Customer CASE STUDY

databricks

Databricksによる、機械学習を活用した リアルワールドデータ解析とデジタルマーケティングの革新

田辺三菱製薬

田辺三菱製薬株式会社は、1678年に創業され、大阪に本社を構える日本の大手製薬会社である。「病と向き合うすべての人に、希望ある選択肢を」という企業理念のもと、ステークホルダーとの信頼関係に裏付けされた、創薬力と育薬力を強みに持ち、多くの画期的な薬を創出している。医薬品産業を取り巻く環境が急激に変化する中、新中期経営計画21-25において、プレシジョンメディシンの実現による、「適切な医療を、適切なタイミングに、適切な患者さんに届ける」を目標とし、さらには、「予防から予後にわたりアラウンドピルソリューションを提供することにより、患者さんとご家族のQOL向上に貢献する」ためのビジネスの変革に取り組んでいる。

育薬本部
データサイエンス部
部長
後川 芳輝 氏

育薬本部
データサイエンス部
進藤 淳 氏

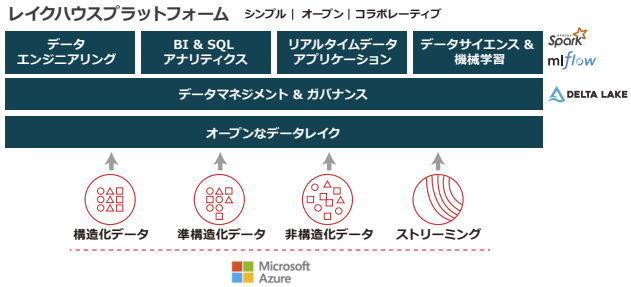
ICTマネジメント室
ICTマネジメントグループ
グループマネージャー
小林 弘幸 氏

ICTマネジメント室
グローバルインフラ
グループ
尾﨑 宏道 氏

ICTマネジメント室
グローバルインフラ
グループ
林 武 氏

ICTマネジメント室
グローバルインフラ
グループ
胥 愷婕 氏

ソリューション



ハイライト

インフラの構築・保守にかかる社内リソースが不要！

従量課金と柔軟な拡張性により健全な投資規模で処理が可能！

各種言語対応とAzure Databricksチームによる手厚いサポート！

課題

従来の環境では処理が難しい膨大なビッグデータ。データサイエンスチームとICTチームの協業モデルの模索。その2つの課題に直面。

2021年3月には、2030年に向けた成長戦略骨子である、MISSION、VISION 30を発表し、2021年から2026年までの5年間において、成長戦略とバリューチェーンの生産性向上に向けたデジタル基盤の構築、及び、デジタルトランスフォーメーションの推進を掲げている。

新薬の開発は全てデータを基にしたプロセスであり、製薬

会社のデータサイエンス機能部門は、従来よりデータの管理、加工、解析を主業務として行ってきた。そのため、データの重要性の認識の高さはもとより、統計解析やデータ管理に精通したチームが存在している。

近年、リアルワールドデータ(ビッグデータ)の登場により、従来の新薬開発のプロセス以外の場面においても、データやAIの活用が加速している。これにより、データを管理、加工、分析するための基盤の構築と運営の重要性が高まっている。

現在、主なビッグデータの活用場面の1つとして、MR向けのデジタルマーケティングへの利用がある。社内と社外のデータソースを活用し、プロモーションやMRの営業活動に役立つインサイトを分析により抽出し、営業効率の向上を実現するという狙いである。もう1つは、処方箋データベースに代表されるリアルワールドデータ(日常診療の環境で収集されたデータ)の活用である。実医療から得られる「処方データ」「診療データ」「臨床検査データ」などの外部データを、新薬の開発および製品の価値最大化に活用する試みである。

これらのプロジェクトを進めるに当たり、大きく2つの課題に直面した。1つ目は、「データの量」である。特にリアルワールドデータは、1ファイルのサイズが1TB(テラバイト)を超えることもあり、レコード数も 100 億件を超える規模となっている。この規模のデータを分析する必要があったが、従来の環境(コンピュータリソースおよびアプリケーション)では、簡単に処理をすることができない状況であった。2つ目は、データサイエンスチームとICTチームの垣根を越えた円滑な協業モデルの模索である。製薬企業が扱う従来の分析においては、ICTチームが必要な環境をオンプレミスで構築し、データサイエンスチームがその環境で分析をする形式で、完全な分業が成立していた。しかし、ビックデータの分析という場面においては、急速に進化し続けるデータ分析のための技術をタイムリーに把握し、実装する必要性が生じた。また、この新しいミッションには、各ビジネスに応じたデータ加工および分析手法が必要となり、ICTチームがその都度、分析環境を必要なデータセットを含めて準備することは現実的ではないと考えていた。

採用理由

今後、データ分析は企業の競争優位性の源泉になると確信しているため、データの管理・活用(分析)においては、内製化を強化する事が重要であった。

これらの課題を解決するためにDatabricksの導入が有効であった。Databricks は、並列分散処理により、高速データ処理を可能にするApache Spark をベースとしたデータ解析プラットフォームである。Delta 形式を活用することで、データレイクのデータに信頼性と管理性を提供する。また、TensorFlow、PyTorch、scikit-learn などのデータサイエンス用のライブラリーを兼ね備えている。Databricks により、膨大なデータを迅速に処理するとともに、様々なビジネスに対応するデータ分析がフレキシブルに行う事が出来る。また、共通のワークスペース上で全てのプロジェクトメンバーが共同作業を行うことにより、データサイエンスとICTチームの協業も促進している。

Databricks利用のメリットは主に4点あげられる。まずは、ハードウェアなどのデータ解析基盤のインフラの構築、保守はクラウドベンダーが対応し、そこにかかる社内リソースが不要となる点。また、従量課金モデル、且つ柔軟な拡張性により、小規模なデータセットも大規模なデータセットも、健全な投資規模で処理ができる点。加えて、Python、Scala、R、Java、SQLといった各種言語も幅広くサポートしている。最後に、Azure Databricks チームによる手厚いサポートを高く評価している。今後、データ分析は企業の競争優位性の源泉になると確信しているため、外部リソースを利用したアプリケーション開発の様な案件とは異なり、データの管理・活用(分析)においては、内製化を強化する事が重要であった。そのためには、社内のスキル向上も継続的に行う必要があり、使い方などに関する丁寧な支援は非常に有用である。

今後の期待

分析部隊とシステム部隊、それぞれの分野の人材が垣根を越え、組織として融合していくのが、これからの目指すべきチームの姿だと考えている。

製薬業界は医薬品の有効性やリスクを、データを以て証明しなければいけないため、データの重要性や活用に関する概念は、他の業界よりも元来進んでいると言える。ただ、膨大且つ多種多様なデータを扱い、時には機械学習の技術を導入した分析という新しい領域においては、まだまだ進化が必要だと感じている。ビッグデータの分析を成功に導くために重要な事は、現場を深く理解した分析部隊(データサイエンティスト)を育成すると共に、データサイエンティストはビッグデータを扱うためのシステムの知識も兼ね備える事、さらにシステム部隊はデータサイエンティストが実施する分析を理解し、必要なデータ管理システムをフレキシブルに提案する事である。自部門の最適化を追求した従来の分業制の枠組みでは、ビッグデータ分析という領域ではスピードが遅く、望む成果を発揮できない。それぞれの分野の人材が垣根を越え、組織として融合していくのが、これからの目指すべきチームの姿だと考えている。2021年度より、データサイエンスとICTの両部門は、同じ管掌役員が管理する組織体制となり、両部門の融合が進む事でさらなる変革を実行していく予定だ。

データブリックス・ジャパン株式会社　TEL.03-6821-1670　E-mail:marketing-jp@databricks.com