos are clear in 1.25x or even 1.5x speed.