

コア事業である中古車事業の価格査定業務においてAIを積極的に活用。従来の価格査定プロセスを大幅革新。



データやAIを活用し、売り手と買い手に最適なマッチングを。

株式会社オークネット(以下、オークネット)は、中古車、中古デジタル機器、ブランド品、花き、中古バイク、中古医療機器などのオンラインオークション、および流通に付随するサービスを提供している。オークションにおいて、売り手と買い手に最適なマッチングを提供することで収益を得るため、ビジネスの拡大にむけて、データやAIを活用し効率的に新規会員を獲得したり、会員属性にあったレコメンデーション(推奨)を提供することで、取引の増加を実現することが大切である。

多数の事業を展開する中で、コア事業の一つである中古車事業の価格査定業務においてAIを積極的に活用している。中古車販売店における、従来の中古車価格の査定プロセスは、担当者による目視ベースの勘や経験に基づいているため、価格の算出に手間がかかることに加え、査定結果にバラツキが発生してしまう。一方で、AIの活用によって、「査定依頼から返信までの時間の高速化」「査定価格の高精度化」「限られた人員の中でのオペレーションの効率化」の3つの価値が提供できる。オークネットが有するビッグデータから、メーカー、車種、グ



株式会社オークネット
DX部門 デジタルサービス開発部
デジマ・データグループ
段 裕之 氏



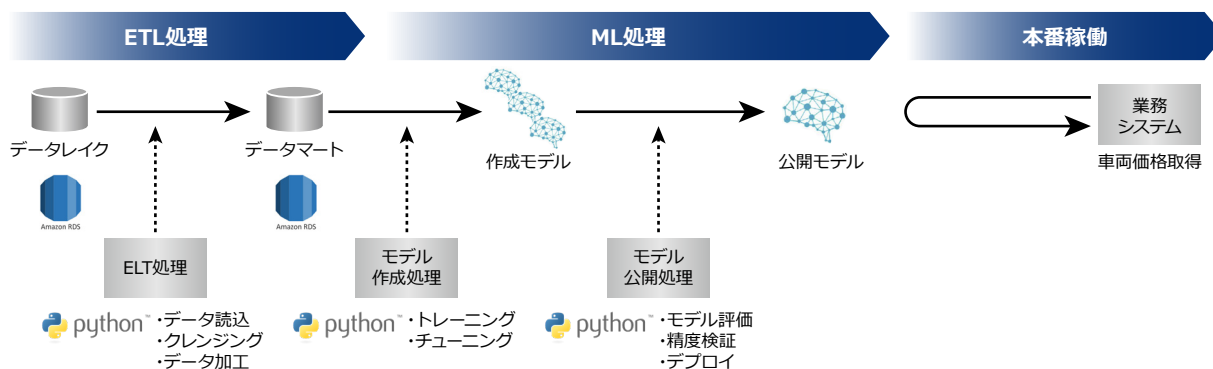
株式会社オークネット
DX部門 デジタルサービス開発部
デジマ・データグループ
林 雄太郎 氏

レード、走行距離、キズの有無や程度などを数値化することによって、一貫した高精度の査定結果をタイムリーに中古車販売業者へ提供している。

従来のアーキテクチャではパイプラインとメンテナンス性に課題

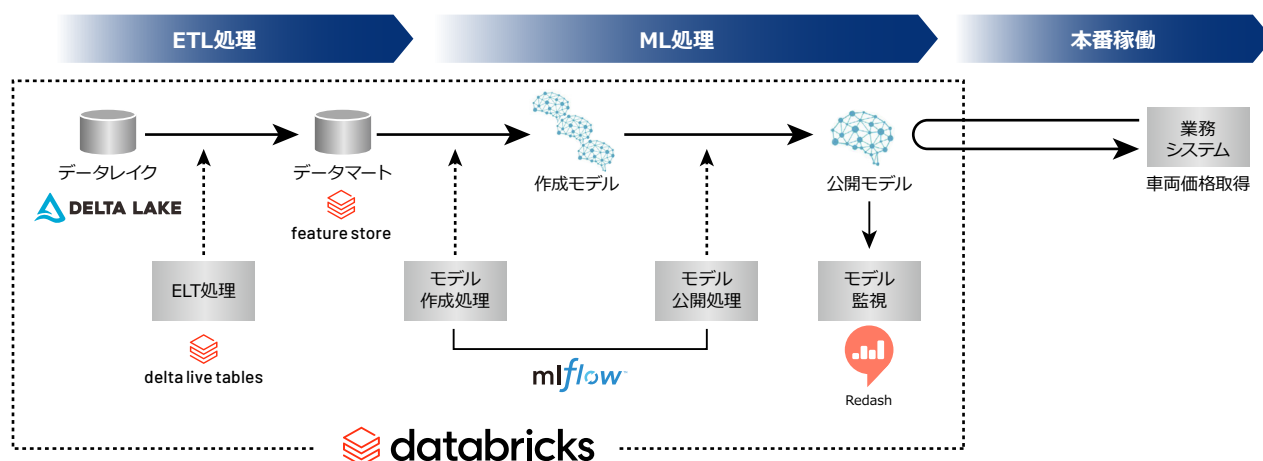
AIベースの価格査定業務を支えるシステムの従来のアーキテクチャ(図1)においては、2つの課題に直面していた。1つは「複数サービスのツギハギによるパイプラインの複雑化、属人化」であり、もう1つは「改修のメンテナンスコストが高く、モデルの評価・改善が困難」という点である。図1内の「作成モデル」に関しては、AutoMLツールを利用していた。

(図1)



- 複数のサービスのツギハギでパイプラインが構築されており、処理が属人化
- 改修のメンテナンスコストが高く、モデルの評価・改善活動が困難

(図2)



- databricksのLakeHouseでAIパイプラインを全体統合
- メンテナンス性を向上し、改善活動を行いやすいアーキテクチャへと転換

「レイクハウス・プラットフォーム」を導入し、課題の解決へ

これらの課題を解決するために導入したのが、データブリックスの「レイクハウス・プラットフォーム」だ。これにより、「AIパイプラインの全体統合」と「メンテナンス性が向上し、継続的なPDCAによるモデルの改善がしやすいアーキテクチャへの進化」が実現した。具体的には、データ保存と管理にはDelta Lake、ETL処理にはDelta Live Tables、データマートにはFeature Store、モデルの作成/公開などのMLOpsにはAutoMLとMLflow、そしてモデルの監視にはDatabricks SQL (旧名称: Redash)の可視化機能をそれぞれ適応し、統合的なレイクハウス・アーキテクチャを実現している。(図2)

データの蓄積・加工に関して、Amazon S3にdelta形式で保存し、DWHライクなインターフェースで運用が可能だ。他のDWH製品と比較して、ファイル形式で扱えるため、非常にコストパフォーマンスが良いと感じている。また、RDSやJSONなどのログデータが保管できる点も魅力的である。ETL処理においては、Delta Live Tableの採用によって、ETL処理の前後関係を可視化でき、処理のエラー箇所の特定が容易になっている。また、モデル運用・監視に関しては、Databricks Workflowsでジョブの管理を行い、モデル公開後もGUI上で1ヶ月毎にモデルを更新し、Databricks SQLのダッシュボードにてモデルの精度を確認している。モデルの精度が一定レベル以下になった場合は、Slackへ自動アラートを上げるような設定も可能だ。

モデルには、「LightGBMというツリーベースモデル」を採用し、目的変数(落札価格)と特徴量(メーカー、車種、排気量等)の2つの要素データを学習データとした。モデルの精度評価指標としては、RMSE(二乗平均平方根誤差)を活用しモデルを作成し、車両の型式等でまとめて正規化したNRMSEを使っている。これを予測値の採用や修正処理に活用している。具体的なモデルの作成、及び精度向上プロセス

は、Databricks AutoMLを使い、プリチューニングモデルを作成後、精度の良いモデルをチューニングする形式を取っている。実際のプロジェクトで、実際にこのプロセスを回した結果、LightGBMがより一番高い精度を達成した。

チューニングには、「特徴量エンジニアリング」と「トレーニングデータの最適割合の探索」を行っている。前者では、追加、間引き、データ形式の変更をしており、後者では、デフォルトで60%を学習データとし、残りの40%を検証データやテストデータとして活用しているが、最適な割合を探索した結果、LightGBMの75%を学習データに回すことで、最も良いRMSE値が出た。

どのようにモデルを生成し、どのモデルを使うべきかの探索では、異なるデータセットを用いて、各モデルの誤差率を特徴量毎に可視化し、各モデルの得意分野や優位性を把握しているMLflowは、一歩踏み込んだモデルの解析も同一プラットフォーム内で完結し、ML機能だけにとどまらないところが、大きなメリットの1つである。

モデルの予測に寄与する指標の確認は、「SHAP値の算出」で行っている。AutoMLのコード内に標準装備されており、どの特徴量がどの程度影響を与えているかを可視化している。この情報を元にユーザーに説明することによって、より説得力のある価値の提供ができていく。中古車価格査定例では、「年式」の影響が非常に大きいということが判明した。その後、実際に手元のデータをインプットすると、どのようなアウトプットが帰ってくるか、という点に関しても、Notebookにデータを直接入れ「Run All」することで、非エンジニアでもパッチテストが可能である。AI検知に不向きなケースが存在する。「希少車種」と「直近の相場変動」である。前者は流通量が少なく、手動での査定になりやすい。これに対しては、該当するモデルを集めた学習データの作成で対応する。後者はモデルの更新タイミング間に発生した突発的な外部要因(例 自然災害による流通量の急低下、及び突発的な需要

増加)によるものである。これに対しては、Databricks SQLを活用して相場変動傾向のダッシュボードを活用して対応する予定だ。

「運用体制の確立」「高精度化」 「ROIの向上」が導入効果

今回のデータブリックスの「レイクハウス・プラットフォーム」の採用の効果としては、「One Platformでの運用体制の確立」、「価格算出の高精度化」、「ROIの向上」の3つあると考えている。AIプロジェクトにはデータの蓄積、加工、分析、機械学習、BIなど、様々なツールを組み合わせる必要があり、各ツールは比較的高価であることが多い。一方で、「レイクハウス・プラットフォーム」は、これら全てのツールを備えているため、特に「ROIの向上」という点が、一番大きなメリットであると感じている。プロジェクトを進める上で苦労した点は、社内の理解を得ることや技術的なキャッチアップがあった。Notebookをうまく活用しつつ、社内のステークホルダーを根気強く説得した。また、データブリックスの技術陣による手厚いフォローにより、今後内製化に関しても更に進められるという確信を得ている。今後も、データ基盤の拡充やAI活用における他部署への展開、運用の高度化に取り組んでいく。