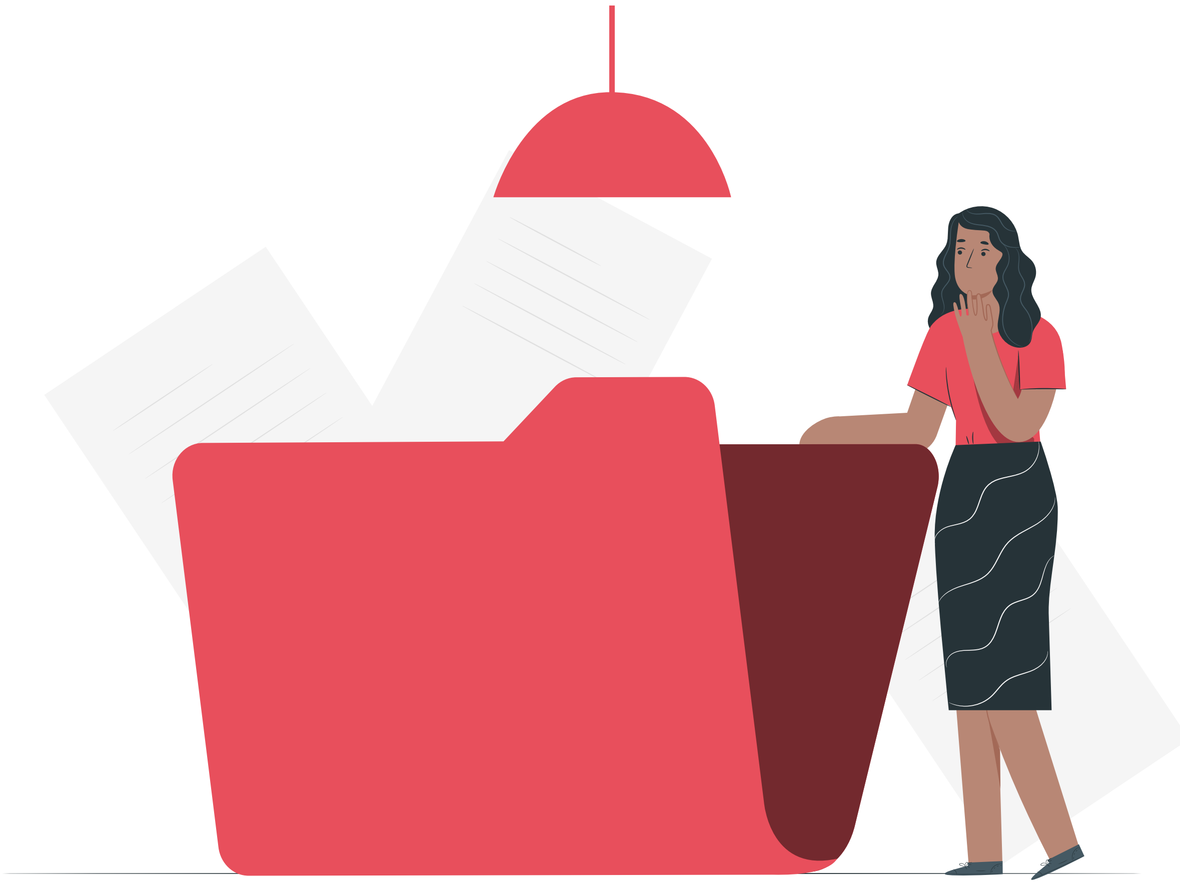
SUPPLYCOPIA

AI-Driven Transformation of Healthcare Supply Chain Budgeting: An In-Depth Analysis

NIRMAL RAYAN, LEAD DATA SCIENTIST, SUPPLYCOPIA



Contents

Abstract

This white paper presents a comprehensive comparative analysis of time series forecasting models tailored for predicting budget spend within the healthcare supply chain. With a focus on enhancing accuracy and efficiency in resource allocation, our study evaluates the performance of traditional models, including Exponential Smoothing and ARIMA, alongside advanced methodologies like LSTM, SVR, and ensemble techniques.

As the healthcare supply chain continues to evolve, our study offers valuable insights into the dynamic landscape of budget forecasting. This knowledge contributes to informed decision-making, optimized resource allocation, and the ongoing resilience of the healthcare supply chain. This white paper serves as a guide for organizations seeking to enhance their forecasting strategies, fostering innovation and adaptability within the intricate and vital realm of healthcare supply chain management.

Keywords— *Healthcare, Supply Chain, Machine Learning, Time Series Forecast, Neural Networks, Prediction Algorithms, Planning*

Introduction

In the dynamic landscape of healthcare supply chain management, accurate budget forecasting is paramount for efficient resource allocation. As healthcare organizations strive to optimize expenditures and enhance operational efficiency, the choice of time series forecasting models becomes a critical decision. This white paper aims to conduct a comprehensive comparative analysis of various time series forecasting models to identify the most effective approach for predicting budget spend in the healthcare supply chain.

A. Background

The healthcare supply chain faces unique challenges, including fluctuating demand, evolving medical practices, and complex procurement processes. Accurate budget forecasts are essential to ensure a seamless flow of resources, avoid shortages, and enhance overall service delivery.

B. Objectives

- Evaluate the performance of different time series forecasting models in predicting budget spend.
- Identify the strengths and weaknesses of each model in the context of healthcare supply chain data.
- Provide practical insights for implementing the most effective forecasting model(s).

Literature Review

The exploration of time series forecasting in healthcare supply chain management has garnered considerable attention in recent literature. Key studies by esteemed authors such as Langenberger et al. (2023), Nagarjuna et al. (2022), Osawa et al. (2020), Rosella et al. (2018), and Emmanuel et al. (2022) have probed into the deployment of sophisticated models for predicting healthcare expenditures.

A. Application of Advanced Models
This section reviews key contributions in this domain, providing insights into the methodologies, findings, and implications.
[1] Langenberger et al. (2023) explored the application of machine learning to predict high-cost patients using healthcare claims data. Their study, published in PLoS One, compares the performance of different models. The findings offer valuable insights into predictive modeling, shedding light on approaches that excel in identifying high-cost patients within healthcare systems.

In their work presented at the 6th International Conference on Computing Methodologies and Communication, [2] Nagarjuna et al. (2022) focused on predicting medicine expenditure using machine learning. The study contributes to understanding the dynamics of medicine expenditure prediction and provides a foundation for optimizing healthcare resource allocation.

[3] Osawa et al. (2020) developed machine-learning-based prediction models for high-need high-cost patients, utilizing nationwide clinical and claims data. Published in NPJ Digit Med, the study emphasizes the potential of machine learning in identifying patients with complex healthcare needs, facilitating proactive and targeted interventions.

[4] Rosella et al. (2018) focused on predicting high health care resource utilization in a single-payer public health care system. Their work, published in Medical Care, introduces the High Resource User Population Risk Tool, contributing to the development of tools for healthcare resource management and allocation.

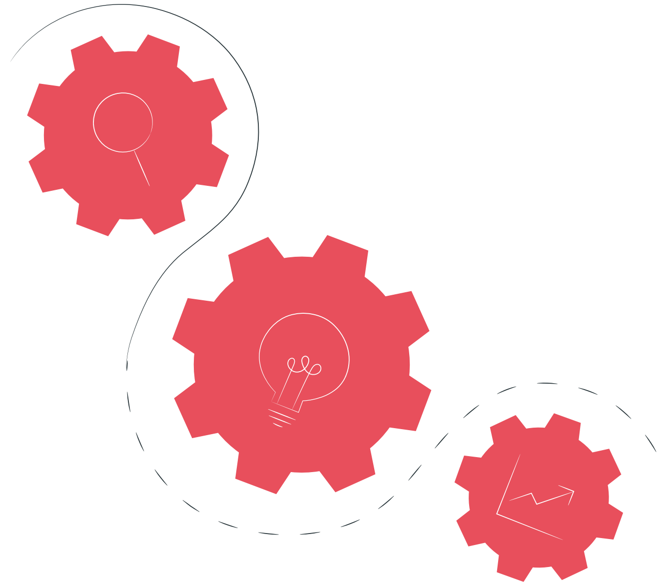
[5] Emmanuel et al. (2022) conducted a literature review on drug supply chain and neural networks. Presented at the 6th International Conference on Information Technology, Information Systems and Electrical Engineering, the review explores the intersection of drug supply chain dynamics and the application of neural networks.

In conclusion, these studies collectively underscore the versatility and potential of machine learning in healthcare forecasting. The diversity of methodologies employed in predicting high-cost patients, medicine expenditure, and health resource utilization reflects the ongoing efforts to optimize healthcare systems, emphasizing the need for tailored approaches to specific domains within the healthcare supply chain.

B. Gaps in Current Research
While existing studies provide valuable insights into specific models' performance, there remains a notable gap in comprehensive comparative analyses that encompass a broader spectrum of forecasting techniques. Few studies have addressed the nuanced challenges posed by the healthcare supply chain, particularly in terms of category-wise, supplier-wise and facility-wise budget predictions.

C. The Need for Comparative Analysis
The evolving nature of healthcare supply chain dynamics necessitates a more exhaustive examination of forecasting models. This white paper seeks to bridge the existing gap by offering a comparative analysis that not only assesses the performance of traditional models but also explores the potential of emerging techniques such as LSTM, SVR, and ensemble methods in the unique context of healthcare budget forecasting.

Methodology



A. Data Collection
The foundation of our comparative analysis lies in a comprehensive dataset sourced from Advantus Health Partners (AHP), a group purchasing organization (GPO) in the US healthcare industry. This dataset encompasses historical records of budget spend within the healthcare supply chain, capturing various dimensions such as time, contract category, suppliers, and facilities.

B. Preprocessing
To ensure the quality and uniformity of the data, a rigorous preprocessing pipeline was employed. This included:

Missing Data Handling: Addressing any gaps in the dataset through imputation or removal.

Outlier Detection: Identifying and addressing outliers that could skew the model training.

Temporal Aggregation: Aggregating data at suitable temporal intervals to align with the forecasting objectives.

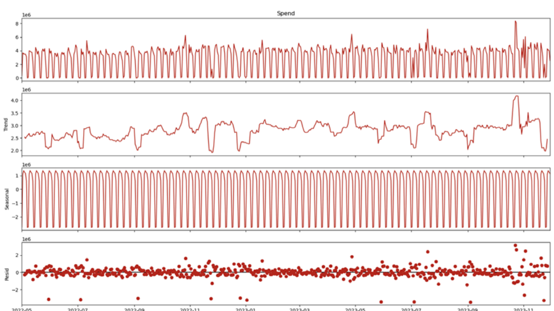


Fig. 1 The decomposition of Spend time series to isolate error, trend, and seasonality

C. Model Selection
A diverse set of time series forecasting models were selected to represent different methodologies and complexities:

Exponential Smoothing: Leveraging the simplicity of exponential smoothing for baseline comparison.

ARIMA (AutoRegressive Integrated Moving Average): A widely-used statistical model for time series analysis.

LSTM (Long Short-Term Memory): A deep learning model well-suited for capturing complex temporal dependencies.

SVR (Support Vector Regression): Exploring the potential of machine learning regression in time series forecasting.

Ensemble Method: Combining predictions from multiple models to harness collective intelligence.

D. Model Training and Evaluation
The dataset was split into training and testing sets to facilitate model training and evaluation. The training set comprised 80% of the data, leaving the remaining 20% for testing. Models were trained on historical data and evaluated on unseen test data.

For each model:

Parameter Tuning: Hyperparameter tuning was performed using techniques such as grid search and cross-validation.

Evaluation Metrics: Performance was assessed using key metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared to capture accuracy and goodness of fit.

E. Ethical Considerations
This research adheres to ethical guidelines in data usage and model development. Privacy and security measures were implemented to protect sensitive information within the healthcare supply chain dataset.

F. Software and Tools
All analyses were conducted using sklearn, tensorflow and keras in Python, ensuring transparency and reproducibility of results.

Model Comparison

The heart of our research lies in the comparative analysis of various time series forecasting models applied to healthcare supply chain budget spend data. Each model was meticulously evaluated using key metrics to discern their efficacy in capturing the intricacies of the dataset.

A. Evaluation Metrics
The performance of each model was rigorously assessed using standard evaluation metrics:

• **Mean Absolute Error (MAE):** Measures the average absolute difference between predicted and actual values.

$$\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

• **Root Mean Squared Error (RMSE):** Penalizes larger errors, providing a more comprehensive view of prediction accuracy

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)}$$

• **R-squared (R^2):** Evaluates the goodness of fit by assessing the proportion of variance in the dependent variable captured by the model.

B. Exponential Smoothing
Exponential Smoothing, while demonstrating simplicity and ease of implementation, exhibits commendable performance in handling the complexity of healthcare supply chain budget data. The model's moderate accuracy, as indicated by an MAE of 502,384 and an R^2 of 0.76, suggests that it may be suitable for scenarios where intricate temporal dependencies and seasonality are less pronounced.

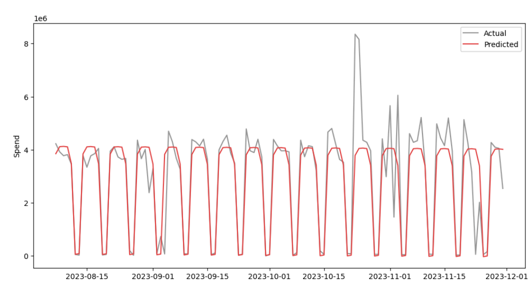


Fig. 2 A line graph to plot actual vs. predicted values using Triple Exponential Smoothing

C. ARIMA (AutoRegressive Integrated Moving Average)
ARIMA, with an MAE of 1,789,181 and an R^2 of -0.02, showcases limitations in capturing linear trends and autocorrelation. However, its struggle with non-linear patterns suggests that it might be less effective in situations where the dataset exhibits complex, non-linear relationships.

D. SARIMA (Seasonal ARIMA)
SARIMA, boasting an MAE of 487,405 and a robust R^2 of 0.77, emerges as a strong performer in capturing seasonality and temporal dependencies. This makes it a suitable choice for healthcare supply chain scenarios with pronounced seasonal patterns.

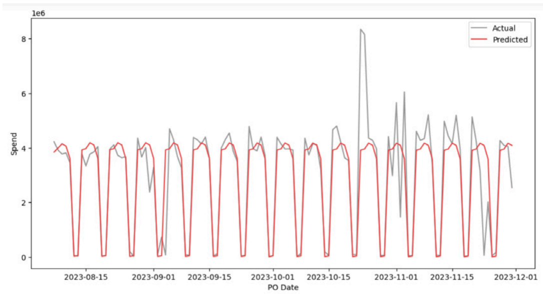


Fig. 3 A line graph to plot actual vs. predicted values using Seasonal ARIMA

E. XGBoost, RandomForest, and Linear Regression
XGBoost, RandomForest, and Linear Regression exhibit competitive performances with MAE values of 953,027, 829,462, and 1,369,936, respectively. These models, with varying strengths, can be considered for scenarios requiring robust predictive capabilities, with Linear Regression offering a balance between accuracy and interpretability.

F. K-Nearest Neighbors (KNN), LightGBM, LSTM, and GRU
KNN, LightGBM, LSTM, and GRU exhibit strengths in capturing complex relationships, as indicated by their competitive MAE values and R^2 scores. These models are well-suited for scenarios with intricate temporal dependencies and non-linear patterns.

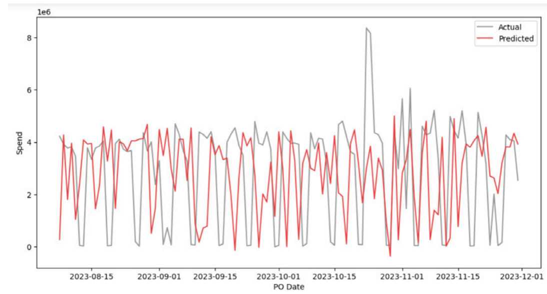


Fig. 4 A line graph to plot actual vs. predicted values using LightGBM

G. DNN (Deep Neural Network)
Deep Neural Network (DNN), with an MAE of 1,477,755 and an R^2 of 0.25, showcases average performance. Its application is recommended in scenarios where a balance between accuracy and model complexity is acceptable.

H. Comparative Analysis

Model Name	Performance Metrics		
	MAE (Mean Absolute Error)	RMSE (Root Mean Squared Error)	R2 (R-squared)
Exponential Smoothing	502,384	971,703	0.76
ARIMA (Auto Regressive Integrated Moving	1,789,181	2,015,945	-0.02
AR (Auto Regression) Average)	745,124	1,145,618	0.67
MA (Moving Average)	1,786,008	2,009,581	-0.01
SARIMA (Seasonal ARIMA)	487,405	964,126	0.77
XGBoost	953,027	1,370,468	0.46
Random Forest	829,462	1,222,360	0.57
Linear Regression	1,369,936	1,554,787	0.31
SVR (Support Vector Regression)	1,385,380	1,980,368	-0.12
K-Nearest Neighbors (KNN)	867,909	1,296,213	0.52
LightGBM (Gradient Boosting)	883,033	1,211,755	0.58
LSTM (Long Short-Term Memory) Network	823,887	1,191,137	0.59
DNN (Deep Neural Network)	1,477,755	1,726,750	0.25
GRU (Gated Recurrent Unit) Network	872,039	1,236,371	0.56

I. Discussion

The results highlight the nuanced performance of each model in forecasting healthcare supply chain budget spend. While SARIMA emerged as the top performer, it's crucial to consider the trade-offs between accuracy, complexity, and computational requirements when selecting the most suitable model for real-world implementation.

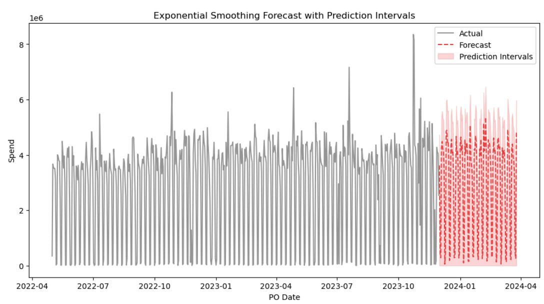


Fig. 5 Time series forecasting of Budget Spend with prediction intervals using Exponential Smoothing.

Practical Implications

The comparative analysis of time series forecasting models for healthcare supply chain budget spend yields valuable insights with direct practical implications for organizations in the sector.

A. Decision-Making and Resource Allocation

Accurate budget forecasts play a pivotal role in the decision-making processes of healthcare supply chain management. The superior performance of the Exponential Smoothing (ES) model and Seasonal ARIMA (SARIMA) suggests that organizations seeking precise predictions should consider leveraging these methodologies. The ability of ES and SARIMA to capture intricate temporal dependencies and seasonality positions them as powerful tools for strategic decision-making and resource allocation.
However, it's important to note that the choice of model should align with the organization's specific requirements and computational capabilities. While ES and SARIMA exhibit remarkable accuracy, their computational efficiency makes them more accessible for organizations with limited resources. In such cases, a balance between accuracy and efficiency should be struck.

B. Category, Supplier and Facility-Level Forecasting

Our analysis focuses on overall budget spend; however, the healthcare supply chain operates at multiple levels, including individual categories, suppliers and facilities. To enhance the granularity of budget forecasts, organizations should consider extending our model comparison to supplier-wise and facility-wise spend predictions.

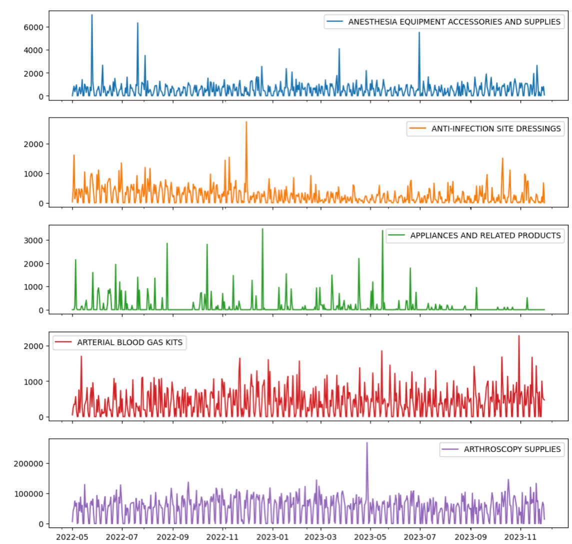


Fig. 6 A sample line graph to visualize trend by contract categories

The nuances of category-specific, supplier-specific and facility-specific forecasting may vary, and different models might exhibit varying levels of efficacy. Tailoring forecasting models to specific entities within the supply chain enables organizations to optimize resource allocation at a more granular level, potentially uncovering insights that can drive operational efficiency.

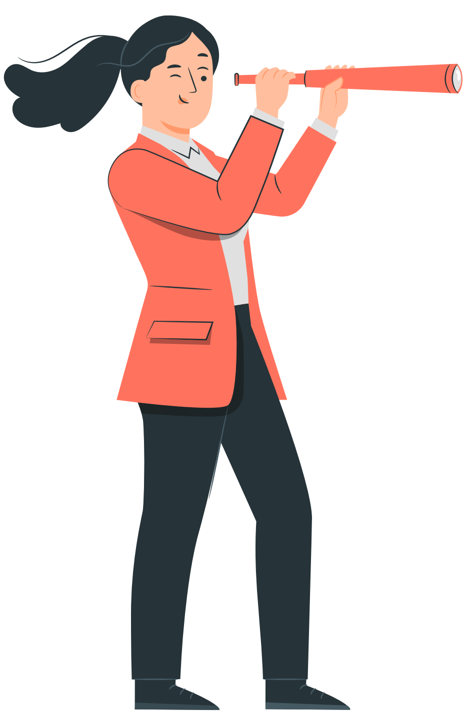
C. Continuous Model Evaluation & Updating

The dynamic nature of the healthcare supply chain necessitates continuous model evaluation and updating. Monthly updates to the forecasting model, as demonstrated in our methodology, ensure that the model remains adaptive to evolving patterns and dependencies within the budget data. Regular assessments of model performance against updated datasets enhance the model's accuracy and relevance over time.

D. Considerations for Model Implementation
When implementing forecasting models within the healthcare supply chain, organizations should consider the following practical aspects:

- **Data Quality and Consistency:** Ensure that the input data remains consistent and of high quality over time. Address any changes in data distribution or characteristics promptly.
- **Interpretability:** While complex models like LSTM and XGBoost may offer superior accuracy, the interpretability of the model outputs is essential for gaining insights into the factors influencing budget forecasts. Consider using models that strike a balance between accuracy and interpretability.
- **Computational Resources:** Assess the computational requirements of chosen models, considering the organization's available resources. Opt for models that align with the computational infrastructure in place.

E. Model Robustness and Adaptability
In addition to accuracy, the robustness and adaptability of forecasting models are crucial considerations. The healthcare supply chain is subject to external factors, such as changes in regulations, public health crises, and market dynamics. Models that demonstrate resilience to unforeseen events and adaptability to changing conditions contribute to a more reliable and responsive forecasting system.



Next Steps & Recommendations

A. Supplier-Wise Spend Forecast
A logical extension of our comparative analysis involves developing models tailored specifically for supplier-wise spend forecasting within the healthcare supply chain. Given the diverse nature of suppliers, each with unique characteristics and demand patterns, specialized models can offer more precise predictions. Organizations are encouraged to embark on the development of supplier-specific forecasting models, leveraging insights gained from our overall budget spend analysis.

B. Facility-Wise Spend Forecast
Similar to supplier-wise forecasting, the healthcare supply chain operates across various facilities, each with distinct requirements and consumption patterns. To enhance resource allocation and operational efficiency, the development of facility-wise spend forecasting models is recommended. Exploring the variations in budget spend at the facility level can unearth insights that may not be apparent in aggregate analyses.

C. Ensemble Models and Hybrid Approaches
Our comparative analysis highlighted the potential of ensemble methods in combining the strengths of multiple forecasting models. Further research into advanced ensemble models and hybrid approaches, combining both traditional and machine learning techniques, can enhance forecasting accuracy and resilience. Investigating the synergy between models with diverse strengths can result in more robust forecasting systems.

D. External Factors Integration
To further refine forecasting accuracy, future research should explore the integration of external factors that may impact healthcare supply chain budget spend. Incorporating variables such as regulatory changes, economic indicators, and public health trends can enhance the models' adaptability to real-world dynamics. Analyzing the interplay between internal and external factors contributes to a more holistic and responsive forecasting system.

E. Continuous Model Monitoring and Improvement
Continuous monitoring of forecasting model performance is paramount for maintaining

accuracy and relevance. Establishing a robust monitoring system that flags deviations and prompts model re-evaluation ensures that the forecasting system remains adaptive. Regularly updating models with new data, as demonstrated in our methodology, contributes to sustained accuracy over time.

F. Collaboration and Knowledge Sharing
The healthcare supply chain is a dynamic ecosystem with shared challenges and opportunities. Encouraging collaboration and knowledge sharing among organizations can facilitate the development of standardized forecasting methodologies and best practices. Industry-wide collaboration fosters collective learning and the evolution of effective forecasting strategies.

G. Ethical Considerations and Data Privacy
As organizations delve deeper into advanced forecasting methodologies, it is crucial to maintain a commitment to ethical considerations and data privacy. Striking a balance between leveraging detailed data for accurate predictions and safeguarding sensitive information ensures responsible and sustainable forecasting practices.

Conclusions

In this comprehensive study, we undertook a rigorous comparative analysis of various time series forecasting models to enhance the precision of budget spend predictions within the healthcare supply chain. The findings provide valuable insights into the strengths and limitations of each model, offering a roadmap for organizations aiming to optimize their forecasting strategies.

A. Key Findings
Our analysis revealed that the choice of forecasting model significantly influences the accuracy of budget spend predictions. While traditional models like Exponential Smoothing and SARIMA offer reliable performance, the advent of deep learning, exemplified by LSTM, introduces a new frontier of forecasting capabilities. The intricate temporal dependencies and non-linear patterns inherent in healthcare supply chain data make Exponential Smoothing and SARIMA potent tools for strategic decision-making.

Moreover, the versatility of machine learning approaches, exemplified by Random Forest and Gated Recurrent Unit (GRU), showcases promising results, demonstrating the importance of exploring diverse methodologies tailored to the unique characteristics of healthcare budget data.

B. Practical Implications
The practical implications of our findings extend beyond model selection. Organizations in the healthcare supply chain should consider the granularity of forecasting, leveraging the reliability of Exponential Smoothing and SARIMA for more targeted resource allocation at the supplier and facility levels. Continuous model evaluation, updating, and monitoring are essential for adapting to evolving patterns and external factors.

C. Recommendations for Future Research
As the healthcare supply chain continues to evolve, our study suggests several avenues for future research. The development of supplier-specific and facility-specific forecasting models, exploration of advanced ensemble methods, and the integration of external factors into forecasting frameworks represent exciting areas for innovation.

D. Ethical Considerations
As organizations embrace advanced forecasting methodologies, ethical considerations and data privacy must remain at the forefront. Striking a balance between leveraging detailed data for accurate predictions and safeguarding sensitive information ensures responsible and sustainable forecasting practices.

E. Moving Forward
Our comparative analysis lays the groundwork for a new era in healthcare supply chain budget forecasting. The collaboration of industry stakeholders, the implementation of advanced forecasting methodologies, and a commitment to ethical considerations collectively contribute to the resilience and adaptability of the healthcare supply chain.

In conclusion, as organizations embark on the journey of enhancing their forecasting strategies, the lessons learned from this study can serve as a compass, guiding them toward more informed decision-making, optimized resource allocation, and ultimately, a more resilient and efficient healthcare supply chain.

Acknowledgment

I extend my sincere gratitude to Advantus Health Partners (GPO) for their invaluable contribution to this research. Their provision of data and insightful guidance significantly enriched the depth and quality of our study, allowing me to derive meaningful insights into healthcare supply chain budget forecasting.

This research would not have been possible without the collaborative efforts of both Advantus Health Partners (GPO) and Supplycopia. I acknowledge and appreciate their commitment to innovation, data-driven insights, and the advancement of healthcare forecasting methodologies.

Special thanks to Mr. Ashok Muttin (CEO of SupplyCopia) for his expertise and assistance throughout all aspects of my study and for his help in writing the manuscript.

References

[1] Langenberger B, Schulte T, Groene O. The application of machine learning to predict high-cost patients: A performance-comparison of different models using healthcare claims data. PLoS One. 2023 Jan 18;18(1):e0279540. doi: 10.1371/journal.pone.0279540. PMID: 36652450; PMCID: PMC9847900.

[2] M. Nagarjuna, P. Pasula, T. Kavyakeerthi and I. Karthik, "Medicine Expenditure Prediction using Machine Learning," 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2022, pp. 948-952, doi: 10.1109/ICCMC53470.2022.9753780.

[3] Osawa I, Goto T, Yamamoto Y, Tsugawa Y. Machine-learning-based prediction models for high-need high-cost patients using nationwide clinical and claims data. NPJ Digit Med. 2020 Nov 11;3(1):148. doi: 10.1038/s41746-020-00354-8. PMID: 33299137; PMCID: PMC7658979. .

[4] Rosella LC, Kornas K, Yao Z, Manuel DG, Bornbaum C, Fransoo R, Stukel T. Predicting High Health Care Resource Utilization in a Single-payer Public Health Care System: Development and Validation of the High Resource User Population Risk Tool. Med Care. 2018 Oct;56(10):e61-e69. doi: 10.1097/MLR.0000000000000837. PMID: 29189576; PMCID: PMC6143224.

[5] G. Emmanuel, Meyliana, A. N. Hidayanto and H. Prabowo, "A Literature Review: Drug Supply Chain and Neural Networks," 2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia, 2022, pp. 559-564, doi: 10.1109/ICITISEE57756.2022.10057939.

SUPPLYCOPIA

SupplyCopia was created to address the critical lack of supply chain intelligence faced by healthcare organizations. This is especially problematic because it can adversely affect quality, costs, and patient outcomes, and the development of more effective relationships among providers and suppliers. SupplyCopia applies the latest data science and software technology to bring maximum transparency to both major constituent groups of the supply chain - to the benefit of both and expense of neither.