



評者



津田塾大学
学芸学部 教授
西川 賢

データサイエンス

ジョン・D・ケレハー、ブレンダン・ティアニー 著

今野紀雄 監訳／久島聡子 訳

ニュートンプレス (2020年1月) 1,800円＋税／304ページ

5 エクサバイト

冒頭の見出しの数字が何を指すか、お分かりになるだろうか。これは2013年以降、人類が毎日生成・記録し続けているデータ量である(1エクサは 10^{18} ＝100京)。これは文字による記録の発明から03年までの5000年間で蓄積されたデータの総量と等しい。現代に生きるわれわれは、かくも多量のデータを日々生み出し、蓄積し続けているのである。

「データは21世紀の石油」「データサイエニティストは21世紀で最もセクシーな職業」などと言われるように、おびただしく生み出されるデータを分析することの重要性は日増しに高まっている。ビジネスはもとより、政府もデータサイエンスを活用する時代である。スマートシティーや医療分野など、データサイエンスの応用で飛躍的に伸びることが予想されるビジネス領域が本書でも紹介されており、こうした分野のビジネスモデル化の必要性も指摘されている。半面、データを分析できる能力を持つ人材の供給は需要に追いつかず、優秀な人材の育成・確保が急務となっている。

だが、データサイエンスがはやっていて、それが利益を生み出しているからといって、

そう簡単にビジネスにできるというものでもない。データサイエンスは理解しておかなければならない前提知識が多いからである。そもそもデータとは何なのか。データの基本的な特徴や性質にはどのようなものがあるのか。データはどう収集すべきか。そして、どのように分析・処理し、分析結果を評価すべきか。例えば、機械学習を用いてデータを分析する場合、どのようなタスクが実行可能なかといった知識だ。

これに付随して、IT分野特有の身に付けるべきスキルも多い。例えば、データを保存・管理するためのストレージなど各種リソースを提供するクラウドコンピューティングサービスを使いこなす技術、分析のために必要な機械学習の技法とそれを実行するためのプログラミング言語の知識、多量のデータ分析を可能とする分析環境の構築といったものが挙げられるだろう。

デジタル化が進む現代においては、データの収集はプライバシーの侵害につながることも懸念されており、倫理上の問題にも十分配慮しなくてはならない。この点も重要である。決して簡単ではないデータサイエンスだが、挑戦するだけのメリットと可能性を秘めていることも事実である。まずは本書から始めてみてはいかがだろうか。初心者でも無理なく学習を進めることができる。