Chapter 25
Meet Industry Needs in the Big Data Era: Data Science Curricula Development

Liguo Yu
Indiana University, South Bend, USA

Yingmei Li
Harbin Normal University, China

ABSTRACT

The potential wide applications of big data analytics have created a high demand for data analysts in various industries, including business, healthcare, bioinformatics, politics, and management. As a result, higher education institutions are capitalizing on this opportunity by offering different data science programs to attract students and cater to industry needs. Over the past decade, there has been a rapid emergence of data science programs both nationally and globally. This chapter will begin by reviewing the impact of big data analytics on different industries. It will then proceed to describe various data science programs, including their curriculum design, course offerings, and target industry sectors for employment. Additionally, the chapter will address the weaknesses of some curricula and propose new teaching areas that are relevant to improve the learning outcomes of students. The aim of the suggestions is to better prepare data science students for the ever-evolving demands of big data analytics in the industry.

INTRODUCTION

Due to the evolution of information technology, including machine learning, statistical methods, and distributed computing, we are now able to analyze traditionally believed hard-to-dissect unstructured data, such as customer shopping history, business communication records, and personal activity records, which include financial activities, communication activities, and social media activities (Qiu et al., 2016; Wang et al., 2016; Liu et al., 2015). Big data analytics not only can help businesses, organizations,

DOI: 10.4018/979-8-3693-0049-7.ch025

institutions, and governments reduce operational costs, improve work efficiencies, and make strategic decisions, but it can also analyze data in real-time and respond to urgent and critical situations. This not only benefits institutions and corporations economically but can also save lives and prevent disasters (Berry & Linoff, 1999; Tansley & Tolle, 2009).

The potential wide applications of big data analytics have created a high demand for data analysts in many areas. Higher education institutions are taking this opportunity to grow their influence by offering different data science curricula to attract students of future data analysts or data scientists to meet the industry needs (De Veaux et al., 2017; Gil, 2014; Lyon et al., 2015). In the past ten years, we have witnessed the fast emergence of data science programs nationally and globally (Anderson et al., 2014; Clayton & Clopton, 2019; Demchenko, 2019). These programs range from bachelor's degree to PhD degree with various of business concentrations, such as finance, healthcare, bioinformatics, social informatics, politics, and management (Dogucu & Çetinkaya-Rundel, 2021).

However, because data science (big data analytics) is an interdisciplinary program, the current curriculum design might not fully satisfy the industry needs. In addition, big data analysis technologies and applications are both fast evolving, which require the data science curricula be reviewed and updated regularly to better fulfill the industries' dynamic needs (De Vaus, 2002; Provost & Fawcett, 2013; Tang & Sae-Lim, 2016). To our knowledge, academically there is a lack of summary of data science programs and their relations to industry needs.

This chapter describes different data science programs, their curriculum design, course offerings, and target industry sectors for employment. In addition, we discuss the weakness of some curriculum design and propose new teaching areas that are relevant to improve the learning outcomes of students. The aim of our suggestions is to better prepare future data science students for the ever-evolving demands of big data analytics in the industry. Our contributions to this field are twofold. First, we summarized current data science programs. Second, we provided guidance for future curriculum improvement to better meet the industry needs.

The remaining chapter is organized as follows. First, we review data science (big data analysis) and its impact on business and society. Next, we describe the common curriculum design of data science programs in the United States and across the world. Then, we analyze the strengths and weakness of some curriculum design and offer our suggestions for improvement. Finally, we conclude the chapter by summarizing the key points and illustrating the future work.

BIG DATA ANALYSIS AND ITS IMPACT

What is data? To answer this question, we need to define the concept of information. Let us say the world is made of entities. We human beings measure the properties of entities and observe the behaviors of entities. We understand the entities through observing and processing the information produced by the entity. The information could be in many different forms: electronic, acoustic, mechanic, chemical, biological, and sociological. It is worth noting that we might only be able to observe a fraction of the information produced by the entity. If we can collect more information, we can have a more accurate understanding of the entity (Seife, 2007; Gleick, 2012).

From information theory, we can define data as records of information, which is the fraction of information emitted from the entity that we observed and recorded. Because data is recorded information, we can analyze it later and reprocess it again and again in order to have a better understanding of the