

演習03: 任意課題

モデル精度向上のために工夫したポイント

1. モデル精度向上のための考え方

- 精度向上に取り組む際には、データの質と量の両方を改善することが重要だと考えました。限られたデータセットでは、モデルが持つ予測能力に限界があるため、データ量を増やす工夫と、より有用な特徴量を生成することで、予測精度を高めることができると判断しました。
- データの増加においては、スライディングウィンドウを使用して時系列データの一部を異なる時間スライスとして取り扱い、実際のデータを「拡張」する手法を取り入れました。また、特徴量の組み合わせや交互作用を取り入れることで、モデルに提供する情報の質を向上させ、過去の売上傾向や、製品、カテゴリ、店舗ごとの特徴をより反映した予測を目指しました。

演習03: 任意課題

モデル精度向上のために工夫したポイント

2. 選択したモデルの型と理由

選択したモデルは **LightGBM** です。LightGBMは、勾配ブースティング法に基づく決定木モデルであり、特に大量のデータに対して高速に学習・予測を行えることが特徴です。

以下がLightGBMを選択した理由です：

- **高速学習・推論**：LightGBMはデータが大規模でも高速に学習できるため、データ量を増やしても効率的にトレーニングが可能です。
- **特徴量重要度の可視化**：LightGBMは特徴量の重要度を容易に確認できるため、不要な特徴量の削除や、追加すべき特徴量の選定に役立ちます。
- **扱いやすい**：汎用性が高く、パラメータチューニングがしやすいこともあり、他のアルゴリズムと比較して短時間で結果を得ることができました。

演習03: 任意課題

モデル精度向上のために工夫したポイント

3. データ拡張と特徴量エンジニアリング

特徴量の選択は、データの時間的な変化や相互作用を捉えることを重視しました。具体的な特徴量の生成に際して、以下の点を工夫しました。

a. スライディングウィンドウでデータ量を増加

スライディングウィンドウ手法を用いることで、異なる時間区間ごとのデータを学習に使用しました。これにより、モデルに与えるデータ量を増やし、過去のトレンドをより多く反映させることができました。特に売上データは季節性や周期性が強いと考えられたため、スライディングウィンドウで過去のデータを使用し、その周期性を学習できるようにしました。

b. 商品、カテゴリ、店舗ごとの特徴量

- **過去22ヶ月の売上個数と価格の合計と平均** : 商品、カテゴリ、店舗ごとの過去22ヶ月間の売上個数の合計と平均、価格の合計と平均を特徴量として取り入れました。これにより、各商品やカテゴリ、店舗の基本的な売上傾向や価格傾向をモデルに学習させ、予測に役立てることができました。
- **複合特徴量** : 商品とカテゴリ、カテゴリと店舗などの組み合わせを使った複合特徴量を生成しました。例えば、商品と店舗やカテゴリ間の相互作用を捉えるために、売上個数や価格の掛け算を行い、商品が特定のカテゴリや店舗内でどのようにパフォーマンスするかを特徴量として導入しました。

演習03: 任意課題

モデル精度向上のために工夫したポイント

3. データ拡張と特徴量エンジニアリング

c. トレンド特徴量

過去のデータから売上や価格のトレンドを捉えるために、以下のトレンド特徴量を生成しました。

- **過去10ヶ月の売上個数と価格の平均・合計**

スライディングウィンドウによって作成した過去10ヶ月の売上個数と価格の平均、および合計を特徴量として導入しました。これにより、短期間の売上傾向や価格動向をモデルが学習できるようにしました。これらの特徴量は、商品やカテゴリの周期的なパターンを捉えるために役立ちます。

- **売上個数・価格の差分**

過去の売上個数や価格の変化を捉えるために、特定の期間間の差分を特徴量として導入しました。特に以下の差分を使用しました：

- **差分 10-9 (個数・価格) :** 過去10ヶ月目と9ヶ月目の売上個数および価格の差分を計算し、短期的な変動を捉えました。
- **差分 10-1 (個数・価格) :** 過去10ヶ月目と1ヶ月目の売上個数および価格の差分を計算し、より長期的な変動を捉えることを目指しました。
- **平均との差分 (個数・価格) :** 過去10ヶ月間の売上個数および価格の平均と最新月の数値との差を取り、トレンドの方向性を捉えました。この平均との差は、商品やカテゴリの全体的な変動を示すため、重要な指標となります。

- **売上傾向フラグ**

過去の売上データから、売上が上昇しているか下降しているかを示す **売上傾向フラグ** を生成しました。具体的には、売上の急激な変化(上昇や下降)をキャッチし、モデルがその傾向を学習できるようにしました。これにより、例えば特定の商品やカテゴリが人気上昇中か、あるいは人気低下しているかをモデルが捉え、予測の精度向上に寄与しました。

演習03: 任意課題

モデル精度向上のために工夫したポイント

3. データ拡張と特徴量エンジニアリング

d. カレンダー特徴量

曜日や祝日などのカレンダー情報も売上に影響を与える可能性が高いため、カレンダー特徴量を追加しました。特に消費者行動が曜日や特定のイベントに大きく左右されることを考慮し、これらの要素をモデルに反映させることが精度向上に寄与しました。曜日や祝日の情報に基づいて、どの期間に売上が上昇するかなどを予測するためにカレンダー情報を活用しました。

演習03:任意課題

- 構築したAIモデルの精度を記入してください。

- RMSE

2.65

- 店舗あたり予実誤差(販売実績あり)

151.3個

- 店舗あたり予実誤差(販売実績なし)

46.0個

- 月間販売個数対比(販売実績あり)

72.0%

- 月間販売個数対比(販売実績なし)

61.6%