Sales Forecasting and Data Analysis of Product Based Company

Ameya Santosh Gidh
Northeastern University
(gidh.am@northeastern.edu)

Abstract: Sales prediction and deriving actionable business insights represent fundamental aspects for companies operating in the product-based sector. The ability to accurately anticipate future sales and supply chain trends is pivotal for maximizing profitability. To attain superior sales forecasting outcomes, innovative technologies are continually explored and implemented across industries. Machine learning techniques provide the means to predict the quantity of products or services to be procured over a specified future timeframe. The forecasting models developed through machine learning offer considerable advantages in augmenting the operational efficiency of traditional product-centric enterprises. This report aims to further our comprehension of the utilization of machine learning and statistical models in sales forecasting. Additionally, it seeks to extract business intelligence and generate various visual representations to enhance data analysis.

Keywords: Warehouse Management; Statistical Models; Machine Learning; Supply Chain Management, Industry 4.0; Product based Companies.

1 Introduction

The global landscape is rapidly advancing into a digital era, with Machine Learning and Deep Learning technologies serving as pivotal drivers across diverse industries. Precise predictive analysis holds immense importance for the progress of product-centric companies, significantly impacting their performance and expansion. Within the realm of supply chain management, a significant challenge revolves around efficiently handling and deriving actionable insights from data. This data serves as a valuable resource for forecasting, optimizing inventory productivity, and facilitating well-informed decisions aimed at maximizing company profits. Utilizing the appropriate technologies for forecasting is paramount, as it can either propel or hinder a company's growth trajectory. In this study, I have harnessed the power of Facebook's Prophet model and the SARIMA model to forecast sales in a specific product-based company, focusing on office supplies and furniture. Furthermore, I have subjected these models to comprehensive metric evaluations, yielding valuable insights. While these models and findings are rooted in the data of a specific company, the techniques and knowledge gained can readily be applied to the data of any other product-oriented organization.

2 Literature Survey

In the 21st century, forecasting has emerged as a vital element in the operations of supply chain companies. Modern product-based businesses are eager to accelerate their sales growth to enhance customer satisfaction and revenue. Small-scale supply chain companies are increasingly turning to the effective utilization of Machine Learning (ML) and Deep Learning technologies to swiftly expand and compete with industry giantsThe evolution of AI and ML technologies has led to transformative changes in warehouse management systems, revolutionizing how companies manage order sending and receiving processes. Traditional systems can pose challenges as a company's order volume grows, requiring adaptable warehouse management systems to accommodate evolving inventory and business dynamics. Ineffectual warehouse management systems and forecasting techniques can severely impact a company's profitability, potentially resulting in financial losses due to inaccuracies in order tracking. The logistics processes encompass activities like package picking, sorting, and dispatching.

3 Data

3.1 Dataset and labels

The dataset utilized in this project was sourced from Kaggle. It consists of 24 columns of data. The key component of this dataset used for forecasting is the 'Order Date,' serving as the time series date column. The primary variable being predicted is the 'Sales' column, representing the total revenue generated by the company. The data is recorded on a daily basis, spanning from January 2014 to December 2017.

`

3.2 Data Preprocessing and Cleaning

The resolution of the data was per day. The data was converted from per day The data initially had a daily resolution. To reduce data volatility and enhance the accuracy of forecasting models, it was transformed from daily sales to average monthly sales. As part of this process, ten days with missing sales data were excluded from the analysis and not factored into sales forecasts.Furthermore, data type conversion was carried out to ensure that the data columns had the appropriate data types. The 'Order Date' column, initially identified as an object type, was converted into the 'datetime' format to prevent potential errors during model fitting. In addition, state names were converted into their respective two-letter abbreviations for mapping purposes. Moreover, a 'month' column was derived from the 'Order Date' data to facilitate sales analysis on a monthly basis.

3.3 Columns for Data Analysis:

Among the 24 columns in the dataset, a subset of them was dedicated to forecasting, while the others served in generating business insights and creating visual representations. The 'Order Date' and 'Sales' columns were specifically utilized for forecasting purposes. The 'State' column enabled the plotting of mean sales per state on a geographical map. 'Customer ID' played a role in identifying the customer with the highest number of orders. Finally, the 'Product Name' column contributed to plotting the product count per state on the map.

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Sales
7981	CA-2014-103800	2014-01-03	2014-01-07	Standard Class	DP-13000	Darren Powers	Consumer	United States	Houston	...	77095	Central	OFF-PA-10000174	Office Supplies	Paper	Message Book, Wirebound, Four 5 1/2" X 4" Form...	16.448

Fig.: 1 Dataset Sample

4 Methodology for forecasting

4.1 SARIMA Model:

The SARIMA model, short for "seasonal autoregressive integrated moving average," is a variation of the standard ARIMA model that incorporates seasonality. To implement SARIMA for the dataset in this project, the 'SARIMAX' function from Python's 'statsmodels' package was employed. For the office data, the chosen SARIMAX parameters were (1,0,1) for the ARIMA component and (1,0,1,12) for the seasonal component. Similarly, for the furniture data, the parameters were set to (2,0,2) for ARIMA and (1,0,[1],12) for the seasonal component. These parameter values were selected based on their ability to minimize AIC, BIC, and error values. The SARIMA model was trained on two years of data, with the remaining one year reserved for testing. The 'auto_arima' function in Python aided in finding these optimal parameters.

The choice of the SARIMA model was justified by the presence of clear seasonality in the dataset, as confirmed by the 'seasonal_decompose' function from 'statsmodels,' which visually depicted the seasonality. While evaluating various statistical models, including ARIMA and ARMA, the SARIMA model consistently outperformed the others in this particular dataset.

4.2 Facebook's Prophet Model:

Prophet is a forecasting tool designed for time series data, employing an additive model to capture non-linear trends. It excels at fitting yearly, weekly, and daily seasonality, along with handling holiday effects. This model is most effective when applied to time series data exhibiting strong seasonal patterns and possessing multiple seasons of historical data. Prophet exhibits robustness to missing data and adaptability to shifts in the trend, making it proficient at handling outliers.

To apply the Prophet model to the dataset, I directly utilized Facebook's open-source Prophet package in Python. Given the dataset's pronounced seasonality, Prophet was the preferred choice for this project. Default parameters were used for model fitting, as Prophet possesses the capability to intelligently identify trends in the data and determine the most suitable parameters. Similar to the SARIMA model employed in this project, Prophet was trained on two years of data, with the remaining one year designated for testing.While exploring model options for this project, alternatives such as LSTM and Deep Neural Network were considered alongside Facebook's Prophet model. Ultimately, the Prophet model emerged as the most effective

choice, delivering the lowest error among all the models tested.

5 Experimental Results and Analysis

5.1 Root Mean Squared Error

The following table shows the error percentage for Facebook's prophet model and SARIMA model for the two categories.

	Prophet Model	SARIMA Model
Furniture	24.5%	32.0%
Office Supply	51.02%	34.8%

Fig.: 2 Percentage Error

From Table it can be concluded that Prophet model should be used for forecasting of furniture whereas SARIMA model should be used for office supply.

5.2 AIC and BIC for SARIMA model

The AIC and BIC value obtained for SARIMA model are in the following table. The model fits when these values are least of different combinations of the model parameter.

	AIC	BIC
Furniture	338.64	346.88
Office Supply	510.44	518.44

Fig.: 3 AIC and BIC

5.3 Mean Sales Per State:

Mean Sales of the State is maximum in the state of Wyoming (WY). The maximum mean revenue generated was in the state of Wyoming which was around 1603.136(in dollars).

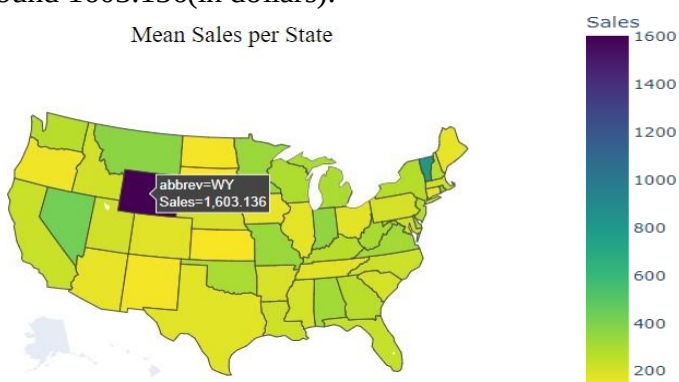

Fig.: 4 Mean Sales Map

5.4 Number of products sold per state:

The maximum amount of product sold is in the state of California. Total number of products sold in California were 2001.

Fig.: 5 Quantity of products sold per state

5.5 Sales per month

November and February are the months with most and least sales respectively.

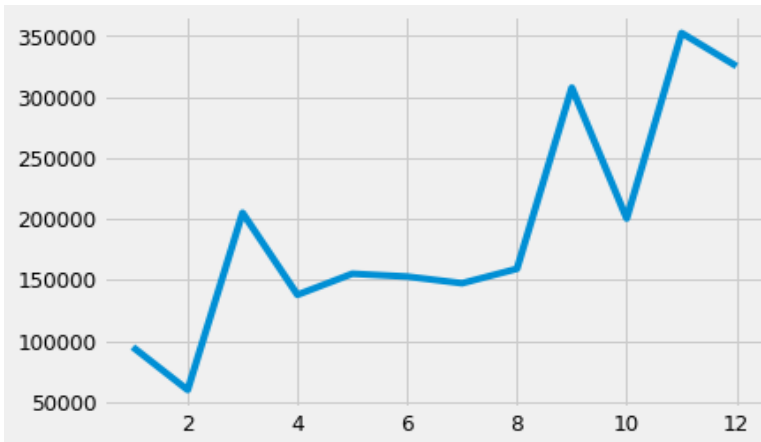

Fig.: 6 Sales Per Month

5.6 Forecasting trend:

Forecasting trend for both the models is increasing until 2024 indicating that demand for both furniture and office supplies will go up. Moreover, the slope for the office supply is much more than furniture indicating office supply will be more in demand in future.

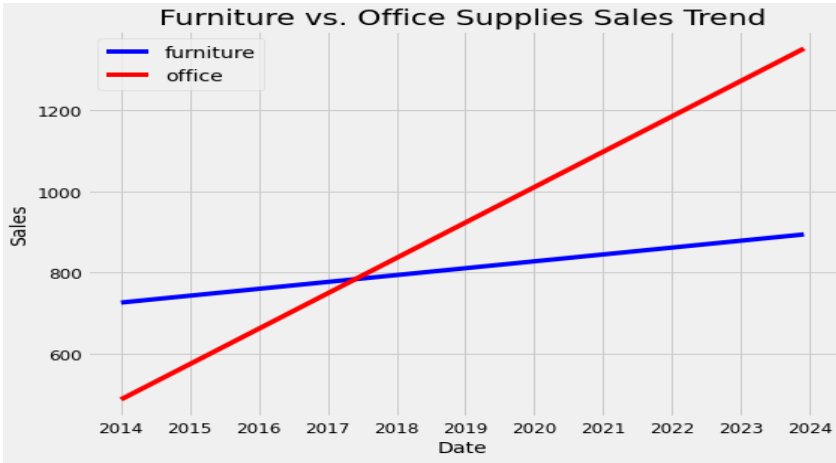

Fig.: 7 Forecasting Trend

5.7 Future forecasts

According to future forecasts obtained by the two models for the office supply and furniture, the estimate sales for office supply is around $1760 and for furniture supply is around $1320

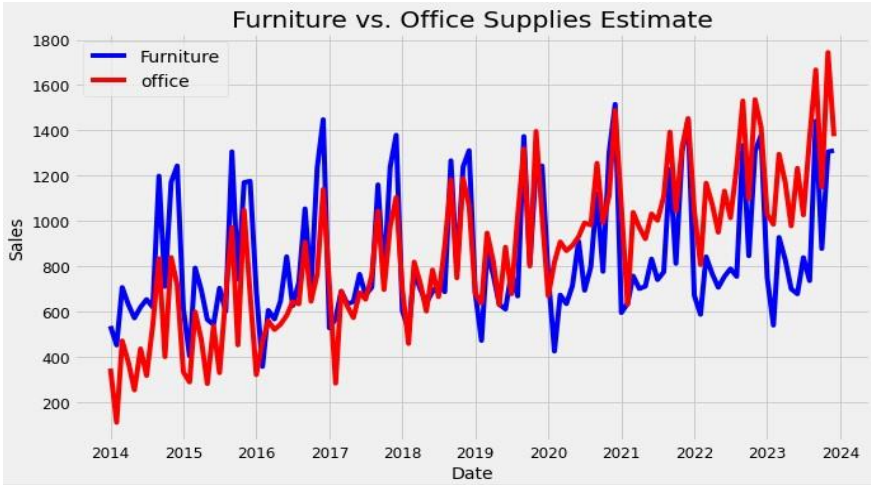

Fig.: 8 Future Forecasts

5 Conclusions

This report highlights the potential of employing Machine Learning and statistical models in the realm of product-based company forecasting. Through problem exploration and an extensive review of both existing systems and Industry 4.0 technologies, it becomes evident that the use of the right Machine Learning models for forecasting can confer a significant competitive edge upon product-based companies. Such models not only contribute to enhancing future sales but also aid in making well-informed decisions regarding inventory management. With a continual influx of innovative models emerging each year, there remains ample room for further advancements and improvements in this field.

Consequently, it is imperative for supply chain companies to consider the integration of Machine Learning into their operations. This move promises enhanced efficiency and rapid market growth, positioning these companies for exponential expansion.

6 References

[1]. Schmidt, A.; Kabir, M.W.U.; Hoque, M.T. Machine Learning Based Restaurant Sales Forecasting. *Mach. Learn. Knowl. Extr.* **2022**, *4*, 105-130. https://doi.org/10.3390/make4010006

[2]. S. Cheriyan, S. Ibrahim, S. Mohanan and S. Treesa, "Intelligent Sales Prediction Using Machine Learning Techniques," *2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*, 2018, pp. 53-58, doi: 10.1109/iCCECOME.2018.8659115.

[3]. Anđelković, Aleksandra, and Marija Radosavljević. "Improving order-picking process through implementation of warehouse management system." Strategic Management-International Journal of Strategic Management and Decision Support Systems in Strategic Management 23.1 (2018).

[4]. Reza Toorajipour, Vahid Sohrabpour, Ali Nazarpour, Pejvak Oghazi, Maria Fischl, Artificial intelligence in supply chain management: A systematic literature review, Journal of Business Research, Volume 122, 2021, Pages 502-517, ISSN 0148-2963, https://doi.org/10.1016/j.jbusres.2020.09.009.

[5]. Facebook's prophet link-https://facebook.github.io/prophet/

[6]. www.kaggle.com/datasets/jr2ngb/superstore-data