Using Time Series Analysis For Sales and Demand Forecasting

Table of Contents

1. Executive Summary

This report presents a comparative analysis of various forecasting models applied to book sales data for *The Alchemist*and *The Very Hungry Caterpillar*. The objective was to develop a robust time series forecasting model to improve inventory planning and sales predictions. Multiple modelling approaches were evaluated, including SARIMA, XGBoost, LSTM, hybrid models, and monthly aggregation techniques.

Findings indicate that **LSTM models** provided the most accurate predictions at the weekly level, outperforming traditional SARIMA and hybrid models in terms of Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). For monthly forecasts, **XGBoost models** significantly outperformed SARIMA, demonstrating the effectiveness of machine learning techniques over classical statistical approaches.

2. Business Problem and Context

Retailers and publishers rely on accurate demand forecasts to optimise stock levels, reduce waste, and capitalise on market trends. Traditional forecasting methods often struggle to account for seasonal variations, external disruptions (e.g., the COVID-19 pandemic), and nonlinear sales patterns. This study aims to evaluate different forecasting techniques to determine which model best captures sales trends and informs strategic decision-making.

3. Data and Methodology

The dataset consists of weekly sales data for *The Alchemist* and *The Very Hungry Caterpillar*, covering more than a decade. Data was cleaned, resampled, and tested for stationarity before applying the following models:

- **SARIMA (Seasonal ARIMA)** – Traditional time series model capturing seasonality and trends.
- **XGBoost (Gradient Boosting Machine)** – Machine learning model using lagged sales as features.
- **LSTM (Long Short-Term Memory Neural Network)** – Deep learning model tailored for sequential data.
- **Hybrid Models** – SARIMA + LSTM models to combine strengths of both methods.
- **Monthly Aggregation** – Forecasting at a monthly level to assess performance over a longer horizon.

Performance was evaluated using **MAE** (measuring absolute prediction error) and **MAPE** (measuring percentage error relative to actual sales).

4. Key Insights

4.1 Model Performance Summary

Model	The Alchemist MAE	The Alchemist MAPE	The Very Hungry Caterpillar MAE	The Very Hungry Caterpillar MAPE
SARIMA	155.08	0.2976	353.14	0.1863
XGBoost	94.99	0.1901	360.24	0.1687
LSTM	**77.69**	**0.1524**	**314.93**	**0.1521**
Sequential Hybrid	169.16	0.3719	476.75	0.2552
Parallel Hybrid	111.38	0.2027	482.81	0.2541
Monthly XGBoost	200.07	**0.0791**	703.80	**0.0723**
Monthly SARIMA	1326.63	0.4878	2863.04	0.3193

4.2 Key Takeaways from Model Performance

- **LSTM was the best-performing model for weekly forecasting**, showing the lowest MAE and MAPE for both books.

- **XGBoost performed well at both weekly and monthly levels,** particularly excelling in long-term trend forecasting.
- **Hybrid models underperformed,** as residual errors compounded rather than improving accuracy.
- **SARIMA struggled with capturing complex demand patterns,** especially for *The Very Hungry Caterpillar*.
- **Monthly forecasting had lower MAPE but higher MAE,** meaning relative accuracy improved but absolute errors increased.

5. Model Analysis and Discussion

5.1 Weekly Forecasting

- **LSTM consistently outperformed all models** due to its ability to model long-term dependencies and non-linear patterns.
- **XGBoost was competitive,** offering strong results with greater interpretability.
- **SARIMA lagged behind,** struggling with non-stationary demand patterns.
- **Hybrid models did not improve results,** as error correction was not effective.

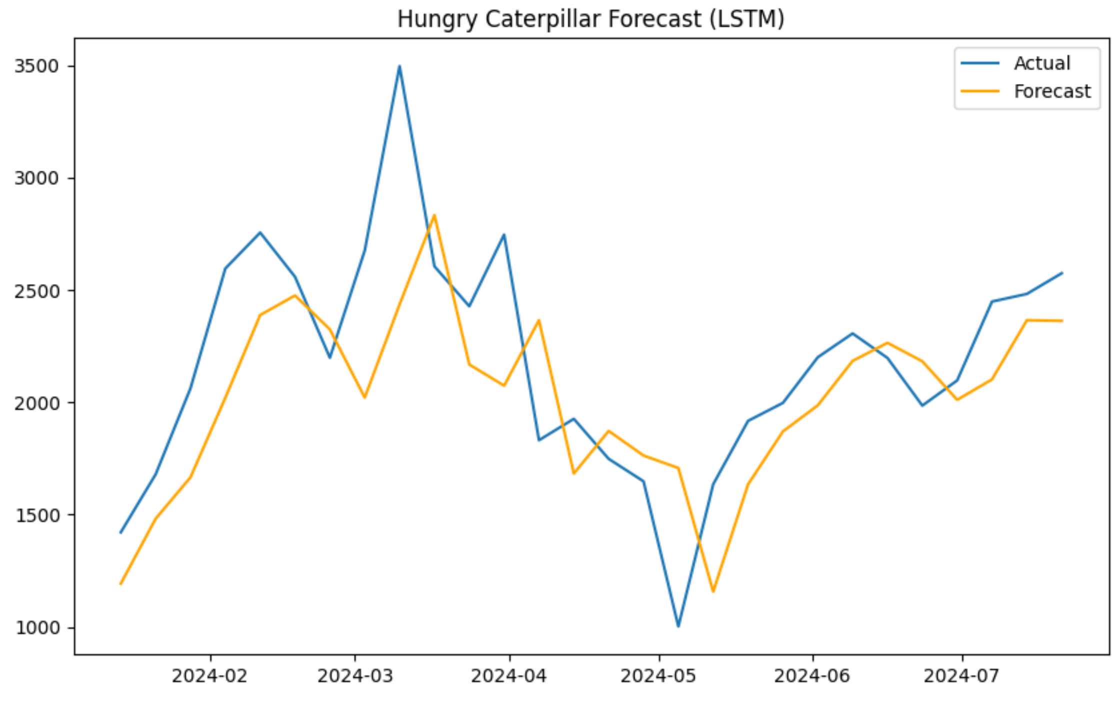

Figure 1. Weekly Forecast for The Very Hungry Caterpillar

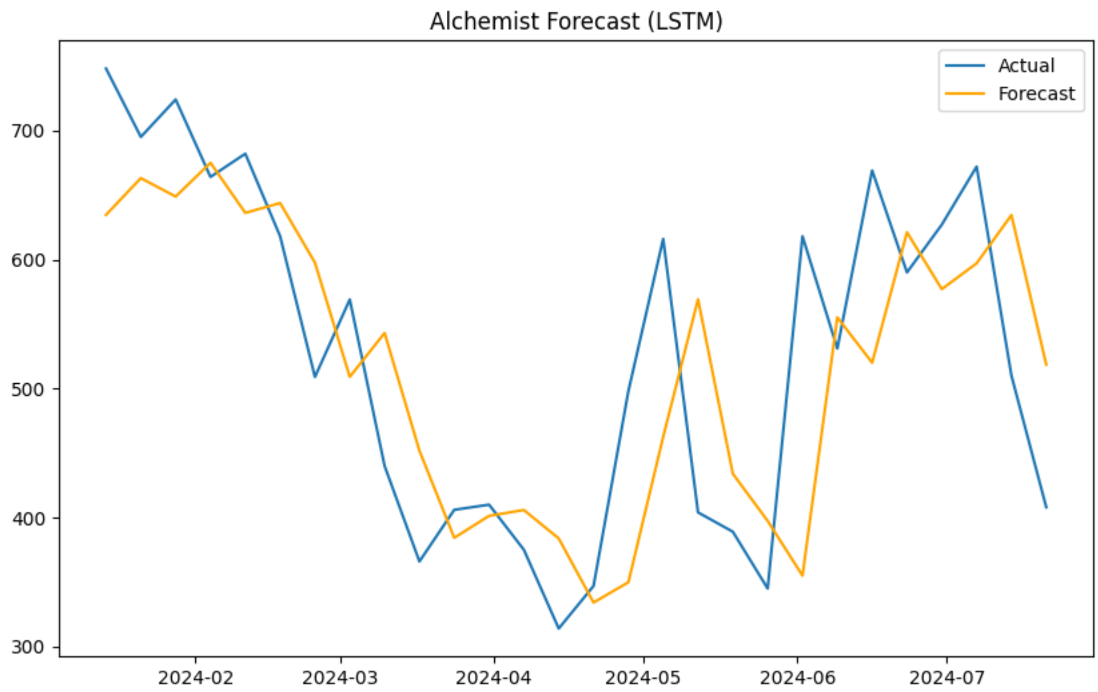

Figure 2. Weekly Forecast for The Alchemist

5.2 Monthly Forecasting

- **XGBoost significantly outperformed SARIMA,** achieving the lowest MAPE for both books.
- **SARIMA struggled,** indicating that traditional time series methods may not be suitable for long-term forecasting.
- **LSTM was not tested for monthly forecasts** due to its focus on short-term sales trends.

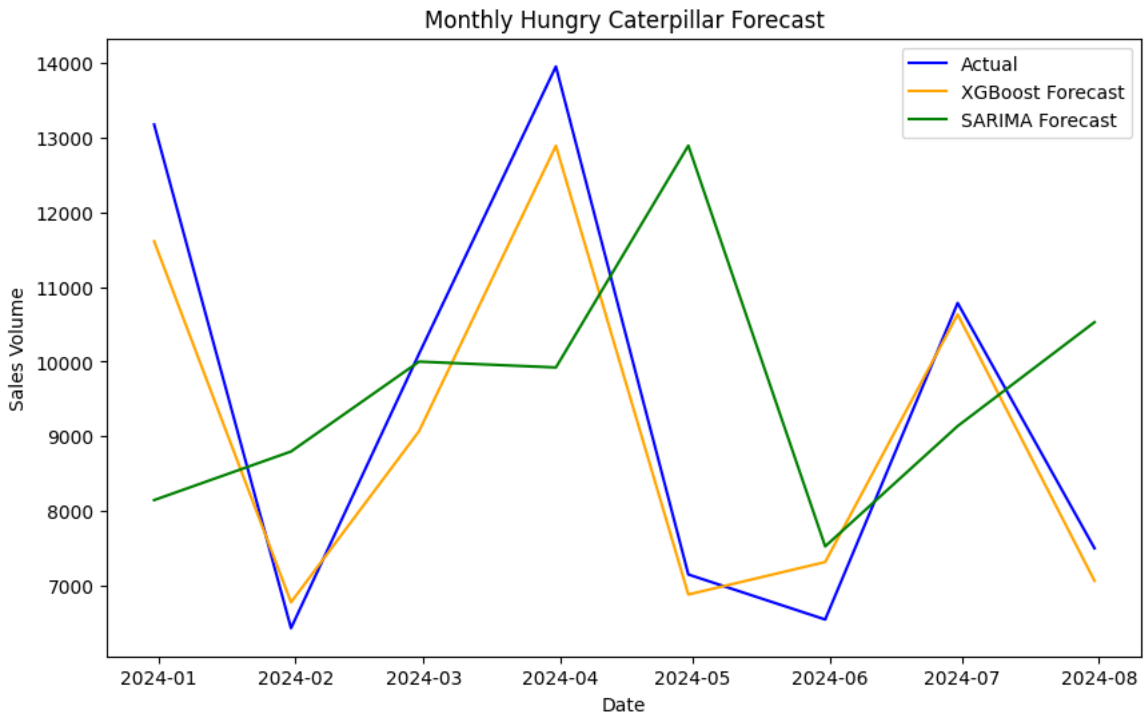

Figure 3: Monthly Forecasts for The Hungry Caterpillar

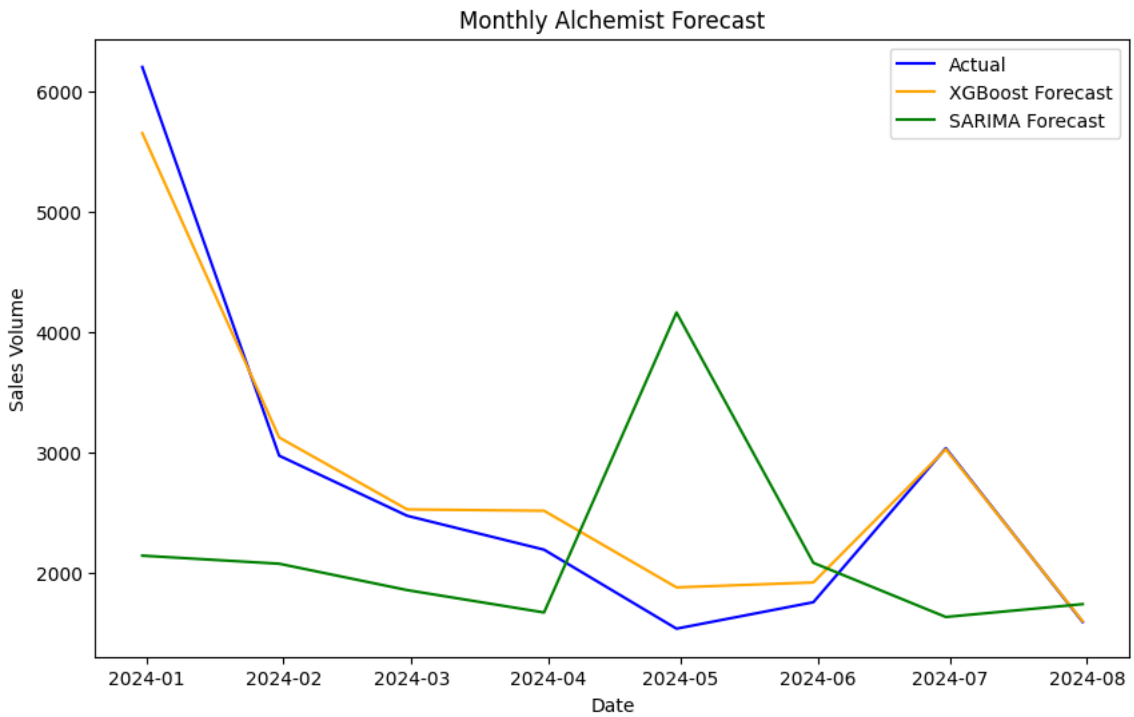

Figure 4: Monthly Forecasts for The Alchemist

6. Strategic Recommendations

- **Use LSTM for weekly forecasting** to optimise inventory for short-term demand fluctuations.
- **Leverage XGBoost for monthly forecasting** to drive long-term strategic planning.
- **Avoid hybrid models**, as they offered no significant accuracy improvements.
- **Monitor external factors** (e.g., promotions, media influences) to refine future models.
- **Automate model retraining** to ensure adaptability to evolving sales trends.

7. Conclusion and Next Steps

This study highlights the effectiveness of deep learning (LSTM) and machine learning (XGBoost) in sales forecasting. While **LSTM excels at short-term accuracy**, **XGBoost is optimal for long-term trend analysis**. The findings suggest that traditional statistical models like SARIMA are **less effective in modern retail forecasting**.

Future work should explore **incorporating external factors** (e.g., marketing campaigns, economic conditions) to further improve predictive accuracy. Implementing these models into inventory management systems can help retailers **reduce waste, prevent stockouts, and increase revenue through data-driven forecasting**.

📊 **Figure: Summary of Model Comparisons** *(Placeholder: Visual representation of weekly vs. monthly model performance.)*