Internship Project Report

Project Title: Demand Forecasting for E-Commerce using Machine Learning and Time Series Analysis

submitted in partial fulfillment of the requirements of Infosys Springboard Internship

By

Name of: Sandipan Rakshit
Email id: sandipanrakshit6@gmail.com

Under the Guidance of
Trapti Singhal

GitHub Link for Code: https://github.com/sandipanrakshit34/Demand-Forecasting-for-E-Commerce

Internship Guide, Infosys Springboard Internship

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to **Ms. Trapti Singhal**, my internship guide, for her continuous support, guidance, and encouragement throughout the course of this project. Her insightful feedback, technical expertise, and valuable suggestions greatly contributed to the successful completion of the internship.

I also extend my heartfelt thanks to **Infosys Springboard** for providing me with this wonderful opportunity to work on the **Demand Forecasting for E-Commerce** project. This internship has significantly enriched my understanding of real-world data analytics, time series forecasting, and the practical application of machine learning in the e-commerce domain.

Lastly, I would like to thank my peers and mentors who supported and motivated me throughout this journey.

ABSTRACT

In the dynamic landscape of e-commerce, accurate demand forecasting is crucial for optimizing inventory management, reducing operational costs, and enhancing customer satisfaction. This project leverages machine learning techniques to develop predictive models for forecasting future product demand based on historical sales data and external factors such as digital marketing metrics.

The dataset includes time-series data enriched with exogenous variables like Google clicks, Facebook impressions, and other relevant KPIs from Google Analytics. Forecasting models used include ARIMA, SARIMA, SARIMAX, and machine learning-based regression models. The project follows a structured approach involving data preprocessing, feature engineering, and model evaluation using performance metrics such as RMSE and MAPE.

The insights derived from this project contribute to efficient resource allocation, cost savings, and improved customer experience in the e-commerce domain.

TABLE OF CONTENTS

Chapter 1. Introduction

1.1 Problem Statement:

E-commerce businesses face significant challenges in predicting future product demand due to dynamic consumer behavior, promotional activities, and seasonal trends. Inaccurate demand forecasting leads to overstocking or stockouts, both of which impact revenue and customer satisfaction.

1.2 Motivation:

The motivation behind this project is to develop a data-driven solution that leverages historical sales and marketing KPIs to improve demand forecasting accuracy using machine learning and statistical models.

1.3 Objectives:

- Forecast future demand using historical and external data.
- Enhance decision-making in inventory and supply chain management.
- Evaluate different forecasting models and assess their accuracy.
- Visualize data trends and predictions effectively.

1.4 Scope of the Project:

The project uses Python for model development and visualization, leveraging libraries such as Pandas, NumPy, Scikit-learn, Statsmodels, and Matplotlib. The scope includes time series analysis, regression modeling, data visualization, and evaluation of forecasting accuracy.

Chapter 2. Literature Survey

Demand forecasting is a widely studied area in operations research, time series analysis, and machine learning. Traditional models like ARIMA have been effective for univariate series, while SARIMA/SARIMAX account for seasonality and exogenous factors.

Machine learning models such as Random Forest and XGBoost have shown high performance in handling multivariate and nonlinear relationships. Studies emphasize combining KPIs from digital platforms (Google/Facebook) to improve predictive capability in e-commerce environments.

References consulted include Kaggle datasets, research papers from ScienceDirect, IJMERR, and academic articles on hybrid forecasting approaches.

Chapter 3. Proposed Methodology

3.1 Workflow:

1. **Data Collection** – Sales data + KPIs (Google Clicks, Facebook Impressions)

2. **Preprocessing** – Handling missing values, outlier detection, normalization
3. **Feature Engineering** – Lag features, moving averages, date-based features
4. **Modeling** – ARIMA, SARIMA, SARIMAX, Random Forest, Linear Regression
5. **Evaluation** – RMSE, MAE, MAPE, R2
6. **Visualization** – Line plots, trend graphs, prediction vs. actuals

3.2 Tools and Libraries:

- Python, Pandas, NumPy
- Scikit-learn, Statsmodels
- Matplotlib, Seaborn
- Jupyter Notebook, Google Colab

Chapter 4. Implementation and Results

- Data merged from multiple Excel files.
- Outliers imputed using IQR and 97th percentile.
- Visual trends observed in Clicks vs. Sales.
- AR Model: Validation RMSE: ~7.46, MAPE: ~11.94%; Test RMSE: ~6.21, MAPE: ~9.25%
- Regression analysis showed positive correlation between KPIs and sales.

Figures:

- Line charts of Clicks/Impressions/Sales
- Boxplots, trendline overlays, prediction comparison plots

Tables:

- Evaluation metrics per model and dataset split (train, val, test)

Chapter 5. Discussion and Conclusion

5.1 Challenges:

- Integrating heterogeneous datasets (sales and KPIs)
- Handling seasonal and non-linear patterns
- Managing time-series data with exogenous variables
- Ensuring accurate forecasting on limited data

5.2 Conclusion:

The project successfully demonstrated a practical application of time series forecasting integrated with marketing analytics for e-commerce. The models built can guide businesses to better manage inventory and customer satisfaction. Future improvements may include integrating LSTM models, weather data, or real-time dashboards for decision support.

References:

1. https://www.ijmerr.com/v4n1/ijmerr_v4n1_10.pdf
2. https://www.kaggle.com/competitions/walmart-recruiting-store-sales-forecasting/data
3. https://www.researchgate.net/publication/354726783_A_Recent_Review_Article_on_Demand_Forecasting
4. https://www.sciencedirect.com/science/article/pii/S2405896318313272