

データサイエンスコンペティションの歩み

—専門知識とデータ科学をつなぐ実践教育—

早稲田大学データ科学センター 野村 亮

1. はじめに

早稲田大学データ科学センターでは、文系・理系を問わず、学生が実際のデータを用いてデータ科学を実践的に学ぶ機会として、2019年度から毎年、データサイエンスコンペティションを開催している。2025年度には第7回を迎え、毎年100名を超える学生がエントリーする全学的な行事へと成長した。

本コンペティションの目的は、単に分析精度や高度な手法を競うことではない。学生が自ら問題を設定し、必要なデータを選び、分析結果を検証・考察し、他者に説明するまでの一連の過程を経験することにある。本稿では、これまでの7回を振り返りながら、コンペティションがどのように発展し、早稲田大学のデータ科学教育に何をもたらしてきたかを考えたい。

2. コンペティションの出発点

第1回は2019年7月、参議院議員選挙の結果予測を課題として、データ科学総合研究教育センター（当時）と政治経済学術院の共催により開催された。参加チームは、政党の得票率や候補者の当落を予測し、実際の選挙結果との差だけでなく、予測モデルの内容についても評価を受けた。現実の出来事を対象とし、結果が明らかになる前に予測を提出するという緊張感のある課題であった。

この第1回には、その後のコンペティションにも引き継がれる二つの特徴が既に表れていた。一つは、現実の課題と実際のデータを用いることである。もう一つは、予測精度や当落予測の正否といった数値上の成績だけでなく、どのような問いを立て、なぜそのデータと分析手法を選び、結果をどう解釈したかを重視することである。

3. データと問題設定の広がり

第2回（2020年度）では、商学学術院との共催により、株式会社ADKマーケティング・ソリューションズから提供された生活者総合調査等のマーケティングデータを用いた。各チームは共通の企業データを起点としながら、自由にテーマを設定して分析と発表を行った。新型コロナウイルス感染症の影響

を受けてオンライン開催となったが、企業が保有する実データに学生が向き合う産学連携型の取り組みとして、新たな展開を示した。

第3回（2021年度）は人間科学学術院との共催で、「SDGs×データサイエンス」をテーマに開催された。学生はSDGsの17目標から自らの関心領域を選び、オープンデータを組み合わせて多面的に分析した。47チーム149名がエントリーし、当日は約120名が参加した。テーマは感染症、食品ロス、子どもの成長、ジェンダー、気候変動、情報の偏りなど多岐にわたり、同じ大きな社会課題に対しても、専門や関心によって異なる問いが生まれることを示した。

第4回（2022年度）は教育・総合科学学術院との共催で、みずほ銀行が提供する統計データサービス「Mizuho Insight Portal (Mi-Pot)」を利用し、各種金融取引等に関する統計データを分析対象とした。37チームが発表会に参加し、防災、公共政策、地域分析、居住地の推薦など、金融データの一般的な用途を越える多様な提案が行われた。オンライン開催でありながら120名以上が参加し、回を重ねるごとに分析と発表の水準が向上していることも確認された。

第5回（2023年度）は理工学術院との共催で、日本経済新聞社の36か月分の新聞記事データを共通の分析対象とした。約30チームがエントリーし、第1回以来4年ぶりに対面で開催された。当日は学生、審査員、関係者等を含む約130名が来場した。ポスター発表やコーヒブレークを通じて、学生同士のほか、学生と審査員との間でも直接議論が交わされた。こうした交流は、他者の分析から学ぶというコンペティションの側面を改めて際立たせた。

第6回（2024年度）は、早稲田大学データ科学センターの単独主催により、「早稲田大学で過ごす中で抱いた興味や問題意識へのデータ科学的アプローチ」を共通テーマとして開催された。学生は、分析対象とするオープンデータを自ら選択した。さらに第7回（2025年度）では、独立行政法人統計センターおよび総務省統計局の後援を受け、統計センターによる国勢調査データの作成に関する講演も実施された。いずれもエントリー数は100名を超えており、第7回の会場には学生を含めて当日約120名が来場した。

こうして振り返ると、初期のコンペティションでは共通課題や企業提供データを分析することが中心であったのに対し、近年は、学生が自ら問いを立て、適切な公開データを探し、分析プロセス全体を設計する方向へと比重が移っている。これは、データを「与えられたもの」として扱う段階から、目的に応じてデータを選び、その限界も含めて考える段階への発展といえる。

4. 専門知識とデータ科学の「二刀流」

データサイエンスという言葉から、理工系の学生が中心の催しを想像するかもしれない。しかし、実際には人文社会系を含む多様な学部・研究科から学生が参加してきた。例えば第7回（2025年度）では、エントリーした学生のうち、理工系の学部および大学院に所属する学生はおよそ16%であった。

学生の発表で特に興味深いのは、所属する学術分野で学んだ専門知識が問題設定に反映される点である。公共経済学を学ぶ学生は政策課題をデータによって検証し、推薦システムを研究する学生はその知見を応用して新たなサービスを提案する。分析対象に関する知識があるからこそ、データから意味のある問いを引き出し、結果を慎重に解釈できる。

同じデータを用いても、チームによって着目点も結論も異なる。他チームの発表を聞いた学生からは、異なる視点からの分析に刺激を受けたという声が聞かれる。総合大学である早稲田大学において全学的なコンペティションを開催する意義は、データ科学を一つの専門領域に閉じず、多様な専門知が交わる共通言語として機能させることにある。

5. 分析手法の先にある学び

ソフトウェアの高度化や生成AIの普及など、分析環境の発展によって、一定の知識があれば高度な手法を試すことは以前より容易になった。一方、複雑な手法を用いること自体が、分析の価値を保証するわけではない。目的や問題設定に合わない手法を用いれば、もっともらしく見える誤った結論を導くおそれがある。

コンペティションで重視してきたのは、課題設定、データ選定、分析手法、得られた知見、プレゼンテーションを総合的に評価することである。基礎的な分析であっても、問いとの対応が明確で、データの性質と限界を理解し、結果を適切に説明できることが重要である。第7回の閉会挨拶でも、松嶋敏泰センター所長から、データを揃えることの難しさに気づくこと、そして「本当にやりたい目的のためにどう分析すべきか」を考えることの重要性が改めて示された。

学生は数か月にわたり、チームで役割を分担しながら、データの収集・加工、分析、検証、資料作成、発表、質疑応答に取り組む。うまくいかなかった分析を見直し、他者の質問を受けて自分たちの前提を再検討する経験も含め、通常の講義だけでは得にくい実践的な学びとなっている。

6. 今後に向けて

7 回の開催を通じて、コンペティションは早稲田大学のデータ科学教育における実践の場として定着した。一方で、今後取り組むべき課題がある。第一は、参加者の裾野をさらに広げることである。高度な手法が、初学者にとって参加の壁とならないようにする必要がある。データ科学教育プログラムの履修者が、学んだ知識を試す次の一步として参加できる導線と支援が必要である。

第二は、専門知識とデータ科学の双方をさらに深めることである。コンペティションをきっかけに生まれた問いや分析を、卒業論文、修士論文、共同研究などへ発展させることもできる。所属分野の専門性とデータ科学の高度な実践力を併せ持つ「二刀流」の人材を育てるためには、コンペティションを一過性の行事に終わらせず、教育科目や研究プログラム、教材と有機的に接続させることが重要である。

その基盤の一つが、センターが継続的に開発しているデータ科学教育プログラムと、その内容に即して刊行を進めているテキストシリーズ「ライブラリ データ科学」である[1]。本シリーズは、記述統計・推測統計・ベイズ統計といった異なる立場からの問題設定を、意思決定という統一的な視点から体系化するもので、『データ科学入門 I～IV』が既に刊行されている。今後は「回帰と分類のデータ科学」「時系列構造のデータ科学」など、コンペティションで学生が実際に用いる分析手法に対応する巻も順次刊行される予定であり、課題への取り組みを通じて生じた関心を、体系的な学びへと接続する足がかりとなることが期待される。

さらに、より研究志向の強い学生に対しては、大学院生を対象とする「データ科学研究力養成プログラム」が受け皿となる。本プログラムでは、参加者が所属研究科での研究とは別に、センター教員の指導のもとでデータ科学そのものに関する研究演習や論文作成に取り組む。参加者はデータ科学の研究者を志す者に限られず、データ科学の研究力を各自の専門分野で活用することが期待されている。コンペティションで芽生えた問題意識を、こうした教育科目・教材・研究プログラムへと段階的につなぐ導線を今後さらに整えていきたい。

第三は、学内外の交流をさらに推進することである。企業や公的機関からのデータ提供、実務家による審査と助言、異なる学術院の学生同士の交流は、分析を現実の課題と結びつけるうえで大きな役割を果たしてきた。特に第 7 回で実現した統計センターとの官学連携は、公的統計に関する専門的知見と大学の教育活動を結びつけるものであり、学生がデータの来歴や特性を踏まえて分析することの重要性を学ぶ貴重な機会となった。また、総務省統計局からも後援を受け、公的機関の支援・協力のもとで実践的な

教育を展開することができた。データの適切な利用と安全管理に留意しつつ、産学官の連携を通じて、学生が社会の課題に向き合う機会を今後も広げていく必要がある。

7. おわりに

データサイエンスコンペティションは、順位を競うだけの場ではない。学生が自ら問いを立て、データと向き合い、専門知識と分析手法を結びつけ、得られた結果を他者と議論する学修の場である。第1回から第7回まで、扱うデータや開催形式は変化してきたが、この基本的な目的は一貫している。

今後も、データ科学を一部の専門家だけのものとせず、多様な学問分野と社会の課題をつなぐための力として育てていきたい。コンペティションが、学生一人ひとりにとって、データによって自らの問題意識を確かめ、新しい知見を生み出す出発点となることを期待している。

参考文献

[1] 松嶋敏泰 監修，早稲田大学データ科学教育チーム 著：『データ科学入門Ⅰ～Ⅳ』，サイエンス社。