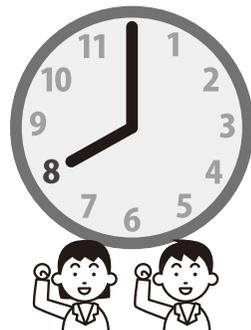


1 時間目 未来を予知するプロダクトを作る

データ・サイエンスの基礎知識



久留 健

[ご購入はこちら](#)

暗号資産(仮想通貨)の基本技術であるブロックチェーンに、NFTと呼ばれるデジタル資産があります。NFTはNon Fungible Token(非代替性トークン)の略称です。2021年に人気が発火し、NFTのアート作品が数千万ドルで売却されたりして注目を浴びました。第4部では、このNFTを例に、データ収集、分析までの一部始終を紹介します。

データ収集には、Web APIを利用するPythonプログラムを使います。データ分析には、データの可視化に特化したBI(ビジネス・インテリジェンス)ツールを使います。(編集部)

データ・サイエンスとは

● 目的…データ・プロダクト作りのための活動

データ・サイエンスは、最終的にデータ・プロダクト(ユーザ・データを活用する製品やサービス)を作ることを目指しています(図1)。

データを使って、未開領域の探索や未来を予想する活動を行います。データを活用したサービスを提供するために、統計や機械学習などの手法を使ってモデルを構築し、コンピュータで独自モデルを処理できるよう、システムを開発します。

代表的な例には、EC(Electronic Commerce)サイトの閲覧履歴や商品履歴から、どのような商品やサービスをレコメンド(お勧め)するかを決めるサービス開発などが挙げられます。消費者がECサイトに滞在する時間をできるだけ引き伸ばして、もっと購入してもらおうとします。

● 目的は異なるけど手段は統計とほぼ同じ

筆者は、データ・サイエンスと統計の違いをよく尋ねられます。日本では、どのような立場で、どのように分析を活用するかによって、統計なのかデータ・サイエンスなのかの議論がよく起きます。

両者は目的が異なり、手段がほぼ同じと言えます。統計は、過去や現在のデータを使って事象を説明するのに主眼が置かれています。データを読み解



図1 データ・サイエンスの目的…データ・プロダクトを作ること

き、モデルを構築して利用するのは、データ・サイエンスと共通した部分になると思います。

● 変化の流れを追い続けることが重要

データ・サイエンスでは、データ・プロダクトを開発するため、複数の分野のスキルが必要です。特にIT技術や機械学習との関わりは強く、日々新技術が取り入れられています。機械学習の論文は、毎日数十本発表されていると言われています。

研究者の間では、新しい価値をいち早く発見するため、新しい手法が開発されればすぐに取り入れられます。新しい手法に乗り換え続けて、より精度の高い探索活動を効率化することが重要です。この変化の流れを追従できないと、データ・サイエンスで何が行われるのか理解しにくいかもしれません。

● 必要とされるスキル領域

一般的にデータ・サイエンスのプロジェクトを推進していくには、次の3つのスキルが必要です。

- ビジネス(経済、ビジネスの仕組み、法律、商習慣など)
- コンピュータ(プログラミング、データ構造、デー