

データ分析によるコンタクトセンタ効率化 DX : いつだれがどこに電話をかければよいか

石井 方邦^{1,2,a)} 榎 俊孝^{1,b)} 大森 久美子^{1,c)} 大槻 知明^{2,d)}

受付日 2020年8月20日, 再受付日 2020年10月7日,

採録日 2020年10月30日

概要: 本稿では, 営業のあり方に関するデジタルトランスフォーメーションの一環として, アウトバウンドのコンタクトセンタの架電に関する改善を行った. 現在蓄積されている顧客情報や過去の訪問履歴・通話履歴を利用してデータ分析を適用することで優先顧客リストを生成し, 電話のオペレータが営業担当者への取次ができる確率を 15% 以上向上した.

キーワード: データ分析, DX, コンタクトセンタ

Contact center DX by data analysis -When, who, where operators call to-

MASAKUNI ISHII^{1,2,a)} TOSHITAKA MAKI^{1,b)} KUMIKO OOMORI^{1,c)} TOMOAKI OHTSUKI^{2,d)}

Received: August 20, 2020, Revised: October 7, 2020,

Accepted: October 30, 2020

Abstract: In this paper, we improved the call of outbound contact centers as part of the sales digital transformation. We generated a priority customer list by data analysis with the accumulated customer information and past visit history/call history.

Using this list improved the success rate that the telephone operators can give propositions to sales representatives by 15%.

Keywords: data analysis, DX, contact center

1. はじめに

デジタル技術が人々の生活のあらゆる側面をよりよく変革していくデジタルトランスフォーメーション (以下, DX) が定義されて 10 数年が経過する [1]. 近年では, あらゆる業種業態において, 業務改革や営業促進に, 保持しているデータが活用されるようになった. データの活用方法については, 総務省が地方公共団体におけるデータ利活

用ガイドブックを掲載するなど, 国家政策としての方針を出している [2]. しかし, 実際のデータ分析の過程や課題, 課題に対する解決手段・工夫, そしてデータを活用した結果については製造業の一部を除き, 公開されていない [3], [4].

本稿では, 営業データを活用した DX 実践の事例を紹介する. 筆者らが, データサイエンティストとして実践した内容の詳細を可能な限りオープンにし, その過程での課題や解決方法, まだ残る改善すべき課題について記述する. 具体例として, 筆者らが実際に行った営業活動を促進, 効率化するためのデータ分析事例を紹介する. 本稿にて分析対象とした営業現場は, アウトバウンドのコールを実施し, 顧客のニーズ情報を聞き出すとともに, ニーズ有で訪問可能な場合は, 後続に控える訪問営業部隊に情報を取り次ぐコンタクトセンタである. コンタクトセンタのオペレータが行う電話営業に対して, データ分析を実施し

¹ 日本電信電話株式会社
Smart Data Science Center, Nippon Telegraph and Telephone Corporation, Minato, Tokyo 108-0023, Japan

² 慶應義塾大学
Department of Information and Computer Science, Keio University, Yokohama, Kanagawa 223-8522, Japan

a) masakuni.ishii.hr@hco.ntt.co.jp

b) toshitaka.maki.ga@hco.ntt.co.jp

c) kumiko.oomori.sz@hco.ntt.co.jp

d) ohtsuki@ics.keio.ac.jp

た。過去の営業データに加えて、外部データも合わせた分析を通して、ニーズが有る可能性の高い（取次可能性が高い）優先顧客リストを作成し、アウトバウンドコールを試みたところ、コール総数に対して、後続の訪問営業部門へ取り次ぐことができたコール数の割合（以降、取次率と呼ぶ）が、15%以上向上した。

本稿では、2章にて、我々のデータ分析の流れについて議論し、3章にて、我々が分析対象としたデータについて、4章にて、我々が実践した分析作業、および分析結果に基づいてアウトバウンドコールを行ったトライアルについて述べる。そして、5章にて、トライアルを通して見えてきた課題について述べ、6章でまとめる。

2. データ分析の流れ

データ分析の流れについては、多くの既存研究や文献が存在する。たとえば、Abuosba 氏ら [5]、Sinaepourfard 氏ら [6]、および Arass 氏ら [7] は、データが発生してから消滅するまでを考慮したデータライフサイクルという観点で、データ分析の流れを定義している。一方で、ビジネス適用という観点から、Kurgan 氏ら [8] や Mariscal 氏ら [9] は、知識発見・データ分析プロセスのサーベイおよび比較を実施している。Mariscal 氏らは、ビジネス分野問わず適用可能な分析プロセスとして CRISP-DM (Cross-industry standard process for data mining) [10] が最も普及していると言及する。CRISP-DM では、データ分析の流れを、ビジネス理解、データ理解、データ前処理、モデル構築、検証、導入の六つのステップに分類している。Barn Raisers [11] や IBM [12] は、CRISP-DM を拡張した実用上のデータ分析の流れをまとめている。本稿でも、データ分析を幅広いビジネス分野へ適用するため、CRISP-DM の汎用的な 6 ステップを拡張した。我々は、実際のデータ分析の営みで汎用的に必要とされた項目を追加し、図 1 のようにクライアントが理解しやすいような 3 ステップにまとめ、データ分析を捉えている。第 1 ステップは、ビジネス理解/データ理解だけでなく、ビジネス的な課題を解くために、業務ヒアリング等を通して問題を定義する。第 2 ステップでは、データ前処理/モデル構築を実施するために、データを利用して知見を抽出し、第 3 ステップでは、データ分析の結果の検証/導入のフェーズとし、実際の業

務へ適用する。

2.1 問題定義

第 1 ステップは、CRISP-DM のデータ分析モデルフローでビジネス理解/データ理解のステップに含有されるが、我々は、実際の課題を発見し、KPI (Key Performance Indicator) を求めるところまでが、実ビジネスでは必要であると考え、問題定義というステップ名を定義した。問題定義は、該当業務の担当者の業務を理解・課題を発見し、課題に対して解決策を考えるステップである。

問題定義には実担当者とのヒアリング（お悩み相談）を経て、具体的な困りごとを突き止める。そして、その困りごとを紐解き、データから導き出す方法を定義し、解決する方法の検討、および課題解決後の目標設定（KPI 設定）を実施する。目標設定は、解こうとしている課題と解決した際の効果を定量的に設定する必要がある。

2.2 データ分析

第 2 ステップは、第 1 ステップで定義した課題をデータから導き出すステップである。既存データだけではなく、課題可視化に必要なと思われるデータの収集および結合も必要ならば実施する。課題をデータで可視化し、分析を通して解決を実践する。つまり、可視化した現状に対して、分析を通して何を導き出せば、課題解決につながるかを導き出すステップといえる。

まず、担当者へのヒアリングにより、既存データの種類と存在場所を特定し収集する。多くの場合、データは、DB だけではなく Excel 形式や紙媒体で保管されている場合もあるため、データを活用できるように加工する必要がある。さらに、多くの場合、データは散在している場合が多く、結合処理も必要となる。活用できるよう、データを加工、結合した後、確認している課題をデータから確認し、解決方法を検討する。途中、データの不足が確認できた場合は、再度ヒアリングを実施したり、データを収集したりすることを繰り返しながら、課題定義・確認を行う。

その際、可視化や収集したデータからでは、直接得ることができない特徴量を加工・生成する。収集したデータは、そのままでは利用できない場合もあり、複数のデータを組み合わせ、特徴量を抽出することで、より高度な分析が可能となる。たとえば、時系列のデータは、各数値をそのまま特徴量として利用しようとしても、学習アルゴリズムによっては時系列的な順序性等が考慮されない。その場合は、時系列的な差分を考慮し時間軸順に整列した組（タプル）を利用することで、時系列を考慮した分析につなげることができる。

そして、解決に向けた仮説を立案し、解決に向けた分析を実施する。解決のためには、加工・生成したデータから、知見を見つけたらルール化する。仮に、ここで変数が

STEP1 問題定義

業務内容理解 課題抽出・目標設定 解決策の検討

STEP2 データ分析

データ収集・結合 可視化 付加価値データ生成 モデル構築

STEP3 業務への適用

業務検証 最適化・業務導入支援

図 1 データ分析の流れ

Fig. 1 Data analysis process.

多く、ルール化した知見がそのままでは業務に利用できない場合、機械学習を用いる。機械学習を利用する際は、その前処理やアルゴリズムの選択に対し試行錯誤を行い、精度向上を試みる。この精度向上のプロセスに関しては、機械学習モデルの構築の精度向上を自動で行う AutoML (Automated Machine Learning) の技術を利用することで、試行錯誤をすべて自動化することも可能である。

2.3 業務への適用

第3ステップでは、第2ステップでルール化した知見、および機械学習モデルを利用した机上検討を行い、業務への適用可能性を検討する。第2ステップのアウトプットが実ビジネスへ、直接適用可能であるかのトライアル的な検証をし、実ビジネスからのフィードバックにより更なる最適化・業務導入支援を行う。業務導入支援の例として、機械学習の予測結果自体は正しいが、運用に資する際に必要な情報が不足している場合などに、アウトプットを利用者向けに改善することなどがあげられる。

3. 利用したデータ

我々は、2章で説明したステップに従って、アウトバウンドのコンタクトセンタの取次率向上を目的としたデータ分析を実践した。詳細については4章で述べるが、アウトバウンドのコンタクトセンタのオペレータがコールした顧客との対応をコール日報として記録し、後続に控えるオペレータも訪問営業部隊へ取り次いだ情報は取次データとして記録し、訪問営業部隊は、顧客訪問した際の履歴を営業日報として保存している。さらに、訪問営業による受注結果は受注データとして記録されている。

我々が分析に利用したデータは、大きく分けると営業データベース（以下、営業DB）とコールセンタデータベース（以下、コールDB）の2種である。営業DBには、顧客データ、受注データ、営業日報、営業担当者データが格納されており、コールDBには、顧客データ、取次データ、コール日報、オペレータデータが格納されている。図2にその構成を示す。データ量は数GB程度の規模となる。

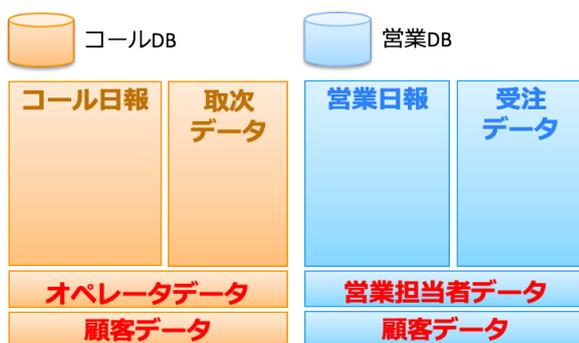


図2 データベースの構造
Fig. 2 Database structure.

また、顧客DBに対して、外部データを利用し、事前に業種を補完することで業種別分析の情報の欠損を13%改善した。なお、IDや電話番号等一部の情報についてはハッシュ関数にて秘匿化して利用しているが、本稿では、その詳細については割愛する。また、それぞれのデータに対して、今回利用したデータ項目（特徴量）については、今後予測に利用することを考え、現状のスナップショットとして利用可能な属性情報、過去の履歴、複数のデータを結合するためのID情報に限ることとした。

3.1 営業DB

営業DBには、訪問営業部隊の営業担当者が、顧客を訪問する際、および顧客を訪問した後の営業関連のデータが格納されている。

3.1.1 顧客データ

顧客データは、顧客情報を格納しており、顧客ID、電話番号、業種、所在地、規模区分、契約商品を分析に活用した。たとえば、ある顧客データのレコードでは、電話番号「03-1234-5678」の業種が「通信業」で、所在地が「東京都港区」で規模区分が「100人未満」の顧客が、「商品A」を契約しているという情報が格納されている。

3.1.2 受注データ

受注データには、受注履歴が格納されている。我々は、営業担当者ID、顧客ID、営業日報ID、受注日、受注商品、受注個数、受注額、利益を分析に活用した。たとえば、ある受注データのレコードでは、ある営業担当者がある顧客を訪問し、受注日「2020年3月3日」に、「商品B」を「1」つ、「20,000円」で受注しその利益が「2,000円」であり、「その際の顧客との対応記録が営業日報に記載されている」という情報が格納されている。

3.1.3 営業日報

営業日報には、営業担当者が顧客を訪問した際の情報が格納されており、営業担当者ID、営業日報ID、顧客ID、提案商品、対応メモ、対応日、対応者、訪問結果、訪問区分を分析に活用した。たとえば、ある営業担当者が、ある顧客に対して、ある実績日「2020年3月3日」に、訪問区分「商品提案」のために、「商品A」を提案し、顧客の対応者が「だれ（役職）」でその結果「受注したかどうか」という情報が格納されている。

3.1.4 営業担当者データ

営業担当者データには、営業担当者の社員情報が格納されており、営業担当者ID、営業ランク、販売実績、入社年度、契約形態を、分析に活用した。たとえば、ある営業担当者に関して、1件間の総利益から導き出す「営業ランク」、「入社年度」および「正社員等の契約形態」が格納されている。

3.2 コールDB

コールDBには、アウトバウンドコールセンターのオペレータが顧客と対応する際のコール関連のデータが格納されている。顧客ID含め、営業DBとは別の管理体系となっている。

3.2.1 顧客データ

顧客データには、顧客情報を格納しており、顧客ID、電話番号、業種、所在地、規模区分を分析に活用した。顧客データは、同じ顧客であっても付与されている顧客IDは営業DBのものとは異なり、それに紐づくその他情報も同一ではない。たとえば、ある顧客データのレコードでは、電話番号「03-1234-5555」、業種「情報・通信業」で、所在地「東京都港区」で規模区分が「80-100人」という情報が格納されている。

3.2.2 取次データ

取次データには、アウトバウンドのコンタクトセンターのオペレータから訪問営業部門の営業担当者へ取り次いだ情報が格納されている。取次管理番号、取次日、対応者、取次商品、コール日報IDを、我々は分析に活用した。たとえば、ある取次管理番号のレコードでは、取次日「2020年3月3日」に、対応者「Aさん」が「商品A」を取り次ぎ、「その際の顧客との対応」がコール日報IDで示されるコール日報に記載されている。

3.2.3 コール日報

コール日報には、アウトバウンドコールセンターのオペレータが実施したコールに関する日報が格納されており、コール日報ID、対応メモ、対応日時、対応結果、対応顧客、提案商品、対応オペレータID、コール目的を、我々は分析に活用した。たとえば、「オペレータA」が、「顧客A」に対して、対応日時「2020年3月3日」に、「ニーズ発掘（2次対応）」というコール目的のために、「商品A」を提案し、その対応結果が「取次」、その際の対応記録が「対応メモ」に格納されている。

3.2.4 オペレータデータ

オペレータデータには、アウトバウンドコールセンターのオペレータの社員情報が格納されており、オペレータID、勤続年数、所属オペレータグループ、役割（後述する1次対応、2次対応、3次対応）を、我々は分析に活用した。たとえば、「オペレータA」が勤続年数「3年」、所属オペレータグループが「地域A」という情報が格納されている。

4. データ分析事例

本章では、2章で紹介したデータ分析の流れに従い、3章で述べたデータを活用して実践した分析事例を紹介する。

4.1 [Step1] 問題定義

4.1.1 業務内容理解

我々が分析対象とする営業部門は、数百万規模の中堅中小の企業に対して、IT商品を訪問営業を通して販売する営業スタイルを持つ。それまで、訪問顧客については、訪問営業担当者の経験と勘に頼る部分が大きく、戦略的な訪問計画が立案できていないという課題を有している。さらに、系列会社が経営しているアウトバウンドコールセンターを有しているものの、コール結果と訪問営業部隊の連携がなされていないという組織上の課題を有している。

我々は、過去の受注状況や、アウトバウンドコールセンターのコール情報を活用していない現状を問題提起し、解決として、訪問営業部隊に対して、アウトバウンドコールセンターが有している情報、および過去の受注状況を活用し、訪問計画立案を試みることにした。我々が目指す営業プロセスを図3に示す。

アウトバウンドコールセンターの業務を深掘りしてみると、オペレータの役割も大きく3種あることが分かった。1次対応で顧客接点を確立し、2次対応でニーズを発掘し、3次対応で商品提案などを行い顧客のニーズをより具体化するという3段階に分かれており、3次対応まで対応が進んだ顧客に対しては、これまでは訪問で商品勧奨を行ってきた。1次対応、2次対応、3次対応については、同じ対応が続く場合もあれば、いくつかの対応についてはスキップされる場合もある。

4.1.2 課題抽出・目標設定

前節の現状から解くべき課題を特定し、解決目標を定量値で設定するために、アウトバウンドコールセンターオペレータ、および訪問営業担当者に対してヒアリングを実施した。ヒアリングの結果、顧客へのコール、訪問営業共に、効果的な対応、訪問方法については検証されておらず、属人的な方法や、就業時の研修程度しか担当者教育もなされていない状況が判明した。そこで、我々は、コールセンター側の課題として効果的なアウトバウンドコールができていないこと、アウトバウンドコールの結果と訪問営業が連携できていないことを課題として設定した。これら課題に対して、訪問営業へのニーズ獲得顧客の取次率向上を、本取組のKPIとして設定することとした。

4.1.3 解決策の検討

取次率向上のためには、過去のコールデータ、および訪問営業による受注データを活用して、どのような顧客がどのようなニーズを有するかという予測モデルを構築し、優

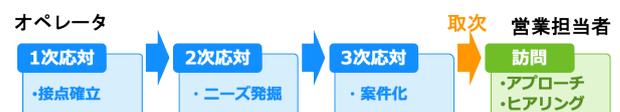


図3 営業プロセス
Fig. 3 Sales process.

先にコールする顧客リストを作成することで解決を図ることを計画した。過去のデータが存在しない新規顧客に対しては、既存顧客の業種や業態を転用することで予測を試みた。

4.2 データ分析

3章で説明したデータを利用して4.1節で特定した課題を表現する。その際、データの特徴を把握し、不足データがあれば収集・加工・生成を試みた。また、加工・生成したデータから、ルールを導き出し、受注を予測するための機械学習モデルを構築した。

4.2.1 データ収集・結合

課題の検証をするために、データの収集・加工を実施した。顧客データであることから、我々が実験に活用できる範囲には制限もあり、必要データは、アクセス権限を有する担当者に取得してもらった。

データを取得した際に露見した課題と解決方法について、いくつか紹介する。まず、データ種別により保存量および期間が異なっていた点、さらには、担当者によって追加されているカラムが存在し、同じデータでも中身が異なるものがあった点については、担当者ヒアリングを実施し手動で補完した。また、文字コードが分析環境と異なっていた点については機械的に変換した。本実践における最大の課題は、コールセンターデータと訪問営業データが異なるデータベース上に存在することであった。これを解決しない限り、課題を特定することができない。

そこで、まず、異なるデータベースに共通のIDを付与することで解決を試みた（これを案件IDと呼ぶ）。案件IDによって、何をきっかけに、どのような流れで、受注・失注に至ったのかを追うことが可能になる。我々は、訪問営業プロセスに対して、コール案件を紐付けるルールを検討し、各データ保持者にヒアリングを行い、ルールに基づいた紐付け結果の共有・検証を繰り返すことで精度を高めた。具体的には図4のような形でのルール検討を実施した。まず、時系列と営業プロセスを考慮してオペレータ側の日報を案件として一つにまとめた。多くの場合は、1次対応、2次対応の次に3次対応がくることに着目し、同じ顧客IDを持つ日報を時系列でまとめる（対応したオペレータはそれぞれの役割で異なる可能性がある）。同様に、営業担当者側の日報を時系列と営業プロセスを元に案

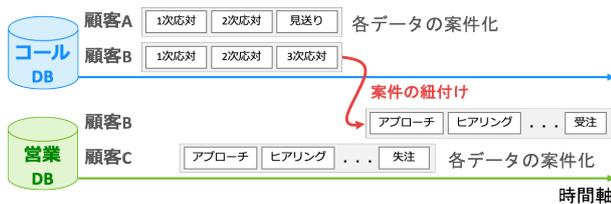


図4 案件ID付与

Fig. 4 Proposition ID assignment.

件としてまとめた。それらを時系列を考慮することで一つにつき、案件IDを付与した。案件IDを付与してみると、取次を行った案件に関しては従来の案件より受注率がおよそ3倍となることが判明した。このことからまず、ある半年の実績を見ると、3次対応を経て訪問営業部門への取次は全体のアウトバウンドコールの11.5%にしか満たないという課題を特定した。さらに、コールセンターから訪問営業担当者への取次率を向上させることが、受注率向上につながるのではないかとという解決に向けた仮説をデータより立証することができた。

4.2.2 可視化

どのような顧客がどのようなニーズを有するのかを予測するために、顧客（業種）、商品、営業担当者、オペレータなど、計500種以上の考えられる特徴量の関係を可視化した。この可視化を通して、ニーズ、受注のルール等の知見のルールを導き出すことを目的とする。

たとえば、顧客（業種）と商品の相関を図5に示す。図5は、横軸が業種、縦軸が商品、もう一つの縦軸に受注率をとり可視化した。顧客ごとでは数が膨大となり視認性が悪いため、その特徴である業種をベースに横軸を規定した。3次元のデータの特徴が分かりやすいヒートマップを利用した。赤いところほど受注率が高く、青いところほど受注率が低いことを表す。図5より、業種ごとに売れている商品に特徴がみられる。このことから、業種という特徴が機械学習の特徴量として有益であることが分かる。

また、別の観点からは、顧客（業種）によるニーズの違いを、オペレータの対応メモを分析することで可視化し導き出した。その結果を図6に示す。業種に対して、横軸が対応記録に出現する単語、縦軸が対応メモ中にその単語が出現した回数を示す。抽出方法は、顧客を業種に置き換え、業種別に対応記録を分類し、業種ごと、出現する単語を数え、頻出するものから降順に並べた。TF-IDF (Term Frequency - Inverse Document Frequency) など、他業種と比較した特徴語を抽出する方式も実験したが、ビジネス利用という観点で、各業種に特徴的なニーズを抽出というより、その顧客に何を提案すべきという観点を重視したた

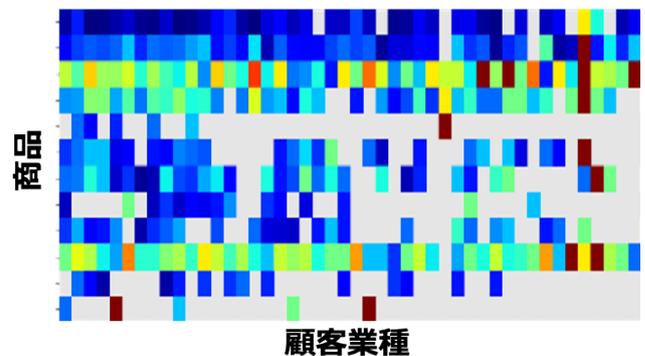


図5 顧客業種・商品別受注率

Fig. 5 Order rate vs customer industry/product.

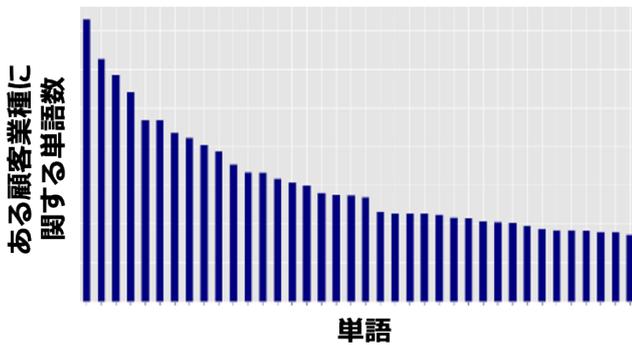


図6 各顧客業種ごとの対応メモ内単語数

Fig. 6 Word count in visit history for a customer industry.

め、よく会話に利用される単語を多く抽出した。

4.2.3 付加価値データ生成

収集・加工データからでは直接得られない特徴量を、複数のデータを組み合わせることで、機械学習に利用するための付加価値情報データとして生成した。生成したデータと利用したデータの一部を以下に示す。

A. 取次重要語

知見のルール化より、コールセンタにおける対応の際に用いる単語が機械学習の特徴量として有益であることが分かった。そこで、オペレータの対応記録から、取次が高い対応に利用された単語を重要語として抽出し、購入動機を表す特徴量として利用することとした。たとえば、「提供、更改、検討、導入、利用」等の導入に関する単語、「セキュリティ、クラウド、PC」等のニーズに関する単語および商品名称に関する単語を重要語として利用した。また、各業種別には、対応メモがない顧客・業種については、欠損値を同業種の平均的なデータで埋めることで精度を保つように補完した。

B. 業種クラスタ

業種別分析を実施するために、業種ごとに利用している単語について対応メモから TF-IDF 値を取得し、頻出上位 20 種の単語を利用してクラスタリングを実施し、ニーズのある業種別クラスタを作成した。アルゴリズムには k-means 法を利用しクラスタ数を調整した結果を人目で確認したところ、7クラスタに分割すると、ある商品に固有のニーズがあるクラスタ（例、クラウドニーズが多い業種等）や特殊なニーズを持つクラスタ（例、商業施設、教育、医療等）に分解ができた。このクラスタを用いて情報が少ない業種に関して情報補完を行った。

C. 役割別の対応回数/成功回数/有効コール回数/コール接続率

各顧客に対してオペレータが対応した回数（コール日報の数）および営業担当者が対応した回数（営業日報の数）、また、それに対して取次・受注が成功した回数、有効であったコール数や、そのコールの割合をカウントして特徴量とした。

D. オペレータ業種偏差値/顧客ステータス偏差値/セールスリード偏差値

オペレータ別分析を実施するために、対応メモを利用して、オペレータが特徴的に利用する単語について TF-IDF 値を取得し頻出上位 20 種の単語を利用して、オペレータごとの偏差値を算出した。オペレータの得意業種を考慮するため偏差値を付与することで、オペレータと業種のマッチングを実現した。さらに、1次、2次、3次の中で得意であるかというステータス偏差値および取次ができるかどうかというセールスリード偏差値を算出することで、オペレータの特徴を表す特徴量を生成した。

E. 営業担当者業種偏差値/提案商品偏差値/受注商品偏差値

受注明細および営業日報より、営業担当者が過去受注した業種および商品を抽出し業種別の偏差値、提案商品、受注商品別の偏差値を生成した。これにより、業種と営業担当者との商品のマッチングが実現する。最適な営業担当者をアサインすることができない場合も、偏差値を利用して近い特性を持つ営業担当者をアサインすることで人員の最適化を可能にした。

4.2.4 モデル構築

4.2.3 項で作成した特徴量を利用して日本電信電話株式会社が開発した AutoML の技術である RakuDA[®][13], [14] を適用し、取次を予測するモデルを構築した。機械学習については、取次できる/できないかターゲットとする 2 値分類問題として、確信度の高いほうから優先顧客とした。RakuDA[®] で比較に利用されたアルゴリズムは KNeighborsClassifier, RandomForestClassifier, LogisticRegressionClassifier, XGBClassifier, LinearSVC, 各種アルゴリズムのアンサンブルであった。RakuDA により F 値最大で選択されたアルゴリズムは RandomForest であり、そのパラメータは決定木個数 200、深さの最大値が 32 であった。

下記が、実際の機械学習モデルとその設定である。

```
RandomForestClassifier(n_estimators=200, max_depth=32, n_jobs=-1, random_state=42, class_weight="balanced")
```

選択されたモデルの中で特徴量の重要度は図 7 のようになった。特に重要度が高いものは、取次重要語であり、オペレータが実際に対応した過去の情報が取次に最も寄与することが分かった。また、実際に対応するオペレータおよび営業担当者の業種偏差値の重要度も高く、ヒアリングした内容を元に、その業種に強い担当者が対応することが取次成功への鍵となることが分かる。

4.3 業務への適用

データを分析して得た機械学習モデルの予測結果を利用して、課題解決の可能性を検証するため、実業務への適用

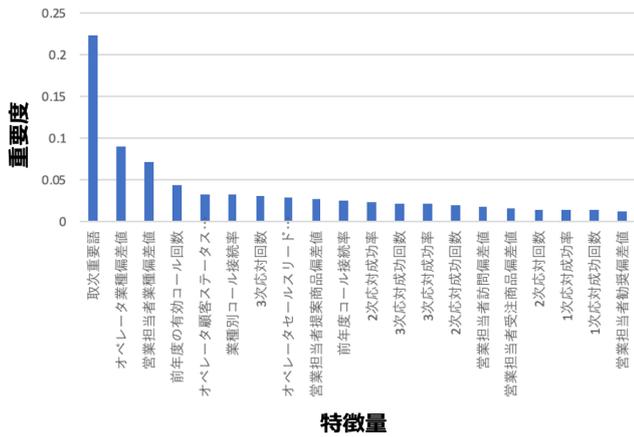


図7 特徴量の重要度

Fig. 7 Feature importance.

トライアルを実施した。トライアルでは、コールセンタオペレータ、訪問営業担当者それぞれに対して、優先顧客リストを作成し、コールを実施し、取次状況を評価した。

4.3.1 業務検証

机上検証にて、2018年4月から2019年3月のデータを利用し学習を行い2019年4月から7月のデータを利用して検証を行った。その結果、構築したモデルを利用すると、機械学習の確信度が0.4以上の顧客のみにコールするとおよそ38%以上の取次率を期待値にできることが分かった。

そこで、2018年4月から2019年7月までのデータを学習させ、2019年7月から11月までコールセンタの実業務にて、優先顧客リストを作成し、トライアルを実施した。その結果、確信度が0.8以上に関しては取次率が48.5%、0.6以上0.8未満に関しては28.1%、0.4以上0.6未満に関しては16.2%ということで、0.4以上顧客への取次率は平均で27.0%となった。これは従来のオペレータから営業担当者への取次率11.5%と比較し15%以上向上した結果となり、十分な有用性が確認できた。

4.3.2 最適化・業務導入支援

優先顧客リストには、顧客およびその確信度・推奨商品のみを掲載していた。しかし、現場へ適用する際、リーズニングデータや売り方に関する情報が必要であるという現場の声を受け、元となる特徴量・情報を優先顧客リストに追加した(表1)。たとえば、抽出した顧客に対する重要単語を加えたトークスクリプト例を追加、また、顧客応対時の情報として最終訪問日や導入済み商品の情報を追加した。追加情報により、オペレータは、予測の理由を理解しながら顧客にコールすることができるようになった。また、現在は、トークスクリプトを利用することで、顧客ニーズに合致しやすい提案で話の焦点を絞ることで、効率よく、より詳細なヒアリングができるようになっている。

表1 改善後の優先顧客リスト

Table 1 Improved priority customer list.

顧客ID	取次確度	コール推奨時間帯	推奨商品	追加したリーズニング部分		
				最終訪問日	導入済み商品	トークスクリプト
5000	84%	10-11	商品A	2018/8/10	商品A	〇〇の更改が…
4000	72%	14-15	商品C	2019/4/8	商品A	XXに関心が高まっている…
8000	15%	12-13	商品F	実績なし	無し	同業種では△△が…

5. 分析するうえでの課題

5.1 データの扱いに関する課題

本稿のデータ分析に関しては、匿名化しているとはいえセンシティブな顧客データを利用している。そのため、分析は、セキュアな環境へデータサイエンティストが向き、閉じられた場所にて実施したため非効率であった。今後は、セキュリティを定義しデータを持ち出せない理由を理論的に説明し、遠隔での分析を可能にする効率的な運用が必要になる。

5.2 可視化の課題

特徴量間の相関を可視化するために、ドメイン知識を利用し、軸を選択し、表現方法を考案した。しかし、この方法では結果として全く相関のないものも可視化され、非効率である。今後は、特徴抽出や表現方法を自動的に選択できるような技術が必須となる。

5.3 名寄せに関する課題

本稿では、詳細については割愛したが、実際にはデータベース間の名前名寄せが必要であった。特徴量名およびその特徴量中の顧客名・商品名等が統一されていなかった。商品名のように数に限りがあるものについては手で補正が可能であるが、数十万規模のデータベースの名寄せについては今後の自動化を検討していく必要がある。

6. まとめ

本稿では、営業データを活用したDX実践の事例として、営業活動を促進、効率化するためのデータ分析事例を紹介した。課題を定義し、定義した課題をデータで表現し、分析を通して解決策を導き出し、実践で検証まで行った。本稿で紹介した三つのステップは、営業に限らず、すべての業種業態で広く利用できると考えられる。今後実践を繰り返すことで、ステップを磨いていきたい。

参考文献

- [1] Stolterman E. and Fors A. C.: INFORMATION TECHNOLOGY AND THE GOOD LIFE, Information Systems Research, Springer Science + Business Media Inc., pp.687-692 (2004).
- [2] 総務省: ICT利活用の促進: データ利活用の促進,

- https://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/bigdata.html (2020年7月29日現在).
- [3] Schuh G. et al.: Data Mining Definitions and Applications for the Management of Production Complexity, 52nd CIRP Conference on Manufacturing Systems, CIRP 81, pp.874-87 (2019).
 - [4] DATAFLAIR TEAM: Data Mining Applications and Use Cases, <https://data-flair.training/blogs/data-mining-applications/> (2020年7月29日現在).
 - [5] K. Abuosba: Formalizing big data processing lifecycles: Acquisition, serialization, aggregation, analysis, mining, knowledge representation, and information dissemination, Proceedings of 2015 International Conference and Workshop on Computing and Communication. pp.1-4 (2015).
 - [6] Sinaeepourfard A., Garcia J., Masip-Bruin X., and Marín-Torder E.: Towards a comprehensive data Life-Cycle model for big data environments, Proceedings of IEEE/ACM 3rd International Conference on Big Data Computing Applications and Technologies. pp.100-106 (2016).
 - [7] El Arass M., Tikito I., and Souissi N.: Data lifecycles analysis: Towards intelligent cycle, Proceedings of 2017 Intelligent Systems and Computer Vision, pp.1-8 (2017).
 - [8] Kurgan L. A. and Musilek P.: A Survey of Knowledge Discovery and Data Mining process models, The Knowledge Engineering Review, Cambridge University Press, Vol.21, Issue 1, pp.1-24 (2006).
 - [9] Mariscal G., Marban O. and Fernandez C.: A Survey of data mining and knowledge discovery process models and methodologies., The Knowledge Engineering Review, Cambridge University Press, Vol.25, Issue 2, pp.137-166 (2010).
 - [10] Shearer C.: The CRISP-Model: The New Blueprint for Data Mining, Journal of Data Warehousing, A 101 Communications Publication, Vol.5, No.4, pp.5-13 (2000).
 - [11] Barn Raisers: 6 essential steps to the data mining process, <https://barnraisersllc.com/2018/10/01/data-mining-process-essential-steps/> (2020年7月29日現在).
 - [12] IBM: Fundamental Methodology for Data Science, IBM Analytics White Paper, IBM Corporation (2015).
 - [13] NTT Software Innovation Center: Automated Data Analysis: RakuDA, <https://www.sic.ecl.ntt.co.jp/e/researchers/team/250.html>. (2020年7月29日現在).
 - [14] NTT Software Innovation Center: データ分析業務を“ラク”にするデータ分析自動化技術「RakuDA」の取り組みを強化・加速, ビジネスコミュニケーション, Vol.56 No.2 (2019).



石井 方邦 (正会員)

2009年慶應義塾大学大学院前期博士課程修了。2010年、日本電信電話株式会社入社。ビッグデータ、機械学習の研究開発に従事。現在、同大学大学院後期博士課程在籍中。

籍中。



槇 俊孝 (非会員)

2017年4月日本学術振興会特別研究員。オープンデータに関する研究開発に従事。2018年3月福岡工業大学博士後期課程修了。同年、日本電信電話株式会社に入社。現在、機械学習に関する研究開発に従事。LOD チャレンジ Japan 実行委員会委員。



大森 久美子 (正会員)

1998年日本電信電話株式会社に入社以来、音声対話処理、ユーザインタフェースの研究開発、およびそれらの経験を活かしたソフトウェアエンジニアリング教育に従事。現在は、NTT ソフトウェアイノベーションセンタにて NTT グループ内の業務分析、業務改善を担当。



大槻 知明 (正会員)

1994 慶應義塾大学大学院理工学研究科博士課程修了。博士 (工学)。現在、同大学理工学部教授。井上研究奨励賞、安藤博記念学術奨励賞、エリクソン・ヤングサイエンティスト・アワード、IEEE Asia-Pacific Young Researcher Award、船井学術奨励賞、国際コミュニケーション基金優秀研究賞、電気通信普及財団賞、CHINACOM'14 Best Paper Award 等 受賞。IEEE ComSoc, SPCE TC Chair, IEEE Wireless Communications Magazine, technical editor 等を歴任。現在、IEEE TVT, Area Editor, IEEE COMST Editor。2019 電子情報通信学会通信ソサイエティ会長、IEICE フェロー、日本工学アカデミー会員。