

喜田昌樹 著

『ビジネス・データサイエンス入門』

白桃書房

2023年、320pp.

柴田淳郎

Atsuro Shibata

滋賀大学 経済学部 / 准教授

近年、ビジネス・データサイエンスと呼ばれる研究分野がアカデミアだけでなく、一般の企業社会においても注目を集めている。そもそもデータサイエンスという概念自体は、1990年代に普及した比較的新しい概念であるが、その基礎となる学問領域の広がりには大きく、その歴史は極めて長い。データサイエンスの学的系譜を辿れば、その根は哲学、言語学、倫理学、歴史学、文献学、認知科学の源流である心理学等、伝統的なリベラルアーツと呼ばれる学問領域にまで及ぶ。近年になって、データサイエンスが注目を集めるようになった理由は、IT技術が進歩し、コンピューターサイエンス、認知科学、人工知能を代表とする機械学習のような先端テクノロジーをその領域に含むようになったからである。データサイエンスが文理融合の先端領域のひとつと言われるのは、このような文系・理系を横断する学問的領域の広がりにある。

本書によると、データサイエンスとは、「データを利活用するための諸領域」とであると定義される。そしてデータサイエンスは、主として5つの領域から構成される。①データマネジメント、②データクリーニング、③データビジュアライゼーション、④統計分析、⑤モデリングである。①データマネジメント及び②データクリーニングは、分析データを作成する領域であり、③データビジュアライゼーションは、作成されたデータ全体を把握する領域である。グラフ化理論などがこれに相当する。④

統計分析は、従来の統計学の議論を中心とする領域である。⑤モデリングは、機械学習、各種アルゴリズム、人工知能などの領域で構成される。

それでは、そもそも本書の主題であるビジネス・データサイエンスとはどのような学問領域なのだろうか。本書では、まず、ビジネス・データサイエンスを「仕事にデータサイエンスを活かすこと」と定義する。その上で、どのようにデータサイエンスを仕事に活かすかについて、CRISP_DM (Cross-Industry Standard Process for Data Mining) を台本としつつ、その意義を論じている。

CRISP_DMとは、1996年にクライスラー社、SPSS社、NCR社を代表する研究者たちにより、ビジネス向けにデータマイニング及びデータサイエンスを活用するために特別に開発されたプロセスモデルである。CRISP_DMのプロセスは以下の5つの段階にモデル化される。第1段階 ビジネスの理解、第2段階 データ理解、第3段階 データ準備、第4段階 モデリング、第5段階 評価、第6段階 展開/共有である。ビジネスにデータサイエンスを応用する際に、これらすべての段階が重要であるが、本書は特に、第1段階ビジネスの理解及び第6段階 展開/共有の段階が重要であると指摘する。なぜなら、データサイエンスをビジネスに活用するためには、経営課題が適切に設定され（第1段階 ビジネスの理解）、第2から第5のデータサイエンスの段階を経て、最終的に付加価値化

(第6段階 展開/共有)を実現しなければならないからである。

本書は、「経験の浅いデータサイエンティストが犯しがちな失敗は、CRISP_DMにおけるモデリングに多大な労力を費やし、その他の段階を軽視することである」と指摘する。もちろん、経営の現場におけるビジネス・データサイエンスの成果物はモデルであり、モデリングに多大な労力を費やすこと自体は悪いことではない。しかし、多大な労力が費やされて完成したモデル自体が適切に経営課題を解決し、付加価値の創造(付加価値化)を実現できなければ、データサイエンスを企業経営に導入する意味がない。本書では、「熟練のデータサイエンティストは、当該のプロジェクトの焦点が定まっているか、適切なデータが揃っているかなどの確認に、より多くの時間を割いている」という指摘からもわかるように、データサイエンスを実際に職場で活用する際に、マネジャーが留意すべき事項が随所にちりばめられている。この意味でも、本書はデータサイエンスのテキストであると共に、それをビジネスに利活用する領域であるビジネス・データサイエンスのテキストでもあると言える。

本書は、上述したビジネス・データサイエンスの各段階に沿って構成されている。

第1章は、ビジネス・データサイエンスとは何か、その定義とプロセス及びその概略が示される。本章では、ビジネス・データサイエンスの流れを第1ステージ 経営課題の発見、第2ステージ データの理解、第3ステージ データ準備、第4ステージ モデリング、第5ステージ モデリングの最適化、第6ステージ モデルの評価の6ステージに分けられること、ビジネス・データサイエンスがデータマイニングとテキストマイニングという2つの方法論があること、そしてビジネス・データサイエンスの企業経営での活用領域の大枠(マーケティング管理やリスク管理等)が示されている。

第2章は、ビジネス・データサイエンスの6ステージのうち、第1ステージ経営課題の発見に必要とされる経営学の理論が紹介されている。本書は本ステージこそがビジネス・データサイエンスを有効に活用するための最も重要なステージであると指摘する。つまり、データサイエンスを仕事に活用するためには、まず経営課題を適切に設定しなければならず、そのためには業務知識も必要であるが、それだけでなく、経営学の理論が必要であることが論じられている。本章では経営戦略論、ビジネスシステム論、マーケティング論、ナレッジマネジメント論の基礎理論及び経営手法とその変化が詳述されている。ここで紹介されている経営学の基礎理論は、この後に本書で展開されるビジネス・データサイエンスの様々なモデルを理解する上で前提条件となっている。

第3章では、企業経営にデータサイエンスを活用する、即ちビジネス・データサイエンスを企業で実践する上で前提条件が議論される。本章では、ビジネス・データサイエンスを実践する前提条件として、第1に、BI(Business Intelligence) ツールに関する知識が必要であること、第2に、「基幹システム(データベース)及びERP(Enterprise Resource Planning:データウェアハウス)の構築」、「データマネジメント」、「分析用データの作成」、「Visualization(可視化、もしくはデータ表現)」が不可欠であることが指摘される。さらに本章ではBIツールのひとつであるIBM SPSS Modelerを活用するための基礎知識が解説され、「基幹システム及びERP」及び「データマネジメント」について詳述されている。

第4章では、第3章に引き続き、ビジネス・データサイエンスを実践する上で、不可欠な前提条件である「分析用データの作成」及び「Visualization」について論じられている。ここで

は、データ理解からデータ準備、分析用データの作成までのプロセスが詳細に説明され、最後にその自動化(データ準備の自動化ツール)についても解説されている。

第5章では、各種アルゴリズムを用いたモデリング(モデルの構築)手法が紹介されている。ここで扱われるアルゴリズムには、ニューラルネットワーク、決定木、クラスタ化、アソシエーション・ルールがあり、それぞれのアルゴリズムの利用要件・長所・短所が詳細に解説されると共に、経営課題の発見・データ選択・アルゴリズムの選択の流れ、即ちフレーム化について説明されている。

第6章では、古くからデータサイエンスが有効に活用できる分野のひとつであるリスクマネジメント、特に金融機関で利用される与信管理のモデリングを中心に解説されている。

第7章では、データサイエンスを活用した顧客行動の理解の方法が詳述されている。ここでは、需要予測やチャーン・マネジメントなどの様々なモデル・その自動化の方法が詳細に紹介されている。

第8章では、顧客理解のもうひとつの方法としてマーケットセグメンテーション(市場の細分化)の理論、その分類軸、小売店の購買活動データに基づく分析が示され、個人属性の合成変数の構築方法についても丁寧に説明が展開されている。

第9章では、ビジネスのロングテール現象の対応策のひとつとしてアソシエーション・ルールを活用したマーケット・バスケット分析及びレコメンドシステムについて、その概念だけでなく、分析実践にまで踏み込んだ解説がなされている。

第10章では、テキストマイニングという手法を活用し、非構造化データを対象としたデータサイエンスの基礎的概念及びその活用方法が紹介されている。テキストマイニングとは、簡単に言えば、

「テキストという非構造化データを定量化・構造化するための手法」であり、ここでは言及頻度分析、ウェブ分析などの手法が、データサイエンスの各段階－ i Visualization (可視化)、ii Analysis (分析)、iii Predictive Analysis (Modeling: 予測的分析)－に沿って詳細に解説されている。

データサイエンスは伝統的なリベラルアーツ分野だけでなく、機械学習やAIなどの先端技術分野をも含む広大な文理融合領域である。本書はビジネス分野へのデータサイエンスの利活用徹底的に焦点化し、企業人がデータサイエンスを実際の企業経営の中で実践するための基礎的な概念や具体的な分析実践を豊富に提示し、さらにデータサイエンス業務の自動化という先端的なトピックにまで踏み込んで、それらをわかりやすく解説している。本書はデータサイエンスと経営学という異なった学問領域を適切に架橋することに成功しており、データサイエンスの企業経営への有効活用やリスクリングに挑戦するあらゆるビジネスパーソンにとって最良のテキストである。

