

Tech Acceleration Program で KPI に即したアーキテクチャを短期間で開発し AI を駆使した広告審査を実現



ヤフー株式会社

<https://about.yahoo.co.jp>

〒102-8282

東京都千代田区紀尾井町 1-3

東京ガーデンテラス紀尾井町 紀尾井タワー

メディア事業に留まらず、「ヤフオク!」や「Yahoo!ショッピング」などのコマース事業、膨大なビッグデータを活かしたデータソリューション事業など、幅広く事業を展開している。従業員数は 7,167 名 (2021 年 3 月末時点)。

インタビュー

写真左から 2 番目

・メディア統括本部 トラスト&セーフティー本部
本部長 一条 裕仁氏

写真左から 3 番目

・メディア統括本部 トラスト&セーフティー本部
プロダクト開発部 部長 伊藤 瞬氏

デジタル メディアの 1 つとして知られるポータルサイト「Yahoo! JAPAN」を筆頭にさまざまなメディア、サービスで広告事業を展開するヤフー株式会社(以下、ヤフー)。日々、膨大な量の広告を取り扱う会社ですが、それだけに不正な広告を審査する労力もとてつもないものになっていました。その悩みを解決すべく Google Cloud 上に構築された新しい広告審査システムについて担当エンジニア 2 名に話を伺いました。

利用している Google Cloud ソリューション

サーバーレス コンピューティング、ストリーミング分析、AI の構築と使用

利用している Google Cloud サービス

Cloud Run、Cloud Functions、Cloud AutoML、Vertex AI (旧称 AI Platform)、BigQuery、Firestore など

Google Cloud プロダクト間の連携性の高さを評価

ユーザーを悪意ある不正な広告から守るため、法律に抵触しているものだけでなく、多くのユーザーに不快感を与えるもの、ユーザーに誤解を与える表現などを独自の広告ガイドラインで禁止しているヤフー。同社が自社での広告事業を開始した 1996 年当時は人力で入稿された広告をチェックするという体制でしたが、その後、リスティング広告、コンテンツ連動型広告の台頭による急成長を受け、人力でのチェックを支援する独自開発の広告審査システムを導入することになったと言います。システムの対象となるのは主にテキストで、広告タイトル、ディスクリプション(概要情報)から問題のある広告をフィルタリングして審査員に通知するというものでした。

「このシステムは審査部門のメンバーがフィルターを日々チューニングするかたちで長く運用されてきましたが、近年ではさらに広告の多様化が進み、限界が目前に迫っていました。また、取り扱う広告件数も年々増加しており、システムに任せられる領域をより大きくしていかないことには人手が足りなくなることが明らかでした。」

そう語るのは同社メディア統括本部 トラスト&セーフティー本部 本部長の一条氏。同社はこれらの課題を解決すべく、2019 年に AI 技術を取り入れた新しい広告審査システムを構築すべく動き始めます。なぜこの仕組みに Google Cloud を採用したのか、その理由を新システムの開発をリードしたメディア統括本部 トラスト&セーフティー本部 プロダクト開発部 部長 伊藤氏は次のように説明してくれました。「この取り組みには 3 つの要件がありました。1 つ目は、どれほどの負荷が発生するのか全く読めない機械学習システムを、既存の広告審査システムに影響を与え

ないかたちで構築することです。具体的には、1 日に何千万件というリクエストを裁く既存システムと同等の環境を、運用レスで構築および並行稼働させる必要がありました。」

当時の開発メンバーがほぼ伊藤氏 1 人であったことから、ここは重要な要件になっていたそうです。

「2 つ目は、データの長期保持です。ヤフーの広告審査で取り扱うデータはあまりに膨大で、一部の審査用データがおよそ 3 日分しか保持できていませんでした。しかし、機械学習モデルを作るにはより長期にわたる過去データが必要ですから、膨大なデータを継続して蓄積および検索できる環境構築が必要です。そして 3 つ目は、機械学習モデルをクイックに継続的に作り続けていける環境の準備です。ゆくゆくは非エンジニア(審査員)でもノンコーディングで機械学習モデルを構築できるようにすることを視野に入れており、実現の可能性を模索していました。」(伊藤氏)



一条氏

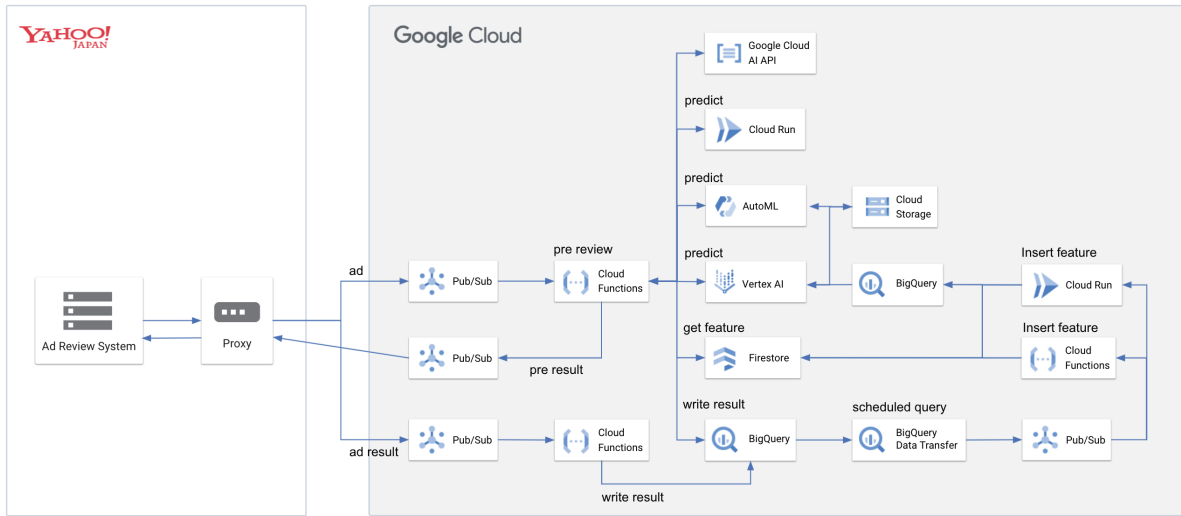


伊藤氏

これらの条件を踏まえてパブリック クラウドを検討した結果選ばれたのが Google Cloud でした。その際、決め手となったのは BigQuery と Cloud AutoML の存在だったと伊藤氏は言います。「BigQuery は、プラットフォーム内のプロダクト連携がスムーズで、データの保

持に制限がなく、検索スピードにおいても秀でており、機械学習を行う上で最も重要な過程であるデータ分析の効率を高めてくれます。Cloud AutoML もデータ連携の柔軟さや扱いやすさという点で優れており、非エンジニアが触る環境として最適だと感じました。マネージドで使えることも要望通りでした。」

Google Cloud の手厚いサポートのもとで生まれた高速システム



上図は伊藤氏らが構築した新しい広告審査システムのシステム構成。既存システムから送られて来た広告データと審査結果を蓄積し、蓄積したデータから機械学習モデルを作成し、広告審査 AI を Google Cloud 上で構築しています。「この仕組みでこだわったのは、Cloud Functions から、Cloud Run や Cloud AutoML、Vertex AI、Cloud Firestore などを活用する構造です。機械学習の前処理や予測に必要な素性を取得し、さまざまな予測を非同期に投げ、パラレルで予測値を集約して、判定、結果を高速に返却する部分ですね。この仕組みのベースは約 2 年前、Tech Acceleration Program の中で作り出されたものが利用されているんですよ。」(伊藤氏)

Tech Acceleration Program とは、Google Cloud のアプリケーション モダナイゼーションチームが KPI に即したシステム構築を短期間で実現するための特別プログラム。要件定義・設計などを検討する 2 日間の事前ディスカッションと、それを踏まえた 3 日間のプロトタイピング、合わせて全 5 日間のプログラムが提供されます。「Tech Acceleration Program の中で最も印象に残っているのはプロトタイピングの部分です。Google Cloud の専門家の方々からのアドバイスを受けて、我々のビジネスニーズや今後の展望をドメイン駆動設計に落とし込んでいくことができたのは本当に得がたい体験でした。これまで Google Cloud を本格的に利用した経験のない我々にとって、実装時の手厚い技術サポートも本当にありがたかったです。プロトタイピングのプログラムはわずか 3 日間で書いたのですが、その 3 日があるから今があるというくらい衝撃的な体験でした。特にプログラムのベースを作ってくれた Google Cloud のサポート エンジニアには本当に感謝しています。」(伊藤氏)

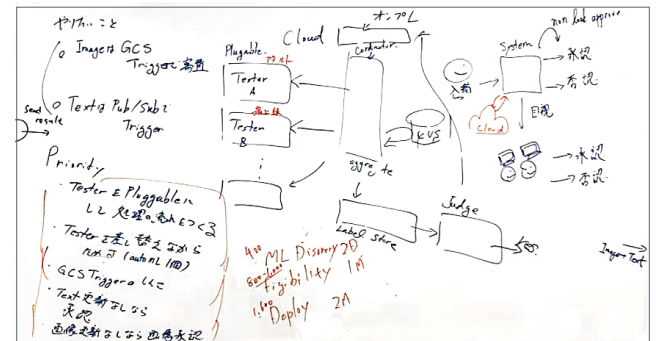
「Tech Acceleration Program に参加した伊藤ら弊社エンジニア陣が一様にホクホク笑顔になっていたのがとても印象に残っています。」(一条氏)

こうして作りあげた新しい広告審査システムは 2019 年冬頃から段階的に投入を開始しました。なお、最初に投入されたのは過去の実績データを元に機械学習モデルを作成した広告の OK / NG 分類機だったそうです。「AutoML Tables で作成した OK / NG 分類機は、審査チームが長年の取り組みの中で蓄積していった知見とノウハウを集めたリスク チェックシステムより、優れた結果を叩き出しました。まだ審査員が見なくて良いというレベルではありませんが、リスク チェックシステムの精度が頭打ちであった現状を打破する結果を残せた功績は大きかったです。今後の審査精度を高める活動に新しい風を感じました。最近では、AutoML Tables が審査結果とその判断に使った情報をインプットするだけ

でモデルを作ってしまうので、非エンジニアの審査員がモデルを作成・検証・導入するという流れも生まれつつあります。実際、OK / NG 分類機以外にも多くのモデルを作成して運用しており、開発当初に想定していた、ヤフーの広告をより安心・安全にしていけるための新しい土台作りができつつあります。」(伊藤氏)

今後もモデルの精度を高める仕組みをどんどん盛り込んでいく予定と語る伊藤氏。中でも今、特に重視しているのは MLOps だと言います。「モデルの精度を向上および維持するには、新しいデータを取り込むことや、取り込むデータに嘘がないかをチェックし、モデルを作り直し、それを比較・評価・導入していく ML Ops の仕組みを作っていかなければなりません。そして広告審査システムの精度を高め、自動で判定できる領域の拡大、機械判定が難しい領域の縮小を目指します。最終的には、審査そのものを行うリソースを、学習データのチェックやラベリング付け、モデル作成を行う業務などにシフトしていく想定です。また、今は広告領域を主軸としていますが、ゆくゆくはヤフーとして AI 審査を外販できるレベルにしていこうと考えています。まだまだこれからなので、この記事を見て面白そうと感じたエンジニアの方がいらしゃったら、ぜひ一緒に作っていきましょう。」(伊藤氏)

「今回の成果を経て、Google Cloud は別の部署でも使えるだろうなと思いました。特にクイックウィンが求められ、少ない人員でスケールの大きな取り組みをしなければならぬところで活躍するのではないかと感じます。」(一条氏)



Tech Acceleration Program での要件定義、設計ディスカッションの様子

Google Cloud を活用することで、ビジネスの将来に注力できるようになります。インフラストラクチャの管理やサーバーのプロビジョニング、ネットワークの構成などに起因する負担を軽減することができます。つまり、インベーターもプログラマーも、自分の本来の仕事に集中することができます。

お問い合わせはこちら
<https://goo.gl/CCZL78>



Google Cloud の詳細については、右記 URL もしくは QR コードからアクセスしていただくか、同ページ「お問い合わせ」よりお問い合わせください。
 © Copyright 2021 Google
 Google は、Google LLC の商標です。その他すべての社名および製品名は、それぞれ該当する企業の商標である可能性があります。

