

# データソン(データ分析ハッカソン)の設計・実施

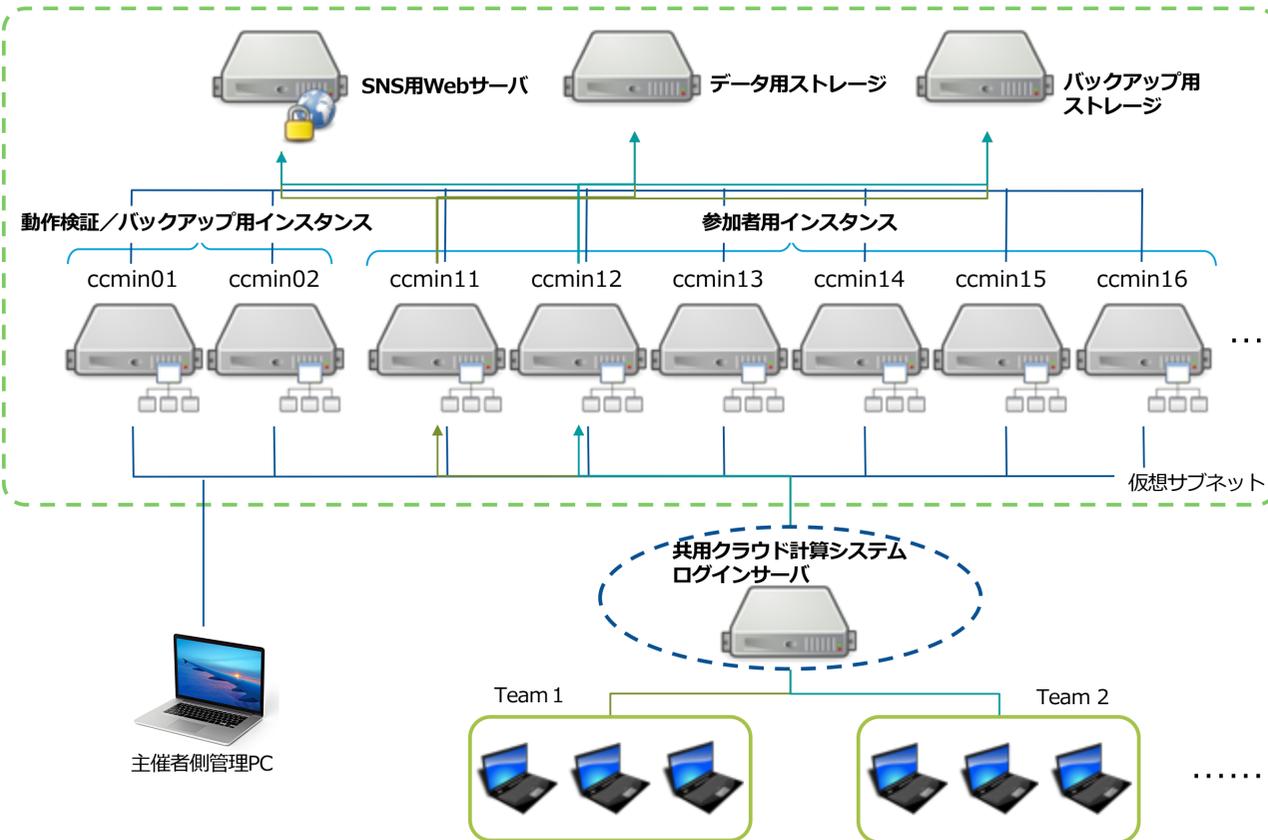
神谷 直樹 統計思考院 特任准教授

## 【はじめに】

データサイエンティストには、データサイエンス力(統計学や統計的機械学習など)、データエンジニアリング力(情報学やソフトウェア工学など)、ビジネス力(ビジネスを理解し推進するスキル)が必要である。これらのスキルを有機的に獲得していくためにはアクティブ・ラーニング型の育成が不可欠で[1,2,3]、例えばデータ分析ハッカソンは、新しい職種としてのデータサイエンティストが注目されるようになった後、インダストリー、アカデミア、非営利組織などにおいて様々なテーマで頻りに実施されるようになった。統計数理研究所では、データ分析ハッカソンを2016年から実施してきた[4,5,6]。

## 【データサイエンティスト育成のためのハッカソン】

データサイエンティスト育成を目的としたハッカソンには、少なくとも以下のような要件があると考えられる。(1)制限されたデータ分析・開発環境で実施する。データサイエンス力やデータエンジニアリング力は分析・開発環境のハードウェア性能に影響されることが容易に想像されるが、学習ツールとしてのハッカソンに注目が集まっている中、データサイエンス力やデータエンジニアリング力のどちらでもない要因が成績に影響しない環境を用意することが必要である。(2)データサイエンスやデータエンジニアリングの最新技術を提供する。技術革新は日進月歩の速さで進んでおり、参加を通じて最新技術が使えるようになることが必要である。(3)国内ではデータやモデルについて制度上の整備が始まったばかりで[7,8]、提供データなどに関して一定の配慮が必要である。



2018年度に実施したデータ分析ハッカソンのネットワーク構成[6]【転載許諾済】

## 【第1回データ分析ハッカソン】

課題: あるエンターテインメント系企業から提供していただいたおよそ1.5年分、合計約1千万レコードのデータから、売上向上の施策を提案すること

結果<sup>注</sup>: 最優秀賞(大学院生、学部生、学部生)、優秀賞(大学院生、大学院生、大学院生)、審査員特別賞(学部生、学部生、学部生)

## 【第2回データ分析ハッカソン】

課題: 株式会社LIFULLから提供していただいた日本全国の賃貸物件データ(1年分、賃料、面積、立地、築年数、間取り、建物構造、諸設備などのデータ)から新しい不動産テック、あるいは現状の不動産テックにおける課題を解決する方法を提案すること

結果<sup>注</sup>: 最優秀賞(社会人、社会人、社会人)、優秀賞(社会人、社会人、社会人)、審査員特別賞(大学院生、大学院生、大学院生)

## 【第3回データ分析ハッカソン】

課題: 株式会社インサイトテックが運営するWebサービス「不満買取センター」に一般ユーザが投稿した様々な不満に関するデータ(約2年3ヶ月分、約130万件の本文とカテゴリデータ、約7万人分の投稿者プロフィール情報も付随)、株式会社日本データ取引所が提供しているデータカタログ内のデータから、データ分析に基づいて、多くのステークホルダーが利益(金銭に限らない)を得られる(相利共生)方法を提案すること

結果<sup>注</sup>: 最優秀賞(社会人、社会人、社会人)、優秀賞(社会人、社会人、社会人)、審査員特別賞(社会人、社会人、社会人)

## 【文献】

注) 赤字は女性、黒字は男性、緑色の下線はチーム内で見たときに所属機関が異なることを表す。

[1] Jonassen, D. (1999). Designing constructivist learning environment. In C. M. Reigeluth (Ed.), *Instructional-Design Theories and Models: A New Paradigm of Instructional Theory, Volume II* (pp. 215-239). Mahwah, NJ: Lawrence Erlbaum Associates.

[2] Reigeluth, C. M., Beatty, B. J., & Myers, R. D. (2017). *Instructional-Design Theories and Models: The Learner-Centered Paradigm of Education, Volume IV*. New York, NY: Routledge.

[3] Briscoe, G., & Mulligan, C. (2014). Digital Innovation: The Hackathon Phenomenon. Retrieved from <http://qmro.qmul.ac.uk/xmlui/handle/123456789/11418>

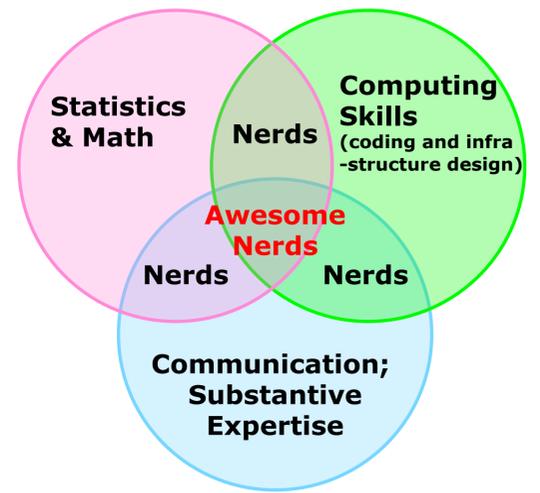
[4] 丸山宏・神谷直樹・宮園法明 (2016). クラウド環境を利用したデータ分析ハッカソンの計画と実施 *Estrela*, 272, 30-37.

[5] 神谷直樹 (2019). データ指向キャリアへの効率的支援プログラムとしてのデータ分析ハッカソンの設計・実施 統計数理研究所公募型共同利用重点型研究「データサイエンス人材育成メソッドの新展開」研究集会.

[6] 神谷直樹・宮園法明 (2019). データソン: 統計数理研究所におけるデータ分析ハッカソン. *日本ソーシャルデータサイエンス論文誌*, 3(1), 2-7.

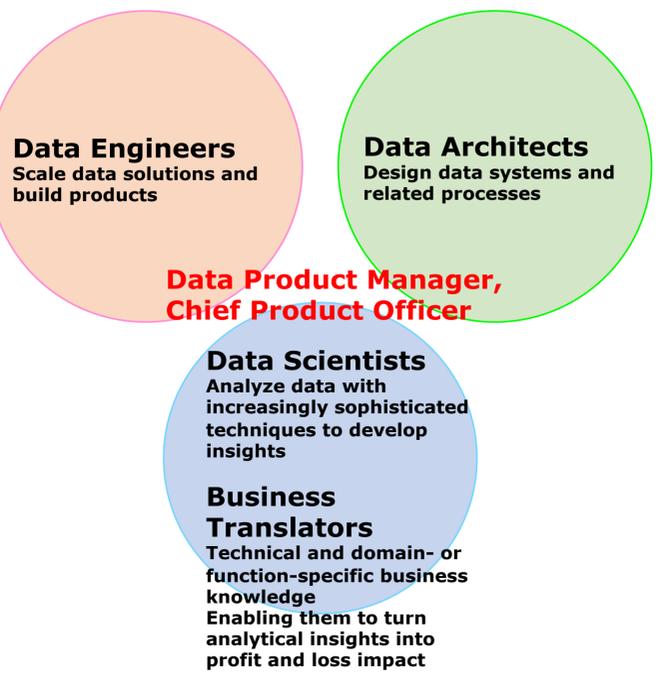
[7] 総務省 (2016). 官民データ活用推進基本法 電子政府の総合窓口 Retrieved from [http://elaws.e-gov.go.jp/search/elawsSearch/elaws\\_search/lsg0500/detail?lawId=428AC1000000103](http://elaws.e-gov.go.jp/search/elawsSearch/elaws_search/lsg0500/detail?lawId=428AC1000000103) (2019年3月29日)

[8] 経済産業省 (2018). 「AI・データの利用に関する契約ガイドライン」. Retrieved from <http://www.meti.go.jp/press/2018/06/20180615001/20180615001.html> (2019年3月29日)



Data Scientists combine several different skills in one profession

参考: Patil and Mason, *Data Driven: Creating a Data Culture*, California: O'Reilly Media, 2015など



データ分析ハッカソンが目指す人材育成像



<https://www.ism.ac.jp/shikoin/training/dac/index.html>