

データ科学

Vol.3

目次

特集：

- 1 データ科学教育コンテンツと認定制度の他大学提供
～大学の垣根を超えたデータ科学人材の育成～
- 2 日立ソリューションズ・データ科学センター共催セミナー
AI とセキュリティの交差点 ～技術革新の中の安全確保～
- 3 産学連携データ科学人材育成フォーラム
～長期インターンシップによるキャリア形成支援の取り組み～

2024 年 7 月

データ科学教育コンテンツと認定制度の他大学提供

～大学の垣根を超えたデータ科学人材の育成～

早稲田大学データ科学センターではデータ科学教育プログラムの開発に力をいれてきた[1][2]。開発したデータ科学教育プログラムには2つの大きな特徴がある。ひとつは、「統計学」「機械学習」「人工知能」といった従来型の科目立てを避け、データ科学を「データを用いた科学的な意思決定のための科学」として再定義し、様々な分析手法を捉え直し、分類、整理した独自内容のプログラムとなっている点である。これにより、一般的には微妙に異なる立場からデータを扱う、統計学や機械学習、人工知能等におけるそれぞれの分析方法を俯瞰的、総合的に捉え、クォーター4科目(データ科学入門 $\alpha \sim \gamma$)でこれらを一通り学ぶことができるように開発されている。もうひとつの特徴は、ほぼ全ての科目をフルオンデマンド科目として開発した点である。これらフルオンデマンド科目は、シラバスから各講義の動画スライド、練習問題など全てのコンテンツを複数教員によるプロジェクト形式で作成している。これにより、担当する教員による内容やレベルの偏りをなくし、カリキュラム全体で整合性のあるコンテンツが作成されている[3]。また、これら開発した教育プログラムの履修を促すとともに、履修学生のモチベーションを向上させるため、プログラムの履修状況に応じて認定を与える「データ科学認定制度」を設置した[2]。2023年度におけるデータ科学教育プログラムにおける履修者は延べ16,697名、これまでのデータ科学認定制度の認定者は延べ2,606名に上り、順調に履修者数、認定者数を増やしている。

また、作成した教育コンテンツは、フルオンデマンドコンテンツである特性を最大限活かし、学内の教員による反転授業用の教材として利用できる仕組み(統計教材モジュール[3])や、正規授業の履修に拘らない大学院生向けの自学自習用のコンテンツ、社会人向け履修証明プログラムの一部コンテンツなど、様々な活用を行っている[4]。データ科学教育コンテンツは多種多様な学部・研究科により構成される早稲田大学において、それぞれの学生に必要なタイミングでデータ科学教育を行うための重要な資源となっている。

他方、2017年に文部科学省が指定した数理・データサイエンス・AI教育強化拠点校6校により設置され、現在多くの大学が参画する「数理・データサイエンス・AI教育強化拠点コンソーシアム」を中心に、現在、全国の大学におけるデータ科学教育の普及・推進が急速に進んでいる[5]。その結果、大学の規模や文理を問わず、データ科学教育の必修化やそれに準ずる科目の設置が進んでいる。しかし、主に人的リソースの制約により、全ての大学が体系的なプログラムを設置できているとは限らないのが現状である。

この状況を踏まえ、早稲田大学は2024年秋学期より、作成したコンテンツのフルオンデマンドという特性を活かし、本学のデータ科学教育プログラムを東京女子大学へ提供する試みを開始する。この

東京女子大学へのデータ科学教育スキーム「初期」提供イメージ



© 2023 Waseda University.

この試みが成功すれば、東京女子大学以外の他大学への教育プログラムの展開も可能となる。特に、人的リソースの制約から十分なデータ科学教育を提供できていない中小規模の大学にとって有効な選択肢となりうる。また、長期的にはこのようなフルオンデマンドコンテンツを活用した大学間連携モデルが、日本の高等教育におけるデータ科学教育の質と量の向上に寄与することが期待される。本学に限らず、各大学の強みを活かしつつ、教育コンテンツやリソースを共有することで、より効果的かつ効率的な教育体制を構築できる可能性があると考えている。

一方で、現段階では課題も多く存在する。異なる大学文化や教育方針の調整、成績評価の基準統一、著作権や知的財産権の取り扱いなど、検討すべき点は多い。これらの課題を一つずつ解決しながら、モデルケースとして成功させることが重要であると考えている。今後はこの取り組みの成果を慎重に評価し改善点を洗い出していく必要がある。また、他大学や産業界との対話を通じて、社会のニーズに合致したプログラムへと発展させていくことも求められる。

データ科学の重要性が増す中、早稲田大学データ科学センターでは、これからも前述のような先進的な取り組みを推進し、日本におけるデータ科学教育、高等教育の新たな可能性を探っていきたい。

参考文献

- [1] 須子統太, 小林学, 堀井俊佑, 安田豪毅, 松嶋敏泰, ” 早稲田大学におけるデータ科学教育の取り組み, ” 日本経営工学会秋季大会予稿集, 2018.
- [2] 野村亮, 小林学, 堀井俊佑, ” 早稲田大学における全学データ科学教育プログラム, ” 第 13 回横幹連合コンファレンス予稿集, 2022.
- [3] 堀井俊佑, 石井雄隆, 須子統太, ” モジュール化されたオンライン統計教材を利用した効率的に学習可能なシステムの構築, ” 第 16 回統計教育の方法論ワークショップ, 2019.
- [4] 堀井俊佑, 野村亮, 須子統太, ” データ科学教育用フルオンデマンドコンテンツの作成と多目的利用, ” 第 13 回横幹連合コンファレンス予稿集, 2022.
- [5] <http://www.mi.u-tokyo.ac.jp/consortium/>

(文責：データ科学センター教務主任 須子統太)

【東京女子大学への提供プログラムの概要】

プログラム名称：データ科学教育スキーム

期間：2024 年度秋学期～

内容：

- ① 早稲田大学データ科学教育プログラムのうち 8 科目の教育コンテンツ
- ② 早稲田大学データ科学 LA による学習支援
- ③ 早稲田大学データ科学オープン認定制度（ノンディグリー）

対象者数：年間 100 名程度

日立ソリューションズ・データ科学センター共催セミナー

AI とセキュリティの交差点

～技術革新の中の安全確保～

1. 飛躍的な進化を遂げる AI 技術 セキュリティの確保は重要な課題に

2024 年 3 月 6 日、早稲田キャンパスにおいて、株式会社日立ソリューションズ・データ科学センター共催のセミナー「AI とセキュリティの交差点 ～技術革新の中の安全確保～」を開催した。日立ソリューションズと早稲田大学は 2020 年 2 月に学術交流協定を締結。同社の寄附により、グローバルエデュケーションセンター設置の提携講座「未来社会を創るセキュリティ最前線」を実施した。

本セミナーには、日立ソリューションズからセキュリティエバンジェリストの扇健一氏とホワイトハッカーの米光一也氏、データ科学センターからは小林学教授と中原悠太講師が登壇。セミナーの前半は小林教授と中原講師による AI の技術動向やセキュリティリスクの事例紹介があり、後半はその内容を受けて 4 名によるパネルディスカッションが行われた。

開会に際し、司会を務めるデータ科学センターの須子統太教務主任から本セミナーの趣旨説明があり、「データ科学とセキュリティは切っても切り離せない関係。データをビジネスの中で活用し、実装していく中で、データ科学とセキュリティは両輪のような役割になっていく」と今回のテーマの重要性が述べられた。また、扇氏は日立ソリューションズの事業紹介とともに、巧妙化するサイバー攻撃に対して早期回復を目指す「サイバーレジリエンス」が重要になっていると現状を説明した。



2. ニューラルネットワークを誰でも手軽に活用できる時代 そこに潜むさまざまな攻撃リスクも顕在化

AI の技術動向とセキュリティリスクに関するセッションは、中原講師がニューラルネットワークや Generative Adversarial Network (GAN) について解説を行い、それぞれに対して小林教授が関連するサイバー攻撃のリスクや事例について紹介する形で行われた。また、会場では実際にセッションの内容を体験できるよう、QR コードを通じてプログラムファイルを配布した。

ニューラルネットワークについて、「平たくいえば ChatGPT などの基礎になっている技術」と中原講師は紹介。特に画像処理の分野における有用性に触れ、Google Vision AI のデモンストレーションを行った。画像に写っているものの判別や認識がニューラルネットワークの得意とする領域であり、医療現場でのガン発見や車の自動運転など、その可能性を説明した。

配布した Python プログラムのファイルをもとに、中原講師は「ニューラルネットワークでは大量の画像の学習、ディープラーニングが重要な技術」と話を進める。求める出力や結果を得るためには、入力に対するパラメータ（関数）の設定が不可欠であり、良いニューラルネットワークをつくるためには、パラメータを少しずつ動かしながらチューニング（学習）していくことが必要、と分かりやすく説明を行った。



中原講師のニューラルネットワークに関する説明を受け、小林教授は「敵対的サンプル」を取り上げ、そのリスクを参加者に提示。入力となる画像への攻撃を行うことで、人間の目には一切変化がなくても AI による出力が変わる、つまり人間が感知できないところで攻撃が行われるリスクがあるとした。「敵対的パッチ」という手法もあり、特定の画像を持ったり、Tシャツにプリントして着用したりすることで、AI から認識されなくなるといった悪用法も紹介した。



続いて中原講師は「GAN」の技術について説明。GAN はニューラルネットワークを活用し、画像など人工的にデータを生成するための技術であり、現実と見分けがつかないほどの画像が、現在生成可能になっている。GAN もニューラルネットワークの一種で、入力に対してあるパラメータを通して出力を返す仕組みである。GAN における入力はノイズおよび実際の画像であり、入力されたノイズを画像として人工的に生成するネットワークと、それを実際の画像か人工の画像かを判別するネット

ワークからなる。この双方のネットワークを組み合わせることで、繰り返し学習することで、パラメータを調整・学習していく、と中原講師は紹介した。

小林教授は、そうした GAN の技術に対して、人間の顔を別人と誤認させる「敵対的眼鏡」の存在を例示。さらに、その技術を発展させ、ウイルス検知を通過してしまう「敵対的マルウェア」もすでに現れていることを紹介した。「攻撃方法はすでに色々と出てきており、対策はイタチごっこになってしまっている」と現状について述べ、前半のセッションをまとめた。

3. AI 技術を悪用した攻撃を防ぐためには、 研究開発側とセキュリティ側の協力が重要

後半のパネルディスカッションでは、小林教授がモデレーター、扇氏、米光氏、中原講師がパネリストとして登壇し、AI とセキュリティに関する議論を行った。



小林教授はまず、AI への攻撃に関する問題提起を行った。攻撃には、学習データを歪ませる「データ汚染」、ニューラルネットワークのモデルを変え、出力を誤らせる「モデル汚染」、そして非公開のニューラルネットワークを模倣し取得する「モデル搾取」があり、すでに技術が公開されていると紹介。「Hugging Face をはじめとしたオープンなプラットフォームで改ざん・攻撃されたニューラルネットワークが公開された場合、利用者がそれに気づくことは難しい」と警鐘を鳴らした。

米光氏はそうした攻撃の事例に対して、危機感を感じた部分と安心した部分があると応答。AI の技術の進歩には驚いたものの、AI に対する攻撃は、現状では針の穴に糸を通すものであり、システム全体で対処していくセキュリティの考えの中で十分対応できる余地があるとした。

ホワイトハッカーの扇氏は、現在も起こっている DDoS アタックを例に挙げ、「どれが攻撃でどれがそうではないか、その境界を我々も常にチューニングしている。AI でもそうした積み重ねが必要になるだろう」と述べ、セキュリティ側だけではなく、AI の開発者によるセキュリティデザインも必要との見方を示した。

データ科学の研究者である中原講師は、見た目にはほとんど変わらないのに AI が別のものだ認識してしまうことが問題であり、その本質はわずかな入力の変化に対してニューラルネットワークの出力を大きく変えてしまう過学習（オーバーフィッティング）の問題であると言及。また、目的に合わせた AI 技術を利用すべきであり、それを判断できるリテラシーを持った人材の育成が必要と提言した。

質疑応答のパートでは、会場から「音声分野での悪用」や「システム運用全体で見た際のセキュリティ」について質問が寄せられ、扇氏からは、「現在も金融や防衛といった重要なシステムはマルウェア検知製品が複数使われている。AI はフリーで提供されているものも多くあり、複数組み込むことで比較的安価にセキュリティを担保できるのでは」と指摘があった。

4. 人工知能の悪用による攻撃はより身近に セキュリティの担保が社会実装の決め手になっていく

パネルディスカッションの2つ目のテーマは、「AI を悪用した攻撃」。小林教授からはAI の悪用による「標的型攻撃」と「情報抽出」の解説が行われた。

米光氏はこうしたAI を使用した攻撃の増加に対し、「便利なツールができれば、皆それを使う。それを活用するか、悪用するかは人間側の問題。総じて抽象的な判断にならざるを得ず、イタチごっこもある程度はやむをえないのではないか」と私見を述べた。



扇氏は、サプライチェーンの脆弱な部分への攻撃リスクを指摘。従来、高度なハッカーでしか見つけられなかった脆弱性を誰でも突けるようになると、サイバーリスクのさらなる深刻化につながることを憂慮している、と述べた。

一方で中原講師は、AI 研究の自由度という観点から、最初から様々な悪用のリスクを全て考慮した上で研究を行うというのはなかなか難しいと説明。現状では人間の詐欺と同じく、リスクを把握したうえで行動することが必要なのではと語った。

ディスカッションの最後のテーマとして、小林教授から、世間でも注目されているディープフェイクの話題が提起された。音声や動画を加工し、別人になりすますことができるディープフェイクでは、すでに詐欺被害も発生している。

米光氏はこうした状況に対して「かなり深刻な状態。テクノロジーを使って人間の心理を攻撃している。攻撃者にとって、AI のコストパフォーマンスが上がってきていて、悪用のハードルが下がっている」と述べ、今後のさらなるリスク増大に危機感を表した。

パネルディスカッションの終わりにあたり、小林教授からは、「AI についてもセキュリティについて考える必要性が生まれてきている。進化のスピードが非常に早いので、人材育成も含めて、セキュリティ側と AI（データ科学）側がタッグを組んで対策を進めていかなければならない」と総括があり、本セミナーは盛況のうちに幕を閉じた。



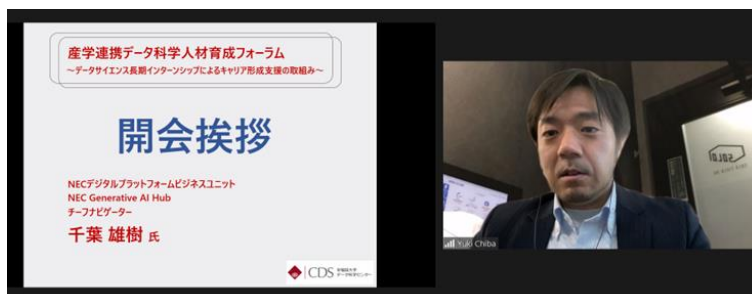
産学連携データ科学人材育成フォーラム ～データサイエンス長期インターンシップによる キャリア形成支援の取り組み～

1. 2023 年度に行われた長期インターンシップの成果を報告 産学連携を通して新たなイノベーションの創出を目指す

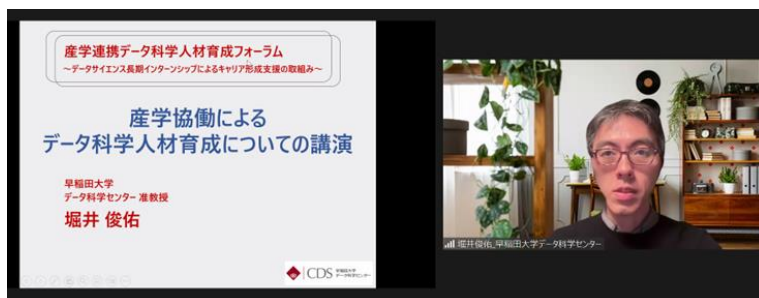
2023 年度に実施されたデータ科学分野での長期インターンシップ。その報告会が 2024 年 3 月 12 日(火)に Zoom 上にて行われました。

データ科学の学修において重要となる実践機会の充実、そして産学連携による新たな人材育成のための仕組みづくりを目的として、データ科学センターは各企業と協力して多様なインターンシップ・キャリア教育プログラムを運営しています。2023 年度に行われた長期インターンシップは、日本電気株式会社（以降、NEC）と GMO フィナンシャルホールディングス株式会社の 2 社と連携し、1 ヶ月以上にわたり実施。報告会では、各社のインターンシップに参加した学生と、受入先企業でのメンターが、それぞれの立場から取り組みや成果を報告しました。

報告会に先立ち開会の挨拶をされたのは、NEC デジタルプラットフォームビジネスユニット NEC generative AI Hub チーフナビゲーターの千葉雄樹氏。日頃から AI を軸とした事業に従事している千葉氏は、海外における学生の長期インターンシップの活況に触れ、「数ヶ月単位の活動を経て論文を提出するような、長期間かつ充実したインターンシップが行われ、それが将来のキャリアパスに直接つながっている。企業サイドとしても、普段のビジネスの中でカバーできない新しいアイデアにチャレンジできる良い機会。両者の好循環をつくっていきたい」と産学協働への期待を語りました。



千葉氏の挨拶に続き、本学の堀井俊佑准教授より産学協働でのデータ活用人材育成の趣旨や、データ科学センターのカリキュラム及び本インターンシップの意義について説明がありました。



2. LLM の理解と実践に取り組んだ NEC でのインターンシップ 柔軟な働き方と最先端のデータ活用環境で、将来のキャリアも具体化

前半は、NEC でインターンシップを行った基幹理工学部応用数理学科 4 年（当時）の山口諒介さんと、先進理工学研究科物理学及応用物理学専攻修士 1 年（当時）の吉村海輝さんからインターン内容に関する報告がありました。NEC でのインターンシップは、主にデモアプリの開発や、LLM のファインチューニングなど、AI 領域での開発・検証・学習・評価に関する一連の流れを体験するものになっています。

数理学を専門としている山口さんは、アプリ開発や LLM のファインチューニングに関して、今回のインターンシップが初めての挑戦となりました。アプリ開発では、企業名とレポート化したい項目を指示することで、簡単かつ見やすいレポートが生成されるシステムを Streamlit で作成。Azure の機能調査では、開発の流れや使いやすさを体感する一方、ユースケースの発見の難しさなど、実践ならではの気づきも得られました。LLM のファインチューニングや RAG の検証フローは、OpenAI の ChatGPT-3.5 も取り入れて最新の環境で実施。インターンシップを通じた学びについて、山口さんは「目的を達成するために課題を分割し、さらにその課題に対してどのようなアプローチを行えばよいか、実践を通して考えることができた。また、限られた期間で成果を出すためには、まずは動くものを作成し、その後機能を追加することも重要だとわかった」と述べました。山口さんは今回のインターンシップで得た技術を生かし、データ分析のコンペティションやプログラミングのコンテストに参加。見事上位で入賞できたと報告しました。



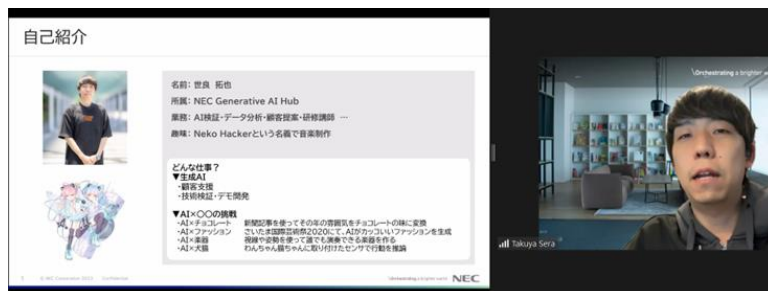
次に報告を行った吉村さんは、データ科学に興味があったものの、自身の専攻は情報系ではないことから自信が持てず、進路の決断に苦慮していました。「実際に企業で働くことで答えが見つかるのでは」と考え、今回のインターンシップへの参加を決意したといいます。

インターンシップ期間内での目標は NEC LLM の優位性の検証。まずは NEC LLM の評価やプロンプ

トの工夫といった基礎的なステップから始めました。10月から11月にかけての学業両立型のインターンシップでしたが、授業やゼミの日は休みを取ることができ、週3日の出勤日のうち2日はリモートワーク可能など、柔軟な働き方がうれしく、メンターのサポートも心強かったと語りました。インターンシップを通してAIのクリエイティブな可能性に魅力を感じ、データサイエンスの道に進むことを決めた吉村さんは、「最初は不安があるかもしれないが、データサイエンスの道に興味がある人は、誰でも自分の専門分野を生かすことができることを伝えたい」と後輩にエールを送りました。



受入先企業からは、NEC デジタルプラットフォームビジネスユニット NEC Generative AI Hub クリエイティブリードの世良拓也氏が登壇。AI 楽器「ANDCHESTRA」や通信困難なシチュエーションにおける生成AI「cotomi」の活用など、NECが進めるAIの社会実装の取り組みを紹介しました。今回のインターンシップについては、「どちらか片方だけが恩恵を受けるのではなく、双方にメリットや刺激があるプログラム」だったと振り返り、今後さらに普及していくAI分野での、データ科学教育の重要性と期待について述べました。



3. GMO で投資に関わるデータ分析と予測モデルの構築を体験 データ分析における前処理と意思決定の重要性を再認識

NEC に続き、GMO フィナンシャルホールディングスでインターンシップを行った教育学部数学科3年（当時）の隆大志郎さんと、先進理工学部物理学科4年（当時）の佐藤響さんが報告を行いました。

隆さんは統計学を専攻しており、データ科学科目を積極的に履修。前回の長期インターンシップの報告会にも参加し、ビジネスの現場で経験を積むために今回の参加を決めました。インターンシップのテーマは、フォワードレートの変動要因の分析と収益予測。トルコリラ・円をターゲットに、課題のキャッチアップやデータの収集・前処理、変動要因の分析と収益予測モデルの構築を進めました。トルコと日本の経済動向の分析では重回帰分析やVIX指数などの指標を用いて調査。収益予測モデルの構築では、XGBoostやLSTMを使用してAIを開発していきました。途中で体調を崩すなどトラブルがあったものの、どちらも目標としていたクオリティに仕上げることで成果に自信を見せました。インターンシップでの学びとして挙げたのは、前処理の大切さ。「データ分析

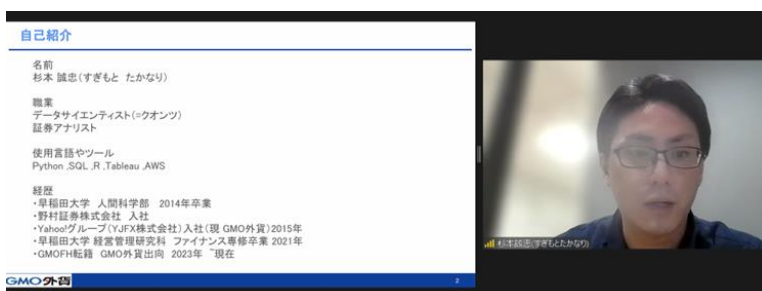
は前処理が8割とよくいわれるが、正確な結果を出力するための準備の重要性を実感した」といいます。ほかにも、報連相や言語化の重要性、課題に対するアプローチの明確化など、ビジネスにおけるデータ科学の実践に必要なスキルを学ぶことができました。さらなるスキルアップに向けて、NECでのインターンシップにも参加している隆さんは、データ科学や機械学習をより深く研究していきたいと今後の目標を語りました。



続く佐藤さんは、「データ科学と金融の双方に興味があり、それぞれの実践的なスキルを身につけられると感じたから」とインターンに参加した目的を説明。インターンシップでは、外国為替市場における投資戦略の妥当性を、重回帰分析やマルコフ連鎖モンテカルロ法という数理モデルによるデータ分析によって検討しました。なかなか思う通りに結果が出ず、何度も分析を繰り返したケースもあったという佐藤さん。その中で、「データサイエンスはあくまで仮定を分析するための道具。結果を出すために闇雲に試行錯誤するのではなく、仮定の見直しや目的意識が重要」という学びを得ました。今回のインターンを通して、計量経済学とデータ科学の未来にあらためて興味を持った佐藤さんは、大学院でデータサイエンスを活用する研究を進めていくそうです。



学生の発表後、受入先である GMO フィナンシャルホールディングスの杉本誠忠氏から企業側の報告がありました。早稲田大学の校友でもある杉本氏は、自身の学生時代を振り返りつつ、今回のインターンシップは GMO にとっても、社内に新たな風を取り込むいい機会となったと語りました。インターンシップで特に伝えたかった点として、データ収集と前処理の難しさ、そしてビジネス上の意思決定を挙げ、「特に意思決定は、その後の作業全体を左右する重要かつ困難なポイントです。もしこの意思決定により何らかの失敗があった場合、大きな責任が伴うことになります。これはインターンシップだからこそ経験できることです」と述べました。最後に、今回のインターンシップが杉本氏自身のモチベーションアップにもつながったとし、産学連携の非常にいい成功例になったと今回のインターンシップの成果をまとめました。



4. カリキュラムを通じた体系的な理解×最前線での実践で 早稲田大学のデータ科学教育プログラムをより深めていきたい

報告会の最後に、小林学教授が閉会の挨拶を行いました。それぞれのインターンシップ内容に関してあらためて触れ、最先端の環境や技術を用いて実践ができる非常に素晴らしい場となったと総括。データ科学科目の履修と、インターンシップ内での実践の両方が組み合わせることでより深い理解につながるとし、協力いただいている企業に対する感謝を述べ、報告会を終えました。



早稲田大学データ科学センター機関紙「データ科学」第3号

発行日	2024 年 7 月 10 日
編集	須子 統太
発行者	松嶋 敏泰
発行所	早稲田大学データ科学センター 〒169-8050 東京都新宿区西早稲田 1 丁目 6-1 1 号館 3 階