

東京大学グローバル消費インテリジェンス寄付講座 GCI 2022 Winter

最終課題

このスライドは機械学習初心者が作成したものです！

参考にしないでください！

状況設定

市場について

対象国：日本（1ドル＝130円とする）

対象商品：移動系通信（携帯電話）

総契約数：2億292万契約

A社について

A社はMNO（自社で回線網を有する事業者）とする

契約シェア：2.5%

契約台数¹：2億292万契約×2.5 %＝507万3000台

契約者数²：507万3000台÷1.25 台/人＝405万8400人

1. 1台につき契約とみなした

2. MMD研究所，調査データ，https://mmdlabo.jp/investigation/detail_1941.htmlより，2台持ちの利用者を12.7%として換算

提案

▶ 提案内容

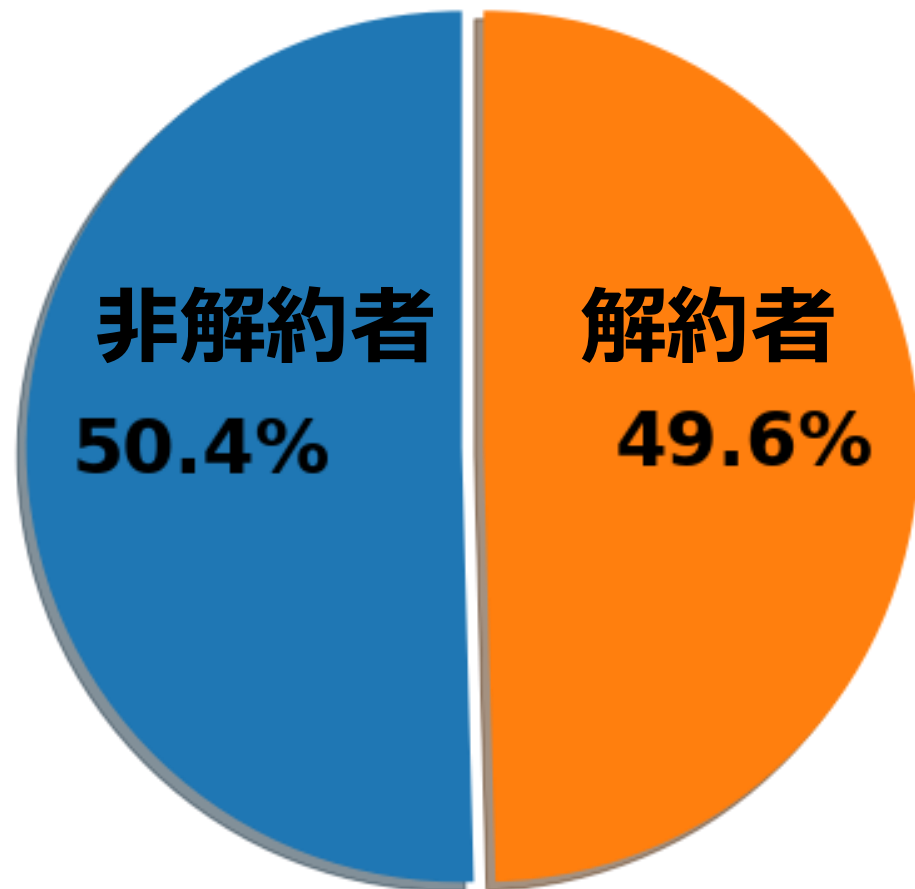
機械学習モデルによって解約者および非解約者を予測します。

予測された解約者の特徴を分析し事業を提案することで、解約者を減少させ、解約による損失額を減少させます。さらに、非解約者の特徴を分析し、支援を行うことで、今後の解約を防ぎます。

▶ 提案の流れ

1. 解約による損失の算出
2. 予測モデルの概要
3. 予測モデルの評価
4. 解約者/非解約者の特徴, 市場の現状
5. 事業の提案
6. 事業導入による効果

非解約者と解約者の割合



ご提供いただいた10万人のデータから、
非解約者 = 50400人
解約者 = **49600人**

全契約者数405万8400人に対しては、
非解約者 = 204.5万人
解約者 = **201.3万人**

であることが分かります。

解約による損失額の算出

ご提供いただいた顧客データから、解約者の過去6ヶ月間の平均月間収益は6227円でした.

$$\begin{aligned}\text{損失額*} &= \text{解約者に対する月の損失額の合計} \\ &= 201.3 \text{ 万人} \times 6227 \text{ 円} \\ &= \mathbf{125 \text{ 億円}}\end{aligned}$$

→解約者を**1%**減らすことで解約による損失約**1.25億円**を防ぐことができます.

*1ドル = 130円として換算

予測モデルの概要

提供いただいたデータに対し、LightGBM, XGBoost, Catboostの3種類のアルゴリズムを用いて予測しました。

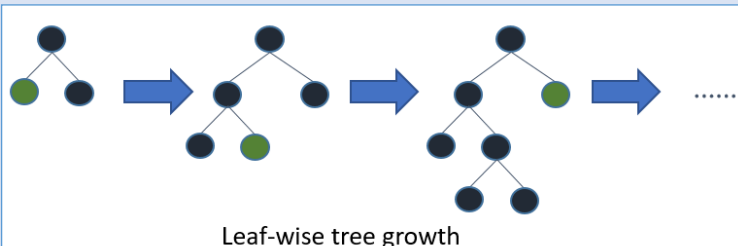
この3種は勾配ブースティング決定木という方法を用いたアルゴリズムです。

勾配ブースティング決定木とは*

決定木と勾配ブースティングを組み合わせた方法です。決定木とは木構造を用い、木の根から条件によって分岐しつつ葉に向かい、特定の葉に到達することで予測を行うものです。勾配ブースティングは、一つ前の決定木の誤りを次の決定木が修正するように繰り返しながら決定木を追加していくものです

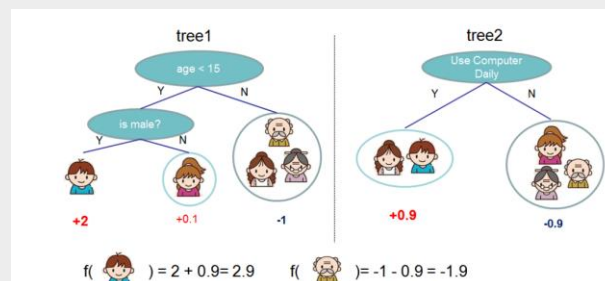
LightGBM^{[1][2]}

- XGBoostよりも訓練にかかる時間が短い。
- 大規模データに適している。
- 過学習しやすい
- Leaf-wiseを採用している（厳密な枝分かれを探さずに予測できる）



XGBoost^{[3][4]}

- 大規模データに対して高速で処理する
- 複数の弱い学習器を用い、前の学習の結果に基づいて重み付けをして強い学習機を作る



Catboost^{[5][6]}

- カテゴリカル変数（男女など性質を区別するための値）が多いデータに対して強い
- 過学習が少ない
- 1つ前のツリーで発生したエラーを減らしツリーを新しく構築する。エラーが少なくなるまでこれを繰り返す



予測モデルの評価

使用するモデルによって精度が異なります。よってモデルの評価をAUC（Area Under Curve）を用いて行いました。

各モデルのAUC	
モデル	AUC
LightGBM	0.697
XGBoost	0.687
CatBoost	0.700

AUCとは

分類のアルゴリズムの性能の良さを表す指標です。

0から1までの値をとり、完全な分類が可能なときは1で、ランダムな分類の場合は0.5になります。つまり、1に近づくほど、優れた分類モデルであるといえます。

*0.7~0.8のモデルは効果があると考えられています。

解約者/非解約者の特徴

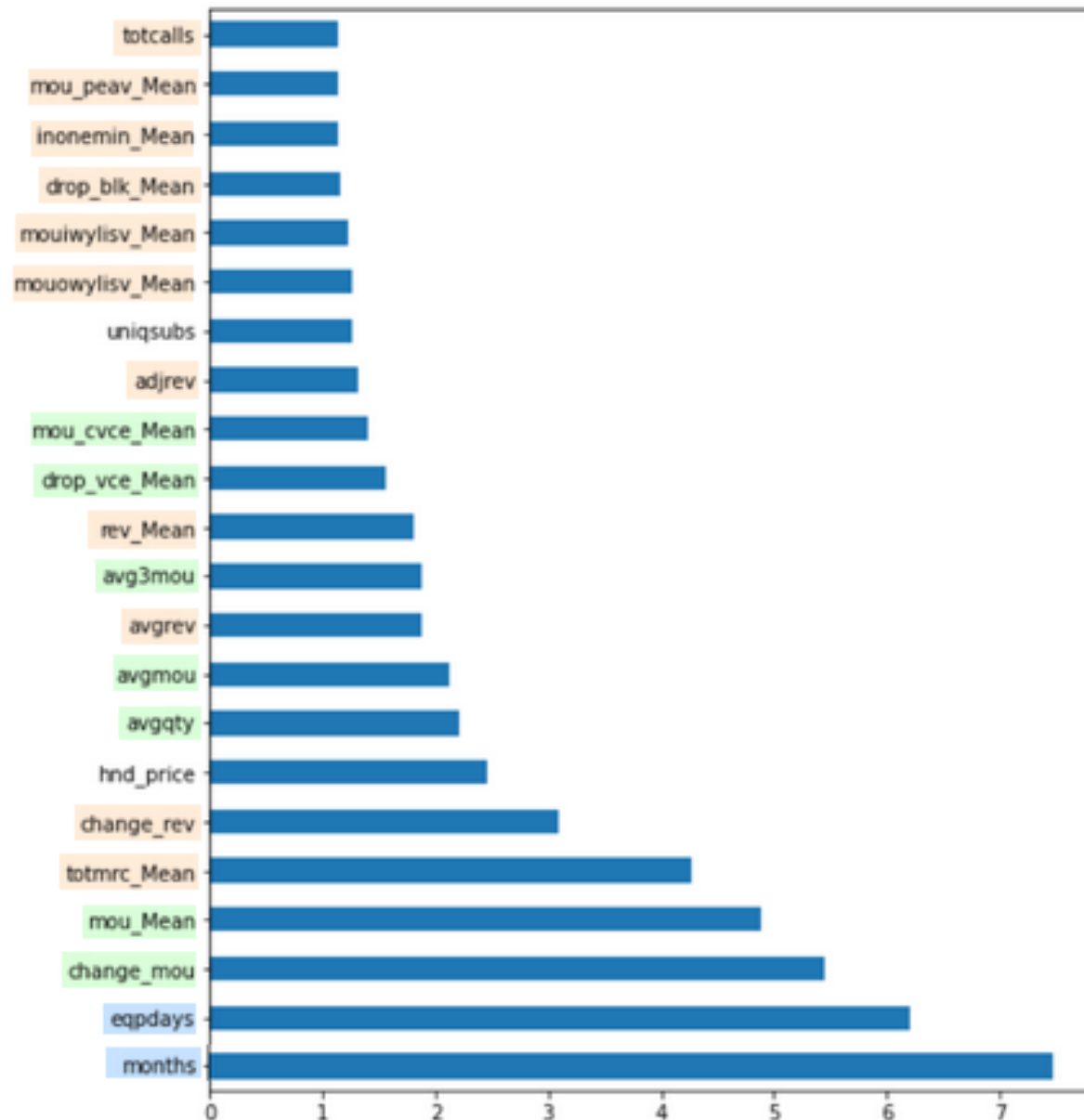
最もAUCが大きかったCatboostを用いて重要度を可視化しました

重要度の大きい項目は以下のように大別できます

- ・携帯電話の使用時間
- ・サービスの利用時間（契約期間）
- ・売上

また、機種価格（hnd_price）も上位10項目に入り、解約者を特徴づけると考えられます。

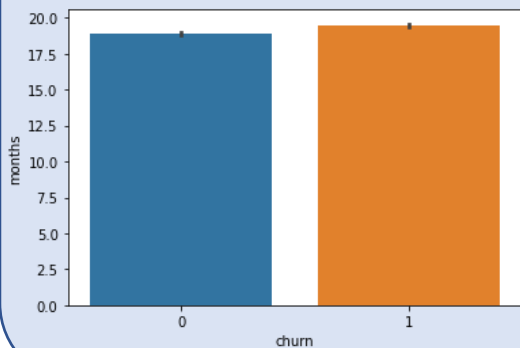
*使用時間は、一つの電話番号から算出されたものとみなします



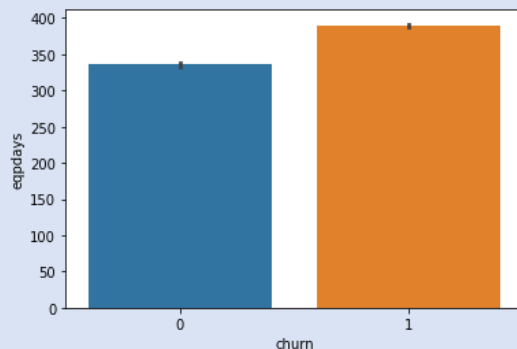
解約者/非解約者の特徴

重要度の高かった上位7項目について、**解約者**と**非解約者**の平均値を算出しました。

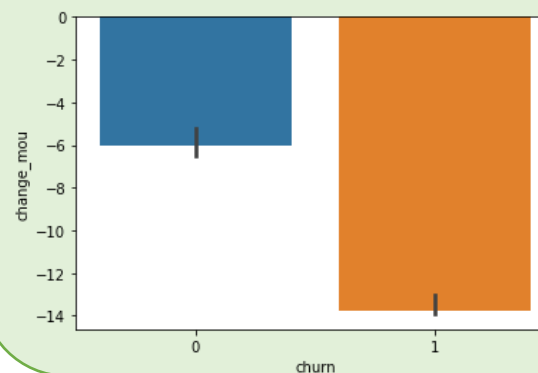
総利用月数



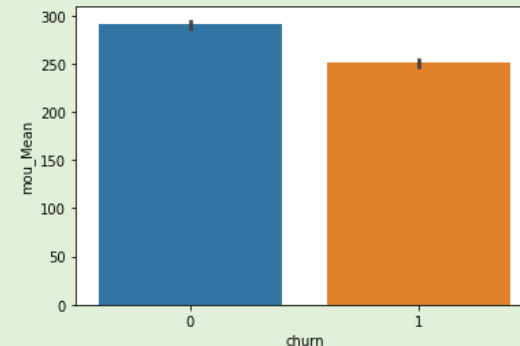
現有機種の日数



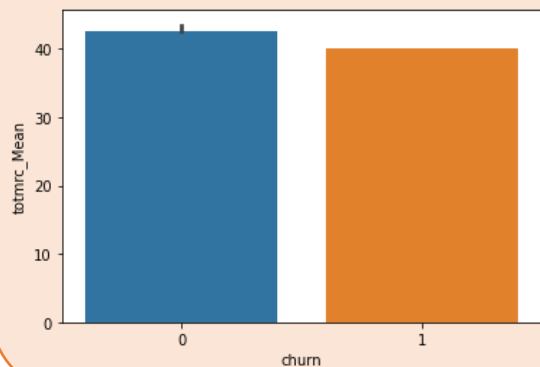
月間使用時間（分）の
対前3ヶ月平均の変化率



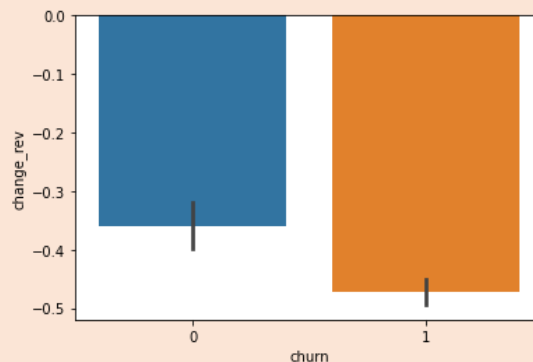
平均月間使用時間（分）



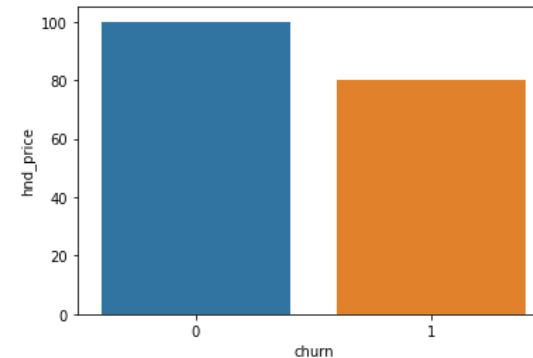
平均月間継続利用料総額



月間売上高対前3ヵ月平均の変化率



現在の携帯電話機の価格

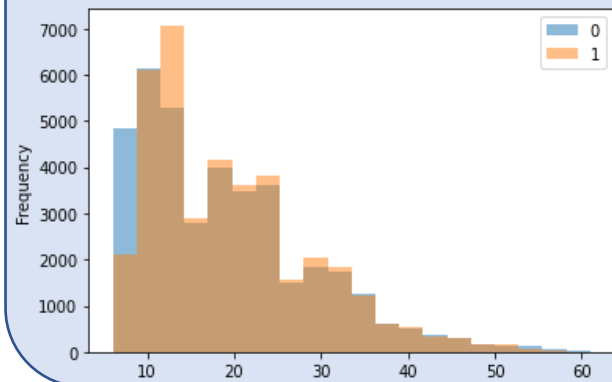


解約者は、使用時間が少なく、機種之年数が大きく、売上げの減少率が大い、継続利用料が安い、機種之値段が安いことが分かります。つまり、古い機種を長く使っており、最近使用しなくなったと考えられます

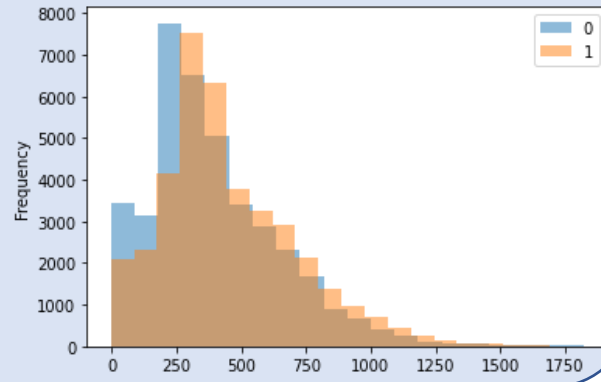
解約者/非解約者の特徴

重要度の高かった上位7項目について、**解約者**と**非解約者**の分布を算出しました。

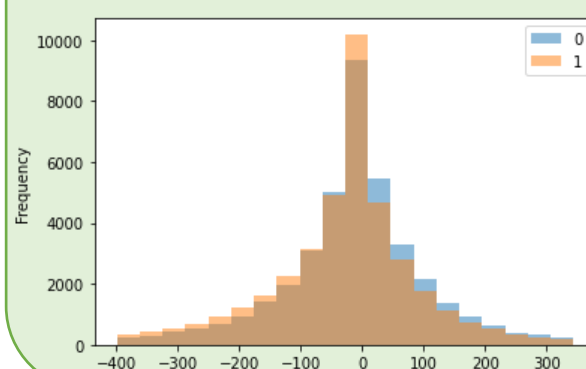
総利用月数



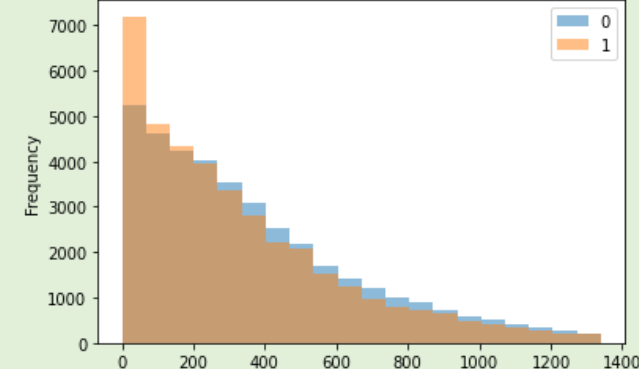
現有機種の日数



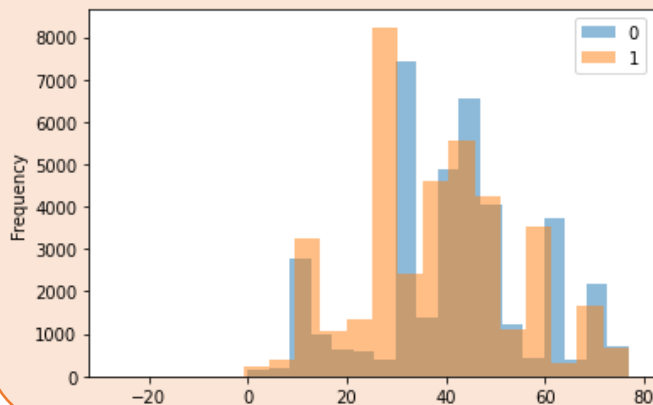
月間使用時間（分）の
対前3ヶ月平均の変化率



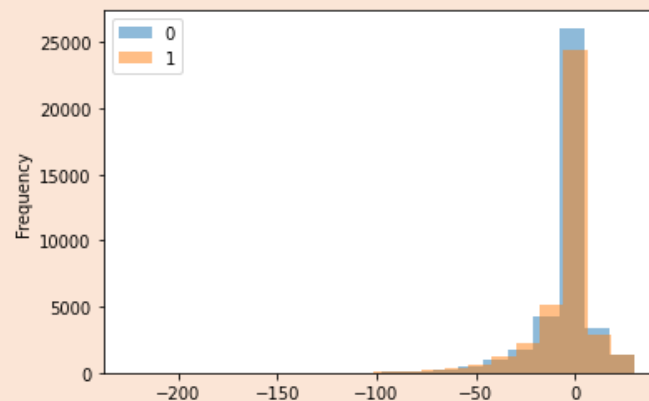
平均月間使用時間（分）



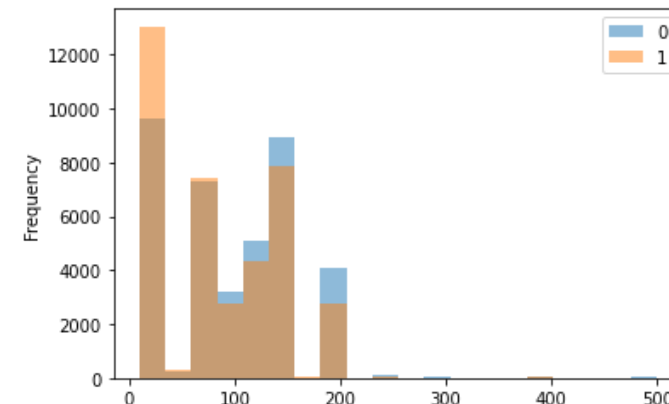
平均月間継続利用料総額



月間売上高対前3ヵ月平均の変化率



現在の携帯電話機の価格



総利用月数，現有機種の日数，平均使用時間についてある値を超えると解約者/非解約者の割合が逆転していることがわかります。

移動系通信市場の現状

契約数

- ・ 移動系通信の契約数は毎年増加している。
- ・ LTEの契約数は継続して減少しているのに対し、5 Gの契約数は、継続して増加している。
- ・ 契約数ベースでは、5 Gへの移行が進展している。

料金

- ・ 低容量低価格のミニ プランの提供が開始
- ・ 基本料金を0円とした上で、利用者が各自のニーズに応じてデータ通信容量 や音声通話のトッピングを組み合わせていることのできるプランの提供が開始。
- ・ 容量増量オプションが提供開始

利用者

- ・ 利用者が事業者を乗り換える動きがさらに活発化している（乗り換えコストの低下）
- ・ 契約解除数が新規契約数よりも大きくなっている（フィーチャーフォンの解約など）

参考：総務省,検討の方向性（案）について, https://www.soumu.go.jp/main_content/000818328.pdf

：総務省,電気通信事業分野における市場検証（令和3年度）年次レポート, https://www.soumu.go.jp/main_content/000835448.pdf

事業の提案

解約者に対する事業

解約者の予測を行い、解約すると予測された顧客に対して、最新機種を機種代金の半額で提供し、解約者を減らします。
解約者の特徴

非解約者に対する事業

非解約者と予測された顧客の中で、月3.5時間以上の使用を1年以上継続すると、5G機種への乗り換えをすると機種代金が30%引きになるプランを提供し、大容量プランを選択肢として追加することによって、解約を防ぎます。
機種料金と継続利用料金の増加を促す
使用時間の変化率と売り上げの変化率の減少を防ぐ

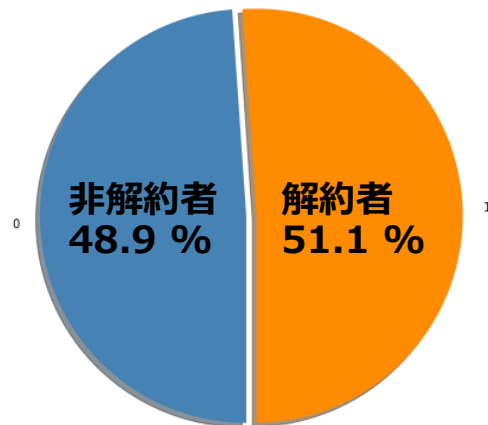
事業の反映

解約と予測された顧客の以下のデータ（重要度の高い7項目）を更新します。

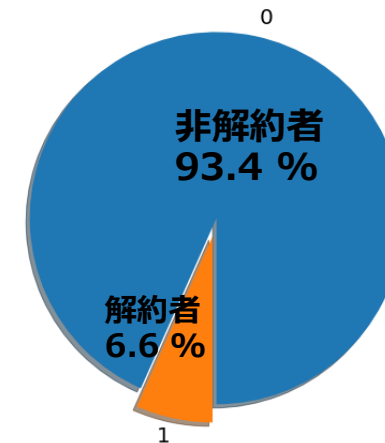
使用日数，総利用月数，現有機種の日数，平均使用時間以外は，非解約者の各項目の75パーセンタイルを用います。

- 総利用月数 5ヵ月
- 端末の使用日数 0日
- 月使用時間の変化率 +43 %
- 月の平均使用時間 523 分
- 月の継続利用料 6500円（50ドル）
- 月間売り上げの変化率 0.36 %
- 端末価格 19370円（149ドル）

事業導入前の予測



事業導入後の予測



*10万件のデータを学習データ8万件，検証データ2万件に分割し，2万件の予測を行いました

事業導入による効果

解約者

$$\begin{aligned}\text{解約者向け事業のコスト} &= \text{解約者と予測された機種代の平均の50\%} \times \text{解約者の人数} \\ &= 11700 \times 0.5 \times 49600 = 2.9 \text{億円}\end{aligned}$$

非解約者

$$\begin{aligned}\text{非解約者向け事業のコスト} &= 5G \text{の機種代の平均}^* \text{の30\%} \times \text{条件の非解約者の人数} \\ &= 100,000 \text{円} \times 0.3 \times 10161 \text{人} = 3 \text{億円}\end{aligned}$$

合計

$$\begin{aligned}\text{事業導入によるコスト} &= \text{解約者向けの事業にかかるコスト} + \text{非解約者向けの事業にかかるコスト} \\ &= \mathbf{5.9 \text{ 億円}}\end{aligned}$$

事業導入後、解約率を6.6%（解約者を1320人）まで減らせるので、

$$\begin{aligned}\text{事業導入によるコスト} + \text{解約による損失} \times A \text{ヵ月} &= \text{解約を防いだ人数分の月の利益} \times A \text{ヵ月} \\ 5.9 \text{ 億円} + (6227 \times 1320 \text{人}) \times A \text{ヵ月} &= 18680 \text{人} \times 6227 \times A \text{ヵ月} \\ A &= \mathbf{5.5 \text{ヵ月}}\end{aligned}$$

→**5.5ヵ月**で回収が可能となります。

*日経クロステック、裏読みスマホ料金, <https://xtech.nikkei.com/atcl/nxt/column/18/01112/101300010/>

参考

可視化

- TH, Python でデータサイエンス, <https://pythondatascience.plavox.info/>

データ分析の方法

- note.nkmk.me,, <https://note.nkmk.me/>
- chatGPT, <https://openai.com/blog/chatgpt/>
- HTOMblog, XGBoostの実装方法と特徴量重要度, <https://htomblog.com/python-xgboost>

アルゴリズム

- [1]Microsoft Corporation., LgihtGBM documentation, <https://lightgbm.readthedocs.io/en/v3.3.2/>
- [2] codexa.net, LightGBM 徹底入門, <https://www.codexa.net/lightgbm-beginner/>
- [3]xgboost developers, XGBoost Documentation, <https://xgboost.readthedocs.io/en/stable/>
- [4] S-Analysis, PythonでXgboost, <https://data-analysis-stats.jp/%E6%A9%9F%E6%A2%B0%E5%AD%A6%E7%BF%92/python%E3%81%A7xgboost>
- [5] Yandex, CatBoost documentation, <https://catboost.ai/>
- [6] S-Analysis, PythonでCatBoostの解説, <https://data-analysis-stats.jp/python/python%E3%81%A7catboost%E3%81%AE%E8%A7%A3%E8%AA%AC/>

事業後の効果の検証

- UCHIDA Tomomasa, https://github.com/tom-uchida/GCI2020_Winter