



製造業 DX を支えるデータ分析基盤とその活用事例 Data Analysis Platform Supporting Manufacturing Digital Transformation and Examples

藤井 北斗 佐々木 誠 宮本 悠矢 鈴木 博順

Abstract

Yamaha Motor has set “Yamaha Motor to the Next Stage” as its digital transformation for further strategic use of the latest digital technology and data. In order to strengthen existing businesses through the use of data, analysis of various types of business data is essential to promote this, and therefore the company is collecting and analyzing manufacturing data, product IoT data, and customer data. One example of this is by analyzing vehicle data from connected motorcycles on the market, it is possible to more accurately understand users’ market levels, which was previously an unknown sector. In addition, by analyzing the user’s activities records on the web page, it is further possible to understand the customer’s journey leading up to the purchase of a product, etc. Using data and these measures, the company aims to create new, and greater customer experiences.

This paper introduces the data analysis platform to conduct data analysis. Specifically, we will introduce a “data analysis platform” that enables data collection, conversion, analysis, reporting, and other processes to be performed by data stores and analysis sites. In addition, examples of data analysis for marketing and manufacturing systems using this data analysis platform will also be introduced.

1 はじめに

ヤマハ発動機は、最新デジタル技術やデータのさらなる戦略的活用を図るための DX として「Yamaha Motor to the Next Stage」を掲げている。その推進にはデータ化された各種業務データの分析が必須となり、既存事業をデータ活用で強くするために製造データ、製品 IoT データ、顧客データの収集と分析を行っている。例えば、市場を走行しているコネクテッドモーターサイクルの車両データを分析することで、従来分からなかったユーザの市場での利用状況を正確に把握することができる。また、Web ページ上でのユーザの行動履歴を分析することで、製品の購買に至るまでのカスタマージャーニーなどを把握することができる。これらの施策によってデータによる新たな顧客体験の創出を目指している。

本稿ではデータ分析を行うためのデータ分析基盤について紹介する。具体的には、データストアサイトおよび分析サイトによって、データ収集・変換・分析・レポートなどの処理が実行可能な“データ分析基盤”について紹介する。また、本データ分析基盤を活用したマーケティング系、製造系のデータ分析事例を紹介する。

2 背景

当社ではクラウドサービスを用いた自動運転運行管理システムなどのクラウドシステムの開発実績や、さらには運行データを使った顧客行動分析の実績がある^{[1][2]}。しかしながら、当社のグローバルに広がる顧客情報や全コネクテッド製品情報のデータ分析を行う場合、複数のデータウェアハウスの統合やビッグデータを分析するための高性能な演算性能を有したデータ分析基盤が必要となる。具体的には製品 IoT データウェアハウスや顧客データウェアハウスなどのデータを統合するために、転送、抽出、加工が必要となる。また、そのデータを AI や統計によって分析する機能が必要であり、さらには分析した結果を共有するためのレポート機能が必要となる。

そこで、社内で育成がすすむデータ分析官が分析で必要とするそれら機能を備えたデータ分析基盤を構築した。

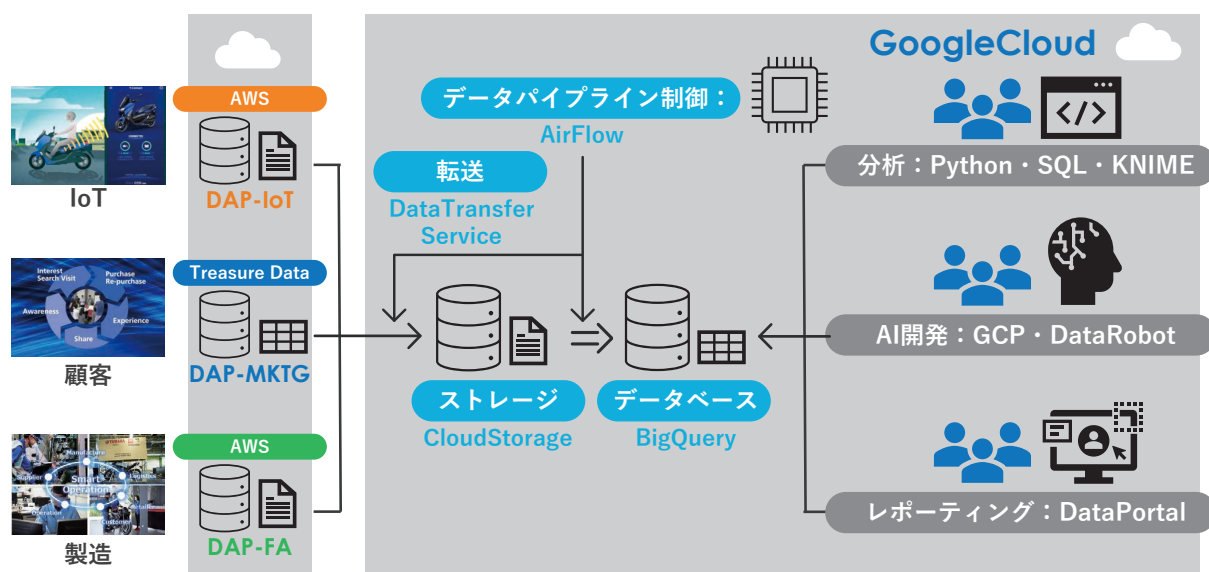


図1 データ分析基盤全体構成

3 データ分析基盤概要

データ分析基盤の概要を図1に示す。本データ分析基盤は“探索的データ分析”、“機械学習”、“分析結果レポート生成”などのデータ分析を行うための基盤であり、図1に示すようにデータソースである“DAP-IoT”、“DAP-MKTG”、“DAP-FA”などのデータウェアハウスと接続することで、それらデータの収集・蓄積・加工・分析・可視化のすべてを行うことができる。具体的にはクラウド IaaS、PaaS サービスである GoogleCloud¹⁾の各種サービスを中心に構築されており、統計解析などを行うための探索的データ分析については、GoogleVertexAI¹⁾（以下 VertexAI）による Python プログラミング環境、ペタバイト級データの抽出が可能な GoogleBigQuery¹⁾（以下 BQ）、ローコードプログラミングツールである KNIME²⁾による分析機能を有している。分析者はビッグデータの中から任意の情報を抽出し、統計解析することができる。

また、多変量解析や予測を行うための機械学習については、構造化データ、画像認識 AI、自然言語解析などの分析が可能な VertexAI や、ノーコーディングでトップデータサイエンティストと同等の予測精度の AI を開発できる DataRobot³⁾を備えている。

さらに、分析結果の共有ツールとして GoogleDataPortal¹⁾（以下 DataPortal）が利用可能であり、グローバル拠点間でのデータ共有を行うことができる。

これらの分析機能を活用することで、顧客の Web 行動解析によるニーズ把握や各種マーケティング施策、製品開発へのフィードバックを行うことができる。

3-1. データ分析基盤詳細

本データ分析基盤は、データを収集・蓄積・加工するための「データストアサイト」と、データを分析・可視化するための「分析サイト」とによって構成される。これらデータストアサイトと、分析サイトについて詳細に説明する。

3-2. データストアサイト（データウェアハウス&データパイプライン）

当社はデータウェアハウスとしての DAP（Data Analytics Platform）の構築を進めている。例えば、DAP-IoT には製品 IoT データが、DAP-FA には製造データが、DAP-MKTG には顧客データが収集されている。本稿で紹介するデータ分析基盤はこれらのデータウェアハウスと接続可能であり、クラウドサービスである GoogleCloud 上でそれらデータの分析を行うことができる。

一例として製品 IoT 向けデータウェアハウスの構成について説明する。図2に製品 IoT 向けデータウェアハウスの構成を、図3に処理の流れを示す。本データ基盤は、データソースからのデータ転送サービスとして GoogleDataTransferService¹⁾（以下 DataTransfer）、データストアサイトのインプットストレージとして GoogleCloudStorage¹⁾（以下 GCS）、データウェアハウスとして BigQuery、データ転送処理を制御するためのオーケストレーションサービスとして AirFlow⁴⁾を採用しており、いわゆるデータパイプライン処理とデータウェアハウスによって構成されている。

本データパイプライン処理によって、一連のデータウェアハウスのデータの取り込み処理、データ加工処理が自動実行される。

具体的には図3に示すように、DAP からデータ分析基盤へのデータの取り込みは、AirFlow によって制御される。AirFlow はデータ転送処理およびデータ変換処理などのフロー制御をつかさどるサービスである。AWS S3⁵⁾に保存されているデータを DataTransfer サービスによって GCS に転送する。転送されたデータは JSON 形式で保存されており、データベースに取り込み可能なデータ形式である Parquet に変換される。この際、データの検索性を向上させるために、データ分割単位であるパーティションの変更を行うことで検索性能を最適化している。また図2および3に示すように、データウェアハウスはマスタ情報としての SCCU カタログと、各車両ごとの走行データ統計値を表す SCCU サマリと、時系列走行データが収納された SCCU ログとを備えている。

これらデータウェアハウスのデータには、後述するデータ分析サイトからアクセス可能であり、データ分析官は図4に示すように分析サイト上で統計解析や AI による多変量解析などを行うことができる。

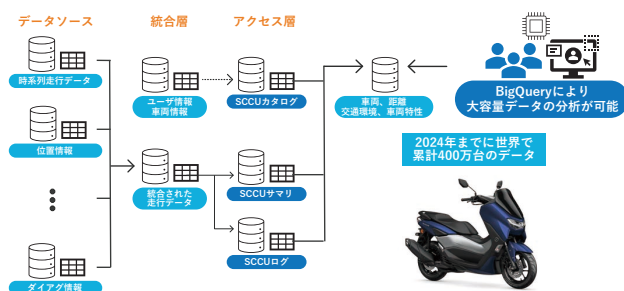


図2 製品 IoT 向けデータウェアハウスの構成

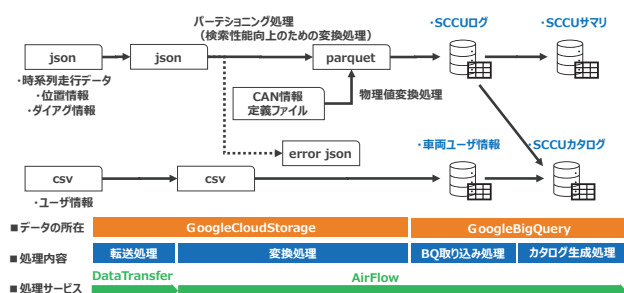


図3 データパイプライン処理の流れ

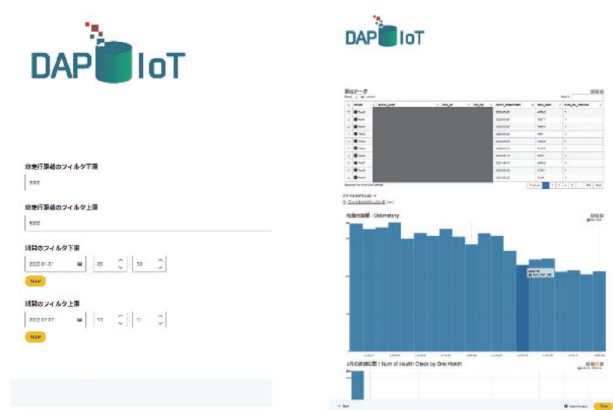


図4 製品 IoT データの検索画面と検索結果

3-3. データ分析サイト(機械学習パイプライン)

3-3-1. 背景

一般に機械学習モデルを業務システム化する際には、データの収集、特徴量の取捨選択、機械学習モデルの作成、作成したモデルを動かすためのコンピューティングリソースの提供など、さまざまな業務を並列に、かつ継続的に行う必要があることが知られており^[3]、それらの管理を容易にするための開発環境としてML-Ops. (Machine Learning Operations)が一般的になりつつある。当社では陸海空それぞれの領域における自動運転や製造現場での異常検知など、さまざまな領域で機械学習システムを業務に導入しているが、導入された機械学習システムの数の増加にともないこれらの業務負荷も増加しており、技術者の作業時間の確保が課題となっていた。

業務負荷が増加する大きな要因として、当社においてはこれらの業務をそれぞれ別のツールを用いて独立して実施していたため、業務間のデータの受け渡しやバージョンおよび組み合わせの管理に非常に大きなコストがかかっていたことが挙げられる。そこでML-Ops. の機能である機械学習パイプラインの整備を進めた。具体的には、前述のデータストアサイトに構築されたデータウェアハウスのデータを分析するための機械学習パイプラインを導入し、データの受け渡しや業務管理を自動化することによりこの問題の解決を図った。

3-3-2. 本機械学習パイプラインの処理

DataRobot などの ML-Ops. 環境も導入しているが、本稿では VertexAI を用いた機械学習パイプライン構築事例について紹介する。図5に構築した機械学習パイプライン処理を示す。

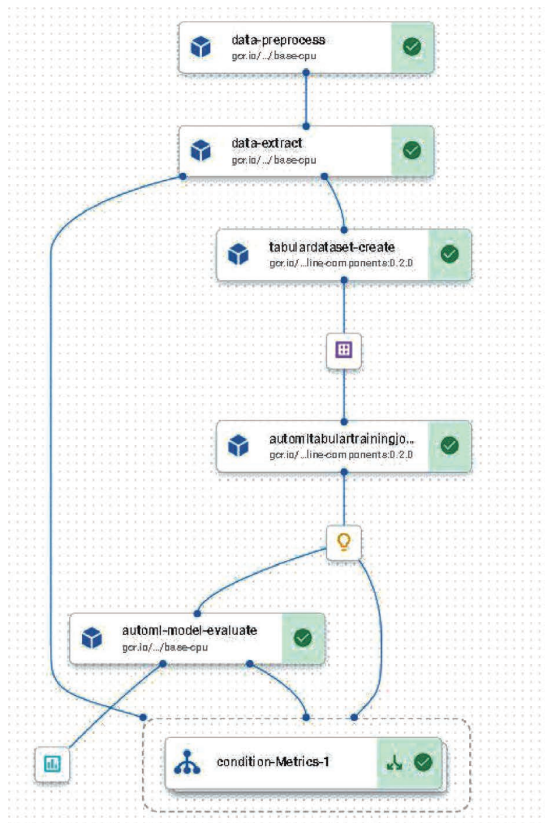


図5 構築した機械学習パイプライン処理

まず、従来は作業者の手作業で行われていた予測に必要な特徴量を含むデータの受信や AI による予測結果の送信などの外部システムとの連携自動化を目的に本プロジェクト専用のデータウェアハウスを BigQuery によって構築し、それを介してデータ連携・システム連携を行うよう設計した。また、従来は手作業によって表計算ソフトなどのツールで処理していたデータ前処理や機械学習トレーニング処理、機械学習予測処理などの複数の処理を Python でプログラム化しており、それらの処理を API (Application Programming Interface) 経由で呼び出し可能とした。これら一連の処理の実行順序やタイミングを制御するためのパイプライン処理(図5)を VertexAI Pipeline によって制御することで、自動化を実現している。またこのパイプライン処理も API 化されており、外部関数やスケジューラなどを用いて API を呼び出すことで定期実行やイベント駆動を実行することができる。

なお、機械学習システムの実運用には予測結果の精度検証を目的とした結果の再現性が求められる。具体的には、AI に投入したデータの履歴や機械学習モデルのトレーニングバージョンなどを管理することで予測結果の再現性を担保し精度検証を行うことができる。VertexAI Pipeline によれば、処理手順を図5のように可視化しさらに、そのパイプライン上の各処理の

実行ログを自動で記録することができる。従来はこういった管理やそれを行うために必要な情報の記録も技術者の業務であったが、それが不要となる点もこのツールを使用する利点である。

機械学習パイプライン処理の導入により、手動での業務記録や業務間のデータ転送が不要になり、またすべての業務を一括して管理できるようになった。これにより、機械学習プロジェクトに必要となるコストを大幅に削減できると考えられる。

4 活用事例

4-1. 製品 IoT データ分析事例

データストアサイトの活用事例として、製品 IoT のデータ分析事例について説明する。図6に製品 IoT データ分析のユースケース例を示す。

図6に示すように、前記データストアサイトの機能を使うことで、分析者は必要なデータに簡単にアクセスできる。例えば、分析者は車種、国、走行距離、ECU 診断回数などの条件でデータをフィルタリングすることで、従来知りえなかった、該当する条件での走行データが市場の車両の何%に相当するかなどの検証をデータに基づき正確に行うことができる。今後は、これらの検討結果を車両開発にフィードバックしていく予定である。



図6 製品 IoT データ分析のユースケース例

4-2. マーケティングデータ分析事例

本節では、前節で説明した機械学習パイプラインを実業務に用いた事例を紹介する。図7にマーケティングデータ分析のユースケース例を示す。

自動車業界を含めた小売業界では、懸賞付きキャンペーンの応募者やイベントの参加者を潜在顧客とみなし、彼らに対して電話営業などのアプローチをとることで購入者数を増やすといったような活動がよく行われる。このような方法の抱える課題の一つとして、潜在顧客の中に、懸賞目当てで申し込み、実際には買う気のない人や転売目的のブローカーなどが含まれてしまい、そういった人へも営業活動が行われてしまうという点がある。そこで、当社では図7に示すようにウェブ行動やアンケートの回答結果などの情報から購入意欲の高い個人利用者とみられる潜在顧客を選定し、そこに営業活動を限定することで効率を上げるという活動を行っている。これをリードスコアリングと呼ぶ。今回、機械学習パイプラインを用いて、ベトナムの潜在顧客を対象に機械学習を用いたリードスコアリングのシステムを開発した。

結果として、データベース形式などが異なるため単純な比較はできないものの、別地域で行った類似のシステム開発と比較して、機械学習システムの開発に必要な時間は増加した一方、モデルを学習させるための特徴量を修正して再び評価を行うといった業務に必要な時間は大幅に減少した。アップデートなどの業務は今後も継続して行わなければならないことを考慮しても、機械学習パイプラインを用いたことでトータルでかかるコストは減少すると考えられる。

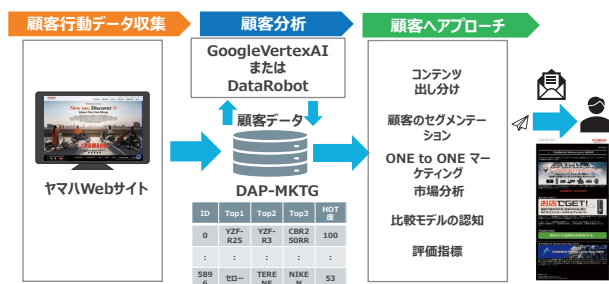


図7 マーケティングデータ分析のユースケース例

5 おわりに

データパイプライン処理と機械学習パイプライン処理とを備えたデータ分析基盤について紹介した。

当社はデータ分析官を含む DX 人材を1200名以上育成することを目指しており^[4]、データ分析人材の活躍にはデータの

運用ルールや体制づくりが必須となる。

そこで、多くのデータ分析官によるデータ活用を可能とするため、図8に示すようなデータマネジメント活動を現在進めている。本データ基盤構築活動に加え、データセキュリティ、データガバナンス、メタデータ整備などの活動を推進している。今後は、秘密情報管理のための行単位でのアクセス権制御機能や、データのアクセスを容易にするためのデータカタログ機能の開発などを行う予定である。

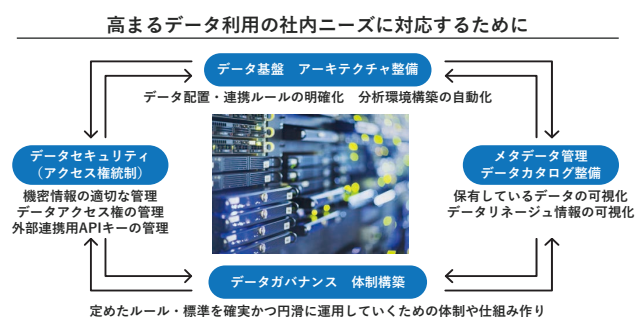


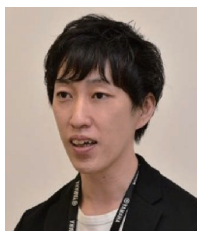
図8 安全安心便利に分析するためデータマネジメント活動

- 1) GoogleCloud [およびかかる資料で使用されるその他の GoogleCloud 商標] は、Google LLC の商標です。
- 2) KNIME は KNIME 社の商標です。
- 3) 「DataRobot」は DataRobot, Inc. の登録商標です。
- 4) AirFlow, <https://airflow.apache.org>
- 5) Amazon Web Services, “Powered by Amazon Web Services” ロゴ、[およびかかる資料で使用されるその他の AWS 商標] は、米国その他の諸国における、Amazon.com, Inc. またはその関連会社の商標です。

■参考文献

- [1] 藤井 北斗, 渡辺 仁:「低速自動走行車両による移動サービスシステム技術紹介」, ヤマハ発動機技報, 2017-12 No.53 (2017)
- [2] 荒木 幸代, 藤井 北斗, 見米 清隆, 渡辺 仁:「リゾート施設における低速モビリティの利用調査と自動運転サービスデザイン」, ヤマハ発動機技報, 2018-12 No.54 (2018)
- [3] D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-François Crespo, Dan Dennison: “Hidden Technical Debt in Machine Learning Systems”, (2015), <https://proceedings.neurips.cc/paper/2015/file/86df7dcfd896fcfa2674f757a2463eba-Paper.pdf>
- [4] ヤマハ発動機株式会社:「新中期経営計画 2022-2024 年」, <https://global.yamaha-motor.com/jp/ir/management/mtp/pdf/2022/2022medium-plan.pdf>

■ 著者



藤井 北斗
Hokuto Fujii
IT 本部
デジタル戦略部



佐々木 誠
Makoto Sasaki
IT 本部
デジタル戦略部



宮本 悠矢
Yuya Miyamoto
IT 本部
デジタル戦略部



鈴木 博順
Hironobu Suzuki
IT 本部
デジタル戦略部