データサイエンスプロジェクト成果物一覧（各項200文字詳細版・全章）

# 目次

1. 1. データ前処理（Preprocessing）
2. 1.1 データ構造と属性の確認
3. 1.2 欠損値と外れ値の処理
4. 1.3 コロナ影響の考慮
5. 2. 探索的データ分析（EDA）
6. 2.1 時系列の可視化
7. 2.2 相関構造の分析
8. 2.3 コロナ影響の可視化
9. 3. モデリング（セグメント別）
10. 3.1 ARIMAおよびProphetモデル
11. 3.2 DeepARモデルの適用
12. 3.3 モデル自動化と管理
13. 4. モデル評価と比較
14. 4.1 評価指標の比較
15. 4.2 実測値と予測値の比較
16. 5. 最終予測結果・成果物
17. 5.1 出力ファイルと形式
18. 5.2 ダッシュボード化
19. 6. ドキュメント類
20. 6.1 報告書の構成
21. 6.2 運用マニュアル
22. 7. 補足説明と活用例
23. 7.1 モデルごとの特性比較
24. 7.2 ビジネス応用例
25. 8. 予測手法の方針
26. 8.1 月別グリッドの時系列モデル
27. 8.2 年別グリッドの回帰モデル
28. 8.3 モデル選定と運用方針

# 1. データ前処理（Preprocessing）

## 1.1 データ構造と属性の確認

本プロジェクトでは、地域、業種、顧客ID、売上、年月などの属性を持つ売上データを用います。データ型やカテゴリ、欠損の有無を確認し、予測に有効な特徴量の選定とデータの信頼性の確認を行います。

## 1.2 欠損値と外れ値の処理

月次・年次売上には一部欠損や外れ値が含まれており、分析の前に適切な処理が必要です。補完には前後値や中央値を使用し、極端な異常値については除外または補正を実施します。

## 1.3 コロナ影響の考慮

新型コロナの影響を強く受けた2020年から2023年5月までのデータは、売上に大きな変動が見られるため、学習に悪影響を及ぼす可能性があります。必要に応じて除外を検討します。

# 2. 探索的データ分析（EDA）

## 2.1 時系列の可視化

各地域・業種の売上推移を時系列で可視化し、季節性やトレンドの有無を確認します。異常変動や構造的変化を早期に発見することで、モデル選定や補完方針の立案に役立てます。

## 2.2 相関構造の分析

自己相関および偏自己相関の分析により、時系列の依存関係や周期性を確認します。これにより、ARIMAやDeepARなど適切な時系列モデルを選定するための基礎情報を得られます。

## 2.3 コロナ影響の可視化

コロナ禍による売上の変動をグラフ化し、通常時と異なる挙動を可視化します。特に飲食・観光業など、影響が大きいセグメントについて重点的に観察を行います。

# 3. モデリング（セグメント別）

## 3.1 ARIMAおよびProphetモデル

ARIMAやProphetは季節性・トレンド成分を明示的に分解し、月次売上データの予測に適しています。Prophetは自動でトレンド変化点を検出でき、実装・調整も比較的容易です。

## 3.2 DeepARモデルの適用

DeepARはRNNを用いた多系列時系列モデルで、地域や業種ごとの売上データを同時に学習できます。系列ごとに異なる傾向を吸収し、信頼区間を含む予測結果を出力できます。

## 3.3 モデル自動化と管理

モデル構築・学習・評価の一連の処理をPythonスクリプトで自動化します。セグメントごとに処理をループで実行し、モデルとログを管理・保存できる仕組みを構築します。

# 4. モデル評価と比較

## 4.1 評価指標の比較

RMSE、MAE、MAPEといった誤差指標を用いて各モデルの性能を比較します。これによりモデルの精度を定量的に評価し、ビジネス要件に合致する最適なモデルを選定します。

## 4.2 実測値と予測値の比較

予測値と実測値を可視化し、モデルがトレンドや季節性を適切に捉えているかを確認します。信頼区間や予測誤差を視覚化することで、実用性を高める判断材料とします。

# 5. 最終予測結果・成果物

## 5.1 出力ファイルと形式

予測結果はCSV形式で出力され、年月・地域・業種・予測値・信頼区間などの情報を含みます。BIツールや分析アプリに取り込みやすい構造で提供されます。

## 5.2 ダッシュボード化

StreamlitやTableauなどを用いて、予測結果を直感的に確認できるダッシュボードを作成します。セグメント選択や期間指定などのインタラクティブ操作が可能です。

# 6. ドキュメント類

## 6.1 報告書の構成

分析プロセス全体を体系的に記述した報告書を作成します。目的、データ内容、処理手順、モデル設計、結果、考察までを含み、説明責任やチーム共有に活用します。

## 6.2 運用マニュアル

モデルの定期更新や再学習、予測失敗時の対応方法などを記載した運用マニュアルを整備します。現場運用者が継続的に対応可能な体制を整えるための資料です。

# 7. 補足説明と活用例

## 7.1 モデルごとの特性比較

Prophetは実装が簡便で短期予測に有効、DeepARは多系列・確率予測に強みがあります。モデルごとの特性を理解し、目的とデータ構造に応じた使い分けが必要です。

## 7.2 ビジネス応用例

予測モデルを活用して売上減少が見込まれる地域・業種を抽出し、販促・営業戦略の優先順位を決定します。リスク予測や人的リソース配分の判断にも貢献します。

# 8. 予測手法の方針

## 8.1 月別グリッドの時系列モデル

月次単位での予測にはARIMAやProphet、DeepARなどを用い、季節性や傾向の変化を捉えます。セグメントごとに予測を行うことで詳細な変動の把握が可能です。

## 8.2 年別グリッドの回帰モデル

2023・2024年の特徴量を基に、顧客単位の2025年売上をLightGBMなどで予測します。データは2年分でも顧客数が多いため、回帰モデルでの予測が実用的です。

## 8.3 モデル選定と運用方針

時系列モデルは短期の変動把握、回帰モデルは個別予測や中長期的な傾向把握に適しています。目的に応じて併用し、柔軟かつ継続的な予測体制を構築します。