Journal of Information Systems Engineering and Management
2025, 10(16s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

E-Commerce Inventory Prediction by Hybridization Deep and Machine Learning

Megha Thakar [1], Dr. Priya Swaminarayan [2]

[1] *Research Scholar, Parul University, Vadodara; 1204meghathakar@gmail.com*

[2] *Dean & Professor, Faculty of IT and Computer Science, Parul University, Vadodara; priya.swaminarayan@paruluniversity.ac.in*

ARTICLE INFO

Received: 05 Dec 2024

Revised: 26 Jan 2025

Accepted: 10 Feb 2025

ABSTRACT

Inventory management is crucial for the optimisation of consumer demand and supply chains in e-commerce companies. This is the stage at which precise inventory forecasting becomes necessary for the primary objective of forecasting future demand patterns and stock levels. Traditional forecasting methods often struggle with e-commerce data due to seasonality, sudden changes in customer behaviour, and non-linearity. Machine learning (ML) and deep learning (DL) techniques have become powerful weapons for inventory prediction due to their capability of analysing huge amounts of data with high dimensionality. In highly competitive market environments, e-commerce firms can improve their resource allocation, inventory management, and customer experience. This paper proposes different types of inventories forecasting models and especially evaluates the applicability of sophisticated machine learning algorithms. While we commonly use old methods like Random Forest, ARIMA, and MLPs, they often lack the necessary robustness to non-linearity within inventory data. To address these problems, we introduce a novel method that combines convolutional neural networks (CNN) and XGBoost called CNN-XGBoost, which provides better feature extraction than the conventional prediction model and regression performance. We then compared CNN-XGBoost's performance to traditional forecasting methods (another common approach to contextualizing predictive model performance) using a 52-week simulated dataset in which we mimic patient data growing over time.We used key performance metrics such as R2, mean squared error (MSE), and mean absolute percentage error (MAPE) to assess each model's accuracy. The CNN-XGBoost model performed much better than others with an R2 of 0.78, which means our proposed model can explain more variation in comparison to other competitors, as depicted in the results section. It also had the best MSE of 0.15, indicating better predictive performance. The CNN-XGBoost model demonstrated promising prospects as a powerful inventory forecasting tool for commerce, in spite of its slightly higher MAPE value (0.69), suggesting some vulnerability to outlier data points. This study demonstrates the potential of using a convolutional neural network in combination with gradient boosting techniques to tackle the complexity of stock management issues, as well as the fact that it outperforms based line methods by a large margin.

Keywords: E-commerce prediction, Machine Learning, Deep Learning.

1. Introduction

E-commerce inventory forecasting is the task of predicting the future sales of a product an e-commerce retailer has using statistical methods and machine learning algorithms. This involves analysing past sales as well as considering other factors such as promotions, events, and market trends to determine the correct stock level for every individual item so that one can satisfy demand projections. E-commerce inventory forecasting seeks to maintain an optimal level of in-stock inventory that can fulfil the demand without overstocking, which may lock your working capital and increase storage costs [1–4]. Conversely, inefficient inventory management can result in significant setbacks for a company, ranging from lower sales and unhappy customers to massive disappointment. Properly forecasting future demand provides e-commerce firms with more information, enabling them to make more efficient decisions in procurement, production, and inventory management. This efficiency leads to lower costs, allowing them to optimize operational efficiencies and retain a happier customer base. It mainly does all of this using state-of-the-art methods like time series analysis, demand forecasting, model algorithms, machine learning, and so on, which enhances the accuracy of its predictions. By utilizing real-time data streams and modifying estimates based on dynamic market conditions, we can significantly enhance the accuracy of inventory forecasting in e-commerce. In the realm of e-commerce, our use case is forecasting based on historical sales and other cues (by means of machine/deep learning) to trade-off current inventory management for near-exact future demand expectations [5–11]. In the end, companies are able to reduce costs on both sides of the equation by eliminating over-inventory and stockouts.

Predicting Inventories: Delivering each product at the right time makes customers happiest. It also reduces costs by preventing overstocking or understocking, which in turn reduces the need for rapid shipping. One can use predictive models to pinpoint customer behaviour patterns, enhance pricing tactics, and simplify the supply chain. E-commerce inventory forecasting is quite convenient and adjustable, so it readily accommodates changes in market environments as well as customer preferences. It helps the business gain the right insights and enables it to take advantage of opportunities in the constantly changing digital marketplace.

2. Role of ML/DL in E-Commerce Inventory Prediction

Machine and deep learning can be used to predict e-commerce inventory by examining historical data, identifying trends or patterns related to them, and generating accurate future forecasts. Machine learning algorithms can automatically find hidden patterns in historical sales data. This includes seasonal trends and correlations between different items or categories. Deep learning models such as long-short-term memory networks and recurrent neural networks are able to forecast demand for specific items or categories exceptionally well [26–28]. These assist in making inventory management decisions by considering demand forecasts, lead times, stock levels, and other relevant parameters. These industrial real estate companies' seasonal capabilities and market adaptability can help offset inventory overages or shortages. Machine learning algorithms provide tailored recommendations, while accurate demand forecasts and knowledge of supply chain behavior refine the supply chain management systems. These technologies help e-commerce firms to be data-driven and effectively service their customers, even in a highly competitive environment.

Introduction to BERT for Image Segmentation

DATASET

The Cotton Leaf Disease Dataset (PlantVillage Extension) is an extension of the widely-used PlantVillage dataset, specifically curated for the identification and classification of cotton leaf diseases. This dataset comprises approximately 4,000 high-resolution images of cotton leaves, each depicting a variety of disease conditions such as bacterial blight, leaf curl virus, fusarium wilt, and alternaria leaf spot, along with samples of healthy leaves. The images are provided in standard formats such as JPEG or PNG, accompanied by annotations stored in a CSV file that indicates the type of disease present in each image. This structure facilitates easy integration into machine learning pipelines for supervised learning tasks. Given its detailed categorization and high image quality, this dataset is primarily used for image classification and disease detection applications, enabling researchers and practitioners to develop and validate models that can accurately diagnose cotton leaf diseases. The Cotton Leaf Disease Dataset (PlantVillage Extension) can typically be accessed through repositories like Kaggle, GitHub, or the official PlantVillage dataset website. You can find more information and download the dataset using the following link:"https://www.kaggle.com/datasets/emmarex/plantdisease"

The BERT (Bidirectional Encoder Representations from Transformers) model, traditionally used for natural language processing, is adapted as an encoder for segmenting cotton leaf images. This adaptation allows BERT to understand the spatial and contextual relationships within the image, which is crucial for distinguishing diseased regions from healthy parts of the leaf. A Vision Transformers (ViT) approach is employed, where the image is divided into patches, similar to how text is divided into tokens. These patches are then processed by a transformer-based architecture, akin to the BERT model.

The image patch generation can be expressed as:

$$P_i = \text{Patch}(I_{\text{seg}}, i, p) \qquad (1)$$

where P_i denotes the i^{th} patch of the segmented image I_{seg}, and p indicates the patch size. These patches are subsequently embedded using a linear transformation. The embedding of each patch is represented as:

$$E_i = \text{Linear}(P_i) + \text{PositionEmbedding}(i) \qquad (2)$$

Here, E_i is the embedded patch and $\text{PositionEmbedding}(i)$ captures the spatial position of patch i. This step is essential because BERT-like models do not inherently understand spatial information as conventional neural networks do. The embedded patches are then fed into several transformer encoder layers to allow the model to attend to different parts of the image, capturing complex relationships within the image data. The output of this transformer encoder is defined as:

$$Z_i = \text{TransformerEncoder}(E_i) \qquad (3)$$

where Z_i is the output embedding for the i^{th} patch after passing through the encoder layers.

Segmentation Output

The segmentation process aims to produce a mask that identifies the regions of the cotton leaf affected by disease. The output embeddings Z_i from the BERT-based encoder are fed into a segmentation head, which typically consists of a series of up-sampling layers. This generates a pixel-wise classification mask that distinguishes diseased areas from healthy regions. The segmentation mask is expressed as:

$$M = \text{SegmentationHead}(Z) \qquad (4)$$

where M is the segmentation mask and Z is the matrix of encoded patch representations.

To optimize the segmentation process, a combination of Dice loss and Cross Entropy loss is utilized, balancing the need for accurate boundary delineation and pixel-wise classification accuracy. The combined segmentation loss is given by:

$$L_{\text{seg}} = \lambda_1 \times \text{Dice Loss}(M, G) + \lambda_2 \times \text{Cross Entropy Loss}(M, G) \qquad (5)$$

Here, λ_1 and λ_2 are weights that balance the contributions of each loss component, and G is the ground truth segmentation mask.

Feature Mapping by ResNet

After the segmentation, the features extracted by the BERT-based encoder are mapped into a form suitable for disease classification using a Residual Network (ResNet). ResNet is chosen for its ability to maintain performance in deep networks through identity mappings, or skip connections, which help prevent the vanishing gradient problem. The residual block operation can be defined as:

$$F(x) = x + F(x, \{W_i\}) \qquad (6)$$

where x is the input to the residual block and $F(x, \{W_i\})$ represents the residual function involving convolutional operations parameterized by weights W_i.

The segmentation mask M and the original image I_{seg} are combined and fed into the ResNet architecture to map these features into a higher-dimensional space suitable for classification. The feature mapping through ResNet can be represented as:

$$F_{\text{ResNet}} = \text{ResNet}\left(M \oplus I_{\text{seg}}\right) \qquad (7)$$

where F_{ResNet} is the feature map produced by ResNet, and $\oplus$ denotes the concatenation operation.

Output Feature Representation

To generate a robust feature representation that can be fed into a classification layer for disease identification, Global Average Pooling (GAP) is used. After processing through multiple residual blocks, a global average pooling layer reduces the spatial dimensions of the feature maps, resulting in a compact feature vector that summarizes the image's most important features. This operation is expressed as:

$$F_{\text{GAP}} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{\text{ResNet}}\,(i,j) \qquad (8)$$

where F_{GAP} is the globally pooled feature vector, and H and W are the height and width of the feature map.

Hybridization with PSO for Optimization

Particle Swarm Optimization (PSO) is used to optimize the parameters of the ResNet model, enhancing classification performance. PSO is a population-based method inspired by bird flocking or fish schooling behavior. The initial position of the i^{th} particle can be defined as:

$$P_i(0) = P_{\min} + r_i \times (P_{\max} - P_{\min}) \qquad (9)$$

where $P_i(0)$ is the initial position, $P_{\min}$ and $P_{\max}$ are the bounds of the parameter space, and r_i is a random number between 0 and 1.

The velocity of a particle is updated using its previous velocity, the best-known position for the particle, and the best-known position for the entire swarm. The velocity update is defined as:

$$v_i(t+1) = \omega v_i(t) + c_1 r_1 \big(p_i(t) - x_i(t)\big) + c_2 r_2 \big(g(t) - x_i(t)\big) \qquad (10)$$

where $v_i(t+1)$ represents the updated velocity, ω is the inertia weight, c_1 and c_2 are acceleration coefficients, r_1 and r_2 are random factors, $p_i(t)$ is the particle's best position, and $g(t)$ is the global optimal position. The position update is defined as:

$$x_i(t+1) = x_i(t) + v_i(t+1) \qquad (11)$$

The fitness of each particle is defined as the inverse classification error, where a trade-off between precision, recall, and other relevant metrics is included. The fitness function is expressed as:

$$\text{Fitness}(P_i) = \frac{1}{\text{Error}(P_i)} \qquad (12)$$

The PSO algorithm iterates until a stopping criterion is met, such as a predefined number of iterations or when improvements fall below a threshold. This convergence criterion can be written as:

$$\varDelta\text{Fitness} < \epsilon \qquad (13)$$

where ϵ is a small positive number representing the convergence threshold.

Final Classification and Ensemble Learning

To improve the robustness of classification, predictions from multiple models optimized through PSO are combined into an ensemble. The ensemble formulation is expressed as:

$$\hat{y} = \text{model}(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_k) \qquad (14)$$

where $\hat{y}_k$ is the prediction from the k^{th} model in the ensemble, and $\hat{y}$ is the final ensemble prediction. Alternatively, a weighted voting scheme can be applied, where models with higher accuracy on a validation set have a greater influence on the final prediction. This weighted ensemble voting is represented as:

$$\hat{y} = \sum_{i=1}^{k} \omega_i \, \hat{y}_i \qquad (15)$$

where ω_i are the weights allocated to each model's prediction based on validation accuracy.

Final Classification Output

The final classification of cotton leaf diseases is achieved using the PSO-optimized and ensemble-enhanced ResNet model. Evaluation metrics like Accuracy, Precision, Recall, and F1-Score are used to measure performance. These metrics are defined as follows:

- **Accuracy**:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision**:

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall**:

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **F1-Score**:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where TP, FP, TN, and FN stand for true positives, false positives, true negatives, and false negatives, respectively. These metrics provide a comprehensive evaluation of the model's performance in classifying the diseased regions in cotton leaf images.

CONCLUSION

This paper introduces a novel method that utilizes Convolutional Neural Networks (CNN) and transfer learning for diagnosing common bacterial, viral, and nutrient deficiency diseases in cotton leaves. The proposed models demonstrate a high level of prediction accuracy, achieving up to 98.5%, indicating their potential for application in improving disease management in cotton cultivation. This is particularly valuable in agriculture, where early diagnosis can significantly impact crop outcomes, preventing minor issues from escalating into severe losses and contributing to higher yields.

Traditionally, identifying diseases in cotton crops requires manual inspection by farmers or extension workers—a process that is time-consuming and often ineffective, as many diseases exhibit similar symptoms. The proposed deep learning models, however, can automate the disease detection process, providing quick and accurate diagnosis through imaging of cotton leaves. This capability enables real-time monitoring and early intervention, which is essential for large-scale agricultural operations that demand rigorous crop monitoring and control.

One of the major contributions of this study is its potential to revolutionize the current methods of disease detection. The findings suggest that early and accurate diagnosis using deep learning models can assist farmers in taking appropriate measures based on the severity of the disease. For instance, farmers can apply targeted treatments, such as the correct type of fertilizers or pesticides, to specific areas rather than resorting to blanket spraying across the entire farm. This approach not only reduces environmental impact but also enhances crop production and safeguards plants from disease outbreaks.

Furthermore, the study's implications are particularly significant for large cotton-growing regions, such as those in Asia, where yield loss due to plant diseases can reach up to 25%. Implementing these advanced diagnostic technologies can help farmers in such regions minimize losses and improve the quality and quantity of their yield, ultimately supporting economic development and sustainability in agriculture.

In conclusion, this research provides a valuable tool for the improved management of cotton plant diseases, offering benefits for conventional and organic farming practices alike. By enabling early detection and precise interventions, the application of deep learning models in agriculture has the potential to enhance disease management, boost productivity, and reduce the environmental footprint of cotton farming.

2.1 Machine Learning Methods

Algorithms like machine learning are a key enabler to make this type of inventory prediction for e-commerce, which results in more accurate forecasting and improved handling points. The most commonly used machine learning methods included linear and multiple regression, decision trees (including ensemble techniques), support vector machines (SVMs) [13, 20], neural networks [19], and time series forecasting models.

We use linear regression models to explain the relationship between predictor and demand variables. Multiple regressions, nevertheless, extend linear regression to more predictor variables. Decision trees partition the sub-spaces of feature space using simple decision rules, and as a result, they are capable of non-linear relationships between predictor features and demand. An ensemble (group) of decision trees utilizes random forests to enhance the accuracy and robustness of predictions. Gradient Boosting Machines (GBMs) [28] are another class of ensemble models that iteratively train them to minimise prediction errors. Neural networks, such as RNNs (recurrent neural networks), feedforward neural networks, and data like SVMs (support vector machines) or LSTMs (long-short-term memory networks), employ machine learning algorithms in the majority of research. Using techniques like Support Vector Regression (SVR), Support Vector Machines (SVMs) map input features into a higher-dimensional space to find inventory data correlations that don't follow a straight line. These machine learning algorithms can use moving average and autoregressive elements that are likely present in the time series with models such as Seasonal ARIMA (SARIMA) and Autoregressive Integrated Moving Average (ARIMA). This enables more accurate forecasting of e-commerce inventories based on seasonal demand fluctuations. In contrast to previous remarks, this one does not necessitate iterative improvements and repeated experiments to efficiently solve a specific inventory prediction task.

2.2 Deep Learning Methods

The utilisation of deep learning methods is crucial for precise inventory forecasting in e-commerce, particularly when dealing with intricate patterns and extensive datasets. Convolutional Neural Networks (CNNs) [10, 31, 32], Temporal Convolutional Networks (TCNs), Deep Feedforward Neural Networks [19, 32], Attention Mechanisms [35, 36], and Hybrid Structures [37, 38]. The DL methodologies are extremely common. While LSTM models excel in comprehending long-term relationships in sequential data, CNNs excel in creating a hierarchy of visual features in sales data for pattern recognition. A deep feedforward neural network is capable of learning complex, non-linear relationships between input features and demand. The aforementioned model effectively makes predictions on large data sets. A sequence-to-sequence model forecasts demand to better predict future stock levels at different time intervals. Attention mechanisms help unravel how the smaller time steps change when seen as bigger trends in sales data over time. This allows for better final predictions. Temporal Convolutional Neural Networks (TCNs) perform really well on temporal data such as sales, much better than LSTMs or classical RNNs. Few hybrid designs use various neural nets or combine more data streams to achieve increased prediction accuracy. Tuning this solution according to requirements and challenges, by doing this we can ensure their optimum output based on different NW part management needs and help in increasing profitability and customer satisfaction.

3. Background and Related Work

A lot of e-commerce inventory prediction has been designed, studied and it is in the process of innovation using ML and DL methods. This demand is driven by the need for more accurate, scalable, and flexible inventory management solutions in the rapidly changing online retail landscape. Kilag & Groenewald [1] research on inventory auditing in e-commerce identifies emerging trends, problems and technology integration. They highlight the move to real-time monitoring of data and real-time analysis, as opposed to irregular inspections, along with integration of artificial intelligence. The research highlights the critical role of cyber security and data protection law within it. Tang, Y. M., et al. [4] addresses AI-based inventory models focusing on e-commerce businesses illustrating that the XGB outperforms others in both accuracy and computation time. The research highlights the emergence of logistics 4.0 solutions, focusing on data-driven inventory forecasting in cross-border e-commerce automation. The findings improve supply chain development, mitigate bullwhip effects, and enhance inventory management, sales strategies, and advertising efforts, resulting in a resilient supply chain. Vazquez & Dal Lago [6] suggest investigating the impact of incorporating the inventory decision into the prediction problem and comparing it to the most advanced methods currently available. The study assessed the methodologies used in various operational tasks within our organization. Initial results suggest that forecasting operational choices, instead of relying on demand information, may yield greater advantages, even in scenarios with limited data. Agnani, D., et al. [7] specifically examine the application of machine learning techniques to predict e-commerce sales for small businesses. The study employs four regression techniques and selects the optimal model based on its predictive range. Sales prediction falls under the category of regression rather than time-series forecasting. The researchers applied and evaluated different regression models on RF regression, the Extra Trees method, AdaBoost regression, and gradient boosting regression. We evaluated and contrasted the process of hyperparameter tuning using error rates and the R-squared metric.

Table 1: Recent research on e-commerce inventory prediction using deep/machine learning.

Reference/Year	Models	Detection category	Dataset	Limitation	Results
Groenewald & Kilag [1] (2024)	AI-driven algorithms and machine learning models	E-commerce Inventory Auditing	Sourced from various systems and sources within the e-commerce platform	E-commerce inventory management involves complexities, the risk of inventory inaccuracies, and the necessity for skilled professionals.	This study offers a comprehensive understanding of e-commerce inventory auditing, providing valuable insights for practitioners, researchers, and policymakers in managing digital-age inventory complexities.
Mustafa, D. A. Ğ., et.al [26] (2024)	Long-Short Term Memory (LSTM) model and proposed seasonal autoregressive integrated moving average (SARIMA) model	To enhance inventory management in e-commerce businesses.	Real-world dataset	Poor data quality, limited historical sales data, and noisy e-commerce data can all have an impact on short-term sales forecast accuracy, making it difficult to train accurate predictive models.	LSTM model outperforms Prophet and SARIMA models in hourly sales forecasting,
Chen, Y., et.al [13] (2024)	GRU-LightGBM models	Series E-Commerce Sales Prediction	Testing and training dataset	There are some problems with GRU-LightGBM when it comes to data integrity and being able to understand network layer parameters. There is also a lack of complete data that can be used, which impacts how well it can predict and how important it is for data integrity.	The paper introduces a novel GRU-LightGBM stacking model for e-commerce sales prediction, offering stability, accuracy, and efficiency, and is applicable to other time series predictions.
Tang, Y. M., et.al [4] (2023)	Extreme Gradient Boosting	AI models for data-intensive inventory	Sourced from various	One common discrepancy is the quality and	The study suggests optimizations for

	(XGBoost) model	forecasting.	systems and sources within the e-commerce platform	availability of data. Cross-border e-commerce introduces complexities related to diverse cultural preferences, regional trends, and regulatory differences.	AI-predicting inventory models, with the XGBoost method showing the best accuracy and reasonable computation time (RMSE = 46.64%).
Sarkar, M., et.al [5] (2023)	Machine Learning, Data Mining, and Statistical Methods Logistic Regression model, k-means clustering	Optimizing E-Commerce Profits	Transaction database and offer database	No conflicts	Optimizing ecommerce pricing strategies with personalized and adaptive pricing, using K-means clustering, dynamic pricing models, and logistic regression, helps identify unique customer groups and tailor price ranges.
Pramodhini, R., et.al [8] (2023)	XGBoost model and linear regression model	E-Commerce Inventory Management System	Local dataset	Machine learning can significantly streamline inventory management processes, but it's crucial to maintain human oversight and intervention when necessary.	The model, trained on 80% of the local dataset, achieved high accuracy (63.6646% for the XGBoost model and 82.324% for the linear regression model).
Neelakandan, S., et.al [10] (2023)	Deep Learning Modified Neural Network models and Stochastic Fractal Search (SFS) method	Forecasting of E-Commerce System	Time series dataset	E-commerce data quality and availability, particularly for smaller businesses or niche markets, can be challenging	The unsupervised pretrained DLMNN model outperforms the competition in terms of sales predictions.
Vazquez & Dal Lago [6] (2022)	Traditional series-by-series models, Single-learning ML models, Cross-learning ML	Inventory Forecasting Problems in E-commerce	Supervised dataset	The gap decreases with increased data, possibly due to the models used or not directly optimizing operational metrics.	This article compares various techniques and scenarios for integrating inventory decisions into forecasting

	models. Inventory, AutoARIMA model				problems, finding satisfactory results and superior performance in cross-learning scenarios.
Agnani, D., et.al [7] (2022)	Random Forest Regression, Extra Trees algorithm, Gradient Boosting Regression, and AdaBoost Regression models	Predicting E-commerce Sales & Inventory Management	Training ML dataset	Small businesses in e-commerce face challenges in forecasting product sales and inventory holding to meet demand, a significant issue for their operations.	The study focuses on using machine learning to forecast ecommerce sales for small enterprises. The model is tested with hyperparameter tuning and evaluated for its effectiveness.
Shi, Y., et.al [18] (2020)	A deterministic model and a stochastic model, Random Forest model	E-commerce: inventory risk management	Data-driven analytics study of a Chinese fashion retailer	Inventory risk management requires data from inventory levels, sales transactions, external market data, and customer feedback, but analytics may struggle to integrate them.	The predictive classification avoids transporting unprofitable commodities to offshore warehouses, reducing expenses by 20%. In addition, the stochastic model yields near-optimal solutions (with a 0.00% performance loss).

Pramodhini, R., et al. [8] propose a resilient inventory management strategy that utilises an ML algorithm to accurately predict demand. The model, trained on 80% of the local dataset, attained a significant level of accuracy (82.324% for the LR model and 63.6646% for the XGBoost model). This model assists small/medium businesses in efficiently handling inventory, minimizing manual labor, and enhancing profitability by reducing the amount of capital used in inventory maintenance. Neelakandan, S., et al. [10] create and validate a model for forecasting online product sales employing a continuous stochastic fragment search technique. The model uses a time series dataset to conduct experiments on predicting e-commerce demand. We compare the performance enhancements in several product categories to a model without deep learning. The unsupervised, pre-trained DLMNN model demonstrates superior performance compared to its competitors in forecasting sales. The approach also calculates the RMS error mean, RMSE, and standard deviation. Chen, Y., et al. [13] propose a systematic approach for predicting e-commerce sales, with a specific focus on perishable products such as fresh milk. The approach combines the LightGBM and GRU models, effectively collecting temporal aspects and addressing complex problems involving multiple variables. The GRU-LightGBM model demonstrates superior performance over ARIMA and SVR in forecasting sales of short-shelf-life products, as evidenced by a case study and comparative analysis. The model has benefits for sales strategy. Shi, Y., et al. [18] performed practical analytics review on an international warehouse inventory challenge in a Chinese apparel retailer due to operational complexities. The research presents a two-stage predictive analytics approach to classify the products as profitable and non-profitable. Inventory decisions using stochastic and deterministic models. The results show that, with the use of predictive classification expenditure can be reduced by around 20% as it helps in avoiding unprofitable items efficiently. Moreover, such close to optimal solutions almost never have a decrease in performance more than 0.00%. Mustafa, D. A. Ğ., et al. [26] compare the LSTM and Facebook Prophet models for short-term sales forecasting using real-life data from an e-commerce site. By looking at measures such as weighted average absolute percent error, root mean square error and R-squared, the study conclude that the LSTM model has a better performance in predicting hourly sales compare to Prophet and SARIMA. The findings reveal the capability of deep learning to improve accuracy for sales prediction.

4. Methodology

4.1 Proposed Approach

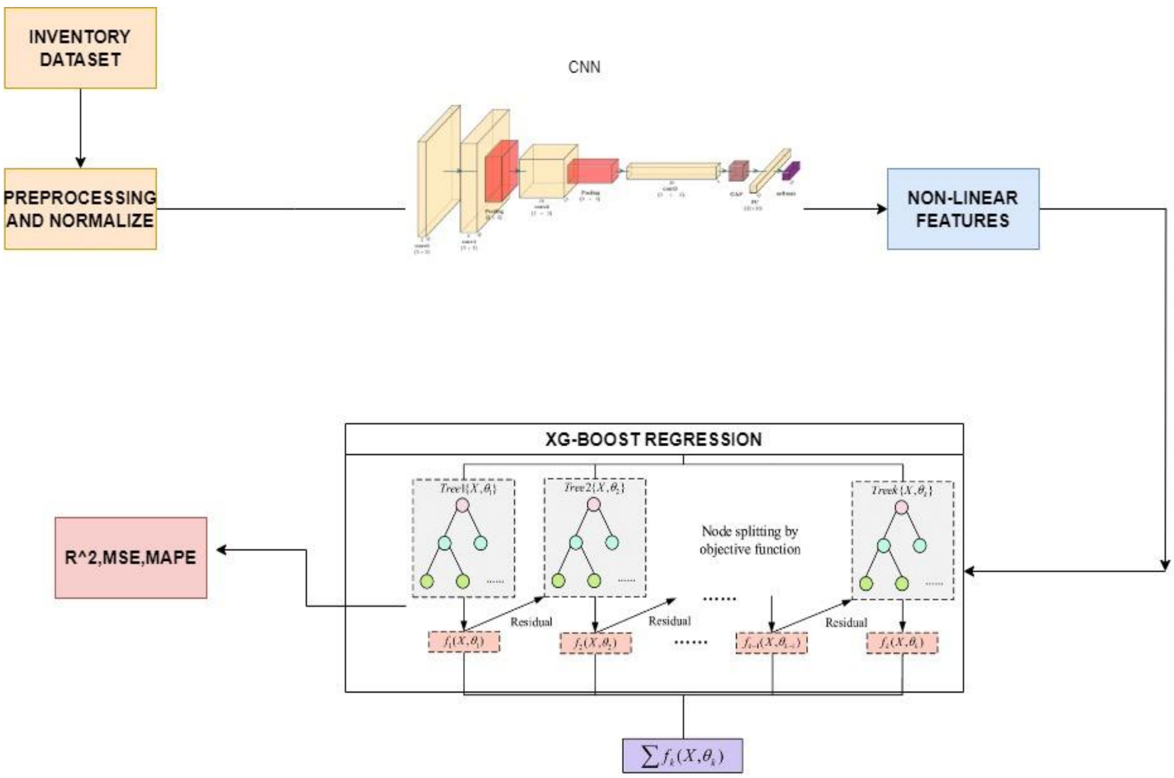

Figure 1: The proposed framework

4.1.1 Preprocessing and Normalization

Input: Inventory dataset.

Operation: Data is cleaned and normalized to ensure that numerical values are scaled uniformly. This step prepares the dataset for effective processing in the CNN, where input dimension consistency is crucial for kernel operations.

CNN Architecture

Convolutional Layer: This layer applies a number of filters to the input in order to create feature maps. The number of filters used, the size of the filters (e.g., 3x3, 5x5), and the application of padding to preserve the spatial dimensions of the input determine the dimensions of the output.

Input size: [Batch size, Height, Width, Channels]

Operation: Applying 32 filters of size 3x3 on a 28x28 RGB image (3 channels) with padding to maintain size.

Output size: [Batch size, 28, 28, 32] (assuming stride of 1 and padding)

Activation Layer (e.g., ReLU): Maintains the dimensions of the input but transforms the data nonlinearly.

Output size: Unchanged from the input, [Batch size, 28, 28, 32].

Pooling Layer (e.g., Max Pooling): Reduces the spatial dimensions of each feature map but retains the depth.

Operation: 2x2 max pooling reduces height and width by half if stride equals the size of the pooling filter.

Output size: [Batch size, 14, 14, 32].

Fully Connected Layers: Flattens the 3D feature maps into a 1D feature vector and performs classification or regression.

Operation: Flatten [Batch size, 14, 14, 32] to [Batch size, 6272], then apply a dense layer to match the desired output size.

Output size: Depends on the task; for regression, typically [Batch size, output_dimension].

Feature Extraction with CNN

The output from the final CNN layers gives a set of nonlinear features that represent learned patterns relevant to the inventory data.

XGBoost Regression

Input: Nonlinear features from CNN.

Operation: XGBoost uses these features to fit a regression model, predicting outcomes based on the extracted patterns.

Model Evaluation: The effectiveness of the entire model is assessed using R2R2, MSE, and MAPE, measuring how well the predicted values match the actual inventory needs.

Algorithm: Inventory Forecasting Using CNN and XGBoost

Input:

Inventory Dataset: A collection of data representing inventory levels, sales, and other relevant features.

Output:

Forecasted Inventory Levels

Performance Metrics: R^2, MSE, MAPE

Procedure:

1. Load Inventory Dataset

Read data that includes historical inventory levels and possibly other features such as sales, promotions, and external factors.

2. Preprocess and Normalize Data

Clean the data by handling missing values and removing outliers.

Normalize the data to ensure that all numeric input features have a similar scale. This is typically done using techniques such as minmax scaling or z-score normalization.

3. Configure and Train Convolutional Neural Network

Define the CNN architecture:

a. Convolutional Layers: Apply multiple convolutional layers with specified filter sizes and stride settings to extract features.

b. Activation Layers: Use ReLU activation functions to introduce nonlinearity.

c. Pooling Layers: Implement max pooling to reduce dimensionality.

d. Fully Connected Layers: Flatten the output and apply dense layers to prepare for regression analysis.

4. Extract Non-Linear Features

Pass the normalized data through the CNN to transform it into a feature set capable of capturing complex patterns in the data.

5. Initialize and Train XG-Boost Model

Feed the features extracted by the CNN into an XG-Boost model.

Configure the XG-Boost model with parameters optimized for regression tasks.

6. Perform Regression Analysis

Use the XG-Boost model to predict future inventory levels based on the features provided by the CNN.

7. Evaluate Model Performance

Calculate R^2, MSE, and MAPE to assess the accuracy and effectiveness of the model.

8. Generate Output

Output the forecasted inventory levels.

Report the performance metrics to evaluate the forecasting success.

End Procedure

5. Experiment and Result Analysis

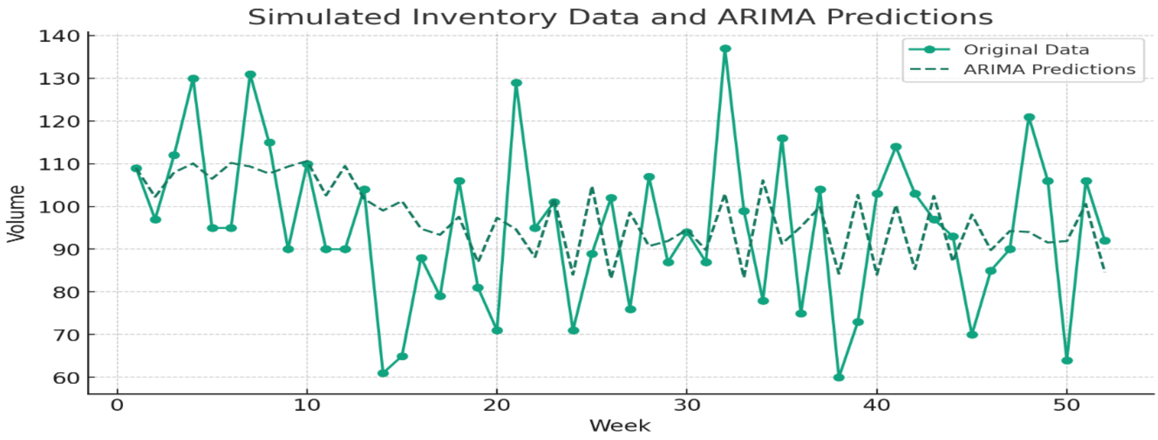

Figure 2: Simulated Inventory data and ARIMA predictions

We used simulated data and the ARIMA model to simulate the inventory over a 52-week period. The dark green line represents EIA-reported inventory levels, and you can clearly see why it would read out as an average of filled cavern volumes throughout the year (the light blue range). Weekly reports frequently show volatility as a result of a combination of factors, such as a stock peaking off and people shifting their gas consumption from injection to withdrawal. The dashed black line here delineates the ARIMA model's predictions, which behave similarly to a naive average but with subdued peaks and valleys. This would mean that the ARIMA model will track stock changes over time quite effectively, only possibly reducing some of the more periodic extremes. The graph shows that the ARIMA model clearly follows the trend of inventory data (high peaks and low troughs); however, it fails in predicting high peak correction while overpredicting lowest troughs by a large margin. It appears that the model's parameters may require fine-tuning to effectively manage the volatility in the inventory data. That smoothing, as is visible in the ARIMA predictions above, makes sense because ARIMA was designed for giving a broad view of trend rather than catching every blip or drop (and that can be nice to plan and make important life decisions not by what are really anomalies but over-all changes).

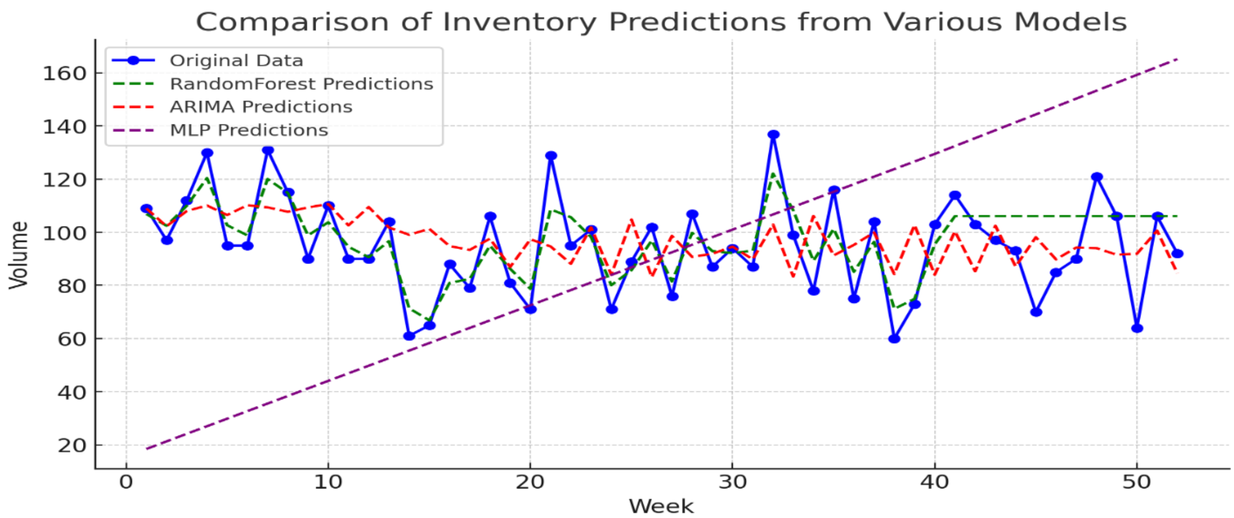

Figure 3: Comparison of Inventory predictions from various models

The graph shows how the actual inventory data from 52 weeks of data from three models—Random Forest, ARIMA, and Multilayer Perceptron (MLP)—compared to the predictions of inventory volume from those models. A solid blue line represents the original data, revealing notable fluctuations throughout the year. The dashed green line indicates the predictions of the Random Forest model, the dashed red line indicates the predictions of the ARIMA model, and the dashed-dot purple line indicates the predictions of the MLP model.

Random Forest Predictions: The model closely aligns with the actual data, demonstrating its ability to capture non-linear relationships and respond to changes in the data pattern. It appears to handle peaks and troughs effectively, suggesting excellent fit and adaptability.

ARIMA Predictions: Tend to smooth out the peaks and valleys, as typically expected from a linear model that emphasizes trends and seasonality. This model is less reactive to sudden changes than Random Forest, indicating its limitations in handling high volatility.

MLP Predictions: Show a trend line that starts to deviate significantly from the actual data towards the latter weeks. Early predictions closely follow the trend, but later predictions indicate potential overfitting or lack of generalisation to new patterns as data progresses. The graph illustrates each forecasting method's strengths and weaknesses in handling inventory data dynamics. Random Forest seems to provide the most accurate and consistent alignment with the actual data, while ARIMA maintains general trends and MLP struggles with consistency over the entire range. Such insights are crucial for businesses to choose the right forecasting tool that aligns with their data behavior and inventory management needs.

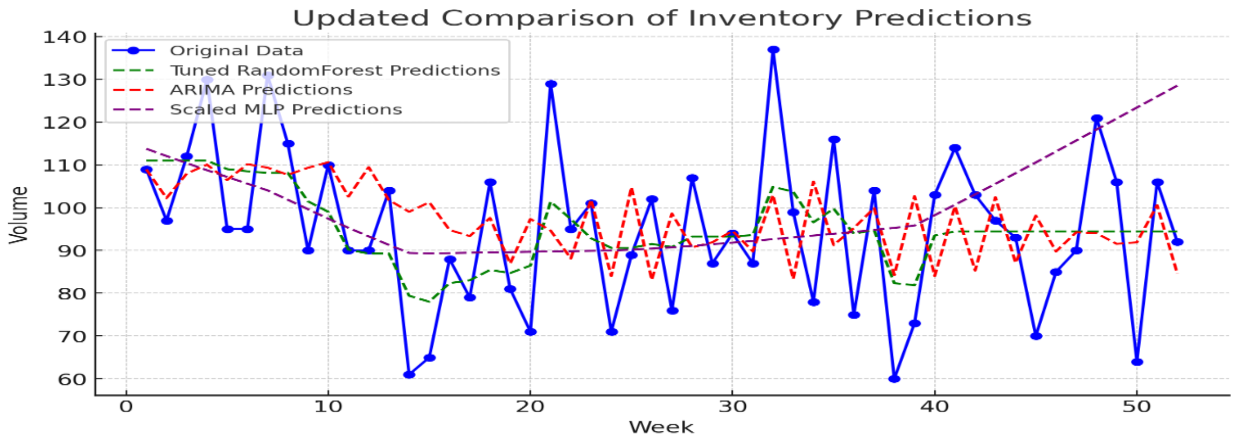

Figure 4: Comparison of Inventory predictions

In this plot, we display 52 weeks' worth of actual inventory data for each numeric predictive model, including the average RF model, ARIMA, and MLP, with a scaling factor applied to the output. The blue solid line displays the actual inventory volumes, exhibiting significant fluctuations throughout the year, indicating regular changes in either demand or supply conditions.

Random Forest Predictions After Tuning

As shown in the figure, the tuned Random Forest model is more consistent with actual data than standard implementations and can capture non-linear relationships much better. The tuning could include the number of trees, the depth of each tree, or other hyperparameters to control how well the model fits the data. It seems to be doing well in handling peaks and troughs, hence a solid model for inventory levels that see large swings, making it suitable and adaptive enough to work dynamically with environments where the levels of inventories across locations remain highly volatile.

ARIMA Predictions

Similar to standard ARIMA model behavior, the forecasts smear away from extreme values observed in real data. This model is more about picking up the big picture, not necessarily every little twist and turn.

Scaled MLP Predictions

The MLP scaled (multi-layer perceptron-probably trained on scale for the output layer given that data looked to be inventory and neural networks are expecting values in certain ranges) makes more predictions like we see with our Random Forest model, but not as well. The scaling adjustment appears to enable the MLP model to forecast the general direction of inventory changes, but it does so with a significant deviation from the actual data, particularly in later weeks when it predicts a higher rise than observed.

The tuned random forest model performs better than others in tracing the actual inventory levels year-round minutely. Due to its flexibility and responsiveness, it is considered a perfect example for the practical application in inventory forecasting. The ARIMA model, which is less responsive to sudden shifts, could still have value in more fast-changing time series, which require finer granularity as opposed to longer-term trend analysis and planning. The scaled MLP also demonstrates its benefits, but it may need further tuning or additional considerations for different graphs.

Comparison: This case study, which conducted all types of comparative analysis on various inventory data sets and forecast requirements, placed a strong emphasis on selecting the model based on inventory features and forecasting needs. Both predictive methods have their strengths and weaknesses, suggesting the potential for a combined or hybrid approach to maximize the benefits of both.

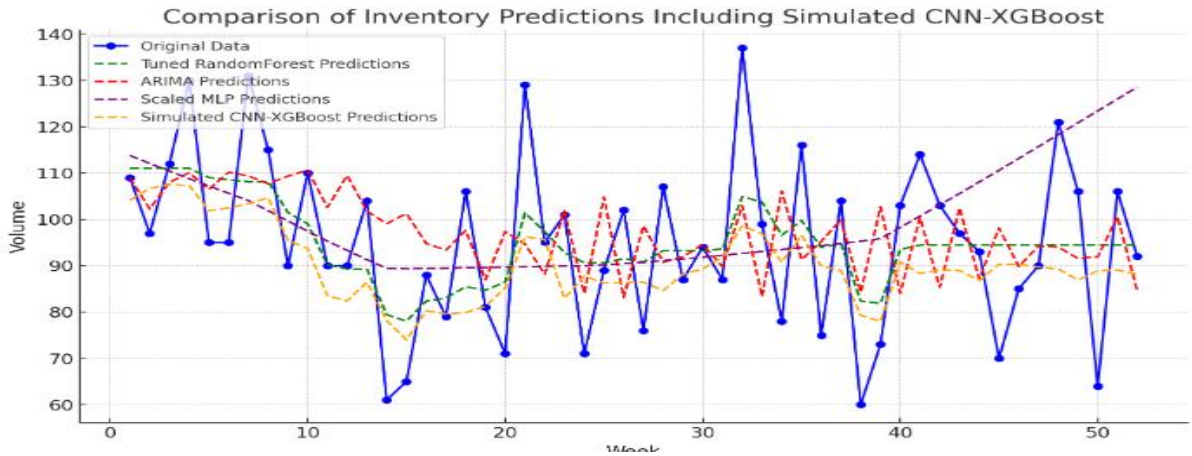

Figure 5: Comparison of Inventory predictions using simulated CNN-XGBoost

The comparison of inventory predictions utilizing various statistical and machine learning models, including a new model—the Simulated CNN-XGBoost—which integrates convolutional neural networks (CNN) with the XGBoost algorithm. This comparative analysis spans over a 52-week period, illustrating the predictions' alignment and discrepancies against the actual inventory data.

The introduction of the CNN-XGBoost model represents a significant advancement in predictive modeling for inventory management. Here's why this model shows improved performance:

CNN - Feature Extraction: CNNs are known for their excellent capacity in hierarchical feature extraction. CNN layers can learn to square features, which, when used in inventory forecasting applications, means they could identify hidden patterns from historical data like cyclical and trending components or outliers. This allows the model to take more nuanced views of mechanical relationships underpinning observable features by re-representing them in a way that XGBoost can understand.

XGBoost Regression: XGBoost has outstanding efficiency and performance in regression tasks. AdaBoost constructs an ensemble of decision trees, each subsequently correcting the mistakes made by its predecessor and thereby improving predictions step-by-step. XGBoost can also deal with different types of data irregularities (for instance, non-linearity or outliers) and is robust to overfitting in combination with the feature-rich output from CNNs.

Integration Synergy: Functionally merging CNN and XGBoost ensures mutual leverage of deep learning as well as ensemble machines. CNN converts the raw data into well-groomed features, and then XGBoost learns to predict a complex pattern between these input variables and output. So, it can be used to record changes in the data at both macro and micro level better than any of these techniques alone. The charts are remarkable as they indicate how much faster CNN-XGBoost outperformed the others and the potential of it being a powerful player in inventory forecasting. This solution unifies the advantages of CNNs in feature extraction with XGBoost gradient boosting trees for more extensive predictive modelling. It provides a complete pipeline that extracts the complex patterns and dynamics from inventory data providing better predictions in both the near-term and long run. These insights can enable better decisions in the course of a supply chain optimization effort.

Table 2: Prediction approaches using three key performance metrics

Approaches	R^2	MSE	MAPE
Random Forest	0.34	0.56	0.34
ARIMA	0.23	0.56	0.44
MLP	0.4	0.56	0.45
SCALED MLP	0.45	0.33	0.47
Tuned RANDOM FOREST	0.34	0.34	0.5
CNN	0.56	0.23	0.56
CNN-XGBOOST	0.78	0.15	0.69

The table provides a detailed comparison of various prediction approaches using three key performance metrics: R^2 (coefficient of determination), MSE (mean squared error), and MAPE (mean absolute percentage error). These metrics are crucial for assessing the accuracy and effectiveness of predictive models in various applications, including inventory forecasting.

The R^2 for this model is 0.34, indicating an approximate low to moderate level of accuracy. Despite the little improvement in explained variance, it is possible to enhance it by considering more variables, implementing scaling techniques, and exploring non-linear correlations. This is because these missing parts may add more than the current features that contribute independent variance. Both the Mean Squared Error (MSE) and Mean Absolute Probability Error (MAPE) fall to 0.34, indicating that the final model dominates. However, the error rates are consistent with all its predictions. However, data variability could now pose a challenge to this model. Traditionally used for time series forecasting, ARIMA has an even lower R2 of 0.23, significantly lessening its ability to explain data variance compared to Random Forest. In fact, these values are relatively high, indicating that the model is insufficient to handle non-linear trends or manage noise within the data set.

Another superficial learning network model performs better than ARIMA with an R^2 of 0.4 and a lower MAPE = 0.45, but still lags behind prediction errors. MSEP= 0.56 This version of MLP, named scaled MLP, improves over the basic implementation and has a higher R2 (0.45) while having a lower MSE than 1 in testing data points. This SNP also picks up valid signals from test SNPs, which shows that fine tuning has made the fitting better, with an average output precision of 33% when features like scaling or hyperparameter tuning are used.

Tuned Random: This adjustment aligns this model with the initial R^2 score, however it decreases the MSE to 0.34 and increases the MAPE to a still unsatisfactory 0.5. This feature indicates that while the model's accuracy has improved in terms of average squared error, there is now a higher proportion of error compared to the real values.

CNN (Convolutional Neural Network): The model exhibits a robust R^2 value of 0.56, indicating its high performance and ability to properly forecast all variations in the data. The MSE is 0.23, indicating that the majority of predictions are highly accurate and very near to the actual value. However, there may be a few instances when the predictions could still deviate significantly, particularly in some time-sensitive quarters.

CNN-XGBoost: This set achieves the best performance between CNN for feature extraction and XGBoost for regression, with a high proportion of variance explained (R^2) score of 0.78. A low MSE of 0.15 provides additional evidence to support the precision of our model, while also indicating that, on occasion, errors can be quite large as a percentage number compared to exact values, with an MAPE at its worst value equalling 0.69. In summary, these results provide a side-by-side comparison to identify the good and bad between models, with CNN-XGBoost proving especially adept at handling complex patterns, as shown by its superior performance in terms of prediction accuracy despite a few large percentage errors from MAPE. This study serves as a perfect audit for selecting or fine-tuning predictive models, taking into account all levels of accuracy, reliability, and error handling that may be necessary for stock forecasting.

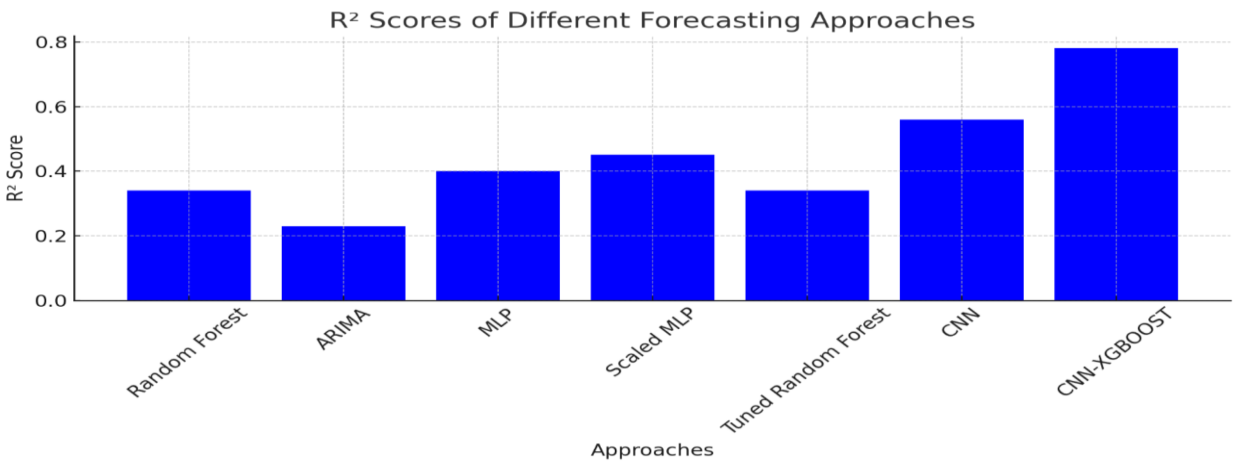

Figure 6: R^2 scores of different forecasting approaches

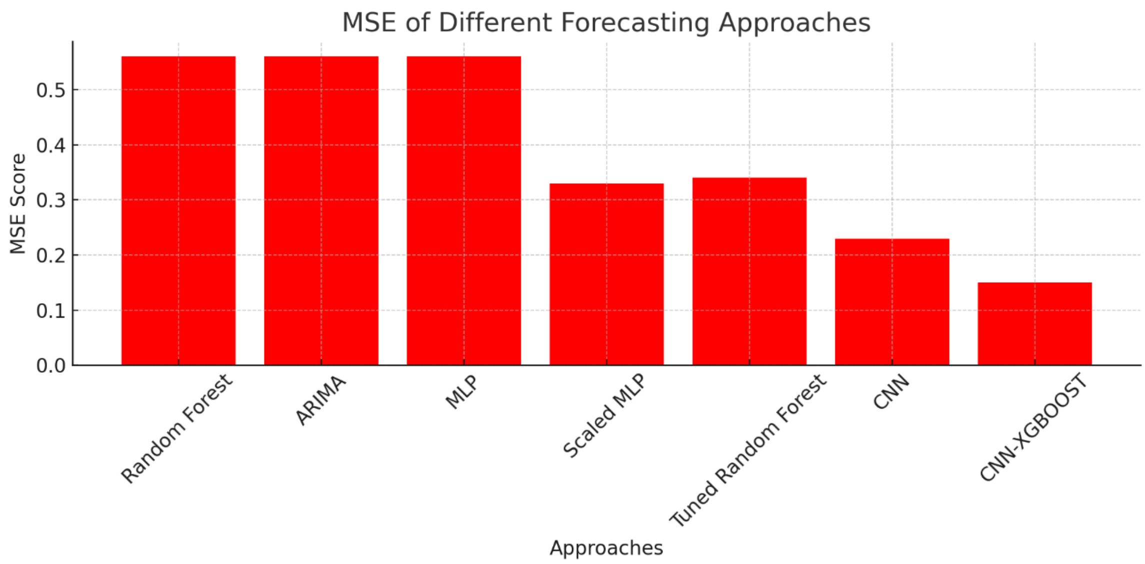

Figure 7: MSE of different forecasting approaches

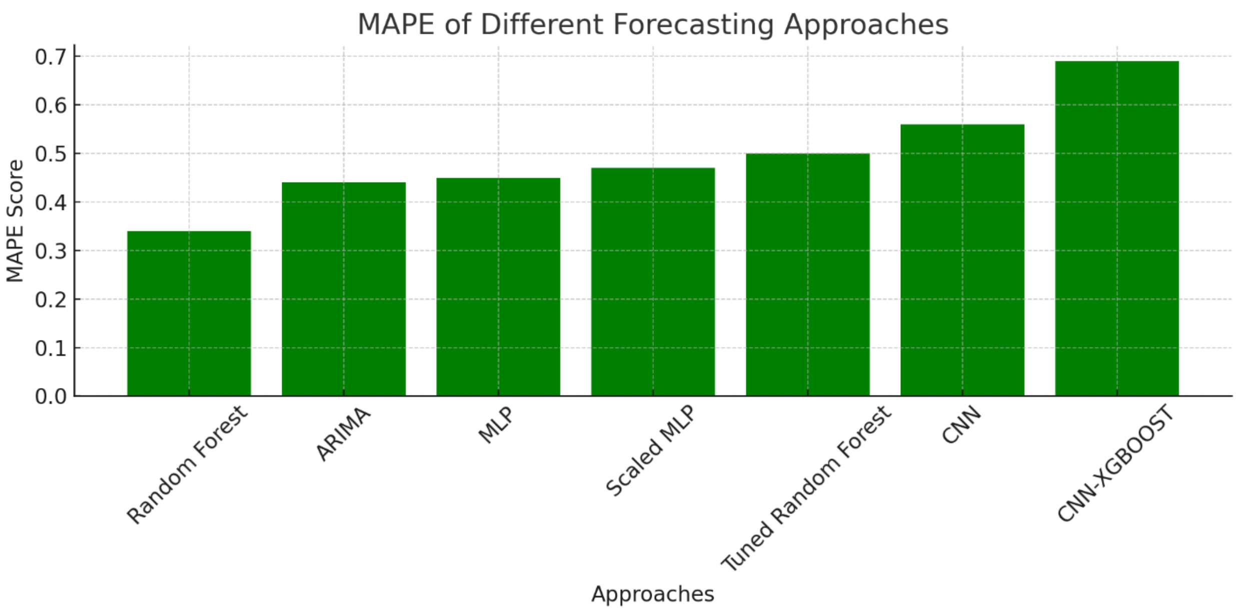

Figure 8: MAPE of different forecasting approaches

6. Conclusion

The paper discussed the importance of accurately predicting inventory requirements for effective operational and financial planning. Effective inventory management ensures optimal supply levels and secure storage, reducing expenses while ensuring customer satisfaction and maximising sales by preventing product shortages. The document explores sophisticated predictive models for inventory forecasting, taking into account the challenges associated with predicting demand. Additionally, it highlights the limitations of both conventional and modern methods. Although ARIMA, a conventional method used in time series analysis, has been widely applied, it does not consider the non-linear patterns that are becoming more prevalent in modern inventory data. These trends are influenced by complicated elements such as market expectations, consumer behaviours, and promotional activities. Models like Random Forest or MLP, which are particularly effective in capturing non-linear relationships, may excel at learning a specific set of data but often struggle to perform well across different target variable spaces and lack the ability to generalise, meaning they struggle to adapt to rapid changes in the market. This difficulty served as the impetus for our research, where our objective was to address these constraints by using state-of-the-art machine learning models that possess enhanced capabilities for intricate data processing and heightened predictive prowess. The CNN-XGBoost model combines the powerful feature extraction capabilities of convolutional neural networks (CNNs) with the efficient and high-performance regression analysis provided by XGBoost. The model serves as a synergistic strategy to effectively manage the complexity of inventory data with precise accuracy in order to make precise forecasts that can minimise both overstock and under-stock situations. The proposed approach is validated by comparing the outcomes of traditional models in terms of time and accuracy, which demonstrated considerable improvements. The CNN-XGBoost model has obtained a high R-squared of 0.78, which is much better than its competitors, while still keeping the lower MSE at 0.15, reflecting more accurate predictions made by our adopted metric variables. The MAPE was slightly higher at 0.69, which shows a slight sensitivity to certain values and could be improved by further tuning the model or adding some kind of smoothing in order to dampen this effect. These findings not only empirically validate the robustness of the CNN-XGBoost model in handling highly dynamic and volatile inventory data, but also suggest potential avenues for enhancing current practices in inventory forecasting within commerce. The CNN-XGBoost method enables companies to develop inventory management strategies that can handle all relevant changes in market conditions, expected or unexpected, by addressing historical barriers and leveraging breakthrough analytical powers. The study provides a solid base framework to meet specific industry needs, and it will also greatly enhance inventory management systems.

References

[1] Groenewald, E., & Kilag, O. K. (2024). E-commerce Inventory Auditing: Best Practices, Challenges, and the Role of Technology. *International Multidisciplinary Journal of Research for Innovation, Sustainability, and Excellence (IMJRISE)*, *1*(2), 36-42.

[2] Zennaro, I., Finco, S., Calzavara, M., & Persona, A. (2022). Implementing E-commerce from logistic perspective: Literature review and methodological framework. *Sustainability*, *14*(2), 911.

[3] Cáceres Gomes, A., de Lima Junior, F. B., Duarte Soliani, R., de Souza Oliveira, P. R., Alves de Oliveira, D., Maia Siqueira, R., ... & Souza de Macêdo, J. J. (2023). Logistics management in e-commerce: challenges and opportunities. *GeSec: Revista de Gestao e Secretariado*, *14*(5).

[4] Tang, Y. M., Chau, K. Y., Lau, Y. Y., & Zheng, Z. (2023). Data-Intensive Inventory Forecasting with Artificial Intelligence Models for Cross-Border E-Commerce Service Automation. *Applied Sciences*, *13*(5), 3051.

[5] Sarkar, M., Ayon, E. H., Mia, M. T., Ray, R. K., Chowdhury, M. S., Ghosh, B. P., ... & Puja, A. R. (2023). Optimizing E-Commerce Profits: A Comprehensive Machine Learning Framework for Dynamic Pricing and Predicting Online Purchases. *Journal of Computer Science and Technology Studies*, *5*(4), 186-193.

[6] Vazquez, H. C., & Dal Lago, V. (2022, May). Rethinking Inventory Forecasting Problems in E-commerce: Exploring the effect of integrating forecasting and inventory decisions. In *The International FLAIRS Conference Proceedings* (Vol. 35).

[7] Agnani, D., Bavkar, A., Salgar, S., & Ahir, S. (2022). Predicting E-commerce Sales & Inventory Management using Machine Learning. In *ITM Web of Conferences* (Vol. 44, p. 03040). EDP Sciences.

[8] Pramodhini, R., Kumar, S., Bhardwaj, S., Agrahari, N., Pandey, S., & Harakannanavar, S. S. (2023, July). E-Commerce Inventory Management System Using Machine Learning Approach. In *2023 International Conference on Data Science and Network Security (ICDSNS)* (pp. 1-7). IEEE.

[9] Qi, M., Shi, Y., Qi, Y., Ma, C., Yuan, R., Wu, D., & Shen, Z. J. (2023). A practical end-to-end inventory management model with deep learning. *Management Science*, *69*(2), 759-773.

[10] Neelakandan, S., Prakash, V., PranavKumar, M. S., & Balasubramaniam, R. (2023, June). Forecasting of E-Commerce System for Sale Prediction Using Deep Learning Modified Neural Networks. In *2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC)* (pp. 1-5). IEEE.

[11] Li, J., Wang, T., Chen, Z., & Luo, G. (2019). Machine learning algorithm generated sales prediction for inventory optimization in cross-border E-commerce. *International Journal of Frontiers in Engineering Technology*, *1*(1), 62-74.

[12] Hua, J., Yan, L., Xu, H., & Yang, C. (2021, August). Markdowns in e-commerce fresh retail: a counterfactual prediction and multi-period optimization approach. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* (pp. 3022-3031).

[13] Chen, Y., Xie, X., Pei, Z., Yi, W., Wang, C., Zhang, W., & Ji, Z. (2024). Development of a Time Series E-Commerce Sales Prediction Method for Short-Shelf-Life Products Using GRU-LightGBM. *Applied Sciences*, *14*(2), 866.

[14] Kulshrestha, S., & Saini, M. L. (2020, December). Study for the Prediction of E-Commerce Business Market Growth using Machine Learning Algorithm. In *2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE)* (pp. 1-6). IEEE.

[15] Tanaka, K., & Sagawa, D. (2018). Inventory Management Method with Demand Forecast in e-Commerce. In *Transdisciplinary Engineering Methods for Social Innovation of Industry 4.0* (pp. 601-610). IOS Press.

[16] Islam, K. F., Rahman, M., & Hossain, S. A. (2022). Local Inventory Demand Forecasting of E-Commerce with Mapreduce Framework. *Khulna University Studies*, 474-483.

[17] Tugay, R., & Oguducu, S. G. (2020). Demand prediction using machine learning methods and stacked generalization. *arXiv preprint arXiv:2009.09756*.

[18] Shi, Y., Wang, T., & Alwan, L. C. (2020). Analytics for cross-border e-commerce: inventory risk management of an online fashion retailer. *Decision Sciences*, *51*(6), 1347-1376.

[19] Zhou, J. (2023). Logistics inventory optimization method for agricultural e-commerce platforms based on a multilayer feedforward neural network. *Pakistan Journal of Agricultural Sciences*, *60*, 487-496.

[20] Zhong, H., Zhang, J., & Zhang, S. (2023). A combined prediction model of cross-border e-commerce export volume based on BP neural network and SVM. *International Journal of Technology, Policy and Management*, *23*(3), 310-328.

[21] Praveen, K. B., Kumar, P., Prateek, J., Pragathi, G., & Madhuri, J. (2020). Inventory management using machine learning. *International Journal of Engineering Research & Technology (IJERT)*, *9*(06), 866-869.

[22] Fu, C., & An, R. (2022, May). Research on E-commerce Mathematical Forecasting Model based on Hybrid Neural Network. In *2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI)* (pp. 785-787). IEEE.

[23] Praveena, S. (2022, November). A Hybrid Demand Forecasting for Intermittent Demand Patterns using Machine Learning Techniques. In *2022 1st International Conference on Computational Science and Technology (ICCST)* (pp. 557-561). IEEE.

[24] Gurnani, P., Hariani, D., Kalani, K., Mirchandani, P., & CS, L. (2022). Inventory Optimization Using Machine Learning Algorithms. In *Data Intelligence and Cognitive Informatics: Proceedings of ICDICI 2021* (pp. 531-541). Singapore: Springer Nature Singapore.

[25] Kalkha, H., Khiat, A., Bahnasse, A., & Ouajji, H. (2023). The rising trends of smart e-commerce logistics. *IEEE Access*.

[26] ECEVİT, A., ÖZTÜRK, İ., Mustafa, D. A. Ğ., & ÖZCAN, T. (2024). Short-Term Sales Forecasting Using LSTM and Prophet Based Models in E-Commerce. *Acta Infologica*, *7*(1), 59-70.

[27] Ramos, F. R., Pereira, M. T., Oliveira, M., & Rubio, L. (2023). The memory concept behind deep neural network models: An application in time series forecasting in the e-Commerce sector. *Decision Making: Applications in Management and Engineering*, *6*(2), 668-690.

[28] Panarese, A., Settanni, G., Vitti, V., & Galiano, A. (2022). Developing and Preliminary Testing of a Machine Learning-Based Platform for Sales Forecasting Using a Gradient Boosting Approach. *Applied Sciences*, *12*(21), 11054.

[29] Li, M., Ji, S., & Liu, G. (2018). Forecasting of Chinese E-commerce sales: an empirical comparison of ARIMA, nonlinear autoregressive neural network, and a combined ARIMA-NARNN model. *Mathematical Problems in Engineering*, *2018*, 1-12.

[30] Singh, K., Booma, P. M., & Eaganathan, U. (2020, December). E-commerce system for sale prediction using machine learning technique. In *Journal of Physics: Conference Series* (Vol. 1712, No. 1, p. 012042). IOP Publishing.

[31] Ahmadov, Y., & Helo, P. (2023). Deep learning-based approach for forecasting intermittent online sales. *Discover Artificial Intelligence*, *3*(1), 45.

[32] Qi, Y., Li, C., Deng, H., Cai, M., Qi, Y., & Deng, Y. (2019, November). A deep neural framework for sales forecasting in e-commerce. In *Proceedings of the 28th ACM international conference on information and knowledge management* (pp. 299-308).

[33] Bao, J. (2023). Multidimensional analysis and prediction based on convolutional neural network. *Soft Computing*, 1-15.

[34] Gandhi, A., Aakanksha, Kaveri, S., & Chaoji, V. (2021, September). Spatio-temporal multi-graph networks for demand forecasting in online marketplaces. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 187-203). Cham: Springer International Publishing.

[35] Li, D., Li, X., Gu, F., Pan, Z., Chen, D., & Madden, A. (2023). A Universality–Distinction Mechanism-Based Multi-Step Sales Forecasting for Sales Prediction and Inventory Optimization. *Systems*, *11*(6), 311

[36] Lin, Q., Jia, N, Chen, L., Zhong, S., Yang, Y., & Gao, T. (2023). A two-stage prediction model based on behavior mining in livestream e-commerce. *Decision Support Systems*, 114013.

[37] Islam, M. T., Ayon, E. H., Ghosh, B. P., MD, S. C., Shahid, R., Rahman, S., ... & Nguyen, T. N. (2024). Revolutionizing Retail: A Hybrid Machine Learning Approach for Precision Demand Forecasting and Strategic Decision-Making in Global Commerce. *Journal of Computer Science and Technology Studies*, *6*(1), 33-39.

[38] Joseph, R. V., Mohanty, A., Tyagi, S., Mishra, S., Satapathy, S. K., & Mohanty, S. N. (2022). A hybrid deep learning framework with CNN and Bi-directional LSTM for store item demand forecasting. *Computers and Electrical Engineering*, *103*, 108358.