

引越しの需要予測

氏名：山内 雄登

発表日：2024年5月17日(金)



やまのうち ゆうと

山内 雄登 30歳

2017年3月東京電機大学中退
2022年前後に遺跡発掘作業員として3現場分就労
2023年8月Neuro Dive渋谷 通所開始
機械学習コンペでは上位10%に入った経験有

診断名

うつ病,発達障害

得意なこと

論理的に考えること。粘り強く取り組むこと。

苦手なこと

音や視界の動きに敏感(ノイズキャンセリングイヤホンで対処)

配慮事項

プログラミング中はイヤホンか耳栓を使えると助かります。
(通所開始～今日まで「うつ症状」は軽く、自己対処もできています)

引越し業者の困りごと

- ・ 繁忙期と閑散期の依頼件数に大きな差がある
- ・ 繁忙期に合わせて料金設定を高くして年間収支を保っている
- ・ 現場スタッフの「緊急出動」や「休日が決まらない」という問題が発生



- ・ 休日予定のスタッフも呼ばないと...
- ・ 短期間アルバイトと単発アルバイトも募集しなきゃ...



この時期は休み
が取れなくてし
んどい...

作成環境

使用環境	OS:Windows11 エディター:VScode
使用言語	Python
使用した 機械学習モデル	Prophet,LightGBM
使用したデータ	SIGNATE「アップル 引越し需要予測」 <u>学習データ：2010年7月1日～2016年3月31日</u> <u>予測期間：2016年4月1日～2017年3月31日</u>

データの分析から機械学習までの流れ

1. データの取得
2. 既存のデータから具体的な傾向を分析する
(データの傾向をグラフとして可視化)
3. 機械学習
4. Prophet
5. LightGBM
6. アンサンブル学習
7. 予測

使用データの具体的な内容と工夫

2010年7月1日～2016年3月31日の引越し件数のデータと、休業日や料金区分等のデータが含まれています。

※予測対象の2016年4月1日～2017年3月31日は下図の「引越し件数」のみが欠けている状態です。

学習データ

- ・ 日時
- ・ 引越し件数
- ・ 法人が絡む引越し日フラグ
- ・ 休業日フラグ
- ・ 午前の料金区分
- ・ 午後の料金区分



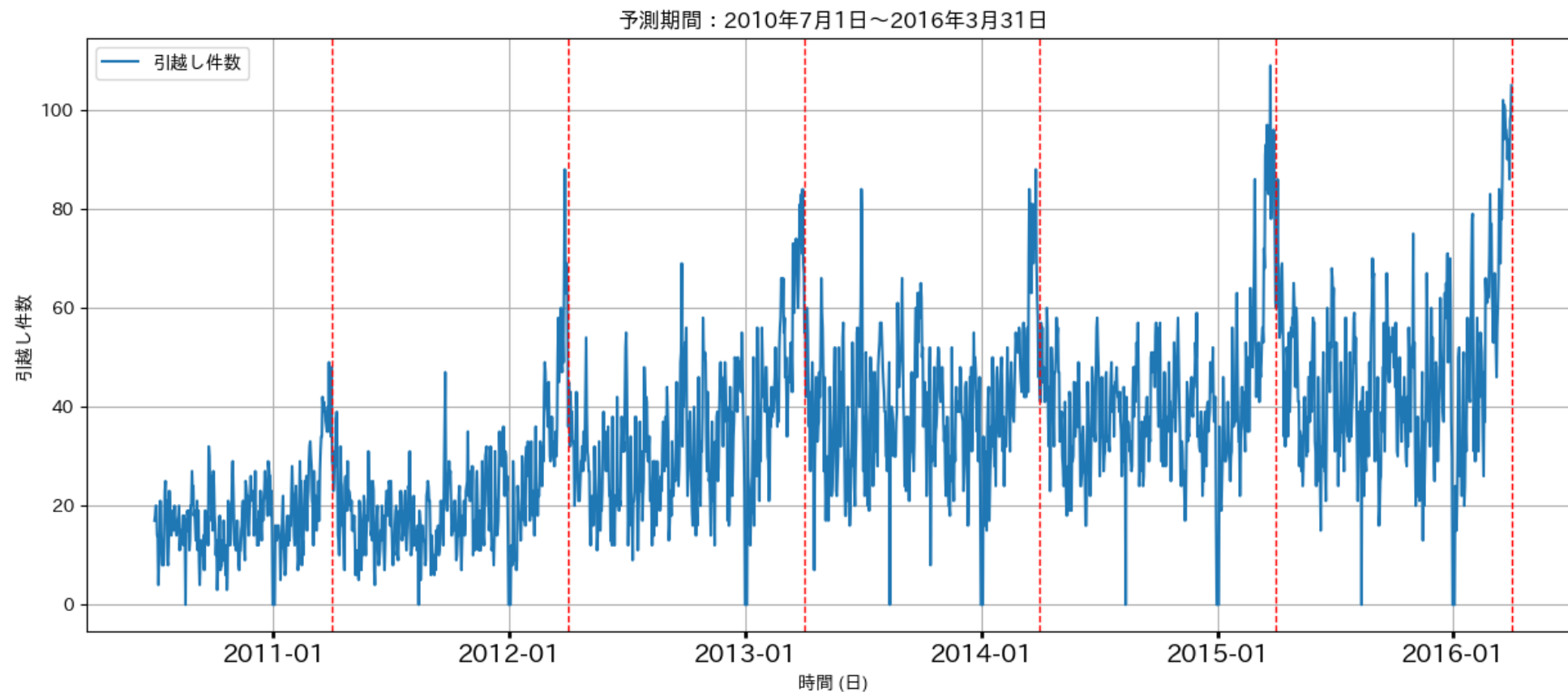
追加したデータ

- ・ 祝日

引越し件数を可視化

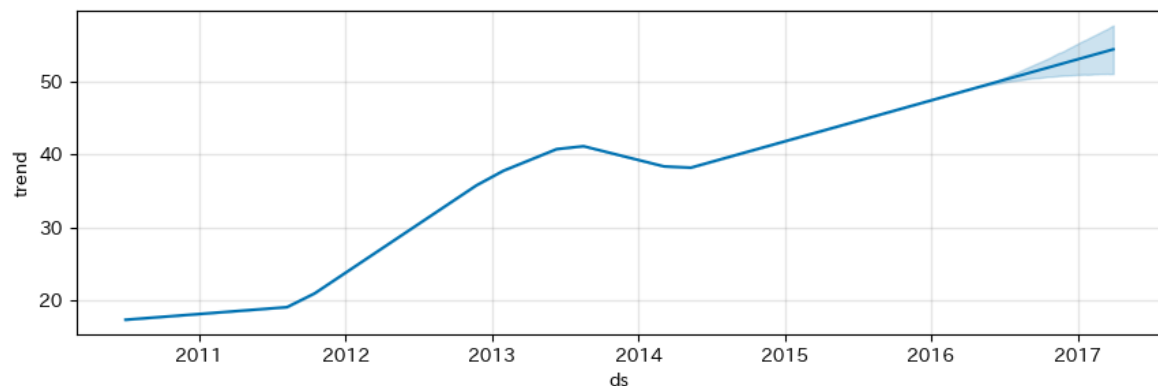
見た目では…

- ・ 引越し件数は毎年徐々に増えている
- ・ 1年毎の波形が似ている (----毎年4月1日)

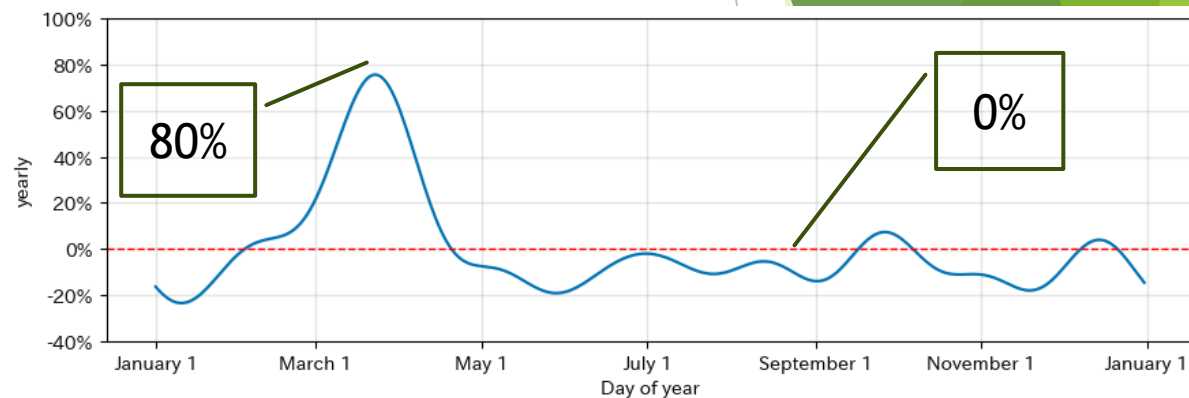


学習データから判明したデータの傾向・パターン

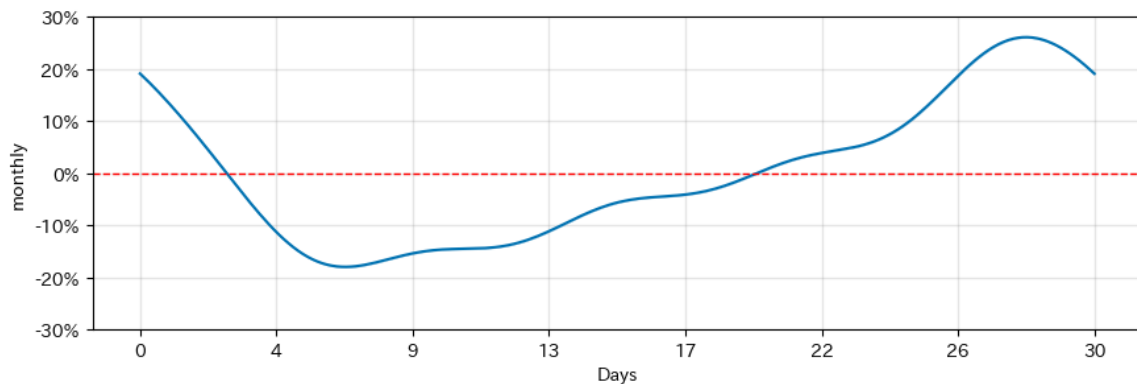
①データの長期的な方向性



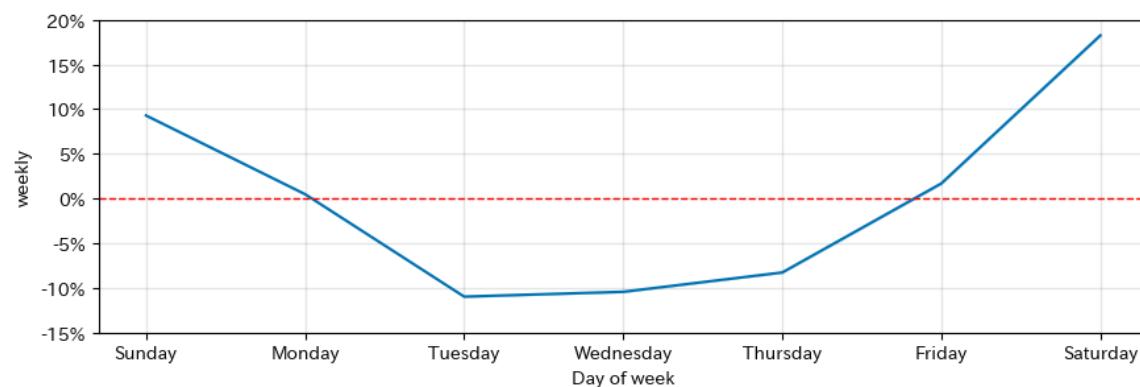
②年間の周期的な変動



③月間の周期的な変動



④週間の周期的な変動



(---0%ライン【0%=平均となり、80%であれば平均×1.8の需要があることになる】)

分析結果から分かること

<u>①毎年増加傾向</u> 2010年～2017年)世界的な金融危機からの回復期 2010年～2017年)若年層の大学進学率・就職率の増加 2010年～2017年)賃貸住宅の割合が増加	<u>②年間だと3月,4月の需要が多い(繁忙期)</u> 他の月の2倍以上の需要(入学,入社シーズン)
<u>③月間だと月末月初の需要が多い</u> 家賃の支払いが月末締め	<u>④週間だと土日の需要が多い</u> 休日の中に引越しを済ませる人が多い

提案

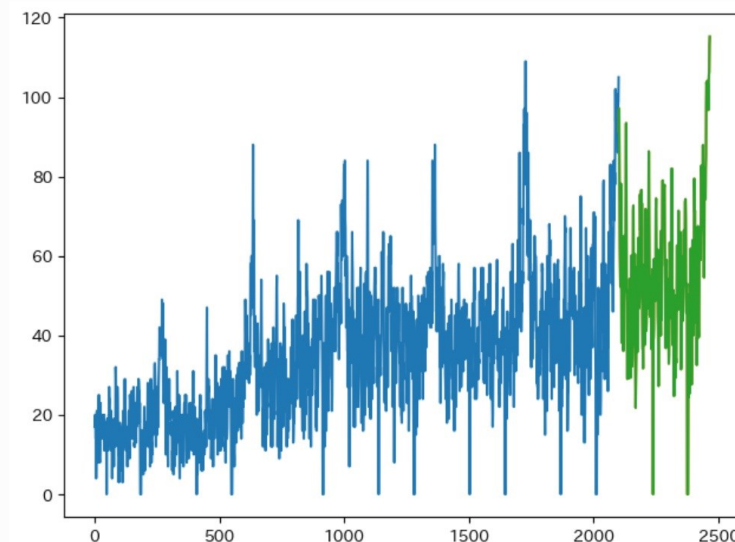
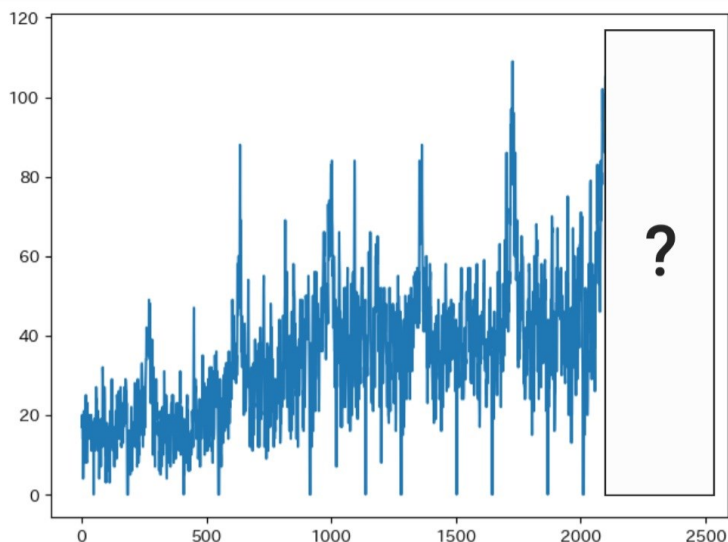
- ・最も需要件数が多くなる3月末に100件以上
- ・3月末の前後2週間は80件以上
- ・3月末以外の月末月初は60~80件の需要があると見込んだ方がいい

データの規則性が顕著に現れているので、AIで詳細な予測が見込めそう

AIで引越し需要件数を予測する

もし、需要予測ができれば...

- ・ 予測に応じて適切に料金設定ができる
- ・ 予測に応じて適切に人員配置ができる
- ・ 現場スタッフが働きやすくなり、人材が定着しやすくなる



予測モデルの説明

	強み	弱み
Prophet	<p>周期的なパターンを持つデータに適している</p> <ul style="list-style-type: none">・長期的な上昇や下降の傾向、週単位や年単位などの繰り返しパターン、祝日や特別な日などの影響を考慮して予測を行います。	<p>データは区切られた区間ごとの関係性のみを反映します</p> <ul style="list-style-type: none">・トレンドや季節性を把握するために、時間ステップごとの細かな変動は一括で処理されます。
LightGBM	<p>複雑な関係性や相互作用を持つデータに適している</p> <ul style="list-style-type: none">・分類や回帰などの幅広い機械学習タスクにおいて、複雑なデータ構造を捉える能力を持っています。	<p>学習データに存在する範囲の値しか予測できません</p> <ul style="list-style-type: none">・LightGBMも過去のデータから学習した長期的な上昇や下降の傾向、季節性のパターンを用いて未来の値を予測しますが、未来の値が学習データの範囲外の場合、モデルがその新しい傾向を正確に捉える事は難しいです。

モデルのアンサンブル

- Prophet ← 予測の大半を担当

時系列データのトレンドと季節性を捉えるが、細かい変動は見逃してしまう

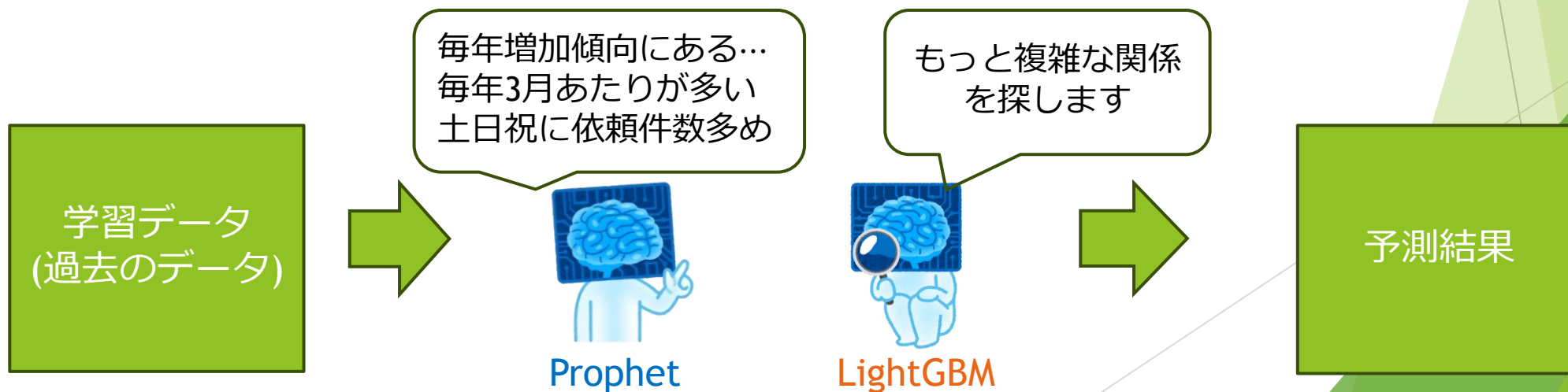
- 残差

Prophetの実測値と予測値の差には見逃した情報が含まれる。

残差は時間が経っても長期的な上昇や下降の傾向がなくなる。

- LightGBM ← 細かい見落としを修正

残差をLightGBMで学習し、Prophetの予測値を補完

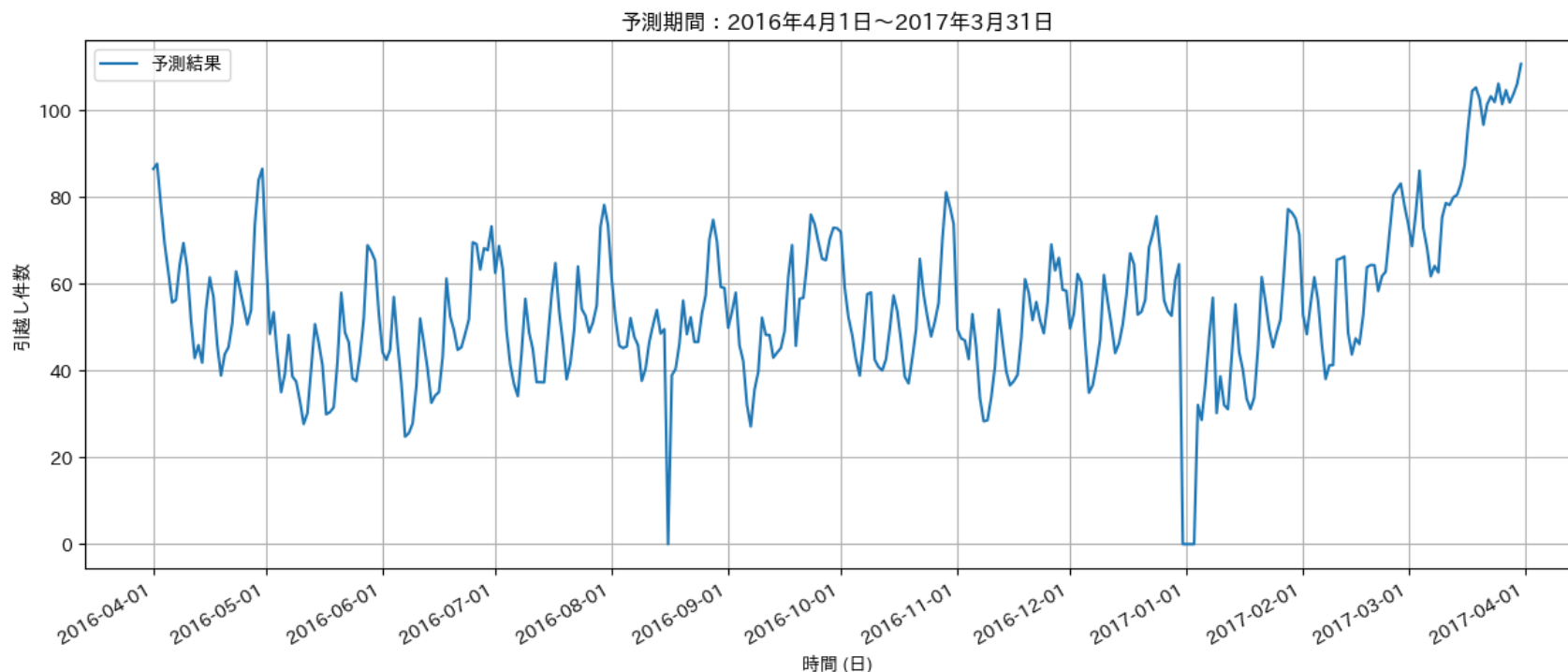


予測結果

未来の1年分の予測が1日あたり約9件の誤差で予測できた

- ・ 暫定評価(MAE) : 9.0784
- ・ 最終評価(MAE) : 9.1547 => 1047人中260位

↪ 上記2つのスコアの差は小さいので予測が安定したモデル



予測精度

予測モデルの精度や信頼性を評価する指標としてはMAE(平均絶対誤差)を使用します。この値が小さいほど予測の誤差が少ないといえます。

実測値	予測値	誤差	誤差の絶対値
100	90	-10	10
120	120	0	0
80	50	-30	30
110	140	30	30
90	100	10	10

誤差の絶対値の総和 = 80
データの数 = 5

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

平均

MAE : 16



MAEは予測の偏りを検出できない点に注意

上記の例だと1データ誤差が30でも0でも「1データ辺りの誤差は16」と結論付けられます。

特に効果があった特徴量

モデル	LightGBM	特徴量の説明
損失関数	l1	
目的変数	[引越し件数]-[Prophetが予測した引越し件数]	
特徴量重要度 (上位11個)	trend dayofyear multiplicative_terms yearly x_dayofyear weekly_lower weekly trend_lower monthly_lower multiplicative_terms_upper day y_dayofyear	データの長期的な変動 1年の何日目か トレンドの大きさに比例して増減する「乗法的な影響」 年次の季節性 週次の季節性の信頼区間の下限 週次の季節性 「データの長期的な変動」の信頼区間の下限 月次の季節性の信頼区間の下限 トレンドの大きさに比例して増減する「乗法的な影響」 1ヶ月中の何日目か

X=-1~1,y=-1~1の円周上のどの点に位置しているかを表している。
(例:○年1月1日と○年12月31日は日数的に離れているが、年の季節的には近いことを表現している)

考察・課題点・今後の施策

・予測データからの考察

支店の場所のデータや担当地域が分かれば駅の降車人数や人口推移のデータを追加する事でより精度の高い予測が出来る可能性が高い。

・Prophetの扱い方を工夫する

MAE以外の評価指標の使用、パラメータの調整、交差検証の間隔調整を組み合わせることで精度向上が見込めると考えています。

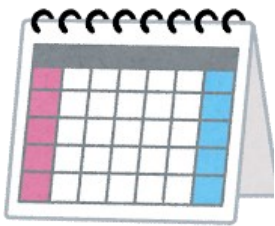
・アンサンブル学習を工夫する

Prophetで得られたトレンド成分を実測値から差分をとり、それに対してLightGBMを使用する等のアンサンブルも有効な可能性が高いと考えています。この時、Prophetがトレンドを捉える事に重点を置いた使い方も有効かもしれないと考えています。

付録 Prophetは他に何ができるのか

何らかの周期性があるデータに対して高い予測精度を発揮します。

- ・ 今回の予測に似た例だと→商品の売上予測(個数,売上金額)
- ・ 他にも →電力需要,商業施設の顧客流動予測



カレンダーに沿った傾向があると予測しやすいイメージ
(日本以外の祝日,手動でのイベント設定可)

これらが事前に予測できれば、マーケティング戦略や人員配置などの意思決定を行うためのデータを得られます。