

# 成果物資料

## 引っ越し需要予測のための機械学習モデル

### 目次

- |       |          |
|-------|----------|
| 1. 概要 | 4. 手法    |
| 2. 背景 | 5. 結果と効果 |
| 3. 目的 | 6. 今後の施策 |



**引越し業者における時期ごとの最適な価格設定**のため  
機械学習を利用した**受注件数の予測モデル**を開発

最終的なモデルは

- 長期的な件数変動のトレンドを捉えるモデル
- 日々の細かい件数の変動を捉えるモデル

**2つの基本モデルを組み合わせて構築**

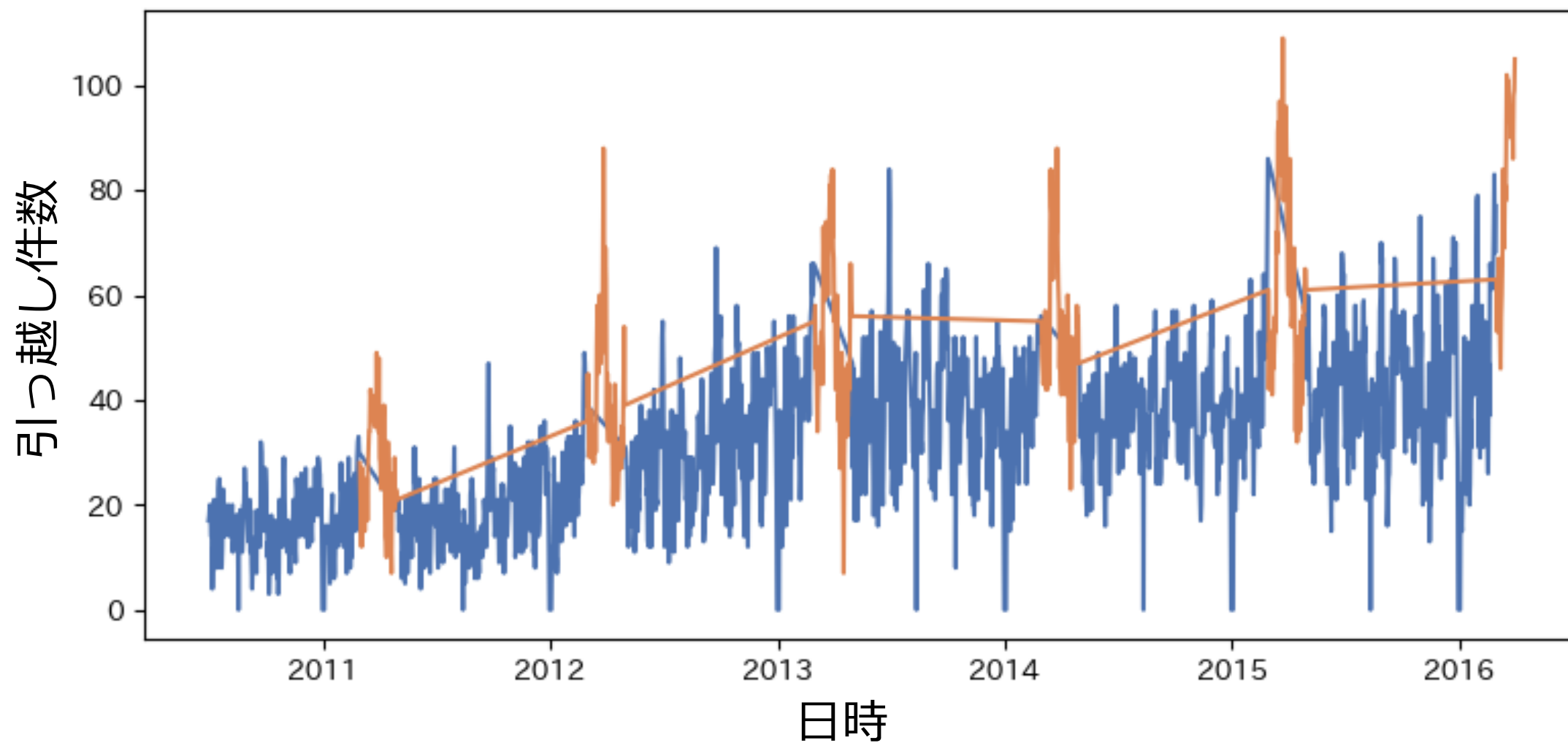
予測誤差は平均して±7件程度と**高精度**

引越し需要の予測に有用であることが示された

今後本モデルを発展させ、さらなる

引越し業界の課題解決に役立つことが期待される

# 背景 – 引っ越し業界が直面する課題<sup>3</sup>



—通常期



—繁忙期

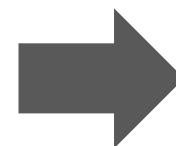


季節や曜日によって需要は大きく変動

人員配置・価格設定で需給バランス調整

予測を誤ると...

- 人員不足 → 受注逃し
- 人員過剰 → 人件費増
- 料金設定ミス → 収益悪化



**的確な需要予測が不可欠**

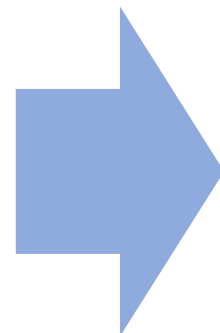
## 機械学習を活用した引越し需要予測モデルの構築

- ✓ 時期・価格設定等によって複雑に変化する需要を  
人の経験や勘ではなく、  
機械学習を用いて**データに基づいて的確に予測**
- ✓ 最適な人員配置・価格設定を実現し、**生産性・利益率を向上**

従来

人の経験と勘による予測

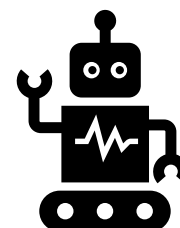
昨年同時期と比べると、  
おそらくこの週は  
1日に50～80件くらい？



本成果物

機械学習を活用した正確な予測

**データA, B, C, Dを勘案し  
モデルから算出すると、  
この日はxx件です**



引越し業者の価格と需要などに関する  
以下のデータからモデルを構築  
(データ分析コンペから引用)

属性名	説明
datetime	日時(YYYY-MM-DD)
y	引越し件数
client	法人契約での引越しに関するフラグ
close	休業日
price_am	午前の料金区分
price_pm	午後の料金区分

全2101件					
datetime	y	client	close	price_am	price_pm
2010/7/1	17	0	0	-1	-1
2010/7/2	18	0	0	-1	-1
2010/7/3	20	0	0	-1	-1
...	...	...	...	...	...
2016/3/29	98	1	0	4	4
2016/3/30	99	1	0	5	4
2016/3/31	105	1	0	5	4

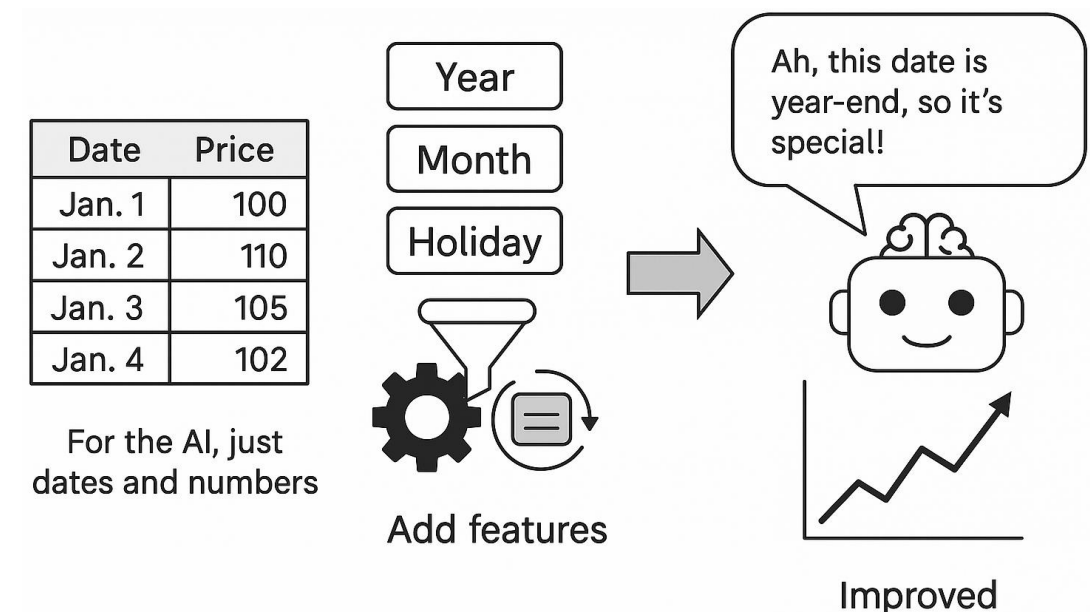
5が最高値  
-1は欠損

# 手法2 – モデルに背景を理解させる<sup>6</sup>

機械学習モデルは「この日は祝日だから特別」と勝手に**背景まで理解してはくれない...** 😞

モデルに背景を把握させるために**データの属性情報を追加**

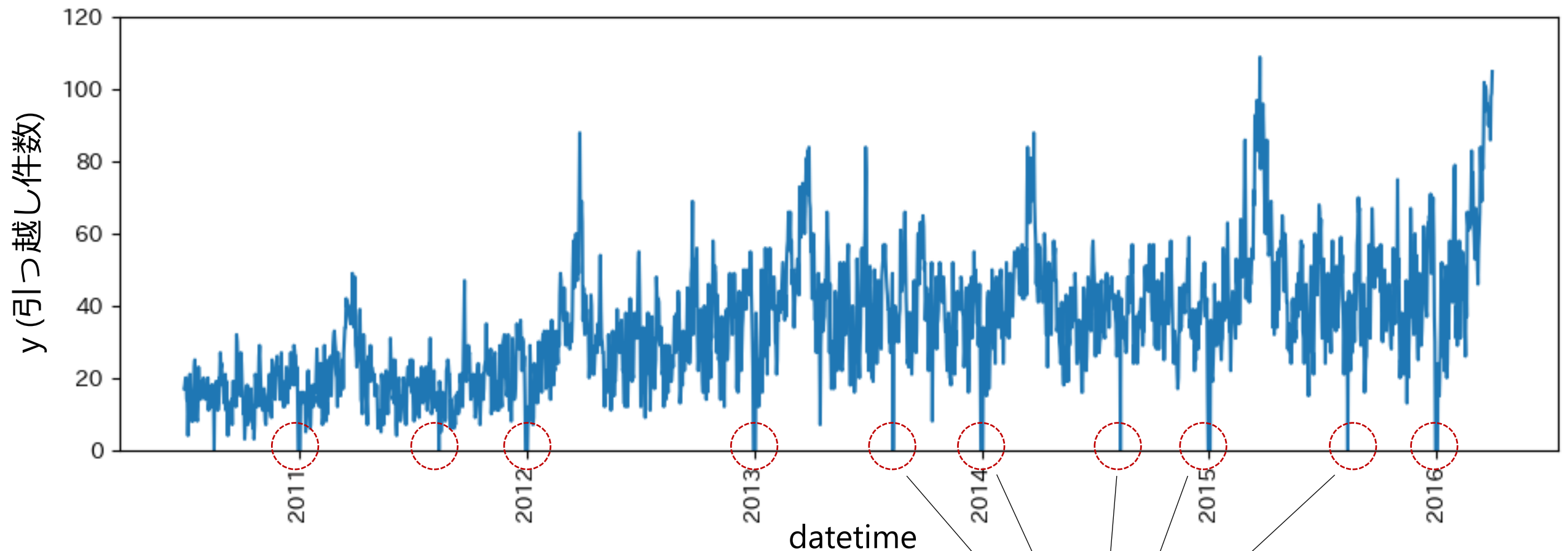
- 年、四半期、月
- 四半期や月の始まり・終わり
- 曜日、土日、祝日・休日
- (午前・午後価格の合計、差分、積)



人が暗黙のうちに感じている季節性やタイミングを認識できるようになり、**予測精度が向上**

# 手法3 – 休業日と引っ越し件数

7



休業日(年末年始やお盆)は必ず引っ越し件数ゼロ(赤丸)といった**明確な法則**によって**機械学習ではなくルールにしたがって出力**

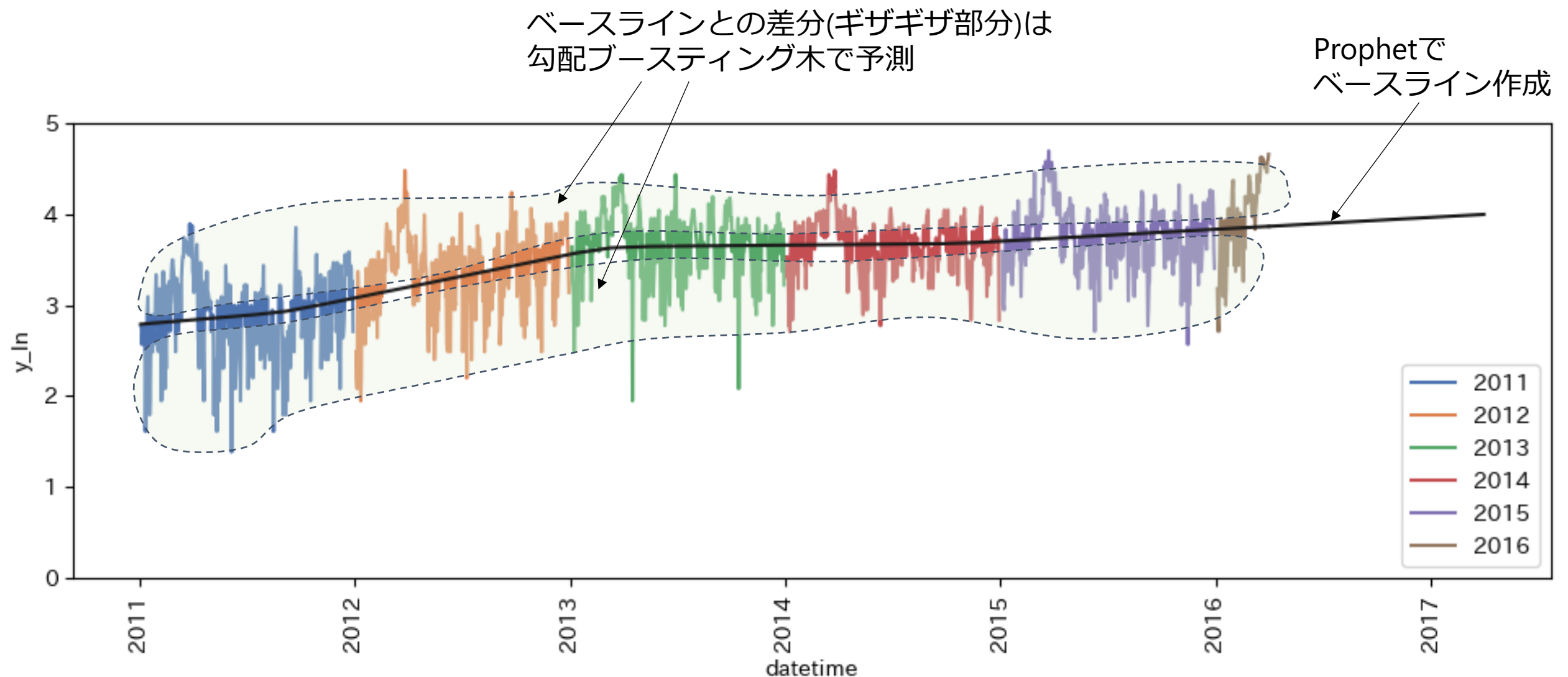


**機械学習モデルの負担を軽減**



# 手法4 – 2種類のモデルを合体

8



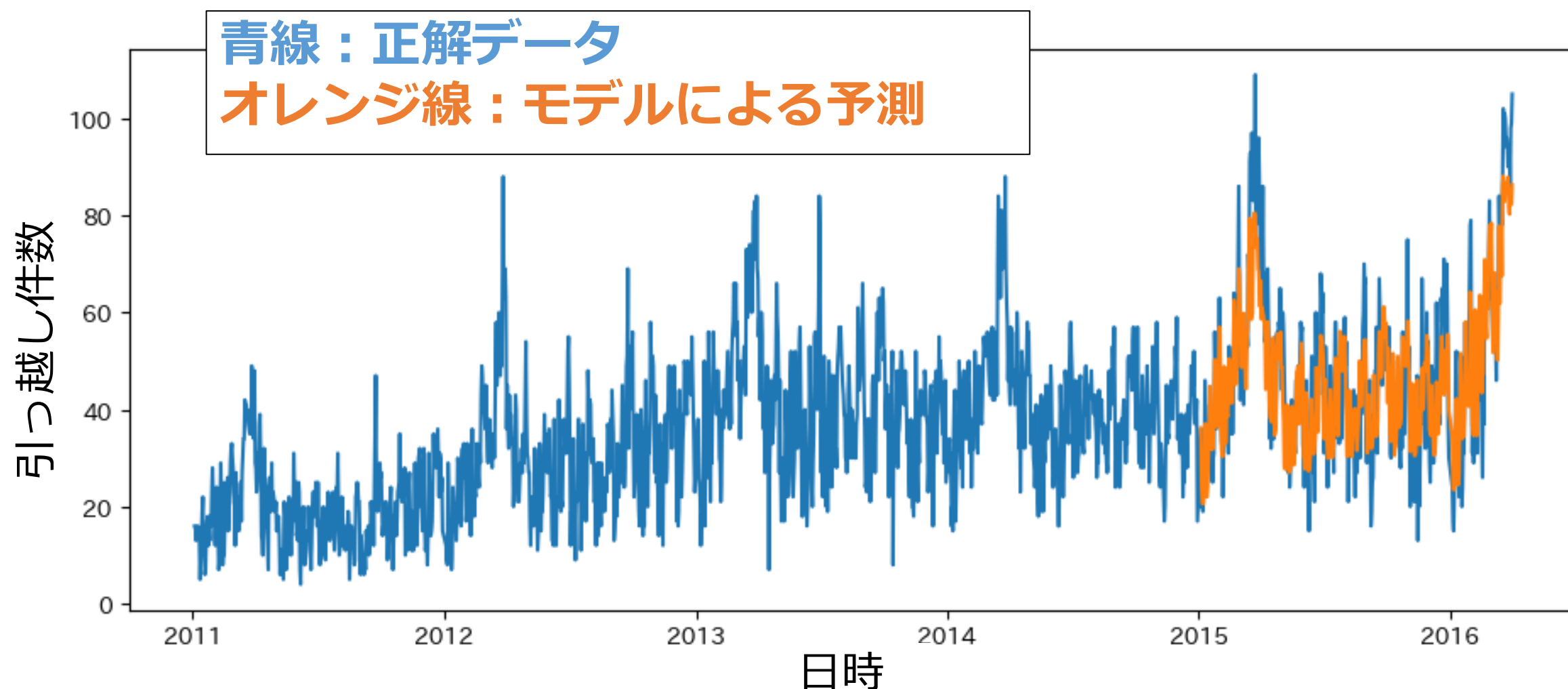
Prophet :  
大まかな全体の傾向を捉えるのが得意

+

勾配ブースティング木 :  
日ごとの細かな変動や  
突発的なギザギザした動きが得意

**2つのモデルの強みを組み合わせ、より高い精度での予測を実現**





誤差：平均±約7件 (コンペ内順位**上位2%**)  
予測にかかる時間：**5~10秒程度** (標準的なノートPC)

- ✓データに基づいた高精度な予測を実現
- ✓最適な人員配置・価格設定のための重要なツールに

# 今後の施策1-データ属性の追加 10

## データ属性を更に追加

- ・ ○○日前の需要といった過去の情報
- ・ 気温・天気などの気象情報
- ・ 景気動向などの経済情報
- ・ 他社の引越し件数などの業界動向（可能なら）

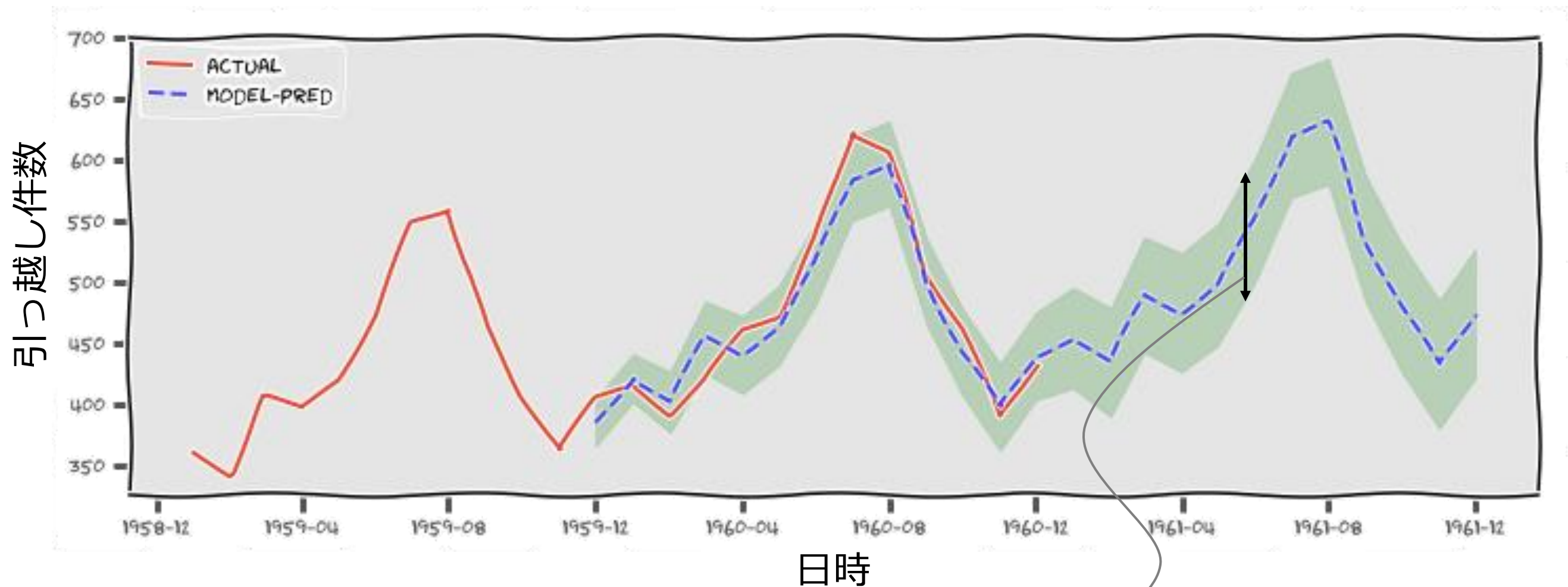
id	datetime	y	y_ln	client	price_am	price_pm	year	quarter	month	week	ordinal_day	day	week_of_month	day_of_week	quarter_start	quarter_end	month_start
str	date	i64	f64	i64	i64	i64	i32	i8	i8	i8	i16	i8	i8	i8	i32	i32	i32
"2011-01-04"	2011-01-04	16	2.772589	0	0	0	2011	1	1	1	4	4	1	2	0	0	0
"2011-01-05"	2011-01-05	16	2.772589	0	0	0	2011	1	1	1	5	5	1	3	0	0	0
"2011-01-06"	2011-01-06	13	2.564949	0	0	0	2011	1	1	1	6	6	1	4	0	0	0
"2011-01-07"	2011-01-07	14	2.639057	0	0	0	2011	1	1	1	7	7	2	5	0	0	0
"2011-01-08"	2011-01-08	16	2.772589	0	0	0	2011	1	1	1	8	8	2	6	0	0	0



**多彩な情報を基に、より高い予測精度を実現**

# 今後の施策2-幅を持たせた予測

11

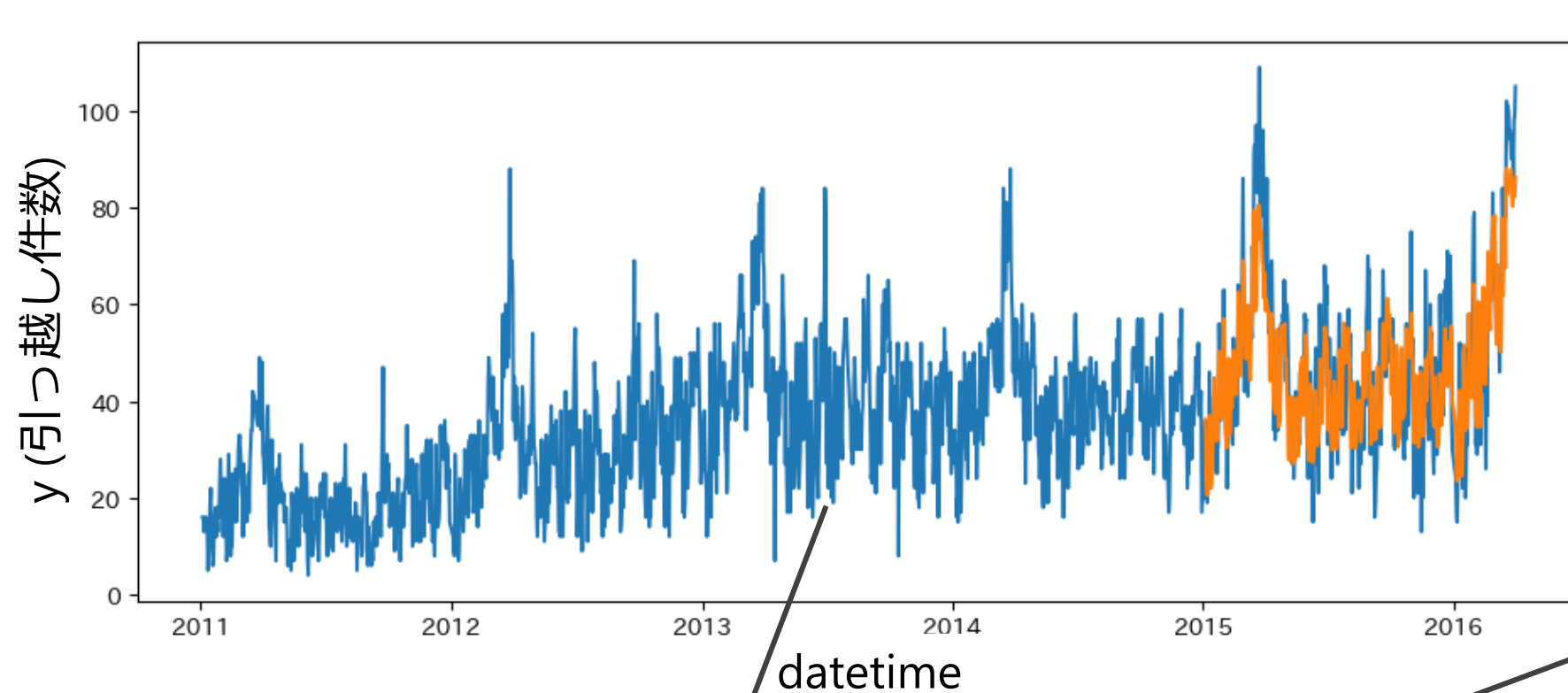


〇〇%の確率で実際の需要はこの範囲内に収まる



**より柔軟性のある施策が打てるように**

# 今後の施策3-価格と需要から利益まで算出<sup>12</sup>



料金設定

日時	価格 (午前)	価格 (午後)
2016/4/1	xx	yy
...	xx	yy
...	xx	yy
...	xx	yy
...	xx	yy

$$\text{利益} = \text{需要} \times \text{料金} - \text{人件費} - \text{その他変動費}$$

**収益性の評価まで実現**

必要な追加データ

- 各料金区分の具体的な金額
- スタッフの時給や作業時間
- 車両関連費、資材費など

ご清聴ありがとうございました。

# (Appendix)苦労点 -納期がある中での工夫<sup>14</sup>

今回の分析は、事業所での疑似就労企画の一環として取り組んだ

## 企画の設定



顧客役：アップル引越センター マネージャー

依頼内容：機械学習を使って需要予測をして欲しい



先輩社員役：モデル構築や資料作成について相談



自分：

モデル構築、仕様書作成、発表資料作成を担当

顧客役、先輩社員役と連絡を取りながら納期内に

顧客の要求を満たすものを作成

## 行ったこと

- 顧客の要求の**ヒアリング**
- モデル作成、仕様書作成、発表資料作成の**スケジュール管理**
- 作成しているものが顧客の要求に沿っているか**中間報告**
- モデル納品。発表**

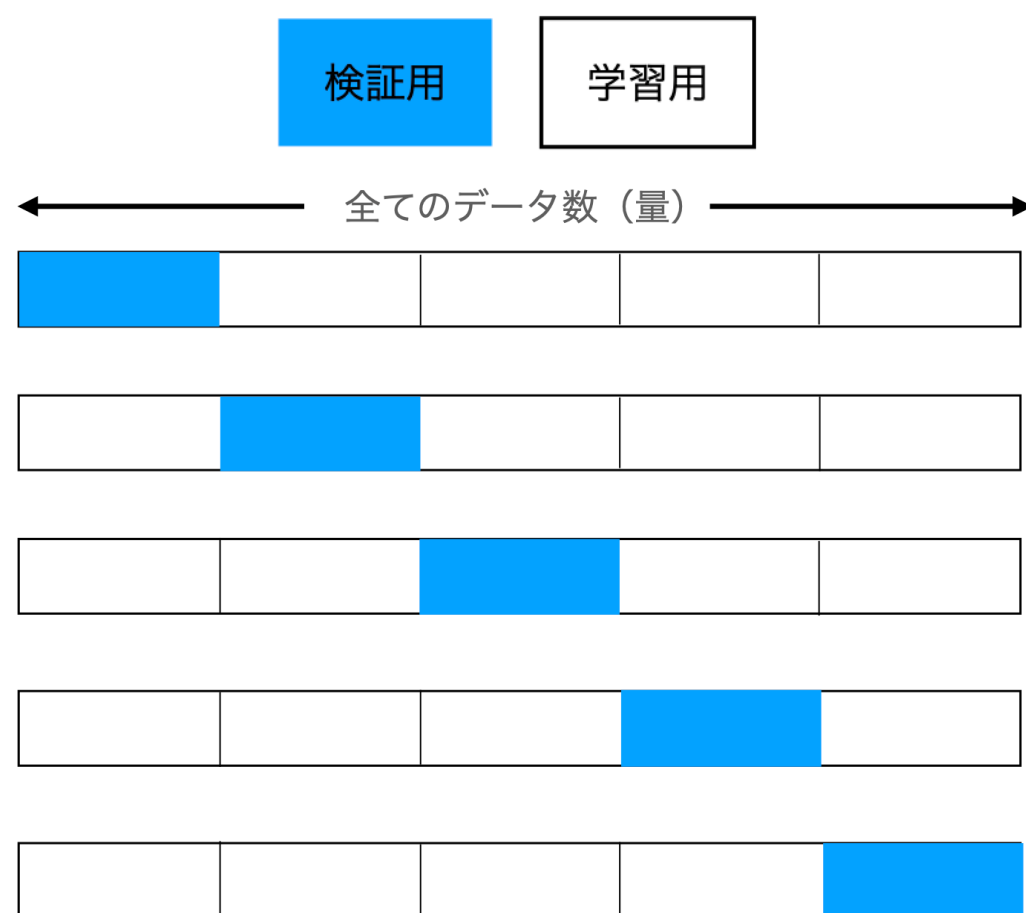
# (Appendix)手法5 – モデル検証時の注意点<sup>15</sup>

既知データの一部を使い、モデルの精度を検証する  
しかし**既知データに過剰適合すると、未知データに対する精度が低下**

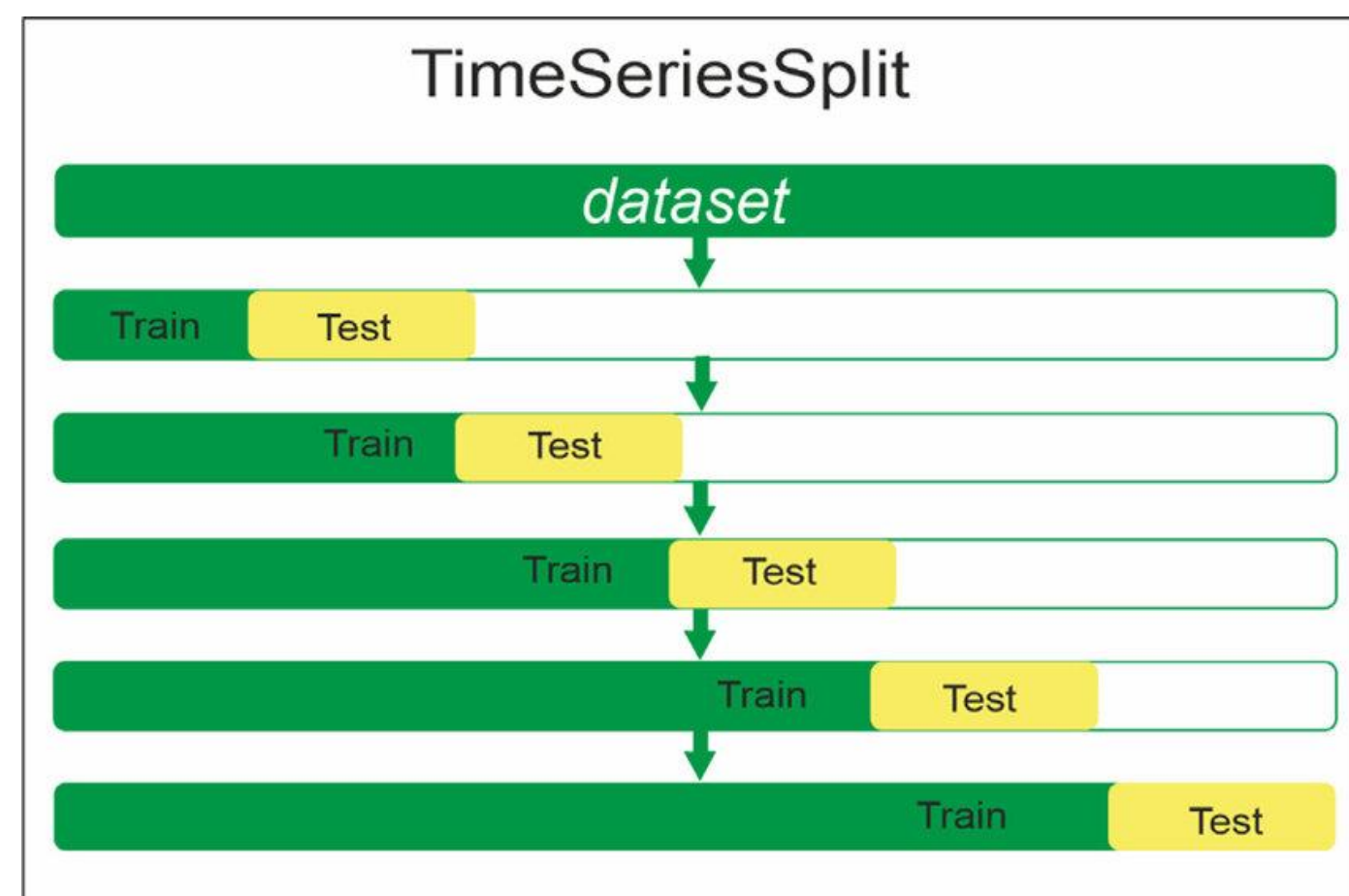
↓

**精度検証のためのデータセットを複数パターン用意**  
することで、検証の信頼性を担保

**ただし、時系列データでは注意が必要**



通常のk-fold(単純な5分割)  
ではなく



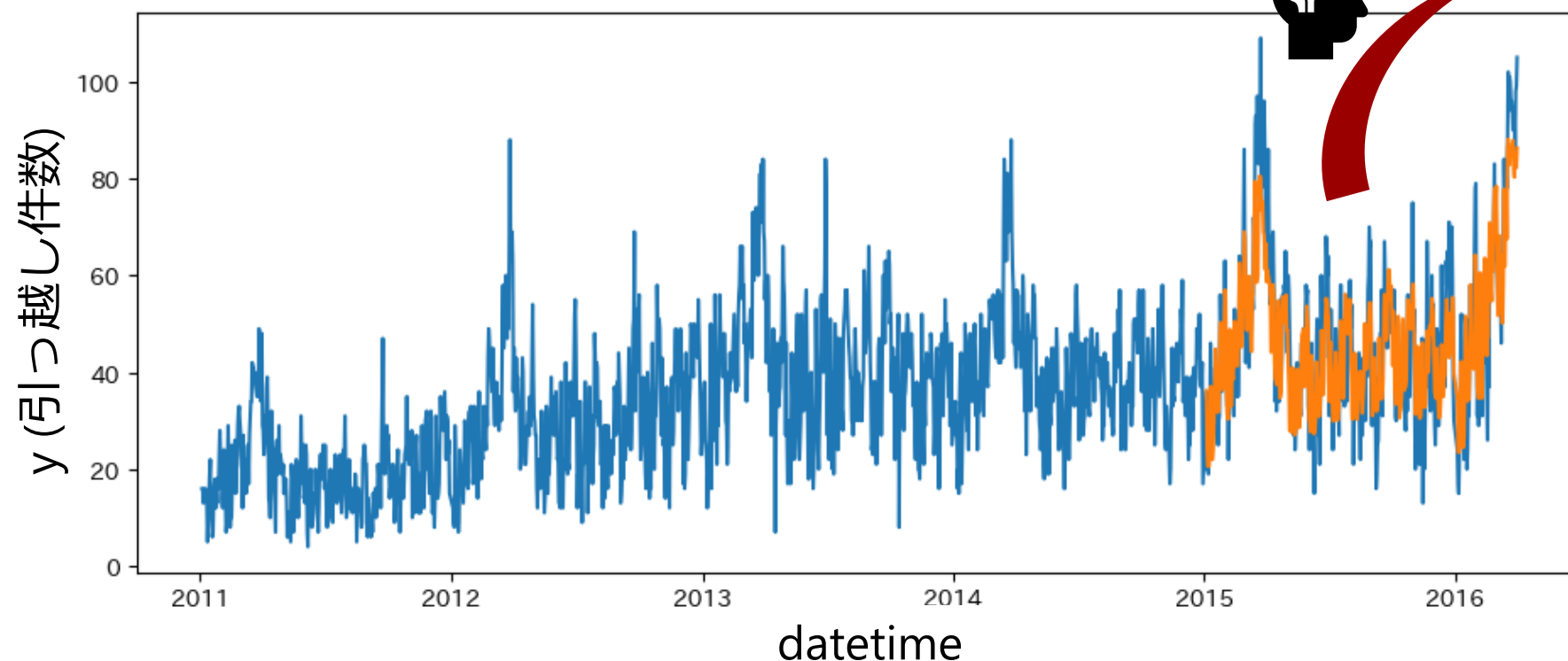
未来の情報を使ってしまわないように注意  
(ズルになってしまう)



# (Appendix)今後の施策4-最適な需要から価格を逆算<sup>16</sup>

内部で繰り返し予測して最適価格を算出

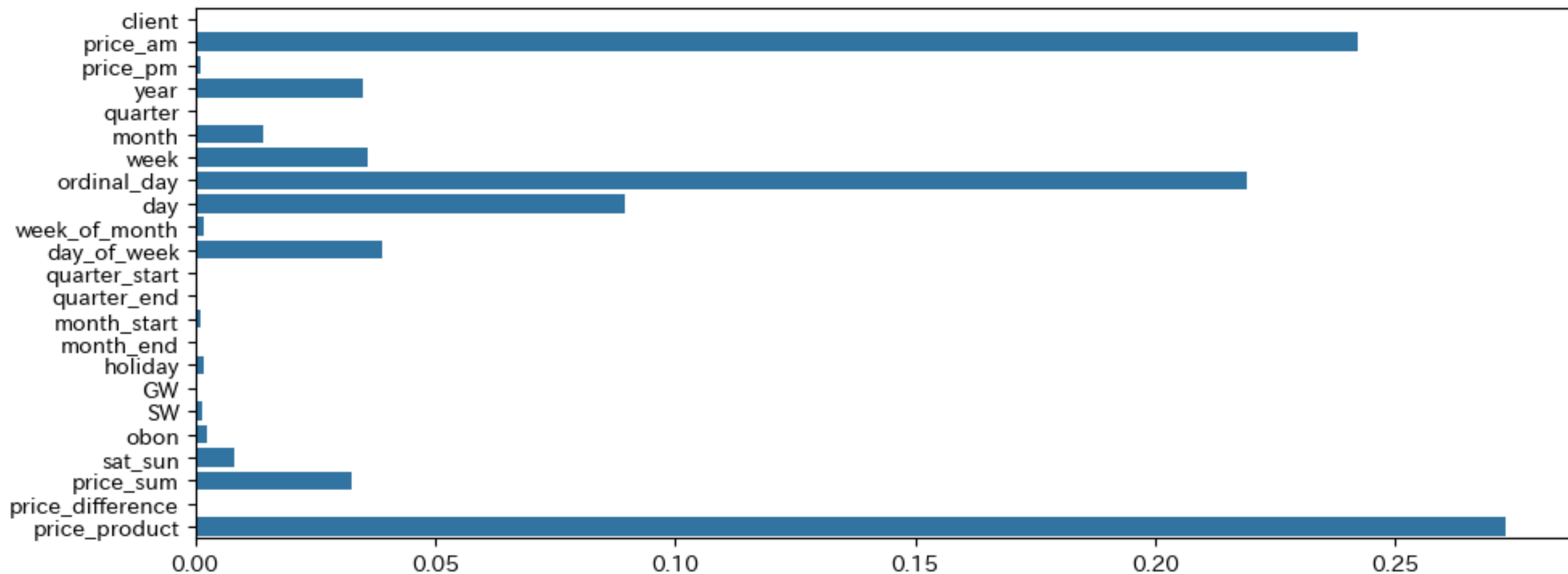
先に目的の需要を決め、  
価格を逆算



日時	価格 (午前)	価格 (午後)
2016/4/1	xx	yy
...	xx	yy
...	xx	yy
...	xx	yy
...	xx	yy

値付けを何度も試行錯誤して計算させる必要なく  
**目的の需要を満たす値付けの逆算が可能に**

## 特徴量重要度



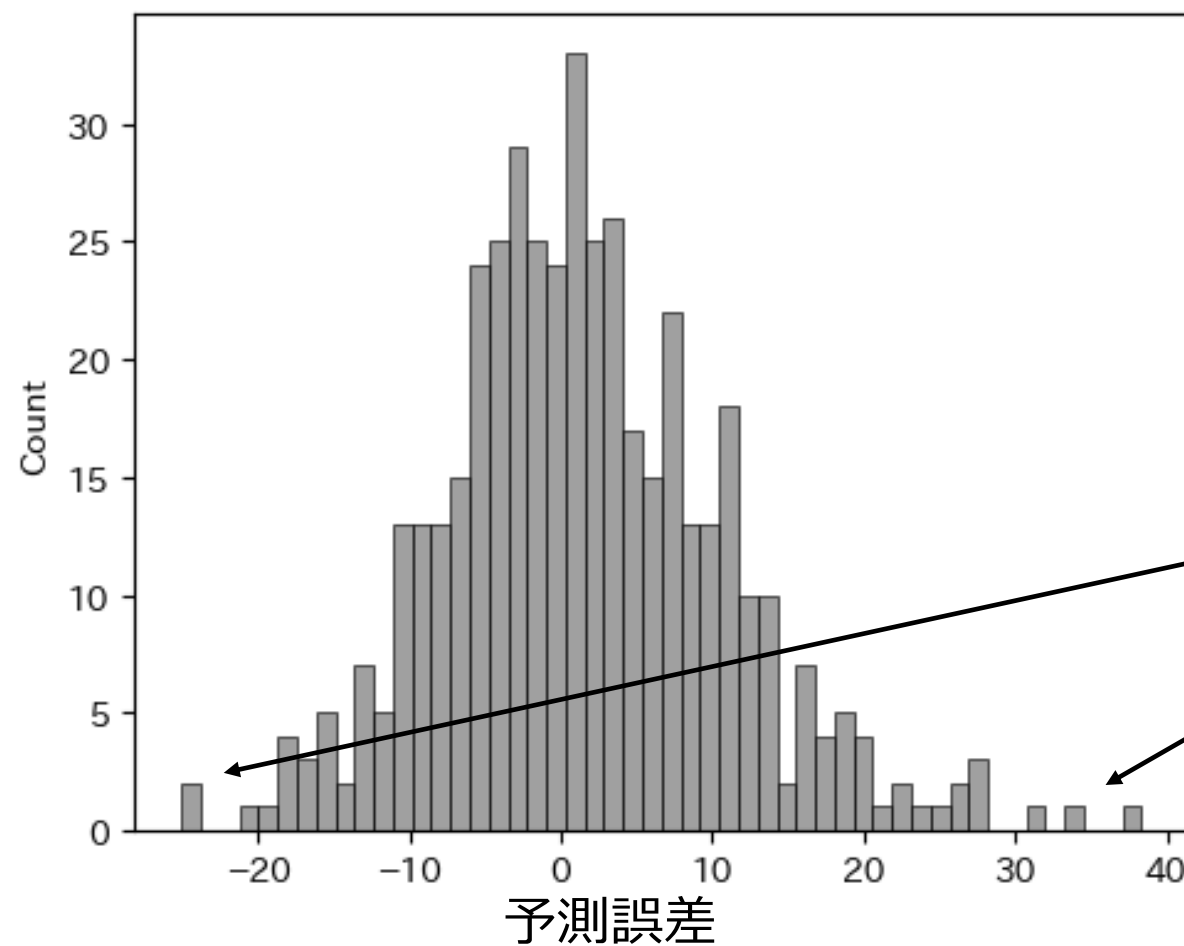
price\_am、ordinal\_day、price\_productの3つが特に重要だという結果に

ordinal\_dayはその日がその年の1月1日から数えて通算何日目であることを示す特徴量

→時期を表す情報としてモデルが活用しているものと推察

price\_productはprice\_amとprice\_pmを乗算した特徴量

## 予測誤差の分布について



平均すると誤差は  $\pm$  約7件 程度  
多くの場合  $\pm$ 10件以内に収まる

しかし、稀に数十件程度の誤差が生じる場合もある

**実務では、予測は一点の決め打ちではなく信頼区間も表示させたほうが良い？**

