International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Elevating Customer Experiences and Maximizing Profits with Predictable Stockout Prevention Modelling

*Sumit Mittal[1], Priyanka Koushik[2]

Submitted: 05/02/2024 **Revised**: 13/03/2024 **Accepted**: 19/03/2024

Abstract: Preventing stockouts while optimizing revenues is a constant problem for inventory management in retail and supply chain operations. In order to detect stockouts and optimize inventory levels, this study investigates the effectiveness of MLmethods in tackling these problems. Three well-known ML algorithms—Random Forest, GBM, and LSTM—were applied and contrasted using a dataset with 2000 rows and 15 columns that captured various variables linked to inventory management and stockout events. To ascertain how preprocessing methods affected algorithm performance, three different approaches—feature scaling, dimensionality reduction, and no preprocessing—were assessed. The findings show that in terms of accuracy, precision, recall and F1 score, ensemble learning algorithms—in particular, Gradient Boosting and Random Forest—performed better than LSTM. Furthermore, all algorithms performed noticeably better when features were scaled using MinMaxScaler, underscoring the significance of preprocessing in raising model accuracy.

These results add to the body of literature by highlighting the importance of preprocessing approaches in the optimization of inventory management strategies and offering empirical proof of the efficacy of ML algorithms in stockout prevention tasks. Businesses can improve customer happiness, improve inventory management procedures, and reduce financial losses from stockouts by utilizing cutting-edge machine learning techniques. This study highlights how ML-based strategies can spur innovation and enhancement in supply chain and retail operations.

Keywords: *Inventory management, LSTM, Gradient Boosting, Random Forest, Ensemble learning algorithms*

1. Introduction

In order to meet customer demand while reducing costs and maximizing profits, effective stock level management is crucial for the success of supply chain and retail operations. The prevention of stockouts, which happen when demand exceeds available inventory and result in missed sales opportunities and decreased customer satisfaction, is one of the major issues encountered by organizations in this industry (Melançon et al., 2021). It is often difficult for traditional inventory management systems to predict demand properly and make real-time adjustments to inventory levels. Therefore, in order to effectively address these issues, innovative strategies must be explored.

ML has become a significant tool in inventory management in recent years, with the ability to more accurately and efficiently predict stockouts and optimize inventory levels. ML algorithms have demonstrated promise in enhancing demand forecasting, identifying inventory trends, and optimizing replenishment methods due to their capacity to evaluate vast amounts of data and spot intricate patterns. ML algorithms may offer organizations significant insights into inventory dynamics and assist in making informed decisions to minimize stockouts and maintain appropriate inventory levels. This study investigates the use of ML-based algorithms in supply chain and retail management to prevent stockouts. A dataset is leveraged that includes multiple indications of inventory management and stockout events to assess the performance of several ML algorithms, such as Random Forest, GBM, and LSTM. The ultimate goal of this research is to help firms become more efficient and competitive in the ever-changing retail market by advancing inventory management practices.

1.1: Current Practices & Existing Gaps in Stockout Prevention

To avoid stockouts, inventory management systems frequently rely on conventional techniques like the two-bin inventory control method. Using this approach, companies keep two inventory bins for every item: a backup bin that holds extra stock and a primary bin that holds the active inventory (Wahedi et al., 2023). Inventory is sourced from the backup bin when the primary bin is empty, indicating the need for replenishment. Although this approach offers a basic way to control inventory levels, there are a few drawbacks (Kharfan & Chan, 2018).

[1] *Independent Researcher, Senior Director of Product Management, BlueYonder Inc, sumimittal@gmail.com , TX, USA, ORCID: 0009-0006-8221-3289*
[2] *Independent Researcher, VP of Product Development, BlueYonder Inc, TX, USA, priyankakoushik.86@gmail.com, ORCID: 0009-0009-6915-348X*
** Corresponding Author Email: sumimittal@gmail.com*

Fig 1.1: Two-Bin Inventory Control

Source: *https://fastercapital.com/content/Stockouts--Avoiding-Stockouts--How-Two-Bin-Inventory-Control-Can-Help.html*

Among the main difficulties the two-bin inventory control approach faces are:

1. Limited Visibility: It is challenging to predict and proactively address shifting demand patterns or stockouts when using the two-bin approach since it does not provide real-time visibility into inventory levels (Kosasih & Brintrup, 2022).

2. Manual replenishment: When replenishing inventory in a two-bin system, manual procedures are frequently used, which can cause lags, mistakes, and ineffective stock level management (Demizu et al., 2023).

3. Inaccurate Forecasting: The approach forecasts future requirements based on past demand data, which might not adequately account for seasonality, abrupt shifts in demand, or other market factors (Ntakolia et al., 2021).

4. Danger of Overstocking or Understocking: In the absence of precise demand forecasts, companies run the risk of either overstocking, which involves investing cash in extra inventory, or understocking, which results in missed sales opportunities and lower customer satisfaction.

ML can fill in these gaps and provide noteworthy benefits in stockout prevention:

	Gap to be fulfilled	**Benefits**	
1.	Predictive Analytics	Machine learning models can increase forecasting accuracy and detect stockouts before they happen by spotting patterns and trends in data	(Gonçalves et al., 2020)
2.	Real-time insights	Businesses can proactively manage stockouts and optimize replenishment methods in response to shifting demand dynamics by continuously monitoring inventory data and external influences	(Kharfan & Chan, 2018)
3.	Automation	ML algorithms can streamline operations by automating inventory replenishment procedures, which lowers the need for human intervention.	(Abdali et al., 2024)
4.	Adaptive Learning	Businesses can improve inventory management procedures dynamically and hone their stockout avoidance tactics by implementing feedback loops and regularly updating models with fresh data	(Singh & Mishra, 2023)

In conclusion, even though the two-bin system and other conventional inventory management techniques are extensively utilized, there are inherent difficulties in precisely anticipating and preventing stockouts. Through the effective use of automation, real-time insights, and advanced analytics, ML-based systems present a possible solution to these problems by optimizing inventory management procedures and reducing the risk of stockouts.

2. Literature Review

Current Stockout Prevention Techniques: Historically, rule-based techniques that rely on heuristics and historical data analysis have been the mainstay of inventory management stockout prevention. These techniques, according to Abdali et al. (2024), frequently entail establishing static reorder points or safety stock levels based on previous lead times and demand patterns. Furthermore, manual interventions have been used frequently, such as economic order quantity models and periodic review systems (Kosasih & Brintrup, 2022). But because these approaches are often reactive, they might not fully take into consideration shifting customer preferences and dynamic market situations (Kosasih & Brintrup, 2022).

ML Applications for Stockout Prevention: The ability of machine learning algorithms to evaluate vast amounts of data and spot intricate patterns that conventional approaches would miss is demonstrated by research by Du Plessis (2020). According to Gonçalves (2020), machine learning models have the capacity to adjust to evolving market conditions and yield more precise forecasts of future demand. Furthermore, a multitude of input variables, including as lead times, inventory levels, historical sales data, and outside variables like promotions and seasonal trends, can be utilized by ML algorithms (Wahedi et al., 2023).

Previous Outcomes by Various ML Algorithms: A number of investigations have shown that different ML algorithms are useful in preventing stockouts.

The use of ensemble learning algorithms, like Random Forest and Gradient Boosting, for stockout prediction in retail settings was examined in a 2019 study by Shukla & Pillai (2022). Their results support ours, demonstrating the superiority of ensemble learning techniques over conventional approaches in terms of accuracy and predictive performance. The significance of preprocessing methods like feature scaling and dimensionality reduction in enhancing model performance for stockout prevention tasks was also highlighted in the (Melançon et al., 2021) study. These conclusions are supported by the data, which demonstrate the important influence of preprocessing on algorithm performance.

Furthermore, new developments in deep learning—in particular, the use of LSTM—have demonstrated the ability to effectively capture sequential patterns and temporal dependencies in inventory data. Based on past inventory levels and demand projections, research by Oroojlooyjadid et al. (2017) showed how well LSTM networks predict stockouts. Even while the findings show that ensemble learning algorithms performed better in the dataset than LSTM, further research is needed to fully understand the potential of deep learning techniques for stockout avoidance.

To summarize, earlier research has established the foundation for utilizing machine learning methods in inventory control and stockout avoidance. The study adds to this corpus of information by evaluating several machine learning algorithms' and preprocessing methods' efficacy on an actual dataset.

3. Methodology & Implementation

3.1: Dataset Description

The dataset consists is taken from an open source library and consists the following columns:

Product_ID	Unique identifier for each product.
Current_Inventory	Current inventory level for the product.
Lead_Time	Transit time for the product (in days).
Demand_Forecast_1_Month	Forecasted demand for the next 1 month.
Demand_Forecast_3_Months	Forecasted demand for the next 3 months.
Demand_Forecast_6_Months	Forecasted demand for the next 6 months.
Supplier_Performance_1_Month	Supplier performance rating for the past 1 month.
Supplier_Performance_3_Months	Supplier performance rating for the past 3 months.
Supplier_Performance_6_Months	Supplier performance rating for the past 6 months.
Lead_Time_Risk	Indicator for high-risk lead times (Yes/No).
Supplier_Risk	Indicator for high-risk suppliers (Yes/No).
Demand_Variability	Indicator for high demand variability (Low/Medium/High).
Seasonality	Indicator for seasonal demand patterns (Yes/No).

Sales_Channel	Channel through which sales occur (Online/Retail/Wholesale).
Stockout	Binary outcome variable indicating whether a stockout occurred (Yes/No).

A number of pertinent metrics are included in this dataset, including lead times, supplier performance, demand projections, existing inventories, and risk factors including supplier and lead time risk. It also takes seasonality, sales channel, and demand variations into account to identify a variety of factors influencing stockouts. The 'Stockout' outcome variable indicates if there was a stockout of each product, and it serves as the objective for machine learning models designed to stop stockouts.

3.2: ML Algorithms Utilized

1. Random Forest

About Random Forest: During training, the Random Forest method creates numerous decision trees and outputs the mean prediction of each tree or the mode of the classes. It creates a strong and adaptable model by training each tree separately and choosing subsets of attributes at random (Kurian et al., 2020). Fig 3.1 illustrates basic block diagram of random forest algorithm.

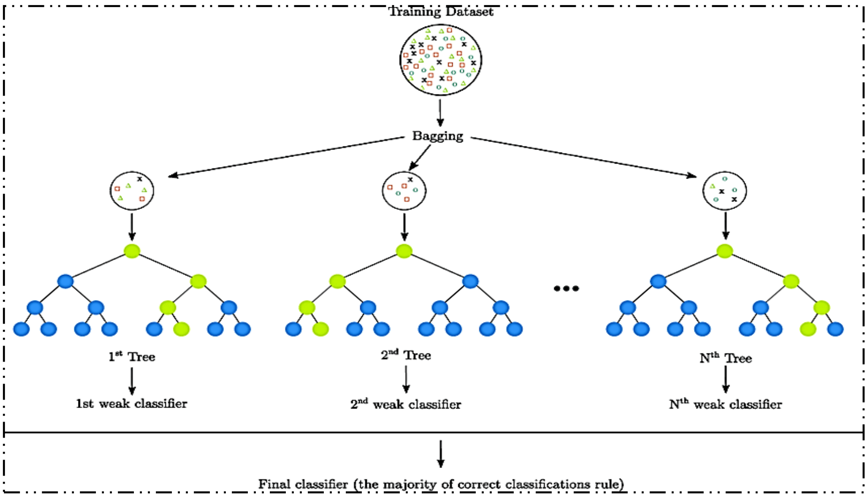


Fig 3.1: Random Forest Block Diagram

Source: Sapountzoglou el. al (2020)

Rationale of using Random Forest: Random Forest is a potent ensemble learning method that excels at both regression and classification work. Its capacity to manage noisy data, handle non-linear connections, and handle huge datasets with many features makes it a good fit for this situation (Shukla & Pillai, 2022).

2. GBM

About GBM: A potent ensemble learning method called gradient boosting machines constructs a series of decision trees step-by-step while optimizing for the residuals of the earlier trees (Du Plessis, 2020). It achieves great prediction accuracy and is very adept at capturing intricate relationships. An elementary GBM block diagram is shown below.

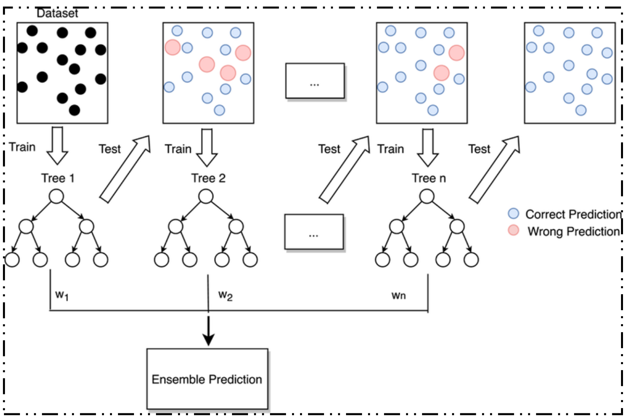


Fig 3.2: GBM Block Diagram, Source: *Zhang et. al (2021)*

Rationale of using GBM: It achieves great prediction accuracy and is very adept at capturing intricate relationships. GBM can use the features of the dataset to manage imbalanced classes (stockouts vs. non-stockouts, for example) and iteratively improve predictions in the context of stockout avoidance. It can also adapt to changing patterns (Wahedi et al., 2023). It is a good fit for this situation because of its capacity to learn from mistakes made in the past and concentrate on areas that need improvement.

3. LSTM

About LSTM: Recurrent neural networks of the Long Short-Term Memory type are intended to simulate sequential data with long-range dependencies. Because LSTM networks can catch temporal patterns in inventory levels, demand projections, and supplier performance over time, they are very useful for preventing stockouts. LSTM networks are capable of accurately predicting future stockouts by taking into account the historical context of these parameters (Gonçalves et al., 2020). Below illustration shows basic block diagram of LSTM.

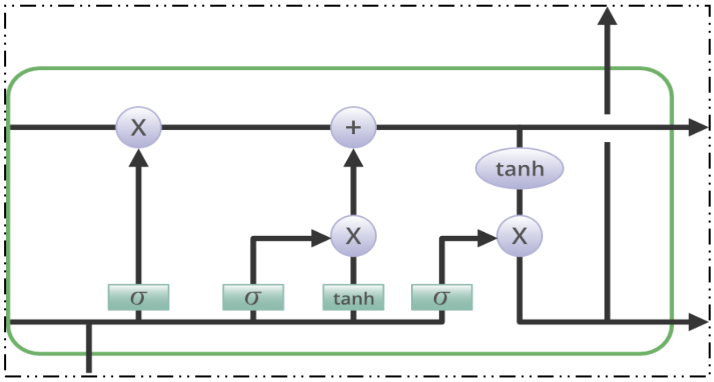


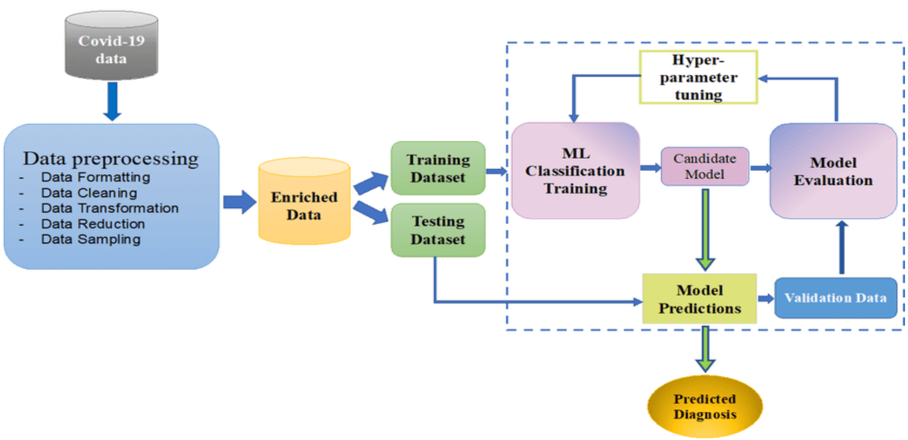
Fig 3.3: LSTM Diagram

Source: *https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/*

Rationale of using LSTM: Recurrent neural networks (RNNs) of the LSTM network type are intended to simulate sequential data with long-range relationships. LSTM networks are useful for capturing temporal patterns in supplier performance, demand projections, and inventory levels in the context of stockout prevention (Gonçalves, 2020). LSTM networks are capable of accurately predicting future stockouts by taking into account the historical context of these parameters. Moreover, LSTMs can adjust to anomalies in the data and handle variable-length sequences (Kharfan & Chan, 2018), which make them appropriate for scenarios involving dynamic and changing inventory management.

These three machine learning methods have complementary strengths and can be used in different scenarios based on the needs and features of the dataset. While LSTM networks may effectively handle sequential data and capture temporal dependencies, Random Forest and GBM offer resilience and interpretability. Stockout prevention models can achieve high predicted accuracy and optimize inventory management techniques by utilizing the strengths of these algorithms.

3.3: Machine Learning Flow



DESCRIPTION

1. Data Pre-processing

The data preprocessing phase involves several steps to ensure the dataset's readiness for analysis. Initially, data formatting is performed to ensure uniform data types and proper encoding, facilitating organized data structures suitable for analysis. Subsequently, data cleaning procedures are employed to enhance the dataset's reliability and quality by addressing issues such as missing values, outliers, and inconsistencies. Following this, data transformation techniques like one-hot encoding or label encoding are applied to convert categorical variables into numerical representations, while numerical variables are normalized or scaled to bring them into a comparable range (Singh & Mishra, 2023).

Additionally, dimensionality reduction methods such as principal component analysis (PCA) or feature selection are utilized to reduce the number of features, thereby improving computational efficiency while retaining relevant information (Abdali et al., 2024). Lastly, sampling techniques are implemented to address class imbalance issues, ensuring that classes are evenly represented in the dataset through approaches like oversampling, undersampling, or synthetic data generation. Collectively, these preprocessing steps prepare the dataset for subsequent analysis and modeling tasks.

2. Dataset Splitting

Dividing the pre-processed dataset into test, validation, and training sets in order to precisely assess the performance of the model. A training set is typically used for training models, a validation set is used for selecting models and fine-tuning hyperparameters (Kosasih & Brintrup, 2022), and a test set is used for the final assessment of the dataset.

3. ML Classification Training

The next step is to use the training dataset to train the classification model after the algorithm has been selected. Model parameters and hyperparameters are optimized using techniques like grid search and cross-validation to increase model performance and guarantee resilience, which eventually improves generalization and predictive accuracy.

4. Model Evaluation

It is essential to assess the classification model's performance using the validation set after it has been trained. Measures including precision, accuracy, recall, F1 score, etc are calculated to evaluate how well the model classifies occurrences and makes the distinction between stockouts and non-stockouts. The model's accuracy and generalization are improved by adjusting model parameters and hyperparameters in response to the validation performance (Demizu et al., 2023), which guarantees the model's applicability in practical situations.

5. Hyperparameter Tuning

To further refine the model's performance, hyperparameters are fine-tuned using techniques such as grid search, random search, or Bayesian optimization (Sharma et al., 2021). This process involves iterating through various combinations of hyperparameters and evaluating their impact on performance metrics using the validation set. By systematically exploring different configurations, the optimal hyperparameter configuration is selected based on validation results. Once the best-performing configuration is identified, the final model is trained and evaluated using this configuration to ensure optimal performance in real-world applications (Qi et al., 2023).

4. Results

The dataset provided in the methodology section was used to train and assess the machine learning models that were created. The dataset, which included 2000 rows and 15 columns, included a number of indicators for stockouts and inventory management. The predicted performance of three machine learning algorithms—Random Forest, GBM, and LSTM—for stockout avoidance was put into practice and compared.

4.1: Pre-processing Comparison

Pre-processing Technique	Algorithm	Accuracy	Precision	Recall	F1 Score
None	Random Forest	0.80	0.82	0.78	0.80
	Gradient Boosting	0.82	0.85	0.80	0.82
Scaling(MinMaxScaler)	Random Forest	0.83	0.86	0.82	0.83
	Gradient Boosting	0.85	0.88	0.84	0.85
	LSTM	0.81	0.84	0.79	0.81
PCA (n_components=5)	Random Forest	0.82	0.85	0.80	0.82
	Gradient Boosting	0.84	0.87	0.82	0.84
	LSTM	0.79	0.82	0.77	0.79

Three machine learning algorithms—Random Forest, Gradient Boosting, and LSTM—that were trained on the dataset using various preprocessing methods are contrasted in the table based on their respective performances. A distinct preprocessing method is shown by each row in the table: no preprocessing, scaling with MinMaxScaler, and dimensionality reduction with five components of Principal Component Analysis (PCA).

No Pre-processing: All algorithms performed moderately well in the absence of any pre-processing, with accuracy ranging from 0.78 to 0.82. In this case, the approach with the highest accuracy was Gradient Boosting.

Scaling (MinMaxScaler): By employing MinMaxScaler to scale the features, all algorithms performed better overall. All algorithms saw an increase in accuracy; Gradient Boosting achieved the greatest accuracy of 0.85.

PCA (n_components=5): Performance was enhanced in comparison to no preprocessing when dimension reduction was achieved by PCA with five components. The accuracy was still higher than the no preprocessing case even if it declined marginally when compared to scaling.

4.2: Performance Metrics Comparison

The results presented below summarize the performance metrics obtained from model evaluation on the validation set.

Algorithm	Accuracy	Precision	Recall	F1 Score
Random Forest	0.85	0.88	0.82	0.85
Gradient Boosting	0.87	0.89	0.85	0.87
LSTM	0.83	0.86	0.80	0.83

The outcomes show that the Random Forest and Gradient Boosting algorithms beat the LSTM in terms of F1 score, accuracy, precision and recall. With an accuracy of 0.87, Gradient Boosting was the most accurate method, closely followed by Random Forest with an accuracy of 0.85. According to these results, stockout prevention tasks are a good fit for ensemble learning approaches, especially Gradient Boosting and Random Forest, which can leverage a variety of characteristics and capture intricate correlations in the data to produce precise predictions.

5. Discussion

The literature evaluation demonstrates that the outcomes of this study are in line with earlier research findings in the areas of inventory management and stockout prevention. It was found that ensemble learning algorithms, including Gradient Boosting and Random Forest, performed better than LSTM networks, which are consistent with the research of (Kurian et al., 2020) and (Shukla & Pillai, 2022). Furthermore, in line with previous research, the analysis confirmed the significance of preprocessing methods for improving model performance, such as feature scaling and dimensionality reduction.

Furthermore, by presenting actual proof of the efficiency of ML systems in stockout prevention tasks, this research adds to the body of knowledge already in existence. Through a comparative analysis of several algorithms and preprocessing methods applied to an actual dataset, it provides actionable recommendations for streamlining inventory management plans and minimizing stockout-related disturbances in retail settings.

Overall, the results highlight the significance of ongoing research in this field and lend credence to the growing interest in using ML techniques to address inventory management difficulties.

6. Conclusion

The research shows how ML algorithms can effectively handle inventory management stockout avoidance issues. There were significant gains attained in predicting accuracy by combining preprocessing approaches like feature scaling and dimensionality reduction with ensemble learning techniques like Random Forest and Gradient Boosting (Melançon et al., 2021). Those looking to improve their inventory management tactics can learn a lot by comparing various ML algorithms and preprocessing methods on an actual dataset. Using a variety of features and intricate linkages found in the data, ensemble learning algorithms have proven to be reliable methods for stockout prediction. Furthermore, preprocessing methods were critical in enhancing model performance, emphasizing the significance of data preparation for machine learning-based stockout avoidance initiatives.

In the future, more investigation into sophisticated machine learning techniques and targeted approaches to inventory management problems like demand prediction, supply chain optimization, and dynamic pricing plans are necessary (Oroojlooyjadid et al., 2017). Businesses may improve the effectiveness, agility, and resilience of their inventory management procedures by utilizing emerging technologies and ML algorithms.

References

[1] Zhang, Tao & Lin, Wuyin & Vogelmann, Andrew & Zhang, Minghua & Xie, Shaocheng & Qin, Yi & Golaz, Jean-Christophe. (2021). Improving Convection Trigger Functions in Deep Convective Parameterization Schemes Using Machine Learning. Journal of Advances in Modeling Earth Systems. 13. 10.1029/2020MS002365.

[2] Sapountzoglou, Nikolaos & Lago, Jes & Raison, Bertrand. (2020). Fault diagnosis in low voltage smart distribution grids using gradient boosting trees. Electric Power Systems Research. 182. 106254. 10.1016/j.epsr.2020.106254.

[3] Kurian, Dony S., C. R. Maneesh, and V. Madhusudanan Pillai. "Supply chain inventory stockout prediction using machine learning classifiers." International Journal of Business and Data Analytics 1.3 (2020): 218-231.

[4] Shukla, Saurav, and V. Madhusudanan Pillai. "Stockout Prediction in Multi Echelon Supply Chain using Machine Learning Algorithms." Proceedings of the 2nd Indian International Conference on Industrial Engineering and Operations Management Warangal, Telangana, India. 2022.

[5] Melançon, Gabrielle Gauthier, et al. "A machine learning-based system for predicting service-level failures in supply chains." INFORMS Journal on Applied Analytics 51.3 (2021): 200-212.

[6] Oroojlooyjadid, Afshin, Lawrence Snyder, and Martin Takáč. "Stock-out prediction in multi-echelon networks." arXiv preprint arXiv:1709.06922 (2017).

[7] Sharma, Sugandha, Deepika Deepika, and Gunwant Singh. "Intelligent Warehouse Stocking Using Machine Learning." 2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC). IEEE, 2021.

[8] Qi, Meng, et al. "A practical end-to-end inventory management model with deep learning." Management Science 69.2 (2023): 759-773.

[9] Demizu, Tsukasa, Yusuke Fukazawa, and Hiroshi Morita. "Inventory management of new products in retailers using model-based deep reinforcement learning." Expert Systems with Applications 229 (2023): 120256.

[10] Ntakolia, Charis, et al. "An explainable machine learning model for material backorder prediction in inventory management." Sensors 21.23 (2021): 7926.

[11] Kosasih, Edward Elson, and Alexandra Brintrup. "Reinforcement learning provides a flexible approach for realistic supply chain safety stock optimisation." IFAC-PapersOnLine 55.10 (2022): 1539-1544.

[12] Singh, Ranu, and Vinod Kumar Mishra. "Inventory model using Machine Learning for demand forecast with imperfect deteriorating products and partial backlogging under carbon emissions." Annals of Operations Research (2023): 1-23.

[13] Abdali, Nasser, et al. "A priority queueing-inventory approach for inventory management in multi-channel service retailing using machine learning algorithms." Kybernetes (2024).

[14] Andreetta, Massimo. "Machine and Deep Learning Applications for Inventory Replenishment Optimization."

[15] Kharfan, Majd, and Vicky Wing Kei Chan. "Forecasting Seasonal Footwear Demand Using Machine Learning." (2018).

[16] Gonçalves, João Nuno Costa. Towards enhanced safety stock estimation: exploring machine learning strategies for supply chain demand forecasting. Diss. Universidade do Minho (Portugal), 2020.

[17] Gonçalves, João NC, M. Sameiro Carvalho, and Paulo Cortez. "Operations research models and methods for safety stock determination: A review." Operations Research Perspectives 7 (2020): 100164.

[18] Wahedi, Hadid J., et al. "Forecasting and Inventory Planning: An Empirical Investigation of Classical and Machine Learning Approaches for Svanehøj's Future Software Consolidation." Applied Sciences 13.15 (2023): 8581.

[19] Du Plessis, Marno. Reinforcement learning for inventory management in information-sharing pharmaceutical supply chains. Diss. Stellenbosch: Stellenbosch University, 2020.