

2020年11月16日

コンサルティングレポート

AIによる退会抑止

シリーズ「事業経営へのAI活用」①

経営コンサルティング部[大阪] チーフコンサルタント 山村 一夫

AI（人工知能）で何ができるのだろうか。AIは事業経営に積極的に活用できるのだろうか。どう活用すればよいのだろうか。シリーズ「事業経営へのAI活用」ではそれらについて考察していく。ぜひ、これらのレポートについて参照されたい。

【シリーズ「事業経営へのAI活用」タイトル一覧】

- ①AIによる退会抑止 ……<本稿>
- ②AIによる来店客数（売上）予測
- ③AIによる商品購買予測
- ④AIによる商品購買予測とアソシエーション分析
- ⑤AIによる新規出店集客数・売上予測

本稿では「AIによる退会抑止」を取り上げる。会員制度が鍵となるビジネスモデルにおいては、新規会員の獲得と同様に退会抑止（リテンション）が大きな課題の1つである。AIを活用して会員データを分析評価し、マス（集団）としてではなく会員一人ひとりの退会傾向を把握し、先んじて個別性の高い対策を講じることができれば大きなビジネス効果が見込める。

本レポートの後半ではサンプルデータを活用したAIでの分析評価の例を紹介する。どの程度の分析ができるのか、施策への活用が可能なのかについて考察するとともに、ぜひ貴社での活用イメージを高めていただきたい。

I. 退会抑止におけるAI活用

1. 要旨

- (1) AIによる退会抑止では、会員情報を利用して、AIで在籍会員の予測退会確率を計算し、退会確率と退会要因の分析などを行い、退会抑止のための施策を検討・立案・実行する。
- (2) 今回は、機械学習用の「通信サービス会員」のデータセット¹を使い、「AIによる退会抑止」について考察した。データセットの分析・調整後、AIの作成・評価を行い、精度を高めたAIを使って各種予測・分析を行うという流れである。
- (3) AI予測と特徴量の関係を示した図表によりAIの予測根拠への理解が深まる。AIによる個々の会員ご

¹ <https://www.kaggle.com/blastchar/telco-customer-churn>（2020年8月25日確認）を参考にして分析・加工（Kaggleを見るためのブラウザについては、Internet Explorerでは閲覧できなかったが、Chromeでは可能だった。以下同じ。）

との退会予測確率を利用すると施策効率が高まる。

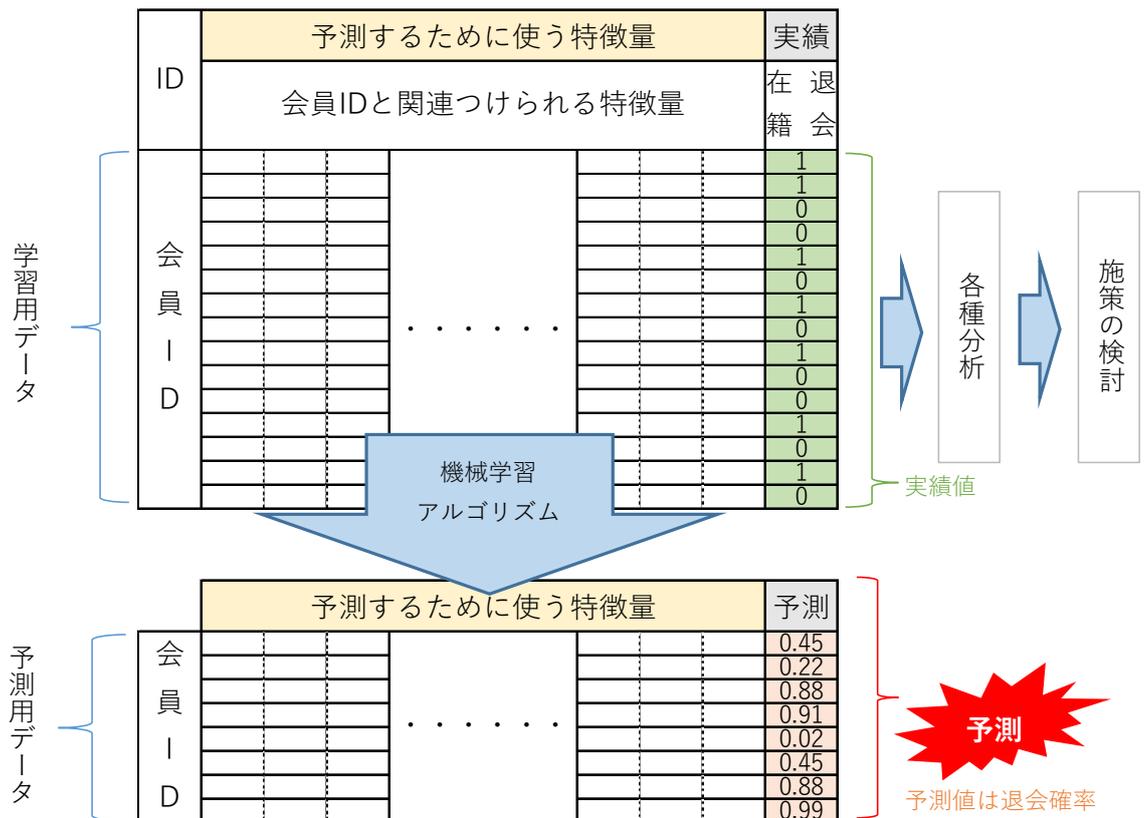
- (4) このデータセットでAI予測精度を高めるには、特徴量（AIが学習・分析の対象とする利用頻度や入会時の出来事などのデータの種類）の増加が必要と思われる。
- (5) 単なる既存データによる会員の退会予測だけではなく、「退会率」を下げるというビジネスの観点で、顧客セグメント分析や施策オプションを組み合わせるとよい。

2. AIによる退会予測・分析

(1) 概要

- ① 会員情報を利用して、AIで在籍会員の予測退会確率を出し、予測退会確率と退会要因の分析を行う。その予測、分析結果を活用して、退会抑止のための施策を検討・立案・実行する。
- ② 分析に必要なデータは主に「個々の会員に関連した情報」「会員の入退会に影響のある情報」と「会員の入会日と退会日（または在籍）の情報」である。
- ③ 一般的に予測・分析の精度が高くなるのは、会員数が多い場合、会員に関連した情報・入退会に影響のある情報が多い場合、データとアルゴリズムの相性が良い場合などである。

【図表 1】 AIによる退会予測・分析イメージ



※学習データの「退会在籍」列の1は退会、0は在籍
 ※予測データの「退会在籍」列は予測された退会確率

(出所) 当社作成

(2) 業種応用例

- ① 学習塾や学校等では成績や出席状況や入学時の学力・試験方法などのデータ特徴量が重要になるであろう。
- ② フィットネスクラブなどの会員制サービスの場合、利用頻度・利用時間などのデータ特徴量は欠かせない。利用頻度や利用習慣性がなくなってくると退会可能性が高まるので、時系列のデータ分析を行う。
- ③ 契約期間内で一定の料金を支払うサブスクリプションサービスなども予測・分析の対象になる。ウェブサイトを使ったコンテンツサービスでは会員の行動に関連したデータ特徴量が利用しやすい。
- ④ 今回は「通信サービス会社の会員」の退会を事例に分析しているが、他の業種でも応用可能である。

(3) プロジェクト成功のポイント

- ① プロジェクトの成功には、「AI の分析力 (データセットの特徴分析・予測精度・予測要因分析など)」・「データセットの拡充」・「ビジネスで有用な施策に結び付けた検討」は欠かせない。
- ② プロジェクトのメンバーには、情報系だけではなく、営業系・企画系など、幅広い人材が必要になる。また、追加した「特徴量」を使い、分析をして、AI の精度を上げ、仮説を検証していくというサイクルも重要になる。
- ③ 情報担当者以外の社内のキーマンが、AI の仕組みを理解し、精度・分析の向上にかかわり、不断に特徴量の充実と精度の向上を進めることが望ましい。

II. サンプルデータによる AI 分析評価

ここからはサンプルデータによる AI 分析評価を行う。ステップ [1] データセットの分析、ステップ [2] AI の作成と評価²、ステップ [3] AI による分析、ステップ [4] 施策の検討の 4 ステップに分けて説明する。

1. 【サンプルデータによる AI 分析評価】ステップ[1]データセットの分析

(1) 内容(特徴量)

- ① 本稿の作成にあたって使用するデータセットは、通信サービス会社の会員約 7,000 人の情報である。このデータセットはウェブ上で IBM コミュニティや Kaggle 等で退会予測の学習用でよく使われるデータセットだ。(実際のデータなのか、学習用に生成されたデータなのかについては、その点の詳しい説明を見つけることはできなかった。)
- ② 会員情報の中身は、「退会したか否か(Churn)、会員 ID(customerID)、性別(gender)、高齢者かどうか(SeniorCitizen)、継続期間(tenure)、契約形態 (月次契約や 1 年契約)、支払方法 (カードや銀行引き落とし)、通信サービスの利用の種類」など 21 項目となる。

² <https://medium.com/analytics-vidhya/using-machine-learning-models-to-predict-customer-turnover-7d52db91d459>、<https://github.com/JNYH/Project-McNulty> (2020 年 8 月 25 日確認)を参考にして分析・加工

【図表 2】 データセット(一部抽出)

customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection
7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	Yes	No
5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes
3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No
7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	No	Yes
9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No

TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
No	No	No	Month-to-month	Yes	Electronic check	29.85	29.85	No
No	No	No	One year	No	Mailed check	56.95	1889.5	No
No	No	No	Month-to-month	Yes	Mailed check	53.85	108.15	Yes
Yes	No	No	One year	No	Bank transfer (automatic)	42.3	1840.75	No
No	No	No	Month-to-month	Yes	Electronic check	70.7	151.65	Yes

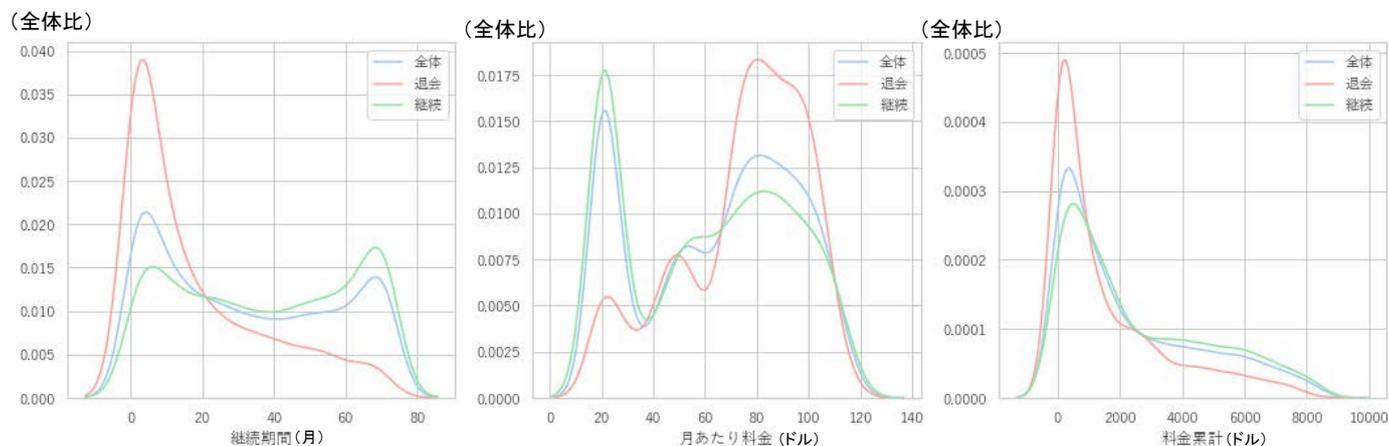
(出所) 「Telco Customer Churn」 Data より³

(2) EDA⁴(探索的データ解析)⁵

① 量的変数の分布

- 退会者が多いのは、「継続期間」グラフでは3~5カ月あたり、「月あたり料金」グラフでは80~100ドルあたり、「料金累計」グラフでは数百ドルあたりである。
- 在籍者が多いのは、「継続期間」グラフでは数カ月と70カ月前後に、「月あたり料金」グラフでは20ドル前後と80ドル前後に、「料金累計」グラフでは退会者よりやや多い所にある。

【図表 3】 量的特徴量の分布と退会者・在籍者



(出所) 「Telco Customer Churn」 Data⁶を使い、「Project-McNulty」 Churn_Prediction.ipynb⁷を参考に当社作成

³ <https://www.kaggle.com/blastchar/telco-customer-churn> (2020年8月25日確認)

⁴ Exploratory Data Analysis 数量の単位は当社推定

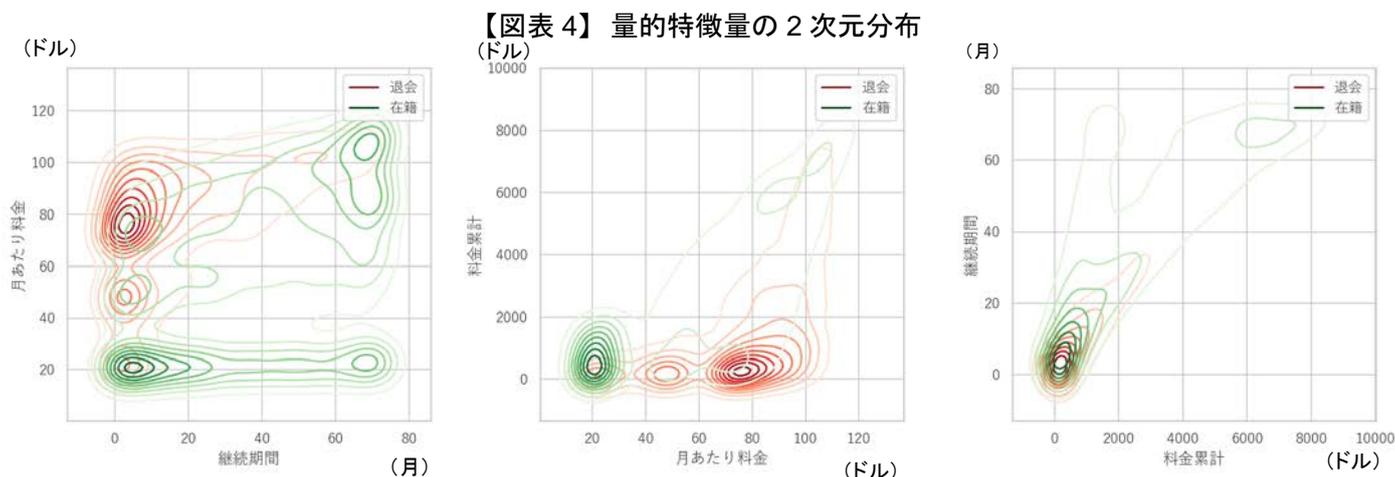
⁵ <https://www.kaggle.com/pavanraj159/telecom-customer-churn-prediction> (2020年8月25日確認) を参考にして分析・加工

⁶ <https://www.kaggle.com/blastchar/telco-customer-churn> (2020年8月25日確認)

⁷ <https://github.com/JNYH/Project-McNulty> (2020年8月25日確認)

② 量的変数の 2 次元分布

- a. 「継続期間」と「月あたり料金」で見ると、退会者は 60～100 ドルの高額者群、50 ドル前後、20 ドル前後の 3 つの塊がある。
- b. 在籍者では、「継続期間」が短期で「月あたり料金」が低い顧客群と、「継続期間」が長期で「月あたり料金」が高額な群と、「継続期間」が長期で「月あたり料金」が低額な群に分かれる。



(出所) 「Telco Customer Churn」 Data⁸を使い、「Project-McNulty」 Churn_Prediction.ipynb⁹を参考に当社作成

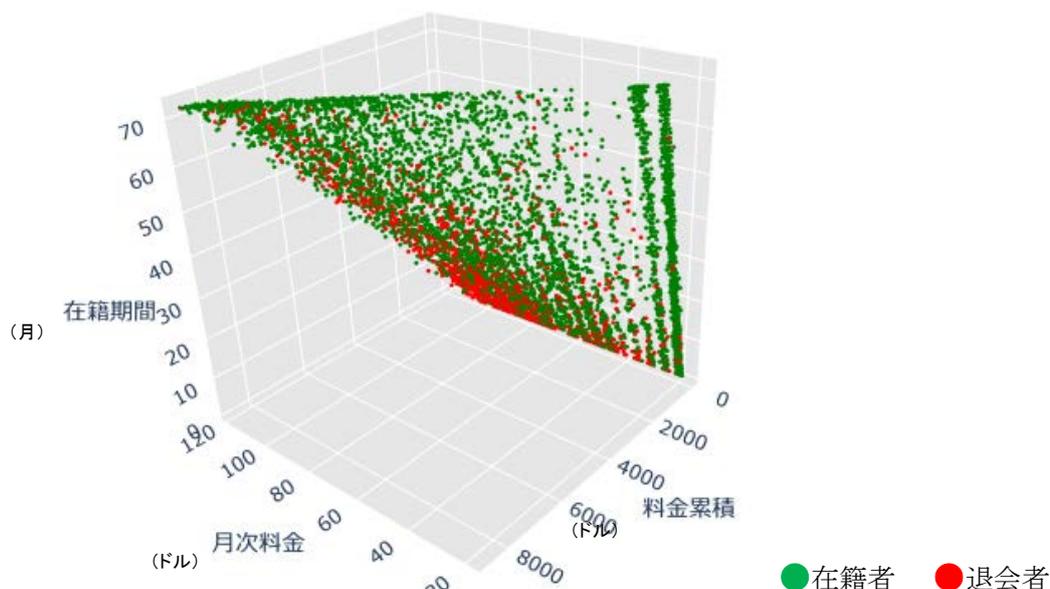
③ 量的変数の 3 次元分布

- a. 退会者は「継続期間」が短い場合にまんべんなく分布しているように見える。また「月あたり料金」が 80～100 ドルでは「継続期間」にかかわらず、退会者が分布している。
- b. 在籍者は全体に分布しているが、「月あたり料金」が 20 ドル程度や、約 80 ドル以下で一定の「継続期間」を経過すると退会者が少ない。

⁸ <https://www.kaggle.com/blastchar/telco-customer-churn> (2020 年 8 月 25 日確認)

⁹ <https://github.com/JNYH/Project-McNulty> (2020 年 8 月 25 日確認)

【図表 5】 量的特徴量の 3 次元分布



(出所) 「Telco Customer Churn」 Data¹⁰を使い「pavanraj159/telecom-customer-churn-prediction」¹¹を参考に当社作成

④ カテゴリ型特徴量の比較¹²

- a. 全ての特徴量のうち、カテゴリ型データ¹³の一部を抜粋した。例えば左上の gender(性別)のカテゴリでは Female(女性)の在籍者が約 2,600 人に対して退会者が約 900 人おり、男性と比較して退会者の割合に大差はない。
- b. 在籍者に対して退会者が多いのは、「インターネットサービス：ファイバー」「オンラインセキュリティ：なし」「オンラインバックアップ：なし」「デバイス保障：なし」「テクニカルサポート：なし」「支払方法：電子チェック」「紙請求：あり」「契約が月次タイプ」である。

¹⁰ <https://www.kaggle.com/blastchar/telco-customer-churn> (2020年8月25日確認)

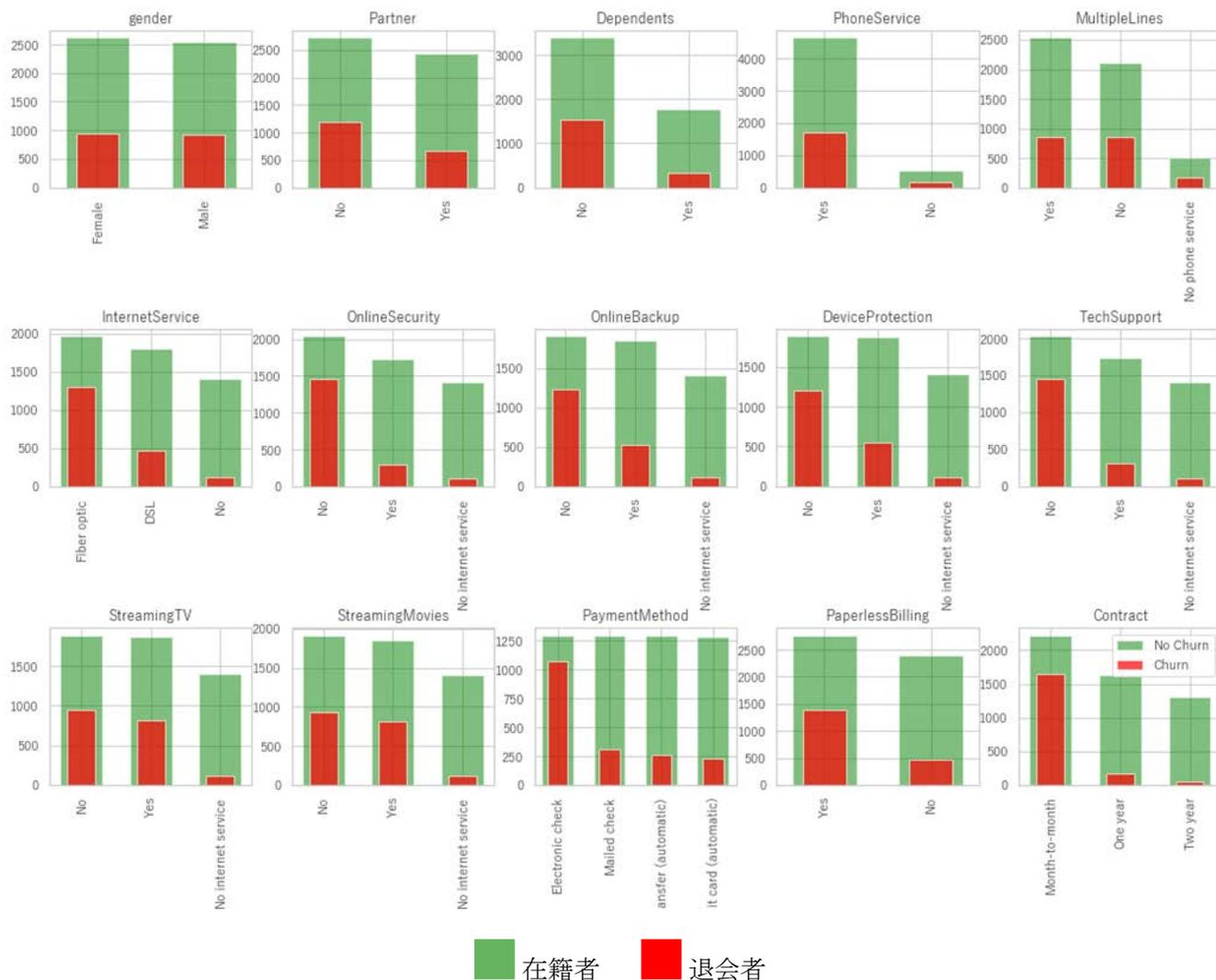
¹¹ <https://www.kaggle.com/pavanraj159/telecom-customer-churn-prediction> (2020年8月25日確認)

¹² <https://github.com/JNYH/Project-McNulty> (2020年8月25日確認) を参考にして分析・加工

¹³ 性別(gender)の男性・女性などのように種類に分けられているデータのこと

【図表 6】 カテゴリ型特徴量(一部抜粋)と在籍者・退会者

(人数)



(出所) 「Telco Customer Churn」 Data¹⁴を使い、「Project-McNulty」 Churn_Prediction.ipynb¹⁵を参考に当社作成

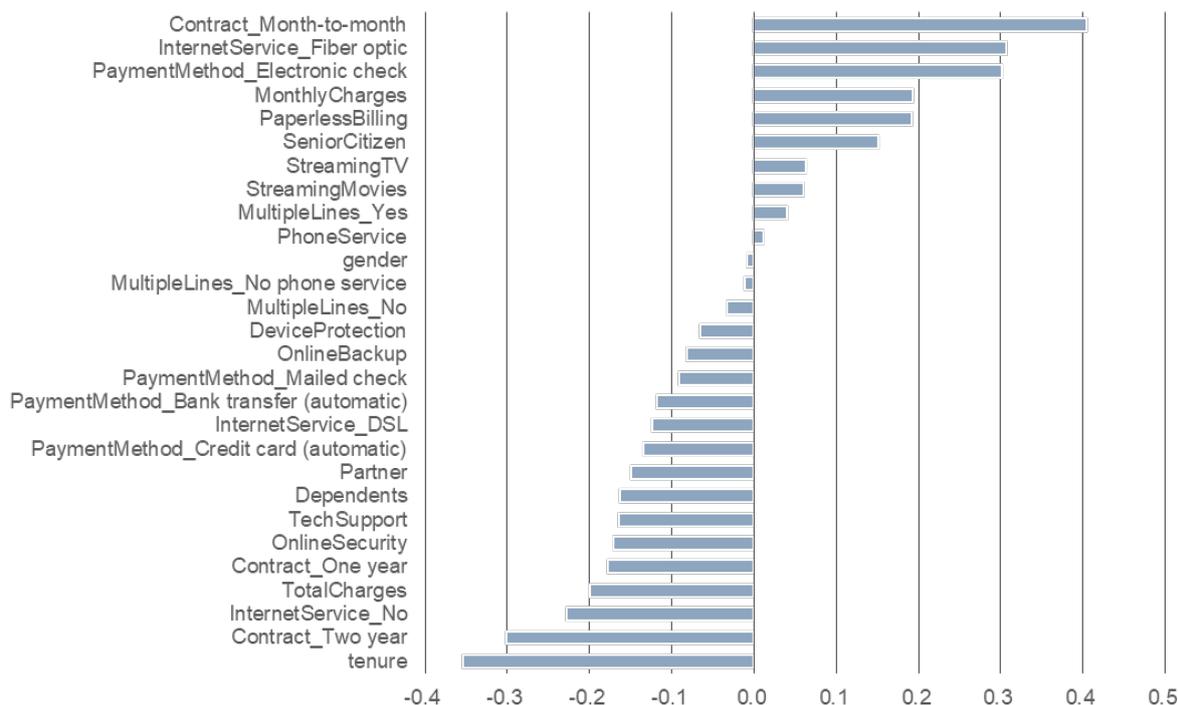
⑤ 退会と相関係数の高い特徴量

- 各特徴量の相関係数の数字が大きいと退会率が高く、小さいと退会率が低い。
- 相関係数の数字が大きいの、「Contract Month to month (月次契約) (0.40)」「Internet Service Fiber optic (光ファイバー) (0.31)」「Payment Method Electric check (電子チェック支払) (0.30)」などである。小さいのは、「tenure(継続期間) (-0.35)」「Contract Two year (2年契約) (-0.30)」「InternetService No (インターネットサービスなし) (-0.22)」などである。

¹⁴ <https://www.kaggle.com/blastchar/telco-customer-churn> (2020年8月25日確認)

¹⁵ <https://github.com/JNYH/Project-McNulty> (2020年8月25日確認)

【図表 7】 特徴量と退会の相関係数



(出所) 「Telco Customer Churn」 Data¹⁶を使い、当社作成

(3) 特徴量の加工

① 新しいカテゴリ変数の作成

アルゴリズムによってはカテゴリ変数との相性がよい場合があり、今回は 3 つの数量データをカテゴリデータにして分析した。

② 均衡データの作成

このデータセットは退会者 (約 0.26) と非退会者 (約 0.74) の比率がアンバランスである。機械学習上データのバランスを確保する方がよい場合があり、SMOTE¹⁷のテクニックで均衡データを作成した。

2. 【サンプルデータによる AI 分析評価】ステップ[2]AI の作成と評価¹⁸

(1) 使用する AI の絞り込み

- ① データセットとアルゴリズムの相性の良さを示す指標を決め、今回はロジスティック回帰、ランダムフォレスト、ナイーブベイズ等を試した。
- ② アルゴリズムと特徴量の相性 (F1 値 (precision と recall の調和平均) のよいもの) から今回は「ロジスティック回帰¹⁹ (複数の変数から 0~1 の間の確率を出す)」をアルゴリズムに選んだ。

¹⁶ <https://www.kaggle.com/blastchar/telco-customer-churn> (2020 年 8 月 25 日確認)

¹⁷ Synthetic Minority Oversampling Technique

¹⁸ <https://medium.com/analytics-vidhya/using-machine-learning-models-to-predict-customer-turnover-7d52db91d459>、<https://github.com/JNYH/Project-McNulty> (2020 年 8 月 25 日確認)を参考にして分析・加工

¹⁹ <https://ja.wikipedia.org/wiki/%E3%83%AD%E3%82%B8%E3%82%B9%E3%83%86%E3%82%A3%E3%83%83%E3%82%AF%E5%9B%9E%E5%B8%B0> (2020 年 9 月 30 日確認)

(2) パラメータチューニング(特徴量の調整)

- ① さらに F1 値 (precision と recall の調和平均) を最大化する C 値 (ロジスティック回帰のハイパーパラメータ) に注目して、ハイパーパラメータのチューニングを行った。C=0.18、F1 値は 0.65 となった。
- ② 退会可能性の高い顧客へのアプローチ数が大切であれば、退会判定の閾値を下げ、再現率 (recall) (退会予測者÷実退会者) を重視することになる。退会予測の精度が大切であれば、適合率 (precision) (退会予測者のうち実際に退会したものの割合) を重視することになる。

(3) AI の全体性能

- ① パラメータチューニング後の指標は、予測確率に対する閾値を 0.32 にした場合、正答率 (accuracy) 0.80、適合率 (precision) 0.57、再現率 (recall) 0.78、F1 値 0.66 となった。
- ② すなわち約 0.80 (80%) の確率で退会するかしないかの予測が当たる。しかし、退会者が約 0.26 (26%) を占めるので、全員を在籍者と予測しても約 0.74 (74%) の正解率となり、このデータセットによる退会予測率の向上は全体で約 6 ポイント (=80%−74%) と考えられる。

3. 【サンプルデータによる AI 分析評価】ステップ[3]AI による分析

(1) 個別の会員ごとの退会率の計算

- ① ロジスティック回帰は予測対象になるすべての会員の退会率を予測する。ある会員の退会可能性は、0.697 などと、確率として算出できる。
- ② 二値分類 (今回は退会「1」、在籍「0」の 2 種類に分類) の場合は、AI の計算する「退会可能性」の値に閾値を設定し、それ以上の場合は「退会」、それ未満の場合は「退会しない」と分類している。例えば、退会可能性が 0.697 のこの会員は、閾値が 0.5 の場合「退会」と判定されるし、閾値が 0.75 の場合「退会しない」と判定される。

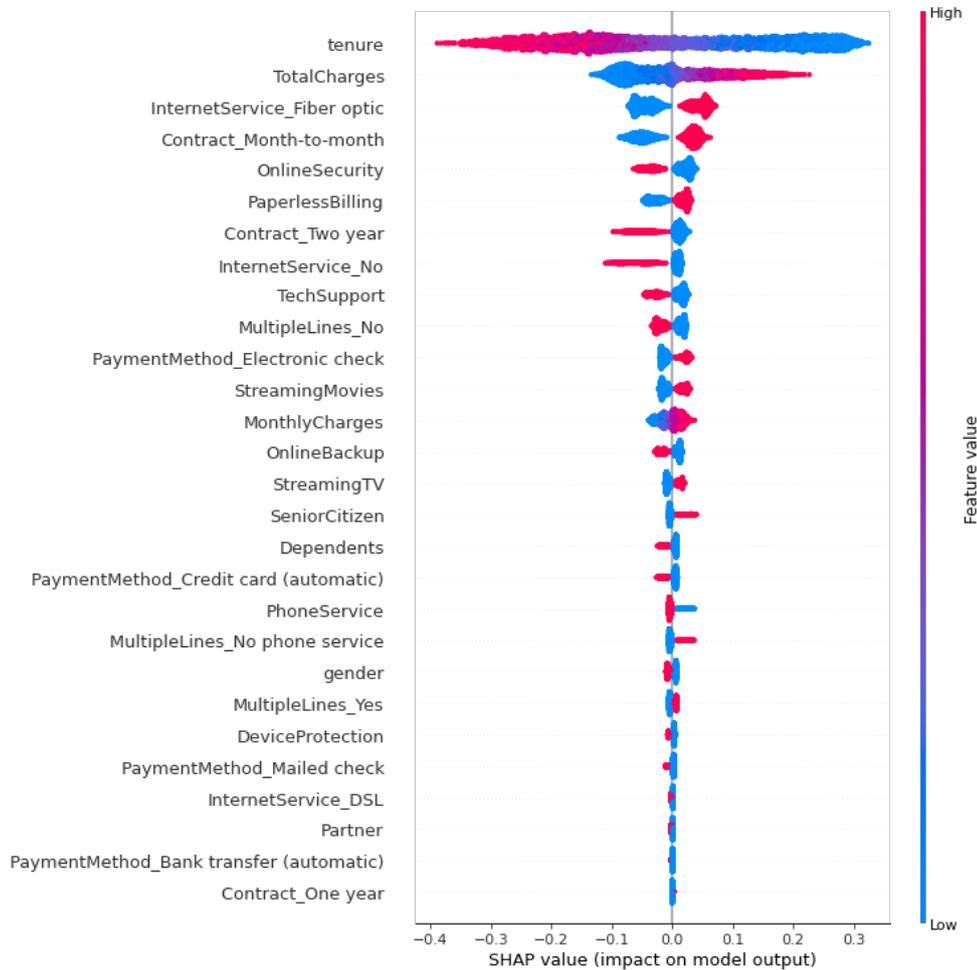
(2) AI が重視する「特徴量」

- ① AI がどのように退会確率を判断しているのだろうか。「SHAP²⁰」を使い、どの特徴量がどのように予測結果に影響を与えるか調べた。「SHAP」は、ゲーム理論を使い、AI 予測に対する各特徴量の寄与度を説明することができる。
- ② 次の図表は「(全体) 特徴量が結果に与える影響」を示している。横軸は退会確率に与えるインパクトで、右側 (プラス) になるほど退会確率が高まる。退会に影響を与えやすい変数は左右に大きく広がる。赤色は、量的特徴量が多いまたはカテゴリ変数が「1」(カテゴリに該当) であることを示す。青色は、量的特徴量が少ないまたはカテゴリ変数が「0」(カテゴリに非該当) であることを示す。グラフの形を見ると各変数と退会確率影響の関係が分布で示されている。
- ③ 今回の分析ケースに当てはめてみると、各特徴量は AI の退会予測確率にどんな影響を与えるのだろうか。影響度の大きい上位 4 つの特徴量を見ると、「tenure(在籍期間)」が短く、「Total Charges(総支払額)」が多く、「Internet Service Fiber optic (光ファイバー)」があり、「Contract month to

²⁰ SHAP <https://github.com/slundberg/shap> (2020 年 9 月 30 日確認)

month（月次契約）」があると、退会確率が上昇するという結果となった。

【図表 8】（全体）特徴量が結果に与える影響



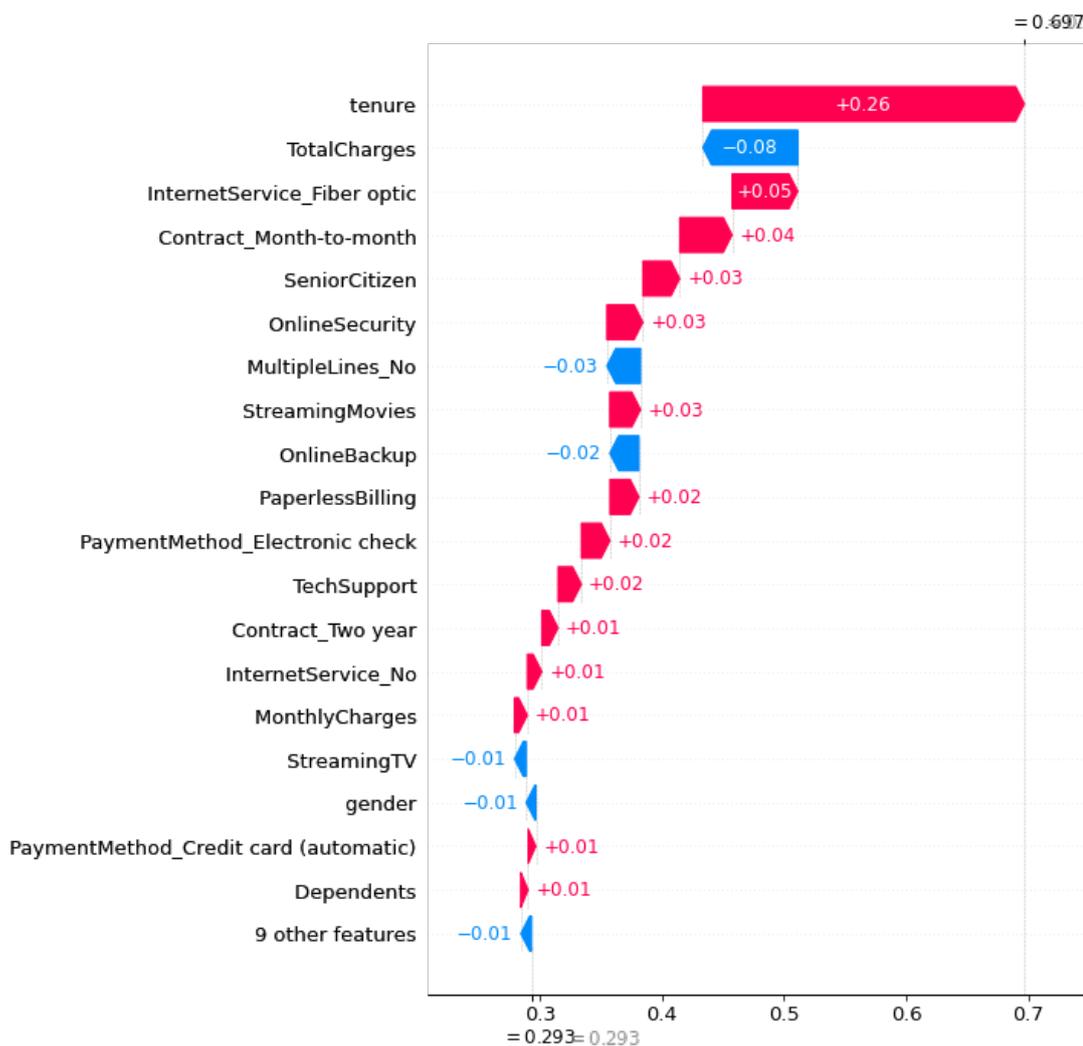
（出所）「Telco Customer Churn」 Data²¹を使い当社作成

- ④ 下記図表は「ある会員の事例で、特徴量が結果に与える影響」である。各特徴量のインパクトを積み上げて、平均確率と予測確率の関係を示している。横棒グラフの特徴量のプラス（赤色）は予測退会確率を上げ、特徴量のマイナス（青色）は予測退会確率を下げる。
- ⑤ 分析サンプル全体の退会確率の平均は約 0.293²²で、この会員の退会確率は約 0.697 である。この会員の退会確率を押し上げているのは、「tenure（在籍期間）の短さ」が 0.26 ポイント、「Internet Service Fiber optic（光ファイバーあり）」が 0.05 ポイント、「Contract month to month（月次契約あり）」が 0.04 ポイントである。押し下げているのは「Total Charges(総支払額)の少なさ」で、 $\Delta 0.08$ ポイント分である。

²¹ <https://www.kaggle.com/blastchar/telco-customer-churn>（2020年8月25日確認）

²² 退会確率の平均値は、今回の SHAP の計算ではサンプルの一部を使っているため、全サンプルの平均とは異なる。

【図表 9】（ある会員の事例）特徴量が結果に与える影響



※ 赤は退会率が上昇、青は退会率が下降

※ 横軸では右に行くほど退会率は上昇

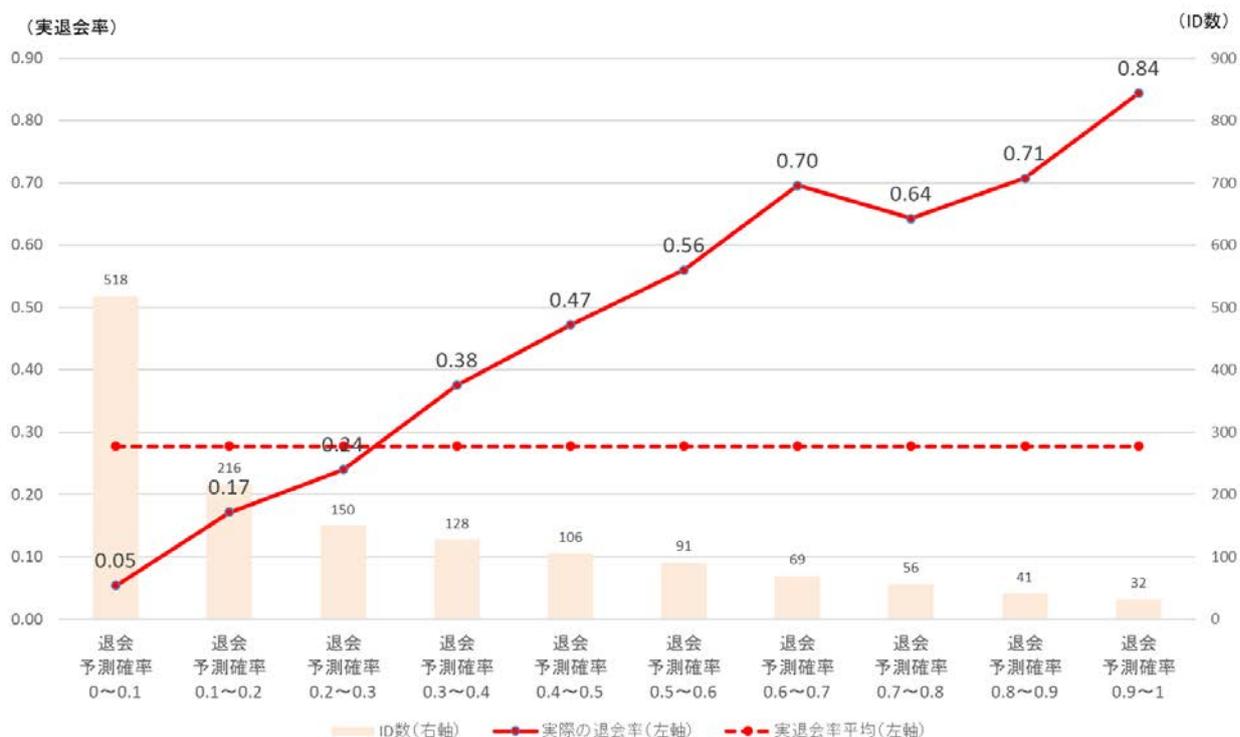
(出所)「Telco Customer Churn」Data²³を使い当社作成

(3) AI の個別退会確率の利用

- ① AI は個々の ID ごとの退会確率を計算する。下記のグラフは「退会予測確率」別に実際の退会率をグラフにしたものである。例えば退会予測確率が 0.9~1.0 となる 32 人の場合、実退会率は 0.84 であり、非常に高い精度で退会が予測されている。逆に退会予測確率が 0.0~0.1 の 518 人の場合、実退会率は 0.05 でありこちらの精度も非常に高い。
- ② 退会予測確率が 0.2~0.3 の場合、実際の退会率は 0.24 である。AI も判断を迷っているようであり、実際の結果も平均退会率の約 0.26 と大きな差がない。
- ③ 個別退会確率を利用すれば、後述するように、似たような退会予測確率を持つ会員ごとにグループングして、各グループに適した施策を打つことは有用である。

²³ <https://www.kaggle.com/blastchar/telco-customer-churn> (2020 年 8 月 25 日確認)

【図表 10】 退会予測確率と実退会率・ID 数



(出所) 「Telco Customer Churn」 Data²⁴を使い当社作成

4. 【サンプルデータによる AI 分析評価】 ステップ[4] 施策の検討

(1) 退会抑止コストの効率化

- ① 施策検討例として、退会確率の高い会員への割引や利用を促す施策などが考えられる。しかし、一般的には退会を引き留める施策に費やすコストや時間資源には限界がある。そこで、AI 予測を使って、退会確率がより高い会員に施策を絞れば、費用対効果を上げることができる。
- ② 例えば、今回のデータセットでの平均退会率は 0.26 であるが、一定の閾値での退会予測の精度が向上した場合、会員一律に割引策を打つよりも、費用対効果が上昇する。おおまかな計算となるが、退会予測確率が 0.9~1.0 の場合、実際の退会率は平均 0.26 から 0.84 に上がる。つまり、 $0.26 \div 0.84 = \text{約 } 31\%$ のコストで退会しそうな会員へのアプローチが可能となるという考え方だ。会員や構成員の退会抑止に人的な働きかけが必要な場合、働き方改革や労働生産性の向上にもつながる。

(2) セグメント分析

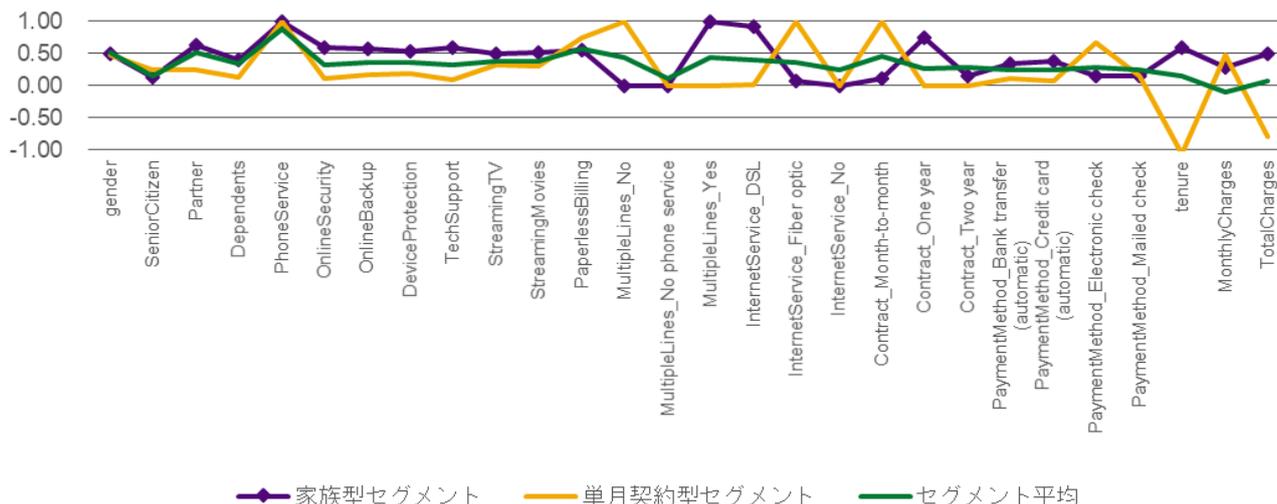
- ① 顧客情報から顧客セグメントを作り出し、その顧客セグメントにあった施策ができる。
- ② 今回は各特徴量の標準化を行った後、PCA（主成分分析）で主成分を 10 まで削減したデータを使い、K-means のアルゴリズムを使って、20 のセグメントを作った。
- ③ 20 セグメントのうち、退会率が高い 2 つと全体平均を下記に図示した。例えば「単月契約型セグメント」は、月次契約で加入年月が短い割に月の料金が低い。また「家族型セグメント」では 1 年契約・複数家族・各種インターネットサービスや電話サービスを利用・自動引き落としの割合が多

²⁴ <https://www.kaggle.com/blastchar/telco-customer-churn> (2020 年 8 月 25 日確認)

い。どちらのセグメントも退会確率が高いが、顧客特性を踏まえた施策にするのがよい。

- ④ AIでセグメントを新たに作れば新しい示唆を得られるかもしれないが、その場合前述のEDA（探索的データ解析）の結果を組み合わせると有用であろう。

【図表 11】 退会率の高いセグメントの特徴量



(出所) 「Telco Customer Churn」 Data²⁵を使い当社作成

(3) 「対処」系と「抜本」系の施策

- ① 退会抑制策を「対処」系と「抜本」系に分けて考える。
- ② 「対処」系施策は、退会しそうな会員に対する即効策で、クーポンや値下げプランやその他特典の付与などが考えられる。上記の「単月契約型セグメント」などが対象となる。
- ③ 「抜本」系は、契約時の提案方法、プランの変更、利用促進、満足度向上策などが考えられる。上記の「家族型セグメント」は、多くのサービスを1年以上利用しているケースが多い。サービスの内容や価格も含め、退会要因をより深耕すべきであろう。

(4) データセットの特徴量の増強

- ① 今回のケースは与えられた20の特徴量を使って、会員の退会可能性を予測している。AIは複数のアルゴリズムの中より精度の高いものを選んでいく。ただ同じデータセットを使っている限り、AI精度の向上には限りがある。分析会員数を増やす方法もあるが、今回は特徴量の種類を増やすことを考えてみる。
- ② 追加候補となる特徴量
 - a. 「単月契約型セグメント」に分類される顧客が、他のセグメントより多くのサービスに加入しているのはなぜだろうか。勢いで入会した人も多いのかもしれない。また入会時特典が大きかったのかもしれない。入会時に注目したデータがあれば分析が深まるのではないか。例えば、「入会の方法」、「入会の動機」、「入会時のキャンペーンの有無」などの特徴量があれば、予測

²⁵ <https://www.kaggle.com/blastchar/telco-customer-churn> (2020年8月25日確認)

精度が高まる可能性がある。

- b. 「家族型」セグメントは、サービスに不満があるか価格が高すぎると感じていないか。例えば、「サービスの利用頻度」・「サービスの利用量」・「顧客満足度アンケート結果」などの特徴量が有用ではないか。

(5) AI 以外の調査

- ① 価格の高さの退会への影響度は高い。セグメントごとに「高過ぎない価格」と「契約形態」・「サービス内容」を探るためにコンジョイント分析を行うのもよいであろう。また、オーソドックスであるが、定期的なサービス満足度調査や、退会者へのインタビューは有用である。
- ② 施策案ができれば、顧客にどの施策がよいかを評価してもらい、その有効度を図る調査も効果的であろう。

－ ご利用に際して －

- 本資料は、信頼できると思われる各種データに基づいて作成されていますが、当社はその正確性、完全性を保証するものではありません。
- また、本資料は、執筆者の見解に基づき作成されたものであり、当社の統一的な見解を示すものではありません。
- 本資料に基づくお客様の決定、行為、及びその結果について、当社は一切の責任を負いません。ご利用にあたっては、お客様ご自身でご判断くださいますようお願い申し上げます。
- 本資料は、著作物であり、著作権法に基づき保護されています。著作権法の定めに従い、引用する際は、必ず出所：三菱UFJリサーチ&コンサルティングと明記してください。
- 本資料の全文または一部を転載・複製する際は著作権者の許諾が必要ですので、当社までご連絡ください。