

データ活用人材の育成・教育 ビジネスデータサイエンス人材

An Undergraduate Program based on CRISP-DM for Data Science for Business

共同研究メンバー

○今泉忠*、出原至道*、久保田貴文* (○代表、執筆者)

1. はじめに

IoT の発展や「ビッグデータの爆発」に伴い、企業において事業課題の発見や解決のためにデータを活用することが必然となっている。クラウドコンピューティングや「ビッグデータ」等が身近なものとなり、データから知見を得る職業人としてのデータサイエンティスト、データアナリスト、データアーキテクト、データアドミニストレーター、データエンジニアなどデータ活用に携わる人材などの人材育成が社会や企業から求められている。例えば、Google 社のチーフエコノミストである Hal Varian が、2009 年にインタビューの中で「今後 10 年間でセクシーな職業は統計家である」と述べている（ニューヨークタイムズ、2009）。従来は、各企業内において蓄積した経験知・暗黙知などをもとに意思決定を行うことが多かった。グローバル化に伴い、ビジネスバックグラウンドが異なる場面での意思決定においてはデータが重要視され、そのデータ内容も経験からは想定ができない様相を示すようになってきた。このようなビジネス環境においては、データに基づき思考し意思決定できる人材の育成が急務で、そのような人材の育成無しには、IoT なども単なる機能拡張や改善として扱われ、イノベーションを生み出すソースの 1 つにはならなくなると考えられる。

本報告では、「ビッグデータ」を扱うビジネス環境での人材育成に関して、プロジェクトチームにおける人材育成のフレームとしての CRISP-DM の活用について報告する。

2. 「ビッグデータ」活用

「ビッグデータ」の特徴としての 4V (Volume, Variety, Velocity, Value) が挙げられるが、重要な特徴として「疎」データで非構造化データであり、従来のスモールデータのように構造化され、「リッチ」データではない。この「ビッグデータ」をもとに着想して意思決定するためには、従来とは異なるスキルを持った人材が必要とされる。学術研究のみならず、経営情報学が特に関係するビジネス関連企業でも「ビッグデータ」活用によるイノベーションを実現するための人材育成プログラムを開発している。提言「ビッグデータ時代における統計科学教育・研究の推進について」では、統計学関連教育について、アメリカでの人材教育（近年では修士

* 多摩大学経営情報学部

課程修了者が年間 2000 人程度) や中国での人材教育(統計学関連科目が第 1 級科目に格上げされ、多くの大学に設置されている。学部卒業生数で年間 8000 人程度) や韓国での人材教育(統計学専攻を有する大学学部は 41 学部、大学院数は 24 研究科ある) が報告されている。日本では、1988 年に設置された総合研究大学院大学複合科学研究科統計科学専攻の 1 か所にすぎなく、滋賀大学が 2017 年にデータサイエンス学部(仮称) の設置に向けて教育プログラムの開発や体制整備を開始している状況である。多摩大学も参加している文部科学省大学間連携共同教育推進事業「データに基づく課題解決型人材育成に資する統計教育質保証」や日本統計学会による統計検定試験の実施などのような新しい動きがあるが、近隣諸国に比べて大幅にデータサイエンス関連教育は遅れており、産業界からのニーズにも対応できていない。ビジネス場面での問題解決のための言語はビジネスデータであり、その文法はビジネスデータ分析であると言えよう。ビジネスデータ分析においては、最終的な意思決定は人であることを考えると一層、最終的な活用場面を想定した適切な結果表現になっている必要があり、それを可能とする人材としては、以下のような人材であることが求められる。

- (1) データに基づいて結果までをデザインできる人材：データから情報を価値として創造できることが必須である。そのためには、データを数値として処理するスキルではなく、データの背後に隠れてる関係を構造化し、記述できる能力が必要である。
- (2) 横断的な知識のみならず横断的な知見を有している人材：データ分析を通じて学べることとして横断的な数理的知識が挙げられることが多い。より価値のある問題解決を行えるためには、数理的にものごとを把握する数理的な思考方法の獲得や、ものごとを結びつけて捉え、それからの知見を記述することができる能力が必要である。

学部教育においては、これらの基本部分について学修することが必須であるが、実際のビジネスにおけるデータを扱うことが少ないなどで、分析課題が不明確なままに単なる手法獲得になり、経験的な分析プロセスの獲得となりがちとなるなどの問題点がある。このような不適切な部分を改善するための教育プログラムが必要となる。

3. データに基づいた課題解決のためのフレームワーク

上の 2 であげた問題点の解決のために、データに基づいて課題解決を行うためのフレームワークを用いることが有効である。それは、データ分析は個人の仕事でなく、チームとしての仕事であり、かつ、アジャイル的分析であるからである。このようなフレームワークとしては様々あるが、ここではデータ分析とビジネスデータ分析人材育成プロセスとしての CRISP-DM を取り上げる。CRISP-DM (CRoss-Industry Standard Process for Data Mining) は、データマイニングのための方法論を規定したものの 1 つで、下図 1 のように 6 つのフェーズから構成されている。矢印は、最も重要かつ高い頻度で発生するフェーズ間の依存関係を表している。これらの依存関係は定型的ではなく、必要に応じてフェーズ間を行き来して作業が行われることになる。各フェーズに関与する人材のもつべきスキルが明確になり、各部門のプレイヤーがどのように CRISP-DM でのフェーズと関係しているべきなのかを表現することや乖離なども表現できるので、人材育成での改善要因や教育プログラムの提供が可能となる。CRISP-DM は多くの場面に適用可能であるが、実際のプロジェクトの遂行場面での人材と対応づけて分析することで、人材配置や育成での改善点が発見できる。ここでは、ある企業のあるプロジェク

トでの作業の流れを調べ、CRISP-DMの各フェーズでの決定場面との対応を比較することで改善について検討する。ある企業でのデータ活用プロジェクトの流れと役割を拾い出して表現したものを表1に示す。この表ではビジネスの理解が付与されている点の特徴である。表1の◎でそのフェーズでは関与している人材、○で副次的メンバー人材を、△で情報共有人材を表している。この表からは、このプロジェクトでの作業での責任者がだれであるかが不明確であることが予想される。どの程度標準的なCRISP-DMのフェーズと関係しているかも評価が困難である。そのために、どのようなデータを活用できる人材を配置すべきか、また、育成すべきかが不明瞭である。この表1のデータを数値化（空白は△を1点、○は2点、◎は3点）として、CRISP-DMの構造に適合するように分析した。その分析結果を図2に示す。

図2での円はステークホルダを示し、その半径は果たしている役割の重要度を示す。この結果から読み取れることは、

- (1) 多くの項目に関係しているステークホルダは中央に位置づけられている。このために、中央に配置されているステークホルダには仕事が集中しているとともに、これらの人材が欠けるとプロジェクト遂行が困難になることを示している。
- (2) フェーズの配置は、CRISP-DMの順序と大きく整合していないフェーズがある。これから、作業分担が明確でなく、これらについて検討することが必要であることがわかる。
- (3) (1) と関係しているが、企画ビジネスリーダーと企画ビジネスの負荷が大きいので、何らかの人的サポートが必要であることを示唆している。このように定量化することでプロジェクトマネージャなどの重要度が評価できる。

4. まとめ

CRISP-DMを用いた場合に、実際のプロジェクトでのどのような役割のプレイヤーが関与しているかをまとめ、それを可視化することで、人材の重要性や役割の評価などを行った。今回は指標化であるか、これを指数化することで、プロジェクト遂行のために人材配置などを評価できよう。

表1 ある企業でのプロジェクトを表現したもの

| 立場 | プレイヤー | ビジネスの理解 | データの理解 | データの準備 | モデルの1次評価 | モデル作成 | モデルの評価 |
|--------|------------|---------|--------|--------|----------|-------|--------|
| データ提供先 | キーパーソン | | | ○ | ○ | | ○ |
| データ活用 | 役員 | ◎ | | | ◎ | ○ | ○ |
| | 本部長 | | △ | △ | △ | ◎ | ○ |
| | ビジネスリーダー | ◎ | ◎ | ◎ | ◎ | ◎ | ◎ |
| | ビジネスパーソン | ○ | ○ | ○ | ○ | ○ | ○ |
| IT部門 | キーパーソン | ○ | ◎ | △ | | | |
| | ITリーダー | ○ | ◎ | △ | | | |
| | ITパーソン | | ◎ | ◎ | | | |
| | セールスパerson | | ○ | | | | |

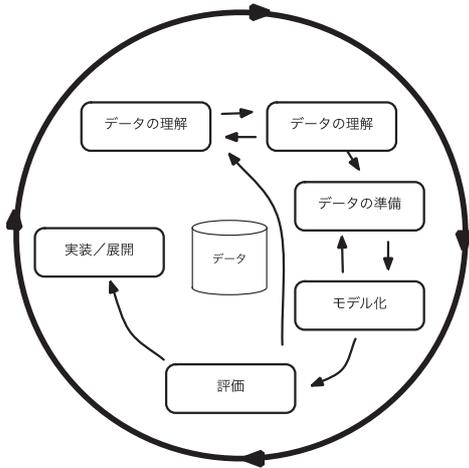


図 1 : CRIP-DM でのフェーズ

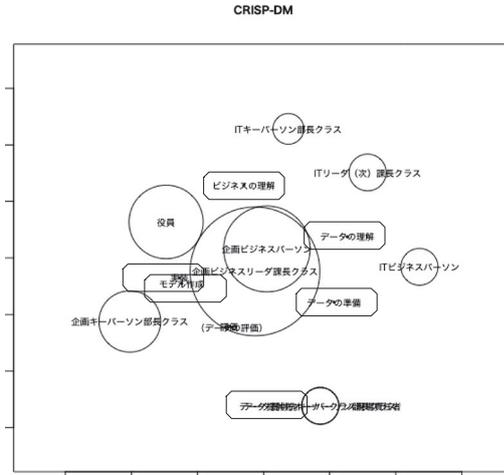


図 2 : 実際の評価データを分析した結果

参考文献

日本学術会議 数理科学委員会数理統計学統計分科会 提言「ビッグデータ時代における統計科学教育・研究の推進について」、2014年8月20日

Steve, Lohr, For Today's Graduate, Just One Word: Statistics,

<http://www.nytimes.com/2009/08/06/technology/06stats.html>