

2020年11月24日

コンサルティングレポート

AIによる来店客数(売上)予測

シリーズ「事業経営へのAI活用」②

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
- ②AIによる来店客数(売上)予測 ……<本稿>
- 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による来店客数予測」を取り上げる。来店客数に顧客単価を乗ずれば売上の予測もできるが、今回は来店客数をメインに検討した。日々の来店客数が鍵となるビジネスモデルにとって、その予測は大きな課題の1つである。AIを活用してデータを分析評価し、マス(集団)としてではなく個々の店舗の傾向を把握し、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。

本稿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. 来店客数予測におけるAI活用

1. 要旨

- (1) 「AIによる来店客数予測」では、予測したい店舗や商品などに関連する情報や過去の来店客数データを利用して、AIで各店舗・商品ごとの予測来店客数を計算し、来店客数と関係の強い要素の分析などを行い、来店客数・売上・利益増加のための施策を検討・立案・実行する。
- (2) 今回は機械学習コンペで有名なKaggleのデータセット¹を使い、「AIによる来店客数予測」を考察する。まずデータセットの分析後、LightGBMで来店客数予測を行い、データセットの分析・修正後、AIの作成・評価を行い、精度を高めたものを使って各種予測・分析を行った。

¹ <https://www.kaggle.com/c/recruit-restaurant-visitor-forecasting> (2020年8月25日確認。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では可能であった。以下同じ。)

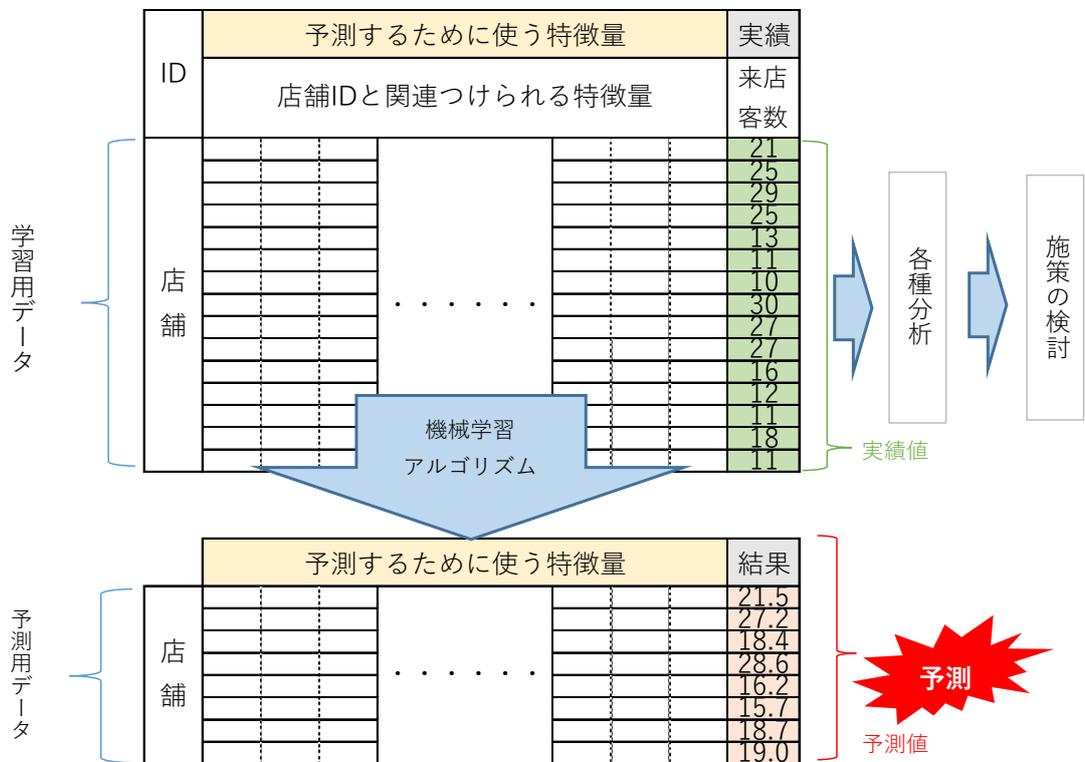
- (3) このデータセットからさらに AI 予測の精度を高めるには、特徴量（AI が学習・分析する顧客情報、時系列の天候や顧客単価などデータの種類）の増加が必要と思われる。
- (4) 来店客数予測の精度が上がれば、「無駄な仕入の削減」や「適切なシフトによる人件費の適正化」につながり、利益が改善する。また来店客数予測に影響を与える要因の分析により、売上増のヒントを見いださう。

2. AIによる来店客数予測・分析

(1) 概要

- ① 来店客数を予測したい店舗の情報や、過去の来店客数等の学習用データを利用する。AI で各店舗・商品ごとの予測来店客数を計算し、来店客数と関係の強い要素の分析などを行い、来店客数・売上・利益増加のための施策を検討・立案・実行する。
- ② 分析に必要な学習用データは主に「来店客数予測対象に関連した情報」、「来店客数に影響のある情報」と「過去の来店客数情報」である。
- ③ 一般的に予測・分析の精度が高くなるのは、学習データ（来店客数予測の場合は店舗数）が多い場合、来店客数に関連した情報・来店客数の増減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による売上(来店客数など) 予測イメージ



(出所) 当社作成

(2) 業種例

- ① 時系列の売上・来店客数データがあれば、多くの業種で同様の予測分析が可能となる。特に Bto (Bto) C の場合、利用できる特徴量が多く、大型受注型の業種と比べて、予測精度や分析効果が高まる。例えば受注・売上とのタイムラグがある特徴量（先行指標）を利用すれば予測精度を上げることができる。
- ② 時系列データを使用した将来予測という枠組みの分析なので、交通事故数や河川水位の年次推移などでも同様に分析が可能である。

(3) AI 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」、「データセットの拡充」、「ビジネスで有用な施策に結び付けた検討」が欠かせない。
- 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・広告系・企画系・出店戦略系など、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である。
- 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、プロジェクトに深くかかわり、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。特に店舗目線での特徴量の創出は鍵となる。
- ④ 評価方法としては、まず「AI 予測」を「既存の予測」と比較し、どれだけ精度が高まるかを試算する。そして、予測精度の向上によるインパクトを営業利益に換算するとわかりやすいであろう。AI 予測対既存予測ではなく、これまでの既存予測のやり方を AI の活用によって一層高めていくという姿勢で協働して予測することが重要である。

II. サンプルデータによる AI 分析評価

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

(1) 内容(特徴量)

- ① 使用するデータセットは、飲食店に関するもので、加盟店舗の場所、来店客数・予約客数、休日情報などからなる。
- ② 今回は 2016 年 1 月 1 日～2017 年 4 月 22 日のデータを使い、2017 年 3 月を中心とした 39 日（閉店日を除く）×829 店舗の来店客数を予測する。
- ③ 店舗情報には、「店舗 ID(store_id)、ジャンル(genre)、地域名(area_name)、緯度(latitude)、経度(longitude)などがあり、店舗別に来店客数・時間、予約客数・時間がわかる。休日情報(holiday_flg)は休日を示す。

【図表 2】 データセット(一部抽出)

index	store_id	hpg_genre_name	hpg_area_name	latitude	longitude
0	hpg_6622b62385aec8bf	Japanese style	Tōkyō-to Setagaya-ku Taishidō	35.6436747	139.668221
1	hpg_e9e068dd49c5fa00	Japanese style	Tōkyō-to Setagaya-ku Taishidō	35.6436747	139.668221
2	hpg_2976f7acb4b3a3bc	Japanese style	Tōkyō-to Setagaya-ku Taishidō	35.6436747	139.668221
3	hpg_e51a522e098f024c	Japanese style	Tōkyō-to Setagaya-ku Taishidō	35.6436747	139.668221
4	hpg_e3d0e1519894f275	Japanese style	Tōkyō-to Setagaya-ku Taishidō	35.6436747	139.668221

index	store_id	visit_datetime	reserve_datetime	reserve_visitors
0	hpg_c63f6f42e088e50f	2016-01-01 11:00:00	2016-01-01 09:00:00	1
1	hpg_dac72789163a3f47	2016-01-01 13:00:00	2016-01-01 06:00:00	3
2	hpg_c8e24dcf51ca1eb5	2016-01-01 16:00:00	2016-01-01 14:00:00	2
3	hpg_24bb207e5fd49d4a	2016-01-01 17:00:00	2016-01-01 11:00:00	5
4	hpg_25291c542ebb3bc2	2016-01-01 17:00:00	2016-01-01 03:00:00	13

index	store_id	visit_date	visitors	index	visit_date	holiday_flg
0	air_ba937bf13d40fb24	2016-01-13	25	0	2016-01-01	1
1	air_ba937bf13d40fb24	2016-01-14	32	1	2016-01-02	1
2	air_ba937bf13d40fb24	2016-01-15	29	2	2016-01-03	1
3	air_ba937bf13d40fb24	2016-01-16	22	3	2016-01-04	0
4	air_ba937bf13d40fb24	2016-01-18	6	4	2016-01-05	0

(出所)「Recruit Restaurant Visitor Forecasting」Data より²当社作成

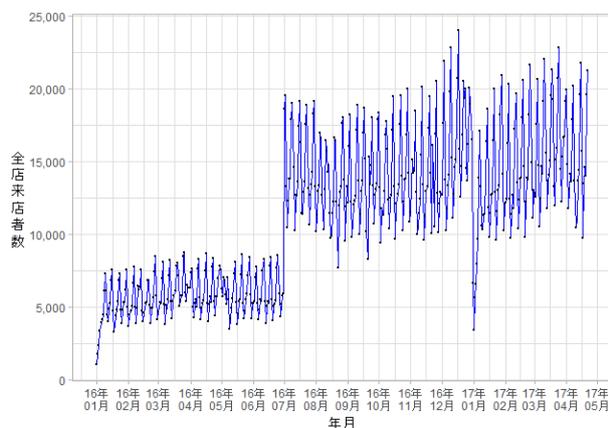
(2) EDA³(探索的データ解析)⁴

① 来店客数の傾向

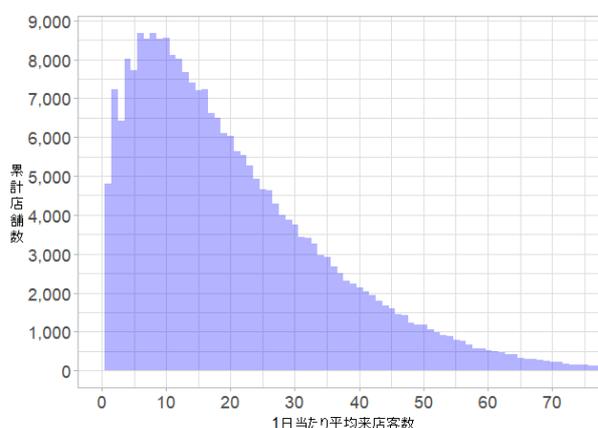
[1] 2016年7月から合計顧客数が増加しているのは、登録店舗が増加していることが要因である(図表3左図)。

[2] 1日あたりの来店客数は、最頻値が10名弱であるが、右側に広がる分布となっている(図表3右図)。

【図表3】(左図) 日次来店客数推移



(右図) 1店舗/1日あたり来店客数分布



(出所)「Be my Guest⁵」を参考に「Recruit Restaurant Visitor Forecasting」Data より⁶当社作成

② 曜日別・月別 来店客数の傾向

[1] 平均来店客数は約20人で、金・土・日の週末の来店者数が多い(図表4左図)。

[2] 3~5月と12月の来店客数が多い(図表4右図)。

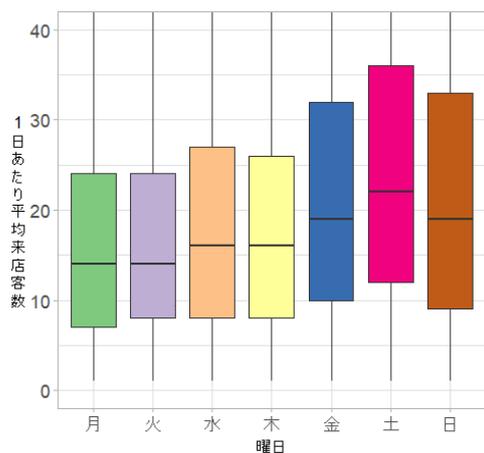
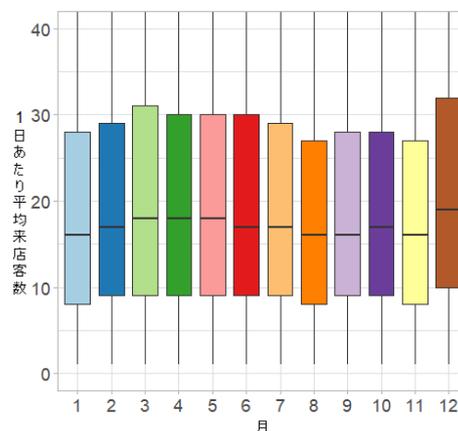
² <https://www.kaggle.com/c/recruit-restaurant-visitor-forecasting/data> (2020年8月25日確認)

³ Exploratory Data Analysis

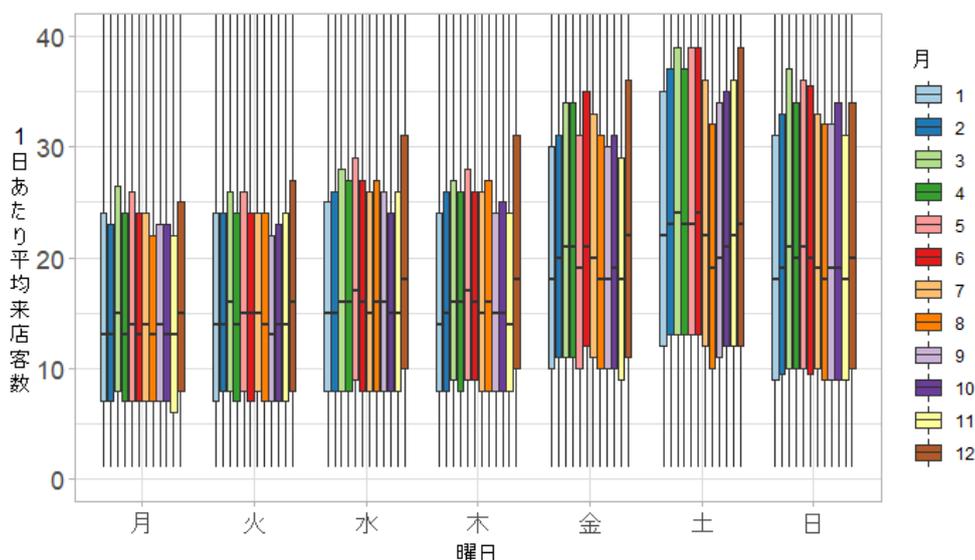
⁴ <https://www.kaggle.com/headsortails/be-my-guest-recruit-restaurant-eda> (2020年8月25日確認)を参照し分析・加工

⁵ <https://www.kaggle.com/headsortails/be-my-guest-recruit-restaurant-eda> (2020年8月25日確認)を参照し分析・加工

⁶ <https://www.kaggle.com/c/recruit-restaurant-visitor-forecasting/data> (2020年8月25日確認)

【図表 4】（左図） 曜日別来店客数 四分位図

（右図）月別来店客数 四分位図


（出所）「Be my Guest⁷」を参考に「Recruit Restaurant Visitor Forecasting」Data より⁸当社作成

【図表 5】 月次・週次 平均来店客数 四分位図


（出所）「Be my Guest⁹」を参考に「Recruit Restaurant Visitor Forecasting」Data より¹⁰当社作成

③ 地域・ジャンル別の来店客数の傾向

- [1] ジャンル別では、日本料理・インターナショナルへの来店客数が多い。
- [2] 地域別では、東京・北海道・大阪・福岡などに偏りがある。
- [3] インターナショナルやカラオケは土・日の来店客数が多いが、日本料理や居酒屋は金・土が多い。

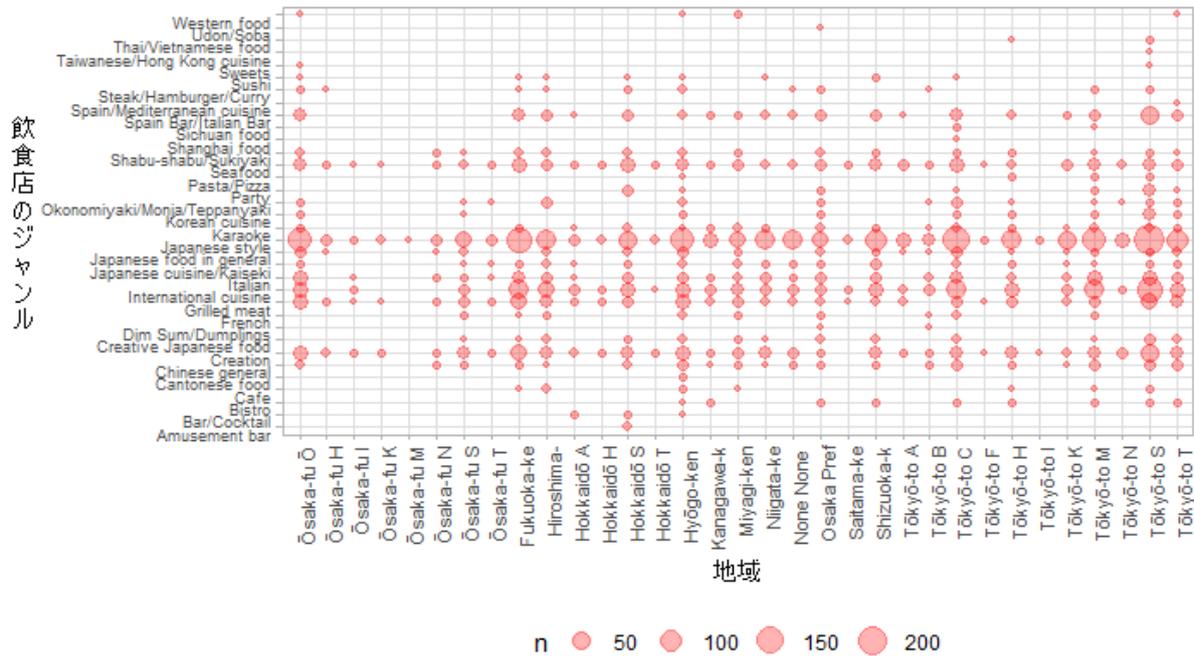
⁷ <https://www.kaggle.com/headsortails/be-my-guest-recruit-restaurant-eda> (2020年8月25日確認)を参照し分析・加工

⁸ <https://www.kaggle.com/c/recruit-restaurant-visitor-forecasting/data> (2020年8月25日確認)

⁹ <https://www.kaggle.com/headsortails/be-my-guest-recruit-restaurant-eda> (2020年8月25日確認)を参照し分析・加工

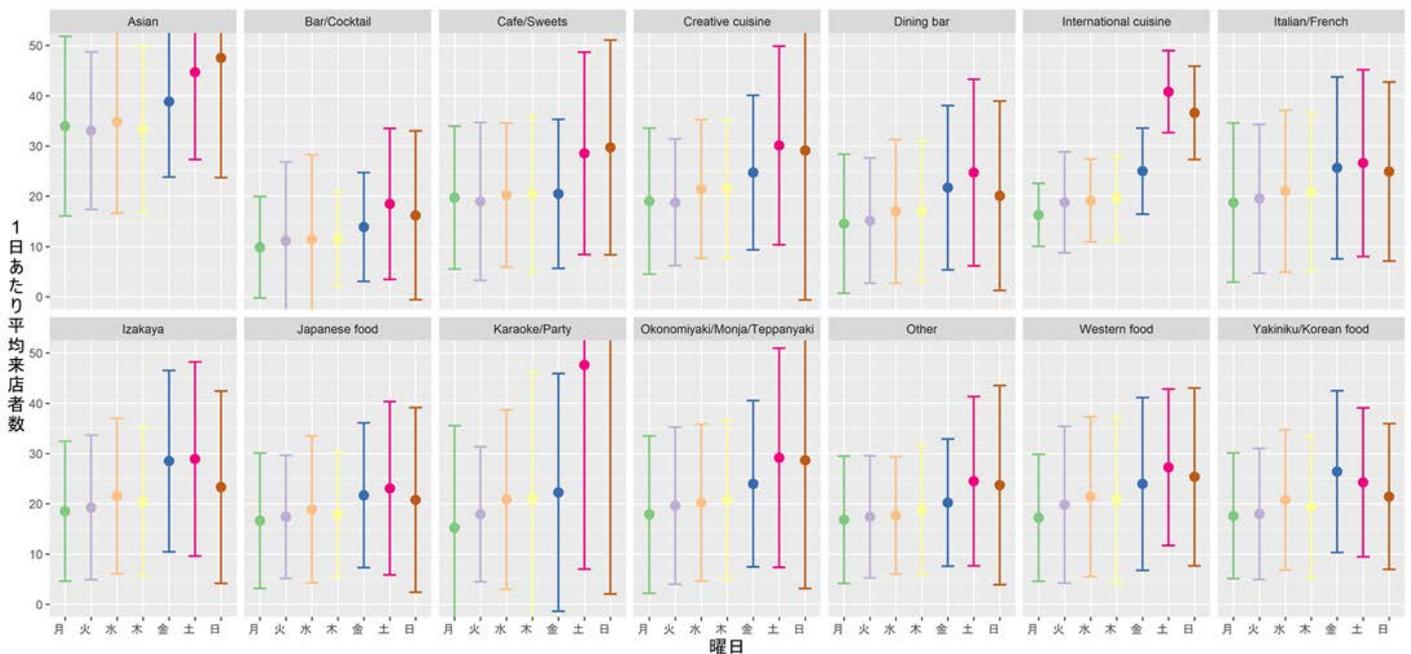
¹⁰ <https://www.kaggle.com/c/recruit-restaurant-visitor-forecasting/data> (2020年8月25日確認)

【図表 6】 ジャンル・地域別 店舗数



(出所) 「Be my Guest¹¹」を参考に「Recruit Restaurant Visitor Forecasting」Data より¹²当社作成

【図表 7】 料理カテゴリと週別来店客数 四分位図



(出所) 「Be my Guest¹³」を参考に「Recruit Restaurant Visitor Forecasting」Data より¹⁴当社作成

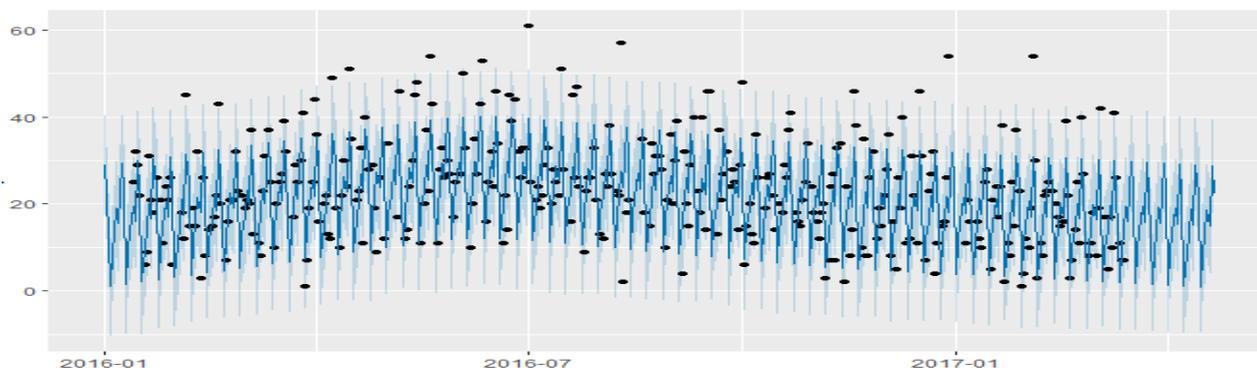
¹¹ <https://www.kaggle.com/headsortails/be-my-guest-recruit-restaurant-eda> (2020年8月25日確認)を参照し分析加工
¹² <https://www.kaggle.com/c/recruit-restaurant-visitor-forecasting/data> (2020年8月25日確認)
¹³ <https://www.kaggle.com/headsortails/be-my-guest-recruit-restaurant-eda> (2020年8月25日確認)を参照し分析加工
¹⁴ <https://www.kaggle.com/c/recruit-restaurant-visitor-forecasting/data> (2020年8月25日確認)

④ 時系列でみたトレンドと周期成分

[1] 来店客数とモデル

下記図表 8 で、来店客数が黒点でプロットされて、データが折れ線グラフでモデル化されている。

【図表 8】 1 店舗あたり来店客数の推移

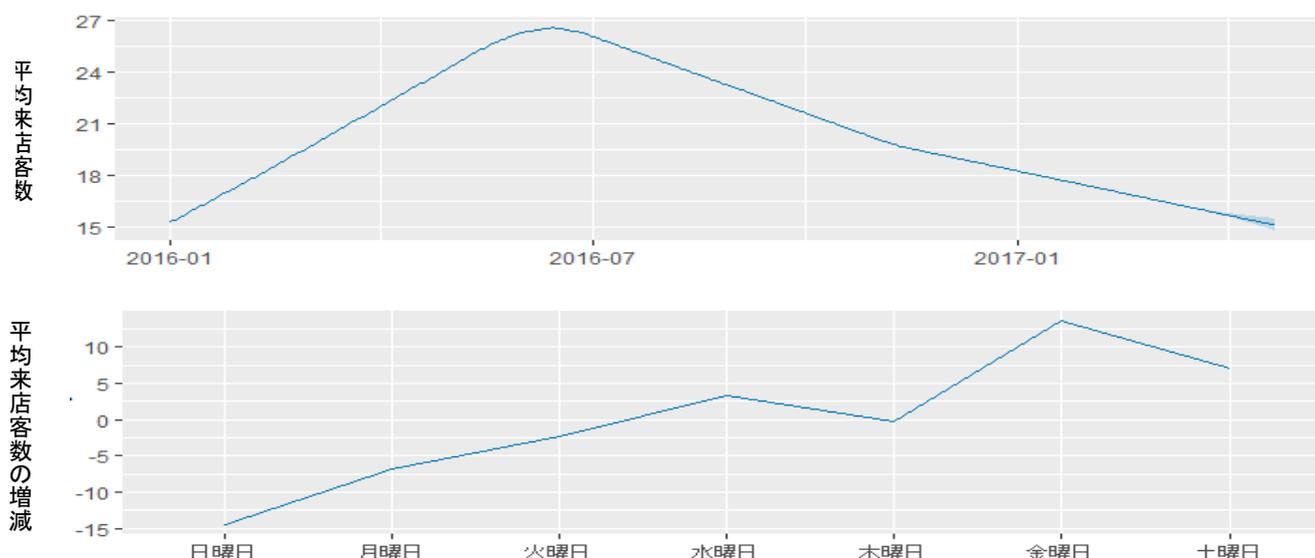


(出所) 「Be my Guest¹⁵」を参考に「Recruit Restaurant Visitor Forecasting」Data より¹⁶当社作成

[2] 「トレンド変動」+「周期変動」への分解

- 上記図表 8 のモデルからトレンド成分と周期成分を取り出した。トレンド成分は大きな増減傾向を示し、周期成分は年・月・週などの単位での変動幅を示す。
- トレンド成分をみると、2016 年 6 月あたりで変化点があり、減少傾向である。
- 1 週間の周期成分をみると、前日比では金・土曜に上昇し、日・月曜に下降している。

【図表 9】 1 店あたり来店客数のトレンド成分(上段)と周期成分(下段)



(出所) 「Be my Guest¹⁷」を参考に「Recruit Restaurant Visitor Forecasting」Data より¹⁸当社作成

¹⁵ <https://www.kaggle.com/headsortails/be-my-guest-recruit-restaurant-eda> (2020 年 8 月 25 日確認)を参照し分析加工

¹⁶ <https://www.kaggle.com/c/recruit-restaurant-visitor-forecasting/data> (2020 年 8 月 25 日確認)

¹⁷ <https://www.kaggle.com/headsortails/be-my-guest-recruit-restaurant-eda> (2020 年 8 月 25 日確認)を参照し分析加工

¹⁸ <https://www.kaggle.com/c/recruit-restaurant-visitor-forecasting/data> (2020 年 8 月 25 日確認)

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^{19,20,21}

(1) 特徴量作成の方針

- ① 予測期間 39 日（閉店日を除く）×829 店舗の来店者数予測をすべて行う。
- ② 時系列データで影響を与えると思われるもの（直近の増減傾向、月次・週次・年次の傾向）を作成する。

(2) 使用する AI の絞り込みと AI パラメータの調整

- ① 評価指標は RMSLE²²（対数平均二乗誤差）であり、数値が低い方が性能がよい。
- ② データセットとアルゴリズムの相性の良さについて RMSLE を基準に選ぶ。来店客数予測や店舗の在庫など予測が下振れすると困るケースの場合、この指標を使うとよいからだ。
- ③ 今回は LightGBM、XGBoost 等を試し、LightGBM をアルゴリズムに選んだ。
- ④ 今回はグリッドサーチにより RMSLE が最小となるようにハイパーパラメータのチューニングを行った。

（参考）

本件データは、Kaggle による AI 活用に関するコンペ（競技会）にて活用されたものであり、その終了後ではあるが Kaggle に結果を投稿すると、パラメータチューニング後の指標は、RMSLE のパブリックスコア 0.47187、プライベートスコア 0.50251 となった。2 年前のスコアとの比較になるが、パブリックでは 20 位、プライベートでは 2 位である。XGBoost で特徴量を減らした場合ではパブリックスコア 0.49375、プライベートスコア 0.55587 であった。

3. 【サンプルデータによる AI 分析評価】ステップ[3]AI の評価と予測実績の差異

(1) 予測と実績の差異

- ① 予測と実績の差異を具体的にみるために、今回は 2016 年 1 月 1 日～2017 年 2 月 22 日までのデータを使って、2017 年 2 月 23 日～4 月 2 日の予測を行い、予測と実績の差異を分析した。
- ② 指標として誤差の絶対値の平均を算定した。平均来店客数約 22.3 人に対して約 7.4 人となった。
- ③ 予測と実績の差異の絶対値の分布
 - [1] 1 店舗 1 日を 1 データとして、約 28,000 件（39 日（但し閉店日を除く）×829 店舗）の予測を行った。
 - [2] 「予測と実績の差異の絶対値」を計算した。
 - [3] 来店者の実績値の平均が 22.3 人である。実績との差異の絶対値が 10 人以下に収まる予測は全予測の約 77.3%で、20 人以上の差異がある予測は約 6.7%（100%-93.3%）であった。

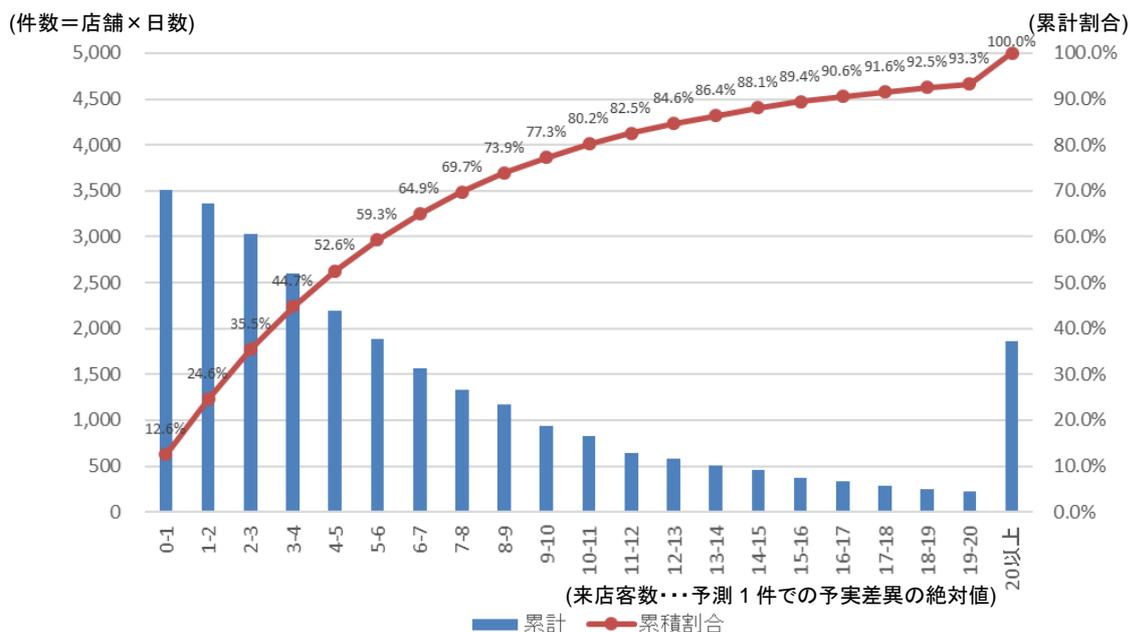
¹⁹ <https://www.kaggle.com/pureheart/1st-place-lgb-model-public-0-470-private-0-502> (2020 年 8 月 25 日確認)を参照し分析・加工

²⁰ <https://www.kaggle.com/festa78/simple-xgboost-lb-0-495> (2020 年 8 月 25 日確認)を参照し分析・加工

²¹ <https://www.kaggle.com/tunguz/surprise-me-2> (2020 年 8 月 25 日確認)を参照し分析・加工

²² Root Mean Squared Logarithmic <https://www.kaggle.com/carlolepelaars/understanding-the-metric-rmsle> (2020 年 8 月 25 日確認)

【図表 10】 予測と実績の絶対値



(出所) 当社作成

(2) ビジネス上の AI の評価

- ① 基本的な考え方は、既存の予測と AI の予測を比べ、改善効果を試算することである。既存の予測は昨年対比や店長判断などいろいろあるが、これらと比較して、AI 予測の有用性は明らかになる。
- ② AI により既存予測より精度が上がった分、仕入や在庫の削減や、人件費の適正化に好影響が出る。それを営業利益に換算すればよい。

4. 【サンプルデータによる AI 分析評価】ステップ[4]施策の展開

(1) 売上・仕入・人員計画への活用

① 仕入・在庫の削減

- [1] 来店客数予測を利用して、仕入・在庫を効率化することが考えられる。AI によって今までよりも来店客数予測の精度が高まれば、仕入の無駄や過剰在庫の低減を実現できる。
- [2] シフトの最適化による利益増加も期待できる。

② 売上計画の客観性

- [1] 店舗が売上計画を立てる場合、昨年対比で 100%以上を目指す場合も多い。減少トレンドの店では目標が未達となり、一方で仕入・人件費が計画通りであれば、利益は減少してしまう。
- [2] AI 予測を売上計画立案の基本要素とすることで、恣意的な売上見込みを抑止し、仕入・人員計画が適正化され、利益増加に貢献できる。

(2) セグメント別施策

- 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を実施できる。
- ② 既存のジャンル別に分析することも有用であるし、集客のピークを「平日型」「土日型」「金土型」に分類したり、売上の増減傾向で分けてセグメント化してみたりするのもよいだろう。

③ AIを使ってセグメントを作ることで、新たな顧客特性を発見できるかもしれない。

(3) 特徴量の増加による AI 精度の向上

① 今回のケースは与えられた既存の特徴量を使って、店舗の来店客数を予測している。AIはより精度の高いものを既に選んでおり、AIの性能による実質的な差は大きくない。分析店舗数の増加や、さらに過去から情報を追加するなどによりデータ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

[1] キャンペーンの有無やキャンペーン金額は、売上に影響を与える。

[2] 顧客情報を店舗情報に関連づけることは有用である。例えば、リピート客数、年齢層、利用頻度、顧客単価、利用人数、店舗からの距離などの情報から、店舗の特徴を抽出することが考えられる。顧客と店舗のセグメント分析は、施策の試行錯誤分析に利用できる。

[3] 一定距離内への競合店舗の新規出店は、売上に与える影響が大きい。

[4] 天気予報の確率によるが、気象庁の天気・降水量・湿度などのデータと売上には相関性がある。

[5] 顧客数だけでなく、売上金額などの特徴量も、精度向上への寄与が期待できる。

[6] SNSサービスの評価推移や顧客満足度調査なども、特徴量として役立つ可能性がある。

[7] 有料となるが携帯電話の位置情報を活用すると精度が上がる可能性がある。

— 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、当社の統一した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。