IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, January 2023
Impact Factor: 7.301

Data Science and its Relationship to Big Data and Data-Driven Decision Making

Mrs. Butala Pooja[1] and Mrs. Ashwini Sheth[2]
Student, M.Sc. I.T., I. C. S. College, Khed, Ratnagiri, Maharashtra, India[1]
Asst. Prof., Department of I.T., I. C. S. College, Khed, Ratnagiri, Maharashtra, India[2]

Abstract: *Companies are realizing that they need to hire data scientists, academic institutions are rushing to develop data science programs, and publications are promoting data science as a hot, even "sexy" career choice. However, there is a lack of clarity regarding the specifics of data science, and this lack of clarity may result in disillusionment as the concept fades into meaningless buzz. We argue in this article that it has been difficult to define data science precisely for good reasons. The fact that big data and data-driven decision making are two other important concepts that are also gaining importance is one reason. Another reason is that people naturally tend to link a practitioner's work to the definition of their field; This can lead to ignoring the field's fundamentals. We do not believe that it is of the utmost importance to attempt to precisely define the boundaries of data science. In an academic setting, we can debate the field's boundaries, but data science can only be of use to businesses if (i) its relationships to other important related concepts are understood and (ii) the fundamental principles of data science are identified. Once we accept (ii), it will be much easier for us to comprehend and precisely explain what data science has to offer. Furthermore, we won't be able to call it data science until we accept (ii).We present a viewpoint that addresses all of these ideas in this article. We conclude by providing a sample list of data science's fundamental principles as examples.*

Keywords: Big Data

I. INTRODUCTION

Companies in almost every industry are focusing on using data to gain an advantage over competitors now that a lot of Data is available. The quantity and variety of data have far surpassed the capacity of manual analysis, and in some instances, conventional databases' capacity has been exceeded. At the same time, computers have become much more powerful, networking is common, and algorithms have been developed that can connect datasets to make it possible to conduct analyses that are more extensive and in-depth than they were previously. The growing use of data science in business is the result of the convergence of these phenomena. Businesses in all sectors have realized that they require more data scientists on staff. Data scientist training programs are in high demand at academic institutions.

Data science is being promoted as a hot and even "sexy" career choice by publications.“1 However, there is a lack of clarity regarding what exactly data science is, and this lack of clarity may well result in disillusionment as the idea becomes just a bunch of meaningless noise. We argue in this article that it has been difficult to define data science for a number of good reasons. The fact that big data and data-driven decision making, both of which are receiving increasing attention, are intricately intertwined with data science is one reason. Another reason is that, in the absence of academic programs that teach otherwise, practitioners naturally tend to associate what they actually do with the definition of their field; As a result, the fundamentals of the field may be overlooked.

At the moment, it is not particularly important to try to precisely define the boundaries of data science. We are able to debate the boundaries of data-science academic programs in an academic setting. However, in order for data science to be useful to businesses, it is necessary to (i) comprehend its connections to these other significant and closely related ideas, and (ii) begin to comprehend the relationship management to manage customer turnover and maximize expected customer value by analyzing customer behavior. Data science is used in the finance industry for credit scoring, trading, fraud detection, and workforce management. Data science is used by major retailers like Wal-Mart and Amazon in every aspect of their businesses, from marketing to supply chain management. With data science, many businesses have differentiated themselves strategically, sometimes to the point of becoming data-mining businesses.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, January 2023
Impact Factor: 7.301

However, there is a lot more to data science than just data-mining algorithms. Data scientists who are successful must be able to analyze business issues from a data perspective. Data-analytic thinking adheres to a fundamental structure and fundamental principles that must be comprehended. Numerous "traditional" fields of study are incorporated into data science. It is necessary to comprehend the fundamental principles of causal analysis. Fundamental to data science is a significant portion of what has traditionally been studied in the field of statistics. methology and methods the fundamentals that underpin data science. Once we accept (ii), we will be able to fully comprehend and articulate the benefits of data science. Additionally, only after we embrace (ii) for visualizing data are vital. There are also particular areas where intuition, creativity, common sense, and knowledge of a particular application must be brought to bear. A data-science perspective should we be comfortable calling it data science.

We present a viewpoint that addresses all of these ideas in this article. We begin by separating this collection of closely interrelated concepts. Data science is emphasized as the link between data-processing technologies, such as those for "big data," and data-driven decision making. We debate the complicated distinction between data science as a profession and as a field. In conclusion, we provide examples of some fundamental data science principles.

II. DATA SCIENCE

Data science is, at its most fundamental level, a collection of fundamental principles that support and direct the methodical extraction of information and knowledge from data. Data mining—the actual process of extracting knowledge from data using technologies that adhere to these principles—may be the concept that is most closely related to data science. There are hundreds of different data-mining algorithms, and the methods used in the field are very well-detailed. We contend that a much smaller and more concise set of fundamental principles lies beneath these numerous particulars.

These principles and techniques are applied broadly across functional areas in business. Probably the broadest business applications are in marketing for tasks such as targeted marketing, online advertising, and recommendations for cross-selling. Data science also is applied for general customer

2.1 Data Science in Action

For concreteness, let's look at two brief case studies of analyzing data to extract predictive patterns. These studies illustrate different sorts of applications of data science. The first was reported in the New York Times:

Hurricane Frances was on its way, barreling across the Caribbean, threatening a direct hit on Florida's Atlantic coast. Residents made for higher ground, but far away, in Bentonville, Ark., executives at Wal- Mart Stores decided that the situation offered a great opportunity for one of their newest data-driven weapons. predictive technology.

A week ahead of the storm's landfall, Linda M. Dillman, Wal-Mart's chief information officer, pressed her staff to come up with forecasts based on what had happened when Hurricane Charley struck several weeks earlier. Backed by the trillions of bytes' worth of shopper history that is stored in Wal- Mart's data warehouse, she felt that the company could ''start predicting what's going to happen, in- stead of waiting for it to happen,'' as she put it.2 of the hurricane would buy more bottled water. Maybe, but it seems a bit obvious, and why do we need data science to discover this? It might be useful to project the amount of increase in sales due to the hurricane, to ensure that local Wal-Marts are properly stocked. Perhaps mining the data could reveal that a particular DVD sold out in the hurricane's path—but maybe it sold out that week at Wal-Marts across the country, not just where the hurricane landing was imminent. The prediction could be somewhat useful, but probably more general than Ms. Dillman was intending.

It would be more valuable to discover patterns due to the hurricane that were not obvious. To do this, analysts might examine the huge volume of Wal-Mart data from prior, similar situations (such as Hurricane Charley earlier in the same season) to identify unusual local demand for products. From such patterns, the company might be able to anticipate unusual demand for products and rush stock to the stores ahead of has already designed a special retention offer. Your task is to devise a precise, step-by-step plan for how the data science team should use MegaTelCo's vast data resources to decide which customers should be offered the special retention deal prior to the expiration of their contracts. Specifically, how should MegaTelCo decide on the set of customers to target to best reduce churn for a particular incentive budget? Answering this question is much more complicated than it seems initially.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, January 2023
Impact Factor: 7.301

2.2 Data Science and Data-Driven Decision Making

Data science involves principles, processes, and techniques for understanding phenomena via the (automated) analysis of data. For the perspective of this article, the ultimate goal of data science is improving decision making, as this generally is the hurricane's landfall.

Indeed, that is what happened. The New York Times reported that: ''. the experts mined the data and found that the stores would indeed need certain products—and not just the usual flashlights. 'We didn't know in the past that strawberry Pop-Tarts of paramount interest to business. Figure 1 places data science in the context of other closely related and data-related processes in the organization. Let's start at the top.

Data-driven decision making (DDD)3 refers to the practice of basing decisions on the analysis increase in sales, like seven times their normal sales rate, ahead of a hurricane,' Ms. Dillman said in a recent interview.'

You have been called in to help understand the problem and to devise a solution. Attracting new customers is much more expensive than retaining existing ones, so a good deal of marketing budget is allocated to prevent churn. Marketing of data rather than purely on intuition. For example, a marketer could select advertisements based purely on her long experience in the field and her eye for what will work. Or, she could base her selection on the analysis of data regarding how consumers react to different ads. She could also use a combination of these approaches. DDD is not an all-or- nothing practice, and different firms engage in DDD to greater or lesser degrees.

The benefits of data-driven decision making have been demonstrated conclusively. Economist Erik Brynjolfsson and his colleagues from MIT and Penn's Wharton School recently conducted a study of how DDD affects firm performance.3 They developed a measure of DDD that rates firms as to how strongly they use data to make decisions across the company. They show statistically that the more data-driven a firm is, the more productive it is—even controlling for a wide range of possible confounding factors. And the differences are not small: one standard deviation higher on the DDD scale is associated with a 4–6% increase in productivity. DDD also is correlated with higher return on assets, return on equity, asset utilization, and market value, and the relationship seems to be causal.

Our two example case studies illustrate two different sorts of decisions: (1) decisions for which ''discoveries'' need to be made within data, and (2) decisions that repeat, especially at massive scale, and so decision making can benefit from even small increases in accuracy based on data analysis. The Wal- Mart example above illustrates a type-1 problem. Linda Dillman would like to discover knowledge that will help Wal- Mart prepare for Hurricane Frances's imminent arrival. Our churn example illustrates a type- industries have adopted automatic decision making at dif- ferent rates. The finance and telecommunications industries were early adopters. In the 1990s, automated decision making changed the banking and consumer-credit industries dramatically. In the 1990s, banks and telecommunications companies also implemented massive-scale systems for managing data-driven fraud control decisions. As retail systems communications company may have hundreds of millions of customers, each a candidate for defection. Tens of millions of customers have contracts expiring each month, so each one of were increasingly computerized, merchandising decisions were automated. Famous examples include Harrah's casinos' reward programs and the automated recommendations of Amazon and them has an increased likelihood of defection in the near future. If we can improve our ability to estimate, for a given customer, how profitable it would be for us to focus on her, we can potentially reap large benefits by applying this ability to the millions of customers in the population. This same logic applies to many of the areas where we have seen the most intense application of data science and data mining: direct marketing, online advertising, credit scoring, financial trading, help-desk management, fraud detection, search ranking, product recommendation, and so on.

The diagram in Figure 1 shows data science supporting data- driven decision making, but also overlapping with it. This highlights the fact that, increasingly, business decisions are being made automatically by computer systems. Different Netflix. Currently we are seeing a revolution in advertising, due in large part to a huge increase in the amount of time consumers are spending online and the ability online to make (literally) split-second advertising decisions.

2.3 Data Processing and ''Big Data''

Despite the impression one might get from the media, there is a lot to data processing that is not data science. Data engineering and processing are critical to support data-science activities, as shown in Figure 1, but they are more general and are useful for much more. Data-processing technologies are important for many business tasks that do not

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, January 2023
Impact Factor: 7.301

involve extracting knowledge or data-driven decision making, such as efficient transaction processing, modern web system processing, online advertising campaign management, and others.

One way to think about the state of big data technologies is to draw an analogy with the business adoption of internet technologies. In Web 1.0, businesses busied themselves with getting the basic internet technologies in place so that they could establish a web presence, build electronic commerce capability, and improve operating efficiency. We can think of ourselves as being in the era of Big Data 1.0, with firms engaged in building capabilities to process large data. These primarily support their current operations for example, to make themselves more efficient.

2.4 Data-Analytic Thinking

One of the most critical aspects of data science is the support of data-analytic thinking. Skill at thinking data-analytically is important not just for the data scientist but throughout the organization. For example, managers and line employees in other functional areas will only get the best from the company's data-science resources if they have some basic understanding of the fundamental principles. Managers in enterprises without substantial data-science resources should still understand basic principles in order to engage consultants on an informed basis. Investors in data-science ventures need to understand the fundamental principles in order to as-incorporated basic technologies thoroughly (and in the process had driven down prices) they started to look further. They began to ask what the web could do for them, and how it could improve upon what they'd always done. This ushered in the era of Web 2.0, in which new systems and companies investment opportunities accurately. More generally, businesses increasingly are driven by data analytics, and there is great professional advantage in being able to interact competently with and within such businesses. Understanding the fundamental concepts, and having frameworks started to exploit the interactive nature of the web. The changes brought on by this shift in thinking are extensive and pervasive; the most obvious are the incorporation of social- networking components and the rise of the ''voice'' of the individual consumer (and citizen).

Similarly, we should expect a Big Data 2.0 phase to follow Big Data 1.0. Once firms have become capable of processing massive data in a flexible fashion, they should begin asking: What can I now do that I couldn't do before, or do better than I could do before? This is likely to usher in the golden era of data science. The principles and techniques of data science will be applied far more broadly and far more deeply than they are today.

It is important to note that in the Web-1.0 era, some precocious companies began applying Web-2.0 ideas far ahead of the mainstream. Amazon is a prime example, incorporating the consumer's ''voice'' early on in the rating of products and product reviews (and deeper, in the rating of reviewers). Similarly, we see some companies already ap- plying Big Data 2.0. Amazon again is a company at the forefront, providing data-driven recommendations from massive data. There are other examples as well. Online advertisers must process extremely large volumes of data (bil- lions of ad impressions per day is not unusual) and maintain a very high throughput (real-time bidding systems make decisions in tens of milliseconds). We should look to these and similar industries for signs of advances in big data and data science that subsequently will be adopted by other industries. for organizing data-analytic thinking, not only will allow one to interact competently, but will help to envision opportunities for improving data-driven decision making or to see data-oriented competitive threats.

Firms in many traditional industries are exploiting new and existing data resources for competitive advantage. They employ data-science teams to bring advanced technologies to bear to increase revenue and to decrease costs. In addition, many new companies are being developed with data mining as a key strategic component. Facebook and Twitter, along with many other ''Digital 100'' companies,5 have high valuations due primarily to data assets they are committed to capturing or creating.{ Increasingly, managers need to man- age data-analytics teams and data-analysis projects, marketers have to organize and understand data-driven campaigns, venture capitalists must be able to invest wisely in businesses with substantial data assets, and business strategists must be able to devise plans that exploit data.

As a few examples, if a consultant presents a proposal to exploit a data asset to improve your business, you should be able to assess whether the proposal makes sense. If a competitor announces a new data partnership, you should recognize when it may put you at a strategic disadvantage. Or, let's say you take a position with a venture firm and your first project is to assess the potential for investing in an advertising company. The founders present a convincing argument that they will realize significant value from a unique body of data they will collect, and on that basis, are arguing for a sub-stantially higher valuation. Is this reasonable?

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, January 2023
Impact Factor: 7.301

subsequently apply data-science methods or to understand the results. However, that does not change the fact that the day-to-day work of a data scientist—especially an entry-level one—may be largely data processing. This is directly analogous to an entry-level chemist spending the majority of her time doing technical lab work. If this were all she were trained to do, she likely would not be rightly called a chemist but rather a lab technician. Important for being a chemist is that this work is in support of the application of the science of chemistry, and hopefully the eventual advancement to jobs involving more chemistry and less technical work. Similarly for data science: a chief scientist in a data-science-oriented company will do much less data processing and more data- analytics design and interpretation.

At the time of this writing, discussions of data science inevitably mention not just the analytical skills but the popular tools used in such analysis. For example, it is common to see job advertisements mentioning data-mining techniques (random forests, support vector machines), specific application areas (recommendation systems, ad placement optimization), alongside popular software tools for processing big data (SQL, Hadoop, MongoDB). This is natural. The particular concerns of data science in business are fairly new, and businesses are still working to figure out how best to address them. Continuing our analogy, the state of data science may be likened to that of chemistry in the mid-19th century, when theories and general principles were being formulated and the field was largely experimental. Every good chemist had to be a competent lab technician. Similarly, it is hard to imagine a working data scientist who is not proficient with certain sorts of software tools. A firm may be well served by requiring that their data scientists have skills to access, prepare, and process data using tools the firm has adopted.

Nevertheless, we emphasize that there is an important reason to focus here on the general principles of data science. In ten years' time, the predominant technologies will likely have changed or advanced enough that today's choices would seem quaint. On the other hand, the general principles of data science are not so differerent than they were 20 years ago and likely will change little over the coming decades.

III. CONCLUSION

Underlying the extensive collection of techniques for mining data is a much smaller set of fundamental concepts comprising data science. In order for data science to flourish as a field, rather than to drown in the flood of popular attention, we must think beyond the algorithms, techniques, and tools in common use. We must think about the core principles and concepts that underlie the techniques, and also the systematic thinking that fosters success in data-driven decision making. These data science concepts are general and very broadly applicable.

Success in today's data-oriented business environment re- quires being able to think about how these fundamental concepts apply to particular business problems—to think data-analytically. This is aided by conceptual frameworks that themselves are part of data science. For example, the auto- mated extraction of patterns from data is a process with well- defined stages. Understanding this process and its stages helps structure problem solving, makes it more systematic, and thus less prone to error.

There is strong evidence that business performance can be improved substantially via data-driven decision making,3 big data technologies,4 and data-science techniques based on big data.9,10 Data science supports data-driven decision making and sometimes allows making decisions automatically at massive scale and depends upon technologies for ''big data'' storage and engineering. However, the principles of data science are its own and should be considered and dis-cussed explicitly in order for data science to realize its potential.

REFERENCES

[1]. Davenport T.H., and Patil D.J. Data scientist: the sexiest job of the 21st century. Harv Bus Rev, Oct 2012.
[2]. Hays C. L. What they know about you. N Y Times, Nov. 14, 2004.
[3]. Brynjolfsson E., Hitt L.M., and Kim H.H. Strength in numbers: How does data-driven decision making affect firm performance? Working paper, 2011. SSRN working paper. Available at SSRN: http://ssrn.com/abstract = 1819486.
[4]. Tambe P. Big data know-how and business value. Working paper, NYU Stern School of Business, NY, New York, 2012.
[5]. Fusfeld A. The digital 100: the world's most valuable startups. Bus Insider. Sep. 23, 2010.
[6]. Shah S., Horne A., and Capella´ J. Good data won't guarantee good decisions. Harv Bus Rev, Apr 2012.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, January 2023
Impact Factor: 7.301

[7]. Wirth, R., and Hipp, J. CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, 2000, pp. 29–39.