

機械学習による売上予測



2023年度プロジェクトリサーチ

プロジェクトリサーチ 報告書

プロジェクト名: 機械学習による売上予測
アドバイザー教員: 樋口三郎
メンバー: 和氣匠汰

1. 概要

本調査は、瀬田キャンパス・青志館にある食堂の売上を予測する時系列モデルの作成を目的としたものである。モデルは未完成であるため、作成にあたり調査を行った食堂の現状、データの状態、気候といった説明変数として扱えるデータセットの入手方法、モデルの学習に用いる事のできる類似のデータセットをここに掲載する。

2. 瀬田キャンパス・青志館食堂の現状

本章では対象の食堂の現状及び保管されるデータセットの形式について述べる。

2.1 食堂売り上げデータの形式、状態

食堂の売上データはシステムの変更により 2023 年 4 月以降のもののみ保管されている。月毎、週毎、日毎の売上をエクセルの形式で出力することが可能であり、「日付、決済時刻、商品名、単価、カテゴリー」といった項目がある。そのため Python の Pandas ライブラリを用いて「日付、商品名、日ごとの売上個数、カテゴリー」といった具合に時系列解析が可能な形式への整形が容易である。

2.2 仕入れの現状

食堂が採用する会計システムに搭載された分析ツールはあるものの、機械学習を含む予測ツールは用いておらず、職員が曜日や過去の売り上げ個数から仕入れ数を決定しているのが現状である。既存の予測ツールがない以上、ユーザーにとって扱いやすいソフトウェアである必要があるだろう。

3. 気象情報の取得方法

売上や来場者数の予測には気象情報が広く用いられる。本調査において気象情報を取得する手法は以下の条件を満たす事が理想である。

1. 商用利用可能である。
2. 有料でない。
3. 滋賀県という地域を指定できる。
4. データセットを手動で更新する必要がない。
5. 訓練データの期間内の気象データを提供する。

以上の 4 つ全てを満たす API やサービスは 2023 年 10 月時点では確認できない。よって訓練データには気象庁で公開されるデータを使用し、逐次的なデータの取得を要求される実用時には Weather API を用いるという手法を採用する。Weather API は商用利用可能な API であり、過去 14 日分の気象データ

にアクセスできる¹。また気象庁のホームページに公開される気象データは商用利用可能であり²，CSVによる出力も可能である。

4. 公開されている類似のデータセット

本章では事前学習やデータセットの拡張に用いることのできる，売上データに類似した二つの公開データセットについて述べる。

4.1 Kaggle の Predict Future Sales

Kaggle のコンペティションである Predict Future Sales はロシアの IC Company により提供された商品の売上データセットである。³商品名はロシア語で提供されており，2013年の一年分のデータが公開されている。「日付，品名，売上個数，カテゴリー」といったパラメータがあり，商品は食料品ではないが，事前学習データセットとして用いることができる。しかしこのデータセットを商用のモデルの学習に用いることが可能かどうか規約に明示されていないため，問い合わせる必要がある。また，Discussion ページに各コンペティション参加者のアプローチが掲載されているため，参考にすることができる⁴。

4.2 日経 POS 情報

日経 POS 情報とは，日本経済新聞社が独自にスーパーマーケット，コンビニエンスストア，ドラッグストアなどから収集する POS データを公開するサービスである⁵。2019年~2022年の食料品の売上が日毎に公開されている。有料であるが，2023年10月現在，龍谷大学のデータベースに公開されており，龍谷大学図書館蔵書検索システムを用いてアクセスが可能である。しかしこちらもデータセットを商用のモデルの学習に用いることが可能かどうか不明のため，別途問い合わせる必要がある。

5. 総括

2章では，瀬田キャンパス・青志館の現状について，保管されているデータは十分時系列解析を行うことのできる形式として出力可能である事，現在2023年4月以降のもののみ入手可能である事，また仕入れについても現状機械学習の技術を用いてはいない事を述べた。3章では，気象データの取得には気象庁のデータセットと Weather API の両方を用いる手法が有効である事を示した。4章では，事前学習やデータセットの拡張に有用な Kaggle の Predict Future Sales と日経 POS 情報を紹介した。しかし前述の通り，この双方のデータセットの使用には商用利用の可否を確認する必要がある。

¹ WeatherAPI. (2023). Retrieved from <https://www.weatherapi.com/>

²Japan Meteorological Agency. (2023). Commentary. Retrieved from <https://www.jma.go.jp/jma/kishou/info/coment.html>

³Kaggle. (2023). Competitive Data Science: Predict Future Sales - Overview. Retrieved from <https://www.kaggle.com/competitions/competitive-data-science-predict-future-sales/overview>

⁴Kaggle. (2023). Competitive Data Science: Predict Future Sales - Discussion. Retrieved from <https://www.kaggle.com/competitions/competitive-data-science-predict-future-sales/discussion>

⁵ Nikkei POS. (2023). Retrieved from <https://nkpos.nikkei.co.jp/>

謝辞

まず、技術的なご指導を賜りました先端理工学部 数理・情報科学課程准教授 樋口教授に対し、深く感謝申し上げます。ご多忙の中、熱心にご指導いただきましたこと、誠にありがとうございました。

次に、調査の前段階において貴重なご意見とアドバイスを提供してくださった先端理工学部教務課の田畑様に感謝申し上げます。また、特に龍谷大学エクステンションセンターの河角様及び関係者の皆様には、心より感謝申し上げます。皆様の協力により、調査の方向性を明確にすることができました。

また、食堂のデータ提供をしていただきました瀬田ショップ SMYLE 店長 若松様並びに青志館食堂店長 佐崎様に感謝の意を表します。皆様のご協力により、実地調査を円滑に進めることができました。

最後に、本調査に携わったすべての方々に対し、心より感謝申し上げます。まだ途中段階ではありますが、本調査は私自身にとって大変貴重な学びとなっております。