RESEARCH ARTICLE OPEN ACCESS

The Interwoven Roles of Data Science, Big Data, and Data-Driven Decision Making

[1]Prof. Suvendu Tripathy, [2]Prof. Rajaswini Mishra, [3]Prof. Bhakta Charan Jena
Assistant Professor, Department of Computer Science & Engineering, Raajdhani Engineering College, Bhubaneswar
Assistant Professor, Department of MBA, Gandhi Institute of Technology AndManagement, Bhubaneswar
Assistant Professor, Department of MBA, Gandhi Institute of Technology And Management, Bhubaneswar

ABSTRACT
The burgeoning field of data science has captured the attention of industry and academia alike, fuelled by the promise of a highly sought-after career. However, the very definition of data science remains elusive, threatening to dilute its significance into a mere buzzword. This confusion stems from its intrinsic connection to other burgeoning concepts like big data and data-driven decision-making, and from the tendency to equate the field with the tasks performed by its practitioners. While precisely defining the boundaries of data science may be a less critical pursuit, a clear understanding of its relationship to related concepts and the identification of its fundamental principles are paramount. By embracing these fundamental principles, we can better articulate the true value proposition of data science and ensure its effective application in a business context. This essay will address these crucial considerations and offer a preliminary enumeration of fundamental principles that underpin the discipline.

Date of Submission: 20-05-2025 Date of acceptance: 30-05-2025

I. INTRODUCTION

As the availability of vast amounts of data continues to grow, businesses across nearly every industry are focused on leveraging this data for a competitive edge. The sheer volume and diversity of data have overwhelmed manual analysis efforts and sometimes even surpassed the capabilities of traditional databases. Simultaneously, computers have become more powerful, networking technologies are everywhere, and advanced algorithms now enable more extensive and deeper analysis than ever before. These developments have led to the widespread adoption of data science in business.

Organizations have recognized the need to hire more data scientists, while academic institutions are racing to develop training programs for this new profession. Publications often herald data science as an exciting and desirable career path. However, there's still confusion surrounding what exactly data science entails. This ambiguity could lead to disappointment once the concept becomes just another buzzword.

One reason it's difficult to clearly define data science is its close connection to other key concepts like big data and data-driven decision-making, all of which are rapidly growing in importance. Another reason is that, in the absence of formal academic delineation, practitioners are often identified by what they do rather than a clear set of principles—potentially overlooking the fundamental aspects of the discipline.

Currently, precisely defining the boundaries of data science isn't the immediate priority. Academic programs are emerging, and discussions about its scope are ongoing within scholarly circles. Still, for data science to be truly effective in business, it's crucial to (i) understand how it relates to these other concepts, and (ii) start to identify the core principles that underpin it. Once we do (ii), we can better explain what data science offers. Only then can we confidently refer to it as a distinct field.

In this article, we aim to address these interrelated ideas. We start by clarifying the connections between data science, big data, and data-driven decision-making. We explore the complex distinction between data science as an academic discipline and as a professional practice. Finally, we provide examples of some of the fundamental principles that form the foundation of data science.

What is Data Science

Data science is essentially a set of core principles that guide the systematic extraction of useful insights and knowledge from data. While data mining—using specific algorithms to uncover patterns—is closely related, it's just one part of the broader field. Underlying all these techniques are fundamental principles that can be summarized succinctly.

These principles are widely applied across different business sectors. In marketing, for instance, they support targeted advertising, online ads, and product recommendations. Data science also helps optimize customer relationship management by analysingbehaviour to reduce churn and boost customer value. In finance, it's used for credit scoring, trading, fraud detection, and workforce planning. Major retailers like Walmart and Amazon leverage data science throughout their operations, from marketing strategies to supply chain logistics—and some even evolve into data-driven, data-mining companies.

However, data science extends beyond simply applying algorithms. Success requires the ability to interpret business problems through a data-focused lens. There's a structured way to think analytically, rooted in essential principles. This field draws on many traditional disciplines—particularly statistics, causal analysis, data visualization, and domain knowledge. Intuition, creativity, and common sense are also key. A data science perspective offers a structured framework that allowspractitioners to systematically approach and solve problems related to extracting meaningful knowledge from data.

Data Science in Practice

Let's consider two brief case studies demonstrating how data analysis can be used to identify predictive patterns, each illustrating different applications of data science. The first case was featured in the New York Times: As Hurricane Frances approached, threatening Florida's Atlantic coast, residents sought safety, but in Bentonville, Arkansas, Wal-Mart's executives saw an opportunity to apply their latest data-driven tool—predictive analytics. A week before landfall, Wal-Mart's CIO, Linda M. Dillman, instructed her team to generate forecasts based on data from Hurricane Charley, which had made landfall a few weeks earlier. With access to the company's extensive customer purchase records, she believed they could start predicting future shopping behaviours instead of waiting for events to occur.

Why would such predictions be helpful here? While obvious predictions—like an increase in bottled water sales—are straightforward, they might not justify the need for complex data analysis. More valuable would be uncovering subtle, non-obvious patterns—such as spikes in demand for specific products that aren't immediately apparent. Analysing past data from similar storms, like Hurricane Charley, could reveal unusual demand for certain items, enabling Wal-Mart to pre-stock stores accordingly. Indeed, the team found such insights: for example, strawberry Pop-Tarts watched a sales surge of seven times the usual rate ahead of the hurricane, and beer was the top-selling item. These insights allowed for better anticipation and preparation.

Now, consider a different business scenario: You've just landed an analytical role at Mega Telco, one of the largest US telecom providers. They're facing a significant customer retention challenge in their wireless division: in the mid-Atlantic region, 20% of customers leave when their contracts expire, and gaining new customers has become increasingly difficult due to market saturation. Customer churn — where clients switch providers — is costly because it involves spending on incentives to retain existing customers and revenue loss when they leave.

Your mission is to develop a step-by-step plan for how Mega Telco's data science team should leverage their extensive customer data resources to identify which customers should be targeted with a special retention offer before their contracts naturally expire. The key question is: How can Mega Telco determine the right set of customers to offer the incentive to, in order to most effectively reduce churn within the constraints of a limited marketing budget? Crafting this decision-making process involves complex considerations beyond simply offering the deal to all customers nearing expiration.

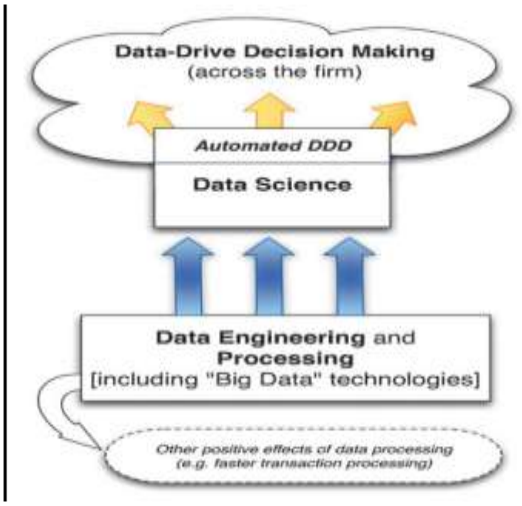
Using Data Science for Informed Decision-Making

Data science encompasses the principles, processes, and techniques used to understand phenomena through the automated analysis of data. For the purposes of this discussion, the primary goal of data science is to enhance decision-making, which is generally of utmost importance to businesses. Figure 1 situates data science within the broader framework of related data-centric processes within organizations.

At the top of this framework is data-driven decision-making (DDD)—the practice of making choices based on data analysis rather than relying solely on intuition. For example, a marketer might choose advertisements based on years of experience and gut instinct, or they might base their decisions on analytical insights derived from data about consumer responses to various ads. Many organizations blend these approaches. DDD is not an all-or-nothing proposition; different companies adopt it to varying degrees. The advantages of data-driven decision-making have been firmly established through research. Economist Erik Brynjolfsson and colleagues at MIT and Wharton developed a metric to quantify how deeply companies rely on data in their decision processes. Their findings show that more data-intensive firms tend to be more productive—controlling for other factors—with a one-standard deviation increase on the DDD scale linked to a 4-6% boost in efficiency. Additionally, higher DDD scores correlate with better return on assets, return on equity, asset utilization, and market value—relationships that appear to be causal.

Our two example case studies illustrate different types of decisions: (1) those requiring the discovery of new insights within data, and (2) those involving repetitive decisions at large scale, where even small improvements in accuracy can lead to significant benefits. The Wal-Mart example demonstrates a type-1 problem, where Linda Dillman seeks to uncover knowledge to help prepare for Hurricane Frances. Conversely, the customer churn example illustrates a type-2 DDD problem, where a telecom company manages hundreds of millions of customers, many of whom are nearing contract expiration. By better predicting which customers are likely to leave and estimating the profitability of targeting each, the company can reap substantial benefits across its large customer base.

This logic applies broadly to many fields where data science and mining are heavily utilized—such as direct marketing, online advertising, credit scoring, financial trading, help desk management, fraud detection, search ranking, and product recommendations. The diagram in Figure 1 shows how data science supports DDD but also overlaps with it. Increasingly, business decisions are being automated by computer systems. Different industries have adopted such automation at varying speeds. Finance and telecoms were early adopters; in the 1990s, these sectors saw a transformation in banking, consumer credit, and fraud management through automated decision systems. Retailers also began automating merchandising decisions as their systems digitized. Notable examples include Harrah's Casino's loyalty programs and the automated recommendations from Amazon and Netflix. Today, we are witnessing a revolution in advertising, driven by the explosion of online activity and the capacity to make rapid, split-second decisions about which ads to serve**.**



Data Handling and the Rise of Big Data

Despite the impression often created by the media, much of data processing is not part of data science itself. Data engineering and processing are vital for supporting data science activities, as shown in Figure 1, but they serve broader purposes as well. These data processing technologies are essential for many business functions that do not directly involve extracting insights or making decisions based on data—for example, efficient transaction processing, modern web system operations, online advertising management, and more.

"Big data" technologies—such as Hadoop, HBase, CouchDB, and others—have gained significant media attention recently. For our purposes, we can define big data as datasets that are too large for traditional data processing systems, thus necessitating new, specialized technologies. Like traditional methods, big data tools are used for various tasks, including data engineering. Sometimes, these technologies are employed directly to implement data mining techniques; more often, they are primarily used for processing data in support of data mining and other data science activities, as depicted in Figure 1.

Economist Prasanna Tambe from NYU's Stern School has studied how the adoption of big data technologies impacts firm performance. His research indicates that, after accounting for various influencing factors, firms that utilize big data technologies experience significant productivity gains. Specifically, a one-standard deviation increase in the use of big data tools correlates with a 1–3% higher productivity compared to the average firm, while a lower level of utilization—one standard deviation below the mean—is associated with a 1–3% decrease in productivity. These differences can translate into sizable productivity gaps between firms at the extremes of big data adoption.

Thinking with Data

One of the most vital aspects of data science is fostering data-analytic thinking. The ability to think analytically with data is crucial not only for data scientists but throughout the entire organization. For instance, managers and employees in various departments will optimize the use of a company's data science resources only if they possess a basic understanding of core principles. Even in organizations with limited data science expertise, managers should understand fundamental concepts to effectively collaborate with consultants. Investors in data-driven ventures also need to grasp these principles to evaluate opportunities accurately.

More broadly, as businesses become increasingly driven by data analytics, having the capacity to interact knowledgeably within such environments offers a significant professional advantage. Understanding key concepts and having frameworks for structuring data-analytic thinking enables individuals to collaborate effectively, identify opportunities for enhancing data-driven decision making, or recognize potential competitive threats related to data. Many traditional industries are leveraging new and existing data resources to gain a competitive edge, employing data science teams to develop technologies that boost revenue and reduce costs. Additionally, numerous new companies built around data mining are emerging as strategic players—Facebook, Twitter, and other "Digital 100" firms, for example, hold high valuations mainly because of their valuable data assets.

As data analytics becomes more integral, managers are increasingly tasked with overseeing analytics teams and projects, marketers need to understand data-driven campaigns, venture capitalists must evaluate businesses with substantial data assets, and strategists need to craft plans that capitalize on data. For example, if a consultant recommends a data-driven initiative to improve your business, you should be able to critically assess whether the proposal makes sense. If a competitor announces a new data partnership, recognizing it as a potential strategic disadvantage is important. Or, suppose you work at a venture firm evaluating an advertising startup, which claims that a unique set of data will generate significant value—can you judge whether their valuation makes sense? With a solid foundation in data science principles, you would be equipped with the right questions to probe these claims.

On a more practical level, data-analytics projects often span all business units. Employees across departments need to interact with data science teams; without a fundamental understanding of data-analytic thinking, they might not fully comprehend how these initiatives impact the business. This misunderstanding can be particularly harmful in data science projects because they directly support better decision making. Close collaboration between data scientists and business leaders is essential; organizations where business personnel lack this understanding risk wasting resources or making poorly informed decisions. A recent Harvard Business Review article highlights this challenge, noting that despite the promises of high ROI from Big Data, many companies find their analytics efforts ineffective—or even detrimental—unless employees are equipped to incorporate data insights into complex decision processes.

II. CONCLUSION

Beneath the wide array of techniques used for data mining lies a smaller set of fundamental concepts that form the core of data science. For the field of data science to develop and grow beyond the hype and fleeting trends, it's essential to look beyond the algorithms, methods, and tools commonly in use. We need to focus on the foundational principles and ideas that underpin these techniques, as well as the systematic way of thinking that drives successful data-driven decision making. These core concepts are broadly applicable across many areas. Excelling in today's data-centric business landscape involves understanding how these fundamental ideas relate to specific business challenges—essentially, thinking data-analytically.

This approach is supported by conceptual frameworks inherent to data science. For example, the process of automatically extracting patterns from data is well-structured, with clearly defined stages. Grasping these stages helps organize problem-solving in a systematic way, reducing the likelihood of errors. There is strong evidence that performance can be significantly enhanced through data-driven decision making, the use of big data technologies, and data science techniques leveraging large datasets. Data science underpins data-driven decisions—sometimes even enabling automatic decision-making at enormous scales—and relies on tools for storing and engineering big data. Nonetheless, understanding and explicitly discussing the core principles of data science is vital for unlocking its full potential.

REFERENCES

[1]. Davenport T.H., and Patil D.J. Data scientist: the sexiestjob 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 innumbers: How does data-driven decision making affectfirm performance? Working paper, 2011. SSRN workingpaper. 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 valuablestartups. Bus Insider. Sep. 23, 2010.
[6]. Shah S., Horne A., and Capella´ J. Good data won’tguarantee good decisions. Harv Bus Rev, Apr 2012.
[7]. Wirth, R., and Hipp, J. CRISP-DM: Towards a standard process model for data mining. In Proceedings ofthe 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining,2000, pp. 29–39.