

金持ちモデル 貧乏モデル

著者 : Pranav Dave, Principal Data Scientist, Teradata EMEA

様々なデータを統合的に活用してモデル構築することでモデルの精度はより上がり、実用的になり、顧客との信頼関係を強化することをも可能にします。

精度の高いモデル構築が、企業の収益にどのような好影響を与えることができるのかを確認してみましょう。

「金持ち父さん 貧乏父さん」（ロバート・キヨサキ、シャロン・レクター著）という本を読んだことがある方は大勢いらっしゃると思います。この本では、金融リテラシーと資産に投資して財を成す事の重要性が提唱されています。これと同じ考え方をデータサイエンスにも適用すると、次のような質問が生じます。「金持ちのデータサイエンス・モデルとはどのようなもので、貧乏なデータサイエンス・モデルとはどのようなものか？」私なら、「金持ちモデルとは企業のために収益をあげるもので、貧乏モデルとは企業の収益にならないものである」と答えるでしょう。

では、企業のために多くの収益をもたらす「金持ちデータサイエンス・モデル」はどのように構築できるのでしょうか？ その答えの1つが、前出の書籍の通り資産への投資です。つまり、この場合は、データが資産に当たります。

本記事では、データサイエンスのビジネスとの関連性を高めてアルゴリズムをビジネスに収益をもたらすものにするために、良質なデータ基盤とデータ統合がどれほど重要な役割を果たすのかを概説したいと思います。

データの統合 = 予測精度の向上 = 年間数億ドル規模の潜在的利益 \$\$\$

予測精度はデータサイエンスの至高の目標であるという人もいます。予測精度の向上は、多くの問題を解決できる可能性を示唆しています。機械故障予測、不正検知、顧客解約予測など、多くのアプリケーションにおいては、予測が正確であることが必要です。

“予測精度はデータサイエンスの至高の目標であるという人もいます。予測精度の向上は、多くの問題を解決できるようになる可能性を秘めています。”

ここでは、通信業界の顧客解約予測を例にとってみましょう。通信業界における年間顧客解約率は10%から67%までの幅がありますが、大部分の通信会社では解約率が20%を超えています（米国での市場統計）。解約率が低い会社は、データを有効に利用して解約を予測し、解約を回避するための措置をとっています。解約しそうな顧客を予測するためには、統合されたデータを使用してデータサイエンス・モデルを開発することが必要になります。

ただし、当たり前のように聞こえるかもしれませんが、顧客の解約を事前に予測して対処するためには、予測モデルが正確であることが必要です。通常、解約の防止には、潜在的解約者にコンタクトしたり何らかの特典や割引を提供したりすることが必要になります。解約予測が正確でなければ、他社への乗り換えを検討していない顧客に対して特典や割引を適用する可能性があり、不要なコストを発生させることとなります。さらに悪いことに、本当に解約する可能性のある顧客に対しては何も提示せずにいることで顧客を失い、会社の収益を減らす恐れがあるのです。

それでは、予測の精度を上げるにはどうすればよいのでしょうか。数多くの数学的アプローチが存在しますが、最も重要なのは様々なデータソースが統合された良質なデータ基盤を整えることです。ほとんどのケースにおいて、解約の予測には以下のようなデータが有用になります。

- **顧客層のデータ**：性別、年齢、配偶者またはパートナーの有無、扶養家族、居住地 など
- **請求および契約関連のデータ**：月額料金、請求総額、ペーパーレス請求、契約期間 など
- **サービス利用データ**：電話サービス、インターネット・サービス、オンライン・セキュリティ など
- **デジタル・エンゲージメント・データ**：Webサイトの利用、モバイル・アプリの利用、メールマガジンのインタラクティブ ション など
- **マルチチャネルでのカスタマー・ジャーニーのデータ**：ナビゲーション・パス、店舗への来店、コールセンターとのやり取り など

予測モデルにおいて使用されるデータに応じた、標準的な解約予測精度を見てみましょう。

使用されるデータ	予測精度*
顧客層のデータ	70%
顧客層のデータ + 請求および契約関連のデータ	74%
顧客層のデータ + 請求および契約関連のデータ + サービス利用データ	81%
顧客層のデータ + 請求および契約関連のデータ + サービス利用データ + デジタル・エンゲージメント・データ	86%
顧客層のデータ + 請求および契約関連のデータ + サービス利用データ + デジタル・エンゲージメント・データ + マルチチャネルでのカスタマー・ジャーニーのデータ	90%

*予測精度の数値は、一般に見られる標準的な解約予測モデルに基づく

ここに見られるように、統合されたデータを基盤として利用することにより、予測の精度は20%近く跳ね上がります。これは、何百万ドルもの収益につながり得る数値です。

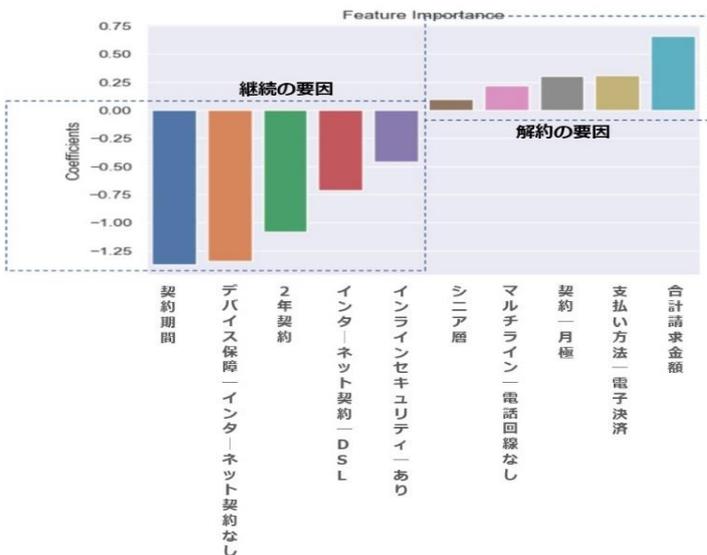
Telecom Tech Newsが最近実施した調査によると、大手通信事業者1社あたりの解約コストは月額で約6,500万ドルに上るとのことです。つまり、予測モデルを20%改善すれば、年間で約1億6,000万ドルのコスト削減を達成できることとなります。

これは膨大な金額です。この金額は、統合型のデータ基盤を持つことの価値とデータサイエンス・モデルの潜在能力がどれほどのものかを明示し、それらが企業1社での巨額のコスト削減につながる可能性を説明しています。

データの統合 = 実用的なデータサイエンス = 年間数億ドル規模の実際のキャッシュフロー\$\$\$

精度の高い予測モデルを持つことは良いスタートとなりますが、それで決して終わりにはなりません。予測精度のパーセンテージが高いのであれば、それを解約回避のための行動へとつなげる必要があります。では、どのようにしてデータサイエンス・モデルを行動へと移せばよいのでしょうか？ 最も一般的な方法の1つは、モデルの特徴量重要度に注目することです。この場合の特徴量重要度とは、解約につながる最も重要な因子としてデータサイエンス・モデルが判定したものを指します。

通信業界における解約の特徴量重要度の典型的な例を示します。



負の係数を示す特徴量は、顧客に解約させない因子です。例えば、契約期間が長ければ長いほど、顧客が解約する可能性が低くなります。すなわち、解約の大部分は新規の顧客によって生じているということです。一方で、正の係数を示す特徴量は、

解約の一因となる因子です。この可視化により、契約月数の短い顧客、および請求総額の高い顧客は解約する可能性が高いことが分かります。これは、解約との関連性が最も高いのは新規の顧客と契約期間の短い顧客であることをはっきりと示しています。これにより、とるべき行動の明確な方向性が分かります。それは、「新規の顧客を固定客へと変える方策を探ること」です。一般的に通信会社は、顧客が契約書にサインした後はその顧客のことを忘れがちです。しかし、特に新規の顧客をターゲットに、顧客と常に連絡をとるマーケティング戦略を遂行すれば、実に有益になるはずで

そのように具体的な行動がとれるのは、予測モデルを意味あるものにするだけの多種多様なデータが存在するからです。予測モデルにおいてさまざまな種類のデータが考慮に入れられていなければ、予測の精度が低くなるだけでなく、結果として起こす行動も間違っただけになる可能性があります。状況を改善するどころか、むしろ悪化させることもあり得るのです。

明確で具体的な行動をとれば、確実に解約を回避し、キャッシュフローが損なわれないようにすることができます。予測の結果として示されたのは、通信事業者1社で年間約1億6,000万ドルのコスト削減の可能性があるということでした。統合されたデータを利用して具体的な行動をとれば、その可能性を現実に変えられるのです。

データの統合 = 顧客のターゲティングの改善 = 顧客の引き付け強化

解約防止策は、事態に対処する「対応型」のものにすべきではありません。解約すると予測された顧客を一人ひとり追いかけることは不可能です。そこで、「先行型」の戦略を策定し、顧客が解約を考えずらしないように図る必要があります。そうした先行型戦略には、定期的に顧客と話をすることやコミュニケーションをとること、何らかの特典を提供することなどが含まれます。特典としては、製品の提供や製品の使用方法のレクチャーなどが挙げられます。

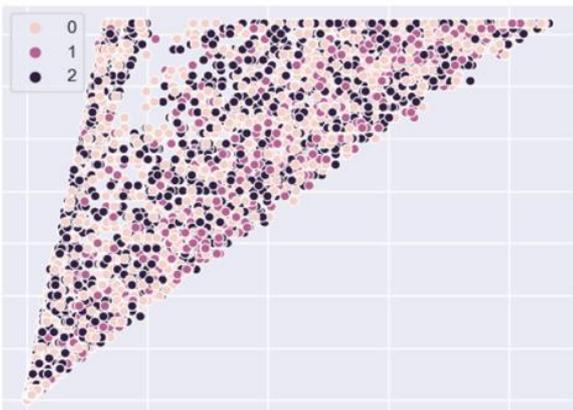
上記の例では、解約の可能性が高いのは契約期間の短い顧客でした。つまり、新規契約の顧客とコミュニケーションをとるという先行型の戦略が非常に重要になり得るということです。ただし、やり取りの際には、すべての顧客に対して汎用のメッセージを送ってはなりません。効果を発揮するのは、顧客に合わせたオーダーメイドのコミュニケーションです。オーダーメイドのコミュニケーションを上げるための1つの方法としては、類似している顧客をセグメントに分類し、そのセグメントに特有のメッセージを送ることが挙げられます。

ここで、1つのセグメントに入れる類似の顧客をどのように見つけ出すのか、という疑問が浮かび上がります。その答えは、統合型のデータ基盤次第で変わってきます。データが多様であればあるほど、同じセグメントに分類される類似の顧客の数が多くなります。これまで見てきた例の場合、顧客層のデータ + 請求および契約関連のデータ + サービス利用データ + デジタル・エン

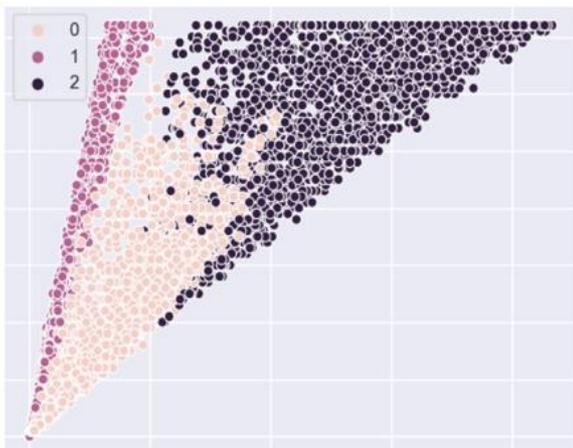
ゲージメント・データ + マルチチャネルでのカスタマー・ジャーニーのデータに基づいてセグメントを作成すれば、各セグメント内の顧客は実際に多くの同じ属性を共有することになるでしょう。もし、顧客層データだけでセグメントを作成すれば、1つのセグメント内の顧客が真の意味で類似していることにはならず、質の悪いセグメントが出来上がってしまいます。類似していないのに、類似しているものと考えてセグメント化された顧客群にメッセージを送れば、解約を減らすどころか、かえって解約を促進してしまう可能性があるのです。

下図は顧客セグメントを可視化したものです。1つ1つの点は1人の顧客を表しています。本当に類似している顧客は同じ色で示されています。①に示されているのは、顧客層データのみを使用して作成されたセグメントです。ここでは、類似する顧客同士が一緒に示されていないために、明確なセグメントがありません。一方②は、顧客層のデータ + 請求および契約関連のデータ + サービス利用データ + デジタル・エンゲージメント・データ + マルチチャネルでのカスタマー・ジャーニーのデータを使用して作成されたセグメントです。こちらは、明確なパターンがあり、真の意味で類似している顧客群によって1つのセグメントが形成されています。

① 統合データ不使用の質の劣るセグメンテーション



② 統合データを使用した質の良いセグメンテーション



統合型のデータ基盤は、より良いコミュニケーション戦略を確保し、事前対応的に解約を回避することを可能にし、将来のキャッシュフローを保証するものとなるのです。

Teradata Vantage を利用すれば、優れた統合型データ基盤を構築することができます。そして、データサイエンス・モデルを金持ち父さんのようにし、企業のために収益を挙げさせることも可能になるでしょう。統合されたデータは、値段の付けようのない貴重な資産となります。Teradata Vantageならその資産形成を実現できます。



ブログ著者 : Pranay Dave

テラデータのヨーロッパ、中東、アフリカ地域 (EMEA) の主任データサイエンティストです。

高度な分析とデータサイエンスを実行し、顧客がビジネス上のメリットを獲得する支援をしています。人工知能および機械学習プロジェクトを運用する多くの企業にて、それらのプロジェクトを主導しており、高度な分析ソリューションからビジネス上のメリットを実現する事に従事しております。また、デジタルメディアにおいて、AIに関する人気ライターとして認知されています。Pranayは、データサイエンス、MBA、およびコンピューターエンジニアリングの学位を取得しています。

テラデータについて

業界をリードするクラウド・アナリティクス・ソフトウェアとサービスを通じて、テラデータはお客様がデータとアナリティクスを活用して、ビジネス上の価値ある答えを獲得することをお手伝いします。

我々の理想は、日々、実現しつつあります。その理想とは、必要なときに簡単にビジネスに必要なインテリジェンスが手に入る世界です。すべてのデータをいつでも参照できる環境を提供し、アナリティクスにおける複雑性を排除することで、データサイエンスをすべての人に提供すること、これが近い将来、企業が競争力を獲得、維持するために必要不可欠になると確信しています。

日本テラデータ株式会社

〒107-0052 東京都港区赤坂2-23-1 アークヒルズフロントタワー
www.teradata.jp

Teradata および Teradata のロゴは、米国テラデータ・コーポレーションおよび/または関連会社の米国およびその他の各国における登録商標です。テラデータでは、新たなテクノロジーおよびコンポーネントの導入に合わせて製品を随時改良しています。本資料に掲載されている情報は予告なしに変更されることがあります。本書に記載された特徴、機能、および運用形態は、地域によっては販売されていない可能性があります。詳細については、テラデータ担当者にお問い合わせいただくか、Teradata.jpをご覧ください。

© 2020 Teradata Corporation All Rights Reserved.

