Capstone Project – Walmart Sales

Table of Contents

Problem Statement

The problem at hand is to analyze and forecast weekly sales data for 45 Walmart stores. This involves exploring the relationship between weekly sales and various features such as temperature, CPI, unemployment, holidays, etc. Additionally, the objective is to build a time series forecasting model for each store to predict sales for the next 12 weeks.

Project Objective

The main objectives of this project are:

- Analyze the impact of different factors on weekly sales.
- Develop SARIMA (Seasonal AutoRegressive Integrated Moving Average) models for each store to predict weekly sales.
- Evaluate the performance of the forecasting models.
- Provide insights and recommendations based on the analysis.

Data Description

The dataset used in this project is sourced from Walmart and contains information about weekly sales, temperature, CPI, unemployment, and holiday flags for 45 stores.

Data Pre-processing Steps and Inspiration

1.Data Importing:

- Imported necessary Python packages.
- Loaded the Walmart dataset using Pandas.

2.Exploratory Data Analysis (EDA):

- Explored the dataset's shape, columns, and data types.
- Checked for missing values and duplicates.
- Visualized relationships between features and weekly sales using various plots.

3.Time Series Analysis:

- Converted the 'Date' column to datetime format.
- Conducted stationarity tests for each store's weekly sales.
- Applied seasonal decomposition to identify seasonality patterns.

Choosing the Algorithm for the Project

The chosen algorithm for time series forecasting is SARIMA (Seasonal AutoRegressive Integrated Moving Average) for each store. SARIMA is suitable for modeling the seasonality and trends observed in the weekly sales data.

Motivation and Reasons For Choosing the Algorithm

- SARIMA is well-suited for time series forecasting with seasonality and trends.
- It takes into account autoregressive, differencing, and moving average components, making it robust for capturing complex patterns.
- The seasonal aspect is crucial for understanding weekly sales fluctuations.

Assumptions

1. The data is assumed to be representative and accurate.
2. The chosen features are assumed to have a significant impact on weekly sales.
3. SARIMA modeling assumes stationarity, and differencing is applied to achieve this.

Model Evaluation and Techniques

1.Model Training:

- Applied SARIMA models for each store using training data.

2.Model Testing:

- Evaluated model performance on testing data.

3.Forecasting:

- Forecasted weekly sales for the next 12 weeks for each store.

Inferences from the Same

1. Identified top-performing and worst-performing stores based on historical data and factors like unemployment, CPI, and temperature.
2. Analyzed the impact of holidays on weekly sales.
3. Investigated the relationship between temperature, CPI, and weekly sales for each store.
4. Built SARIMA models for each store and forecasted sales for the next 12 weeks.

Future Possibilities of the Project

1. Incorporate additional external factors for a more comprehensive analysis.
2. Fine-tune hyperparameters for SARIMA models to improve forecasting accuracy.
3. Explore advanced time series forecasting models such as Prophet or LSTM.
4. Integrate real-time data to enhance the model's predictive capabilities.

References

pandas: powerful data analysis tools for Python. Retrieved from https://pandas.pydata.org/

Matplotlib: A 2D Graphics Environment. Retrieved from https://matplotlib.org/

Seaborn: statistical data visualization. Retrieved from https://seaborn.pydata.org/

Statsmodels: statistical models for Python. Retrieved from https://www.statsmodels.org/

scikit-learn: Machine Learning in Python. Retrieved from https://scikit-learn.org/

pmdarima: ARIMA and Seasonal ARIMA forecasting. Retrieved from https://pypi.org/project/pmdarima/