Sales Forecasting

Muhammad Ammar Ali Khan, Rohan Javed, Zain-Al-Abidin, Mahad Kashif

May 12, 2024

Abstract

Sales forecasting is a crucial aspect of business planning, enabling organizations to make informed decisions regarding inventory management, resource allocation, and strategic planning. In this project, we aim to forecast sales based on historical data from a Superstore sales dataset.

1 Introduction

Sales forecasting is a crucial aspect of business planning, enabling organizations to make informed decisions regarding inventory management, resource allocation, and strategic planning. In this project, we aim to forecast sales based on historical data from a Superstore sales dataset.

2 Motivation

The motivation behind this project stems from the importance of accurate sales forecasting in optimizing business operations and maximizing profitability. By leveraging advanced analytical techniques and machine learning models, we aim to develop robust forecasting models that can provide actionable insights for business stakeholders.

3 Data Set Description

The Superstore sales dataset contains detailed information about orders, including order date, ship date, ship mode, customer details, product details, sales amount, quantity, profit, returns, and payment mode. The dataset spans multiple years and consists of thousands of entries.

4 Data Preprocessing

- Handling Null Values
- Standardization (e.g., Standard Scaler, MinMax Scaler)
- Handling Categorical Variables
- One-Hot Encoding
- Dealing with Multicollinearity

5 Models Used

We utilized various machine learning models for sales forecasting, including:

- ARIMA
- SARIMA
- Holt-Winters
- LSTM
- RNN
- ANN
- K-Means
- Ridge Regression
- XGBoost Regressor
- Random Forest

Table 1: Model Performance Metrics

Model	MAE	MSE
ARIMA	516.42	418921.81
SARIMA	283.70	313578.65
Holt-Winters	298.61	312231.98
LSTM	257.92	299447.62
RNN	258.42	303995.86
ANN	162.77	106283.69
K-Means	58.96	35637.80
Ridge Regression	268.97	309042.61
XGBoost Regressor	244.44	286566.56
Random Forest	227.21	242187.92

6 Graphs and Plots

Graphs and plots depicting the trends, seasonality, and dependencies in the sales data were generated using PowerBI, we created an interactive dashboard with respect to different states.

Figure 1: General

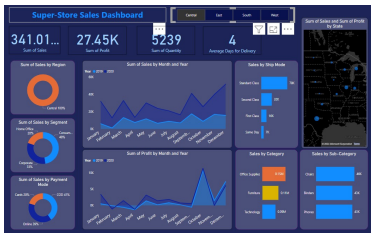

Figure 2: Central

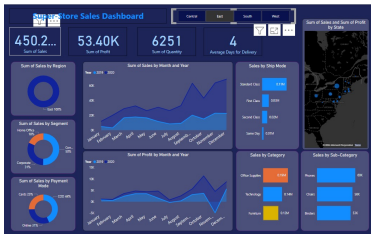

Figure 3: East

Figure 4: South

Figure 5: West

7 Problems Faced

One of the major challenges encountered during the project was fitting the data to the selected models. Additionally, data preprocessing tasks such as splitting the data and encoding categorical variables posed challenges.

8 Solutions

To address the challenges faced, we adopted a systematic approach, including thorough data preprocessing, experimentation with different model architectures, and fine-tuning of hyperparameters .Also used K-means clustering to segment data into distinct groups to find varied sales patterns within the dataset. Collaborative problem-solving and peer review sessions also played a crucial role in overcoming obstacles.

9 Feature Extraction

Used wrapper-based methods to select the most relevant features for accurate forecasting.

10 Contributions

Each team member contributed to different aspects of the project, including data preprocessing, model implementation, experimentation, and result analysis. Muhammad Ammar Ali Khan focused on data preprocessing and model evaluation. Rohan Javed led the model implementation efforts and conducted extensive experimentation. Zain-Al-Abidin contributed to data visualization and result interpretation. Mahad Kashif played a key role in problem-solving and fine-tuning model parameters.

11 Bonus Marks

We used lime to perform comparative analysis on the regressions models.

12 Conclusion

In conclusion, the sales forecasting project provided valuable insights into the application of machine learning techniques in business analytics. Despite encountering challenges, the collaborative efforts of the team resulted in the development of accurate forecasting models. Moving forward, further refinement and deployment of the models in real-world scenarios can enhance business decision-making processes and drive organizational success.