Journal of Excellence in Management Sciences (2024) 4(1): 22–48
ISSN (E): 2755-3787

The Future of Retail Forecasting: A Comparative Study for Enhanced Store Performance

Mian Usman Sattar[1]

Abdul Ghaffar[2]

How to cite this article:

Sattar, M. U., & Ghaffar, A. (2025). The future of retail forecasting: A comparative study for enhanced store performance. *Journal of Excellence in Management Sciences*, *4*(1), 22-48.

Received: 8 February 2025 / Accepted: 27 July 2025 / Published online: 25 Sep 2025

Abstract

Accurate sales forecasts are crucial for retailers to efficiently plan their inventory, refine marketing strategies, and meet regional sales targets. Traditional forecasting methods are deficient in capturing complex patterns and seasonal changes, particularly in an increasingly data-driven market. This research employs advanced machine learning models, namely ARIMA (Autoregressive Integrated Moving Average), LSTM (Long Short-Term Memory), and its Hybrid PSO-LSTM (Particle Swarm Optimization-LSTM) model, to forecast retail sales in the period 2014-2021. The study finds that although ARIMA is easy to use and quick to model and implement, LSTM and PSO-LSTM (especially the latter) are more accurate and robust. PSO-LSTM is more time-consuming, but it is a more accurate and structural forecasting model, with detailed analyses that could help develop long-term plans for the company. In terms of accuracy, the PSO-LSTM model outperformed others, achieving a Mean Absolute Error (MAE) of 15,282 and a Root Mean Square Error (RMSE) of 21,459. The LSTM model followed closely with an MAE of 15,441 and an RMSE of 21,792. The ARIMA1 and ARIMA2 models had slightly higher errors, with ARIMA1 reporting an MAE of 15,460 and an RMSE of 21,557, and ARIMA2 reporting an MAE of 15,547 and an RMSE of 21,652. These results confirm the superiority of machine learning models, with PSO-LSTM providing the best overall performance. These findings highlight the potential of advanced analytics for retail to enhance operational efficiency. This research conducts a comprehensive analysis of forecasting models and provides valuable guidelines for applying these models in a highly competitive business environment.

Keywords: Machine Learning, Deep Learning, Inventory Management, Marketing Strategies, Forecasting

1 Introduction

In today's retail environment, data analytics plays a crucial role in gaining customer knowledge

[1]College of Science and Engineering, University of Derby, United Kingdom
[2]Assistant Professor, Dr Hasan Murad School of Management, UMT
Corresponding Author: abdul.ghaffar@umt.edu.pk

(Kitchens et al., 2018), optimizing inventory management, determining market pricing, and designing effective marketing campaigns (Kumar & Venkatesan, 2021). Most of the activities of organisations now rely on data analysis, and this has shaped corporate behaviour (Mustapha & Sithole, 2025). Due to breakthroughs in data collection and analytics technologies, some retailers are taking a step ahead in making more informed decisions (Brau et al., 2024), which can also lead to improved bottom lines. Retailers collect vast amounts of transactional data from more obvious sources, such as point-of-sale (POS) systems, online sales platforms, and customer loyalty programs (Kholod et al., 2024). The records capture details on sales volumes and prices for products, the time/date of their sale, the customer's identity, the payment method, and (sometimes) the point of purchase (e.g., a particular outlet restaurant) (Yu, 2023). Staying at the forefront requires a business to meet customer demand effectively, and failing to do so could risk losing customers (Kunz & Wirtz, 2024). In the retail industry, customers usually expect rapid service (Cui et al., 2024). If a store does not have an in-demand product on hand for a customer who is willing and able to buy it immediately, the store will likely lose that sale to a competitor rather than hold the sale for a later restock (Yan et al., 2022). This has immediate implications for a store's sales, and it also has lingering consequences for the customer experience, customer loyalty, and the store's reputation (Stanelytė, 2021). This means that forecasting customer demand accurately is central to developing high-precision inventory replenishment models and functional replenishment algorithms (Kumar & Mahapatra, 2021), but this could be tough since factors such as marketing, promotion, and government policies lead to a change in consumers' preferences (Shultz et al., 2022), and therefore what consumers demand is not static but instead undergoes change even rapidly (Angeletos et al., 2024). In this regard, it is often very challenging for retail stores to meet the diverse needs of consumers, as they struggle to anticipate what will be trendy in the market soon. Consequently, this puts them at a disadvantage when business decisions are made. Another reason retailers struggle to meet demand is the scarcity of future-oriented information necessary for strategic planning (Lens et al., 2024). Suppose retailers fail to know what market conditions will be like in the future. In that case, they run the risk of standard demand forecasts conflicting with actual sales quotas (Bilal et al., 2024), which leads to poorly optimized inventory management practices (overstocking or understocking), that lock up capital and incur more substantial storage costs or lower profits (Rana et al., 2023).

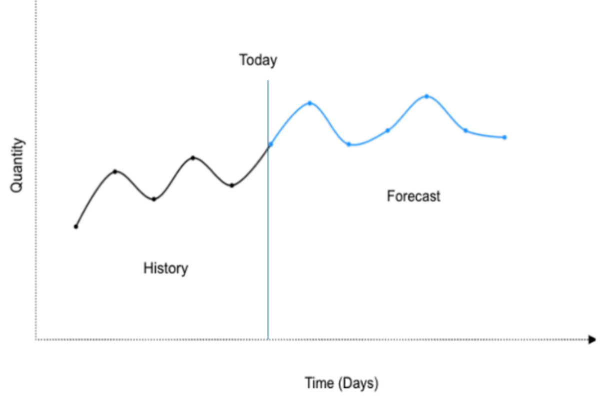

Figure 1: Sales forecasting

The infographic in Figure 1 illustrates the extent to which history plays a determining role in shaping the future. If historical parameters are used as data, then we can make the most optimized decisions to maximize performance and efficiency in all fields, including retail operations (Wang & Aviles, 2023). Without information about future patterns of market trends, decision-makers cannot adjust their strategies in advance (Nordin & Ravald, 2023). They may fail to capitalize on opportunities arising from the upswing of trends and may not be able to mitigate the impact of risks associated with a downward trend (Cherepovitsyn et al., 2025). This approach to reacting to market changes, rather than anticipating them, will be inefficient in terms of both profitability and business operations. Due to these challenges, identifying the best practices in retail store operations has become more difficult for retailers.

Figure 2 illustrates the importance of sales forecasting. There are numerous benefits that sales forecasting provides for a business. These include helping to accurately estimate a company's revenue. It helps the company set realistic financial goals and make informed investment decisions. It also helps reduce costs by optimizing production operations. Moreover, reduces unnecessary expenses. By this, sales forecasting also guides the changes in distribution channels and ensures the goods are available at the outlets that are needed. Meanwhile, sales forecasting helps to improve the targeted marketing strategies for a better return on investment. It determines how to optimize resources, enabling the company to align its inventory, staff, and production capacity with its expected sales volume. Accurate demand predictions enable a company to improve its production planning and budgeting, ensuring it can make informed, profitable decisions. Sales forecasting enables companies to enhance operational efficiency, minimize costs, and make informed, strategic decisions to achieve long-term success.

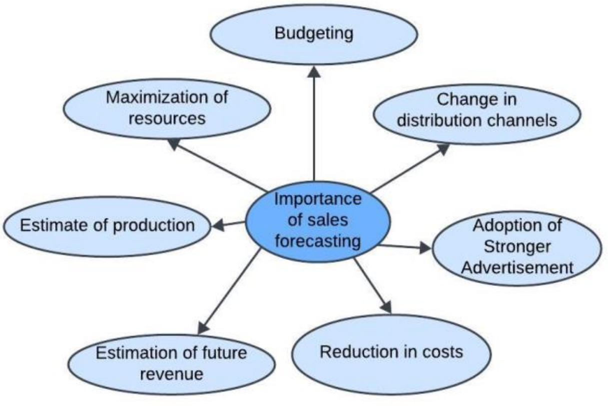

Figure 2: Mind Map for the Importance of Sales Forecasting

The employment of advanced data analytics and forecasting models that can analyze consumer behavior using past sales with predictive algorithms such as ARIMA (AutoRegressive Integrated Moving Average), LSTM (Long Short-Term Memory), and PSO-LSTM (Particle Swarm Optimization-LSTM) (He et al., 2022) which can help us gain an understanding of consumer purchase preferences and behaviors, identify sales trends, and improve demand forecasting, which can better help with inventory levels, price management, and marketing strategies (De Almeida et al., 2022). Thus, the use of these predictive analytics models can enhance retailers' operational efficiency, enabling them to serve consumers better and increase their profits (De Almeida et al., 2022).

Existing retail sales forecasting models face several limitations. Traditional models, such as ARIMA, are effective for linear data but struggle with non-linear patterns often found in retail. Many models fail to account for external factors, such as seasonal variations and promotions, which significantly impact sales. Handling large datasets and dealing with noisy or incomplete data pose challenges for these models. Moreover, extensive manual tuning of parameters is often required, making them time-consuming and not always yielding the best accuracy for practical applications.

1.1 Motivation and Gap

This study is motivated by my desire to find a more efficient way to address the daily challenge of accurately predicting future sales faced by retailers. My team experienced firsthand the frustrations of traditional forecasting methods, which are unable to capture the full complexity of real-world sales data due to its constant shifts, trends, and patterns. A key gap requiring the integration of machine learning methods in retail demand forecasting has not been thoroughly explored, particularly in terms of improving both accuracy and efficiency (Fildes et al., 2022). This presents an opportunity for this research, which aims to fill this gap by providing retailers with advanced tools to better manage inventory, plan effective marketing strategies, and ultimately grow their businesses with greater confidence. The research compares LSTM and PSO-LSTM advanced models with ARIMA to go beyond the limitations of current methods.

1.2 Research Contribution

This study makes important contributions to retail analytics by comparing ARIMA, LSTM, and PSO-LSTM models for sales forecasting. It highlights the strengths and weaknesses of each model, demonstrating how advanced methods, such as LSTM and PSO-LSTM, outperform traditional models in complex scenarios. The research provides practical insights that can improve inventory management, marketing strategies, and regional performance. It introduces methodological advancements by demonstrating the effectiveness of PSO in optimizing LSTM models, leading to improved forecasting accuracy in retail and other time-series applications.

2 Related Works

Over time, retailers have been increasingly utilizing data analytics to optimize various areas of their business, including inventory planning, customer segmentation, pricing, marketing, and other functions. The feasibility of leveraging retail transaction data for optimization purposes stems from the explosion of generated data, along with the mathematical and computational advancements in the field of data analysis. Sales data analysis is now the backbone of retail strategy. Retail analytics provides valuable insights into customers, products, sales and marketing performance, business operations and financial management, product performance, price trends and direction, as well as the strength and growth of markets. In the words of Chen et al. (2012) 'Retail analytics involves analyzing large amounts of data to identify relationships, regularities, anomalies, and drivers that can guide related decisions and actions.' Businesses are increasingly utilizing a range of techniques and tools to gain insights from large datasets. These include simple descriptive analytics that help show the overall picture of sales (e.g., moving average or exponentially smoothed average) against a target, such as sales over a week for a confectionery brand. More complex methods include clustering, also known as segmentation, which uses data science to analyze consumer behaviours (El-Adly & Eid, 2016). Some researchers suggest that a focus on consumer behavior is crucial because rigorous analysis of shopping patterns can lead to more accurate inventory management and market-specific marketing strategies. After all, the retailer could stock up on popular items and run promotions that target a specific segment of customers (Macas et al., 2021).

ARIMA (AutoRegressive Integrated Moving Average) is one of the most popular old-fashioned statistical methods used for time series forecasting, first proposed by George & Gwilym, 1970), They have been widely used for the prediction of future levels of sales for retail sectors based on past or historical data. ARIMA models are efficient and straightforward in explaining the linear time series data. Beaumont et al. (1984) suggest that seasonal ARIMA models can effectively forecast the sales of the past season and inform inventory management accordingly. Models such as Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing are still widely used in research and industry, despite the simplicity of their underlying concepts; they often deliver competitive performance. They have been shown to achieve or even outperform more sophisticated machine learning (ML) approaches, such as Artificial Neural Networks, in several publications and competitions, including the M-Competitions (Kolassa, 2022; Makridakis et al., 2022). These methods were used in several papers as a more sophisticated version of Moving Average (MA) for example, Ramos et al. (2015) compared these methods as a way of approaching the problem of forecasting the retail sales of women's footwear with product patterns having repeating fluctuations, like the demand for boots in winter.

Machine learning and deep learning are emerging methods for time series forecasting that utilize sophisticated yet straightforward models. Together, these methods have revolutionized time series forecasting by utilizing advanced computational power and algorithms. Many historical time series are characterized by simple, regular patterns that basic functions, such as exponential, polynomial, and sine curves, can represent. For these situations, models such as ARIMA and exponential smoothing are ideal choices because they are very easy to understand and apply. There are other, more complex time series that still follow general patterns but are not consistent enough to be described by simple functions. Finding the correct time series model for these situations can be challenging, if not impossible, for humans. Fortunately, ML and DL can help with these types of data, and their computational power continues to increase over time. Given vast amounts of data and high-speed computing, ML and DL can go even further. By using sophisticated algorithms and many available data sources, algorithms can capture complex relationships within time series data that would otherwise be overlooked. A particular type, known as the LSTM (Long Short-Term Memory) network, is a subtype of Recurrent Neural Network (RNN) that avoids the vanishing gradient problem, which has an advantage from a modeling perspective. Essentially, this means that LSTMs are adept at learning over a very long period (Hochreiter & Schmidhuber, 1997).

Machine learning can also aid in time-series forecasting in various ways. One of the key benefits of using machine learning to model time series is that its models can capture far more complex patterns than classic statistical approaches. Models such as Random Forests, Gradient Boosting Machines, or Support Vector Machines can uncover complex interactions among the variables and, therefore, potentially achieve more accurate forecasts than simpler models. ML techniques facilitate a lot of feature engineering, that is, the creation of additional explanatory variables. Economists can include distinct features such as lagged values, moving averages, and calendar variables to create models with better predictive performance (Elsayed et al., 2021). Popular Machine Learning models for time series forecasting include Random Forests which combines multiple decision trees into one that has higher accuracy and is less likely to overfit, Gradient Boosting Machines (GBMs), the high-performing statistical models build up over time, in which each model corrects the flaws of the last one, and generally producing high predictive accuracy and Support Vector Machines (SVMs) which are good for regression tasks, where we try to identify the ‘line’ that best predicts the output variable from all the other input variables, particularly when the variables have complex relationships with each other. In a comprehensive evaluation of forecasting techniques for horticultural sales, a study compared the performance of nine advanced machine learning algorithms with three traditional forecasting methods. The findings indicated that machine-learning approaches consistently outperformed their classical counterparts. Notably, the gradient-boosted ensemble model XGBoost emerged as the leading performer, achieving the best results in 14 of 15 comparisons (Haselbeck et al., 2022).

Recent advances in time series forecasting have been made possible by deep learning models that enable feature extraction and handling of large datasets to be accomplished automatically. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, both deep learning models that can effectively capture relations in time-series data, are particularly suited for modelling temporal dependencies and long-term trends. Network architectures known as LSTM (short for Long Short-Term Memory) and GRU (short for Gated Recurrent Unit) don't suffer from the vanishing gradient problem and can indeed learn long-term causal dependencies, they build on RNNs to avoid those issues. CNNs are often combined with RNNs (or LSTMs) because they can model complex spatial relationships in time series data that can improve forecasting accuracy. Other advantages of DL models include their ability to automatically select the relevant features of raw data, thus avoiding manual feature engineering; they can also scale to analyze large-scale, high-dimensional datasets, which lends itself to a host of real-world applications across different sectors (Calzone, 2022).

The application of ML and DL models in time series forecasting spans multiple industries. The finance industry takes advantage of these models to predict stock market prices, manage risks, and try to detect fraud. Sezer et al. (2020) demonstrated how deep learning techniques can be used for financial time series forecasting. One of the key findings of the research is that DL models can capture long-term dependencies and temporal patterns that are not always easy to catch by expert traders, who usually use simpler models. In this way, DL techniques can be especially helpful in banking and finance for forecasting purposes and determining better ways to manage risks. Zeroual et al. (2020) compared the performance of RNN, LSTM, BiLSTM, GRU, and VAE deep learning models in predicting COVID-19 case numbers from Italy, Spain, France, China, the USA, and Australia, and observed that most deep learning techniques show promising results for correctly predicting the trends of case numbers with the VAE yielding the best results, illustrating the potential of advanced machine learning techniques in public health crises management. ML and DL models might require good-quality data, for example, with complete time stamps for consumers and shops that make up a local market – the reliability of which is not always guaranteed. Developing and adapting ML and DL models is challenging because the models are often 'black boxes', not revealing insights into how these models make precise predictions, although new developments are quickly addressing these problems. Moreover, ML and DL are highly data- and computation-intensive endeavors, and the capacity of algorithms and processors might be approaching a limit when it comes to memory. Nevertheless, the rapid growth of AI technology demonstrates the capacity of ML and DL to continue revolutionizing the quality of time series forecasting (Soori et al., 2023).

LSTM networks have been shown to uncover patterns in retail sales that other algorithms struggle to identify and that are valuable for predicting demanding, highly variable, and non-linear time series. Bandara et al. (2020) found that LSTM models forecast more accurately than ARIMA (autoregressive integrated moving average), a class of predictive models used in econometrics, partly because they can also capture complex sales patterns. For example, they can absorb seasonal fluctuations in the sales pattern, as well as changes in trends over time. This makes these models particularly suitable for ever-changing retail environments. Research shows that LSTM is suitable for handling time series that contain large numbers of gaps and prolonged delays and can effectively analyze temporal series data for forecasting. Here, the LSTM is used to transform, extract the attribute value, analyze the data, and build the forecasting model, and accuracy can be improved by modifying the model (Liu et al., 2018).

PSO-LSTM (Particle Swarm Optimization-LSTM) further improves the predictive ability of LSTM networks. The LSTM model's hyperparameters are optimized through a population-based optimization algorithm known as PSO (Eberhart & Kennedy, 1995). A study report shows that an innovative method combining Long Short-Term Memory (LSTM) networks with Particle Swarm Optimization (PSO) was able to enhance sales forecasting for an e-commerce company. It optimizes hidden neurons and training iterations of LSTM to achieve better forecast performance. Similarly, this innovative method, along with experiments on real systems and benchmark data, demonstrates that PSO-LSTM achieves even better results than nine other methods, as its accuracy is also reported to be significantly higher than that of some competitors. The above discussion demonstrates that optimizing the neurons of an LSTM network can enhance the predictive performance of sales within e-commerce businesses (He et al., 2022).

Comparative studies between various forecasting models illustrate their strengths and weaknesses. For example, a study by Makridakis et al. (2018) compared traditional statistical methods and modern machine learning approaches, concluding that while traditional methods such as ARIMA are still helpful for linear and stationary time series, the ML (machine learning) models, particularly DL (deep learning) models such as LSTM that handle complex, non-linear data are much more successful. Extending this analogy, Fischer and Krauss (2018) compared the performance of both LSTM and PSO-LSTM networks against traditional models in the domain of financial time series forecasting, finding that deep learning models not only outperform traditional models but also significantly improve predictive accuracy. LSTM models can utilize their capability to model long-term dependencies more effectively and accurately represent complex temporal dynamics. By introducing PSO, they can optimize network parameters more efficiently, thereby further improving predictive performance. Studies implementing comparative approaches to forecast nickel prices investigate what is lacking in existing methods and what can be achieved by new techniques. An improved PSO-LSTM neural network model for predicting the nickel price is established, aiming to enhance the reliability of the prediction results. A nonlinear declining assignment method coupled with a sine function is proposed for parameter optimization of the PSO and introduced to the LSTM to enhance performance. Empirical analysis based on London Metal Exchange data shows that the predictive performance of the improved PSO-LSTM model is much better than traditional LSTM and ARIMA models. The prediction error is reduced by 9% and 13% points, respectively, compared to those of the LSTM and ARIMA models, which is beneficial for investors in mitigating the risk in the nickel market (Shao et al., 2019).

Optimizing a retail store, for example, with advanced forecast models is better than the traditional way and guarantees better accuracy and performance. The grocery producers' market has achieved a new level of accuracy and efficiency in making decisions on pricing and distribution due to advanced models, which can forecast daily replenishment and pricing. Accurate forecasts enable the company to make better decisions in vital areas such as maintaining products in stock without overstocking. Hence, forecasting helps businesses optimize their operations by adjusting stock allocation at set time intervals to better accommodate customer needs, thereby reducing stockouts and overstocking. To achieve this, grocers and retailers leverage the advanced analysis of market demand for better business operations. For instance, to analyze correlations and relationships in sales, algorithms such as the Apriori model as well as linear fitting methods are adopted, and LSTM time series prediction models that take into consideration the PSO algorithms are used to forecast the growth pattern of fruits and vegetables, hence making it effective in pricing and replenishment schedules to meet the demand while building the profitability of these product lines that involve fresh fruits and vegetables (Yan & Deng, 2024). Other studies claim that using models such as LSTM for retail forecasting can improve demand forecast. By comparison of three methods, ARIMA, SVM (Support Vector Machine) and RNN (recurrent neural network), and LSTM (long short-term memory) for retail data on performance with five dimensions (prediction accuracy, flexibility, running speed, cost and convenience), researchers found out that the performance of the kinds of products in the retail industry requires different forecast technologies. LSTM performs best on the performance of demand prediction task, which realized the matching of model and product kinds is important for the process of forecast (Wang et al., 2019).

3 Research Methodology

3.1 Ethical Consideration

This research, approved by the University of Derby (ETH2324-4156), uses anonymized Kaggle data to ensure privacy and minimize risks. Strict data security measures were followed, and the methodology was conducted transparently without conflicts of interest. The findings will be presented with a balanced analysis, acknowledging any limitations or biases.

3.2 Proposed Methodology

Given the requirements outlined in our study, we propose a multistep methodology to achieve the aims and objectives not addressed in the study. Accordingly, a set of advanced analytics and an ML model-based approach will be leveraged to achieve more plausible results using the retail sales data as training data. This process will include steps such as data collection, preprocessing, selecting an appropriate model and analysis, validation, and evaluation. Figure 3 illustrates the systematic methodology, which comprises data collection, processing, and analysis.

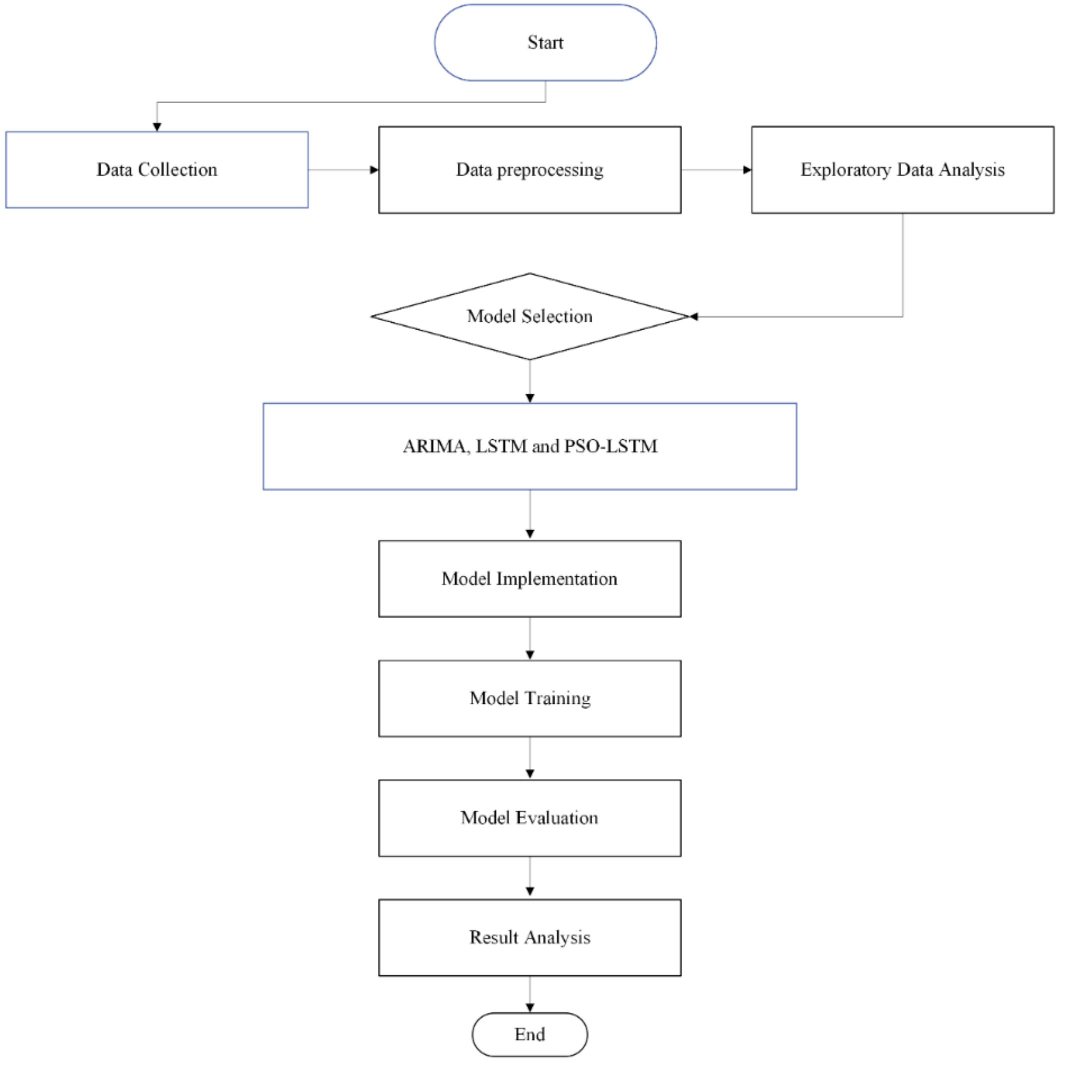

Figure 3: The Proposed Methodology

3.3 Data Collection Method

The data used is from the Kaggle open-source data repository, titled "ECommerce Data Analysis"

(Islam, 2023). It involves a set of several tables and covers 7 years of sales transactions done at more than 700 retail outlets with attributes that are important for carrying out the analysis. This makes it easier to conduct an analysis that covers sales patterns and trends across different periods and geographical locations.

3.4 Data Analysis Method

3.4.1 Selected Tools

We chose PySpark as a primary tool to implement this proposed methodology. PySpark is an integration of the Python programming language and Apache Spark, a large-scale data processing system set up in a distributed computing environment. PySpark enables the handling of datasets in a robust computing environment, allowing for the processing, joining, and analysis of data with big data in mind (Apache Spark, 2025).

3.4.2 Data Preprocessing

The data preprocessing stage involves cleaning the data, i.e., filling gaps and correcting anomalies, transforming the data to the required shape for analysis, and normalizing the data to ensure a comparable scale.

3.4.3 Algorithm Selection

For retail store sales forecasting, three primary algorithms will be used: ARIMA, LSTM, and PSO-LSTM. These algorithms were selected based on their strengths and applicability to time series forecasting.

1. ARIMA (Autoregressive Integrated Moving Average)

ARIMA is among the most used statistical methods for predicting a series of temporally distributed values. It is well-suited for short-term forecasting and cases where the data is stationary. It is built on top of the combination of three components: Autoregression (AR), Integration (I), and Moving Average (MA). As you can imagine, all these components play a particular role in making the ARIMA model so well-suited to forecasting time-series data. The architecture and the mathematical notation of the ARIMA model are illustrated in the section below.

i. Autoregressive (AR) Component

The Autoregressive (AR) component consists of regressing the variable on its own lagged values. In an AR(p) model, the series' future value is calculated as a linear combination of its prior p values.

$$y_t = \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} \ldots + \emptyset_p y_{t-p} + \epsilon_t \quad (1)$$

where in equation 1,

y_t is the value at time t

$\emptyset$ represents the autoregressive coefficients

ϵ_t Is the white noise error term.

ii. Integrated (I) component

The Integrated (I) component involves differencing the data to make it stationary, which is crucial for the efficient application of the AR and MA blocks. The value of d indicates the order to which the data has to be differentiated to make it stationary. If d = 1, the series is different each time.

$$y'_t = y_t - y_{t-1} \quad (2)$$

where in equation 2,

y'_t represents the differenced series

y_t Is the value at time t

iii. Moving Average (MA) Component

The MA component models the error term as a linear combination of past errors. In an MA(q) model, the series' value is expressed.

$$y_t = \in_t + \theta_1 \in_{t-1} + \theta_2 \in_{t-2} - 2 + \cdots + \theta_q \in_{t-1} - q \quad (3)$$

where in equation 3,

y_t is the value at time t

θ_1 represents the moving average coefficients

$\in_t$ Is the white noise error term.

iv. ARIMA Model

When these three components are combined, the ARIMA model becomes ARIMA (p, d, q), where p represents the order of the AR part, d represents the number of different steps, and q represents the order of the MA part.

$$y'_t = \phi_1 y'_{t-1} + \phi_2 y'_{t-2} + \cdots + \phi_p y'_{t-p} + \in_t + \theta_1 \in_{t-1} + \theta_2 \in_{t-2} - 2 + \cdots + \theta_q \in_{t-1} - q \quad (4)$$

where in equation 4,

y_t is the value at time t

ϕ represents the autoregressive coefficients

θ represents the moving average coefficients

$\in_t$ Is the white noise error term.

v. Model Selection and Parameters

a) Choosing the suitable values for p, d, and q requires: Using techniques such as autocorrelation function (ACF) and partial autocorrelation function (PACF) plots to determine potential orders of AR and MA components.

b) Estimating model parameters using methods such as Maximum Likelihood Estimation (MLE).

c) Evaluating the model with diagnostic tools to guarantee that the residuals are white noise (that is, they have no autocorrelation).

The ARIMA model's flexibility and ability to represent a wide range of time series data make it an excellent tool for forecasting. Its comprehensive technique captures multiple aspects of the data, resulting in solid and dependable forecasts, particularly for short-term projections.

vi. ARIMA Model Architecture

Figure 4 shows the Arima Model Architecture

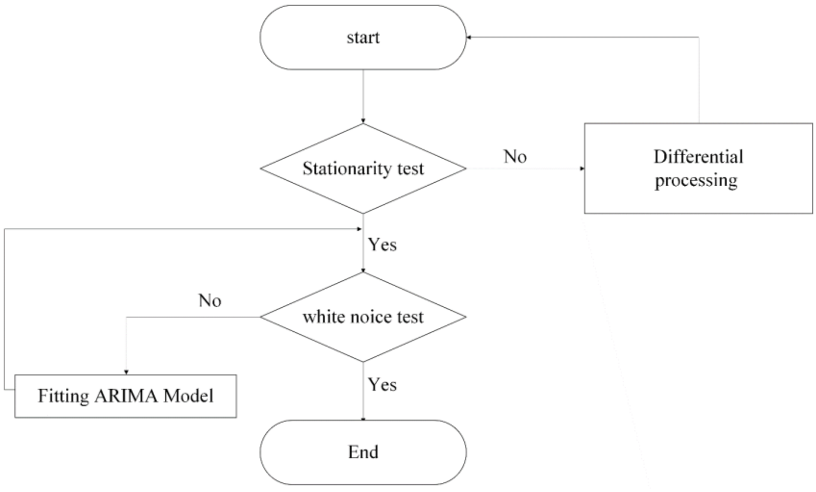

Figure 4: ARIMA Model Architecture

2. LSTM (Long Short-Term Memory)

LSTM is an RNN network that can learn long-term dependencies and is well-suited to sequence or time series data. The design of LSTM helps address issues that traditional RNNs encounter when dealing with certain types of long-term inputs. The memory cells in LSTM remember information for a longer period. In a Long Short-Term Memory (LSTM) network, there are three key gates involved: the Forget Gate, the Input Gate, and the Output Gate. Each of these gates operates as an information selector, supplying an input and output selector vector to filter and manipulate the flow of information through the LSTM cell. Below is a brief description of the working process of these gates:

i. Selector Vectors

Each of these three gates creates a selector vector, that is, a vector consisting of numbers between zero and one that controls how much of the old information should be remembered or discarded in each data-processing step:

Values close to 0: Eliminate the information. Values close to 1: Retain the information.

ii. Gate Mechanism

Each LSTM gate is a neural network producing time-evolving selector vectors utilizing the sigmoid activation function, which ensures that the output code always lies in the 0-1 range, which in turn forces precise control over the information flow:

$$\text{Sigmoid } ((\varkappa) = \frac{1}{1-e^{\varkappa}} \tag{5}$$

iii. Input to the Gates

All three gates use a combined vector of the current input vector (X)and the previously hidden state vector Ht-1. This combined vector serves as the input for the gate computations.

iv. Gates and their Functions

The forgetting gate removes unnecessary information from the cell state, focusing only on relevant historical data. The input gate introduces new data into the cell state, striking a balance between maintaining past information and integrating new insights. The output gate generates the final output by selecting which parts of the cell state will be visible in the subsequent step.

v. Cell State

The main memory of the LSTM unit is the cell state; it retains and transfers relevant information through time. The current cell state is updated by combining the old cell state, scaled by the forget gate, and the candidate values, scaled by the input gate.

vi. LSTM Unit Architecture

The Architecture of an LSTM unit is shown in Figure 5.

LSTM consists of a series of gates that control the flow of information through the memory cell, and the mathematical equation is given as:

$$\begin{cases} f_t = \sigma(\mathcal{W}_f.[h_{t-1}, \varkappa_t] + b_f \\ i_t = \sigma(\mathcal{W}_i.[h_{t-1}, \varkappa_t] + b_i \\ C_t = \tanh\,(\mathcal{W}_i.[h_{t-1}, \varkappa_t] + b_C \\ C_t = f_i.C_{t-1} + i_t.Ct \\ o_t = \sigma(W_{o.}[h_{t-1}.x_t] + \ b_o \\ h_t = o_t.\tanh(\,C_{t)} \end{cases} \tag{6}$$

Where in equation 6

f_i is the forget gate

i_t Is the input gate

C_t Is the cell state

o_t Is the output gate

h_t Is the hidden state

σ Is the sigmoid function

Tanh is the hyperbolic tangent function

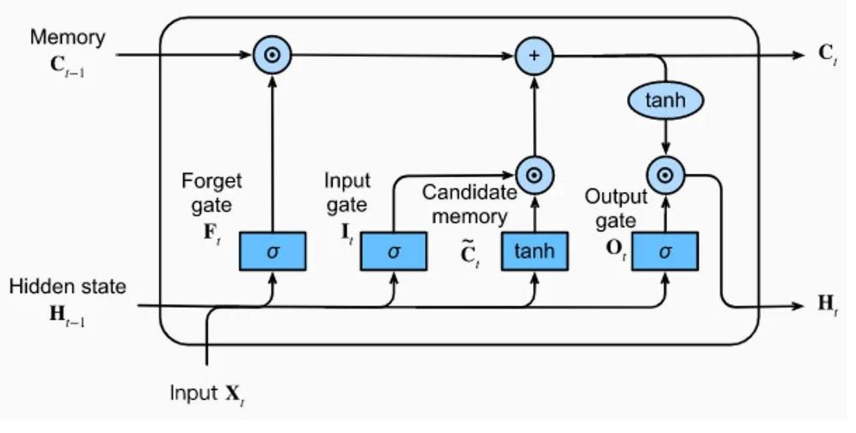

Figure 5: LSTM Unit Architecture

vii. The implementation process of the LSTM Model

The implementation process of the LSTM model is shown in Figure 6.

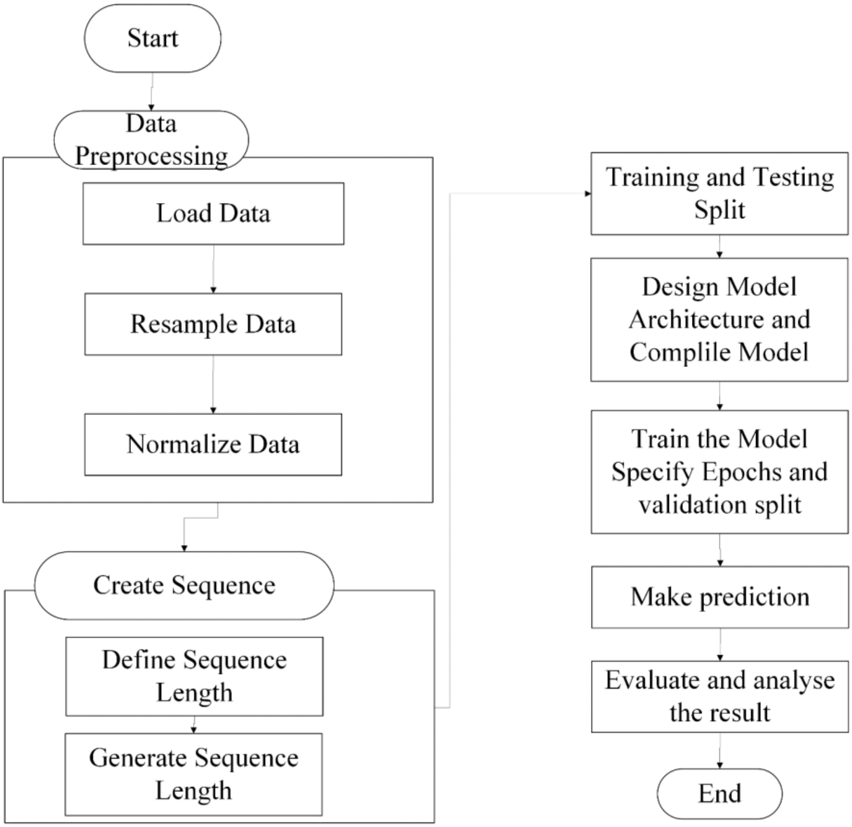

Figure 6: Implementation process of the LTSM Model

1. **PSO-LSTM (Particle Swarm Optimization Long Short-Term Memory)**

PSO-LSTM combines LSTM with Particle Swarm Optimization (PSO), an optimization method that mimics bird flocking or fish schooling behaviour. The algorithm is based on finding the best set of hyperparameters for the LSTM network, which in turn improves the forecasting accuracy. The PSO algorithm iterates over potential solutions (particles) to find the optimal parameters, and the mathematical equation is given as follows:

$$v_i(t+1) = \omega . v_i(t) + c_1 . r_1 . (p_i - x_i(t) + c_2 . r_2 . (g - (x_i(t)) \qquad (7)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \qquad (8)$$

Where in equation 7 and 8,

vi is the velocity of the particle i

x_i is the position of particle i

p_i is the best-known position of particle i

g is the best-known global position

ω is the inertia weight

c_1 and c_2 are acceleration coefficients

r_1 and r_2 are random numbers

Figure 7 shows the implementation process of the PSO-LSTM Model

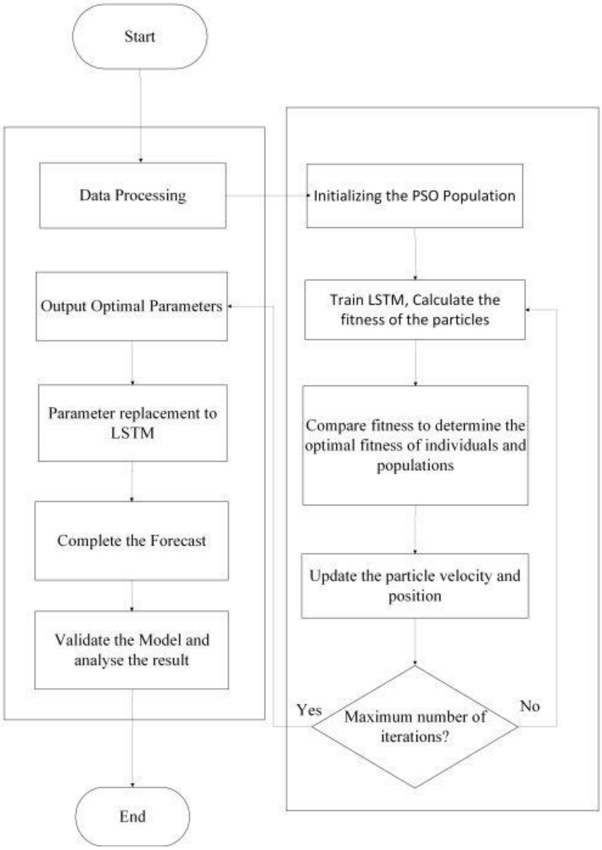

Figure 7: Implementation process of the PSO-LTSM Model

3.4.4 Model Evaluation

Every selected model will be trained in historical sales. The performance of models will be evaluated with metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and cross-validation techniques to become more robust and reliable. The formula for MAE and RMSE is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (9)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (10)$$

Where in equation 9 and equation 10,

n is the number of observations.

y_i is the actual value.

$\hat{y}_i$ is the predicted value.

$|y_i - \hat{y}_i|$ is an absolutely error for each observation.

$(y_i - \hat{y}_i)^2$ is the squared error for each observation.

3.4.5 Data visualization

The visualization will cover various aspects of EDA, including the distribution of sales trends over time, customer segmentation, top-performing and underperforming items, and prediction plot and model loss. Once the models have been trained and evaluated, the results will be visualized to provide insights into sales trends and forecasts. As PySpark does not directly support visualization, the processed data will be converted into Pandas Data Frames and visualized using libraries like Matplotlib and Seaborn. By implementing this comprehensive data analysis method, we aim to enhance our understanding of retail sales patterns, uncover hidden insights, and improve the accuracy of demand forecasting models.

4 Results and analysis

4.1 Data Division

The data was converted from daily to weekly frequency and split into training and testing sets to build and evaluate the forecasting models. First, the data was resampled from daily frequency to weekly frequency and then split into two subsets with 80 percent of the data used for training and 20 percent of the data for testing and A training set of 290 weeks and a test set of 73 weeks were obtained. For robust evaluation, cross-validation techniques were employed. ARIMA Models used 5-fold cross-validation on the training set, which helped prove the model's stability concerning a differently affected subset of the training set. In contrast, the LSTM and PSO-LSTM Models performed rolling cross-validation to preserve temporal order, evaluating the models' performance on sequentially increasing windows of data at every step of the rolling window.

4.2 Feature Engineering

Feature engineering is necessary to optimize a model's performance. In this research, the steps were as follows.

- **Resampling and Aggregation:** Raw data arrived as daily, but we resampled it to weekly after averaging over population total, to reduce noise and smooth over short-sighted fluctuations, as well as to ensure uniformity across the engineered models.
- **Normalization:** Normalized the data using MinMaxScaler, which scaled the values between 0 and 1 to make the LSTM and the PSO-LSTM perform better and maintain stability.
- **Sequence Generation:** Data was organized into sequences of past sales per three consecutive weeks to predict future sales, maintaining temporal order and providing sufficient historical background.
- **Hyperparameter Optimization (PSO-LSTM):** particle swarm optimization (PSO) tuned hyperparameters, including the number of units and sequence length, to improve the number of times the model learned correctly.
- **Feature Selection and Model Inputs:** The Models use the sales sequence over the past periods as the features; the target is the next sales. We also used an additional optimization step to the PSO-LSTM to improve feature selection (PSO-LSTM).

4.3 Models

ARIMA1 Model

An ARIMA (Autoregressive Integrated Moving Average) model is fitted to the training data. We use the auto-arima function to determine the optimal parameters for our best-fitting model. The chosen model is ARIMA (0, 0, 1) with the intercept. We test the model on the test dataset and derive the performance metrics (MAE and RMSE) on the test. The best-fitting model for the forecasting task yields a Mean Absolute Error (MAE) of 15,460 and a Root Mean Square Error

(RMSE) of 21,557, which are informative metrics that indicate the average absolute prediction errors and the standard deviation of the predicted errors, respectively. A 5-fold cross-validation has been made to see how robust our model is. The MAEs were: 13,367, 15,274, 15,407, 16,201, 16,219, and the RMSEs were 17,208, 19,653, 19,572, 21,023, and 21,342, which indicates the stability and consistency of the model when presented with various sub-samples of data.

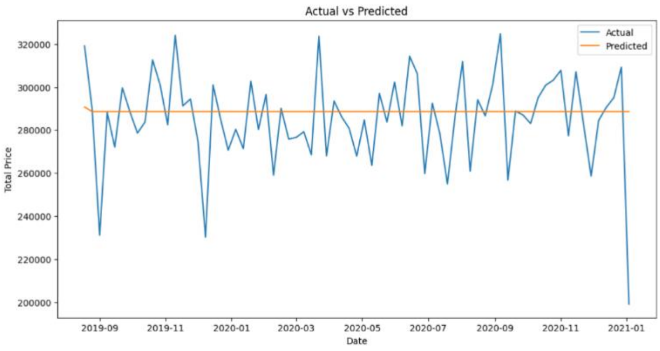

Figure 8: ARIMA1 plot showing the Actual vs Predicted weekly sales

ARIMA2 Model

To determine the optimal values for p and q for the ARIMA model parameters, the ACF and PACF plot for the training data was examined and can be seen in Figure 9. The residuals from the model were checked for stationarity using the Augmented Dickey-Fuller (ADF) test, and it was found that the residuals were stationary. An ARIMA (1,0,1) model was fitted to the training data. The Mean Absolute Error (MAE) was 15,547, and the Root Mean Squared Error (RMSE) was 21,652. The MAE is the average (mean) of the absolute errors, whereas the RMSE is the average of the squared prediction errors. In addition, a 5-fold cross-validation was performed to further verify the adequacy of the model fit. This demonstrates consistency in the performance of the model across the different data subsets of the cross-validation. Additional autocorrelation (ACF) and partial autocorrelation (PACF) plots were created for the residuals to check for any remaining autocorrelation, as shown in Figure 10. The residual plots were examined, and there was not much autocorrelation, which confirmed the adequacy of the models. The fitted ARIMA (1,0,1) was used the model to forecast the test period, and a plot of the predicted and actual values shows the model.

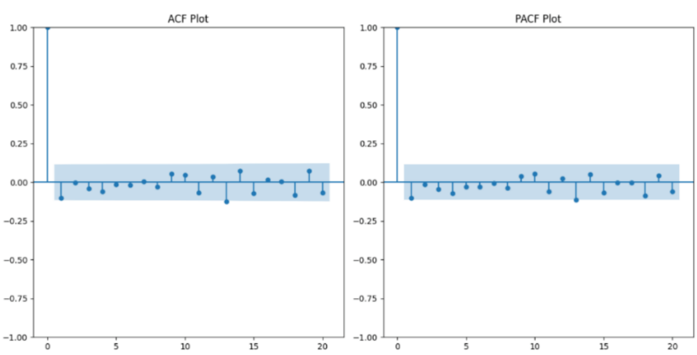

Figure 9: ACF Plot and PACF Plot to determine the p and q

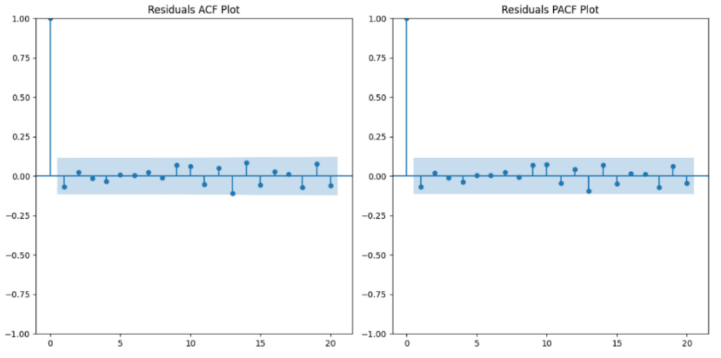

Figure 10: Residuals ACF Plot and Residuals PACF Plot

LSTM

The number of sales each week was normalized with MinMaxScaler to put the data in the range [0,1] to make the LSTM model work better. For each week's sales in the sample, there is a sales order for three weeks. The sequence of 3 weeks can be represented in the LSTM model as input. In the function create$_$sequences, there are X train sequences and X test sequences, where they are sequences in a time series. With a function creating$_$sequences (), sequences were made, and they were split into train and test sets with a ratio of 80 percent and 20 percent respectively maintaining the same time order. This model is a Sequential LSTM built with just one LSTM layer of 50 units followed by a Dense output layer. This model was compiled with the Adam optimizer and the MSE loss function. This model was trained over 100 epochs with an 80-20 split of the training data being used in validation. After which, the model makes predictions on the test set. These predictions are also reverted to the original scale by inverse transforming both predictions and actual values. The final performance of the model on the test set is presented below, along with specific data for both the training and test sets. For this model, the performance was a Mean Absolute Error (MAE) of 15,441 and RMSE of 21,792.

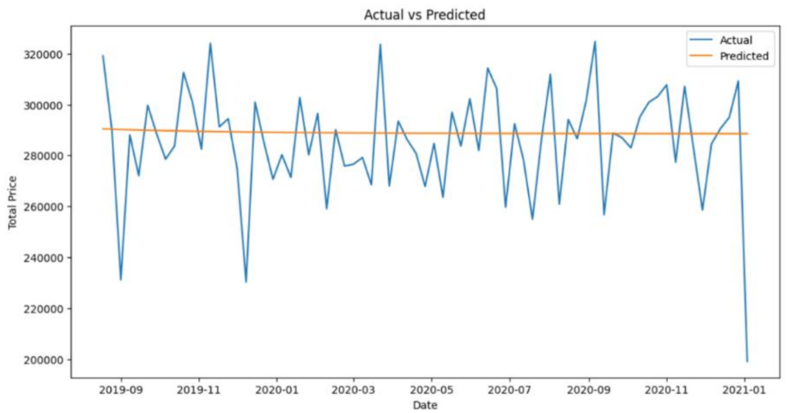

Figure 11: ARIMA2 Plot showing the Actual vs Predicted weekly sales

A 5-fold rolling cross-validation was used to evaluate the robustness of the model. For each split, a LSTM model was built, trained, and tested. The result of each fold was as follows:

Fold 1: MAE = 17025, RMSE = 21193

Fold 2: MAE = 16650, RMSE = 21254

Fold 3: MAE = 13874, RMSE = 17787

Fold 4: MAE = 16305, RMSE = 21070

Fold 5: MAE = 15397, RMSE = 21635.

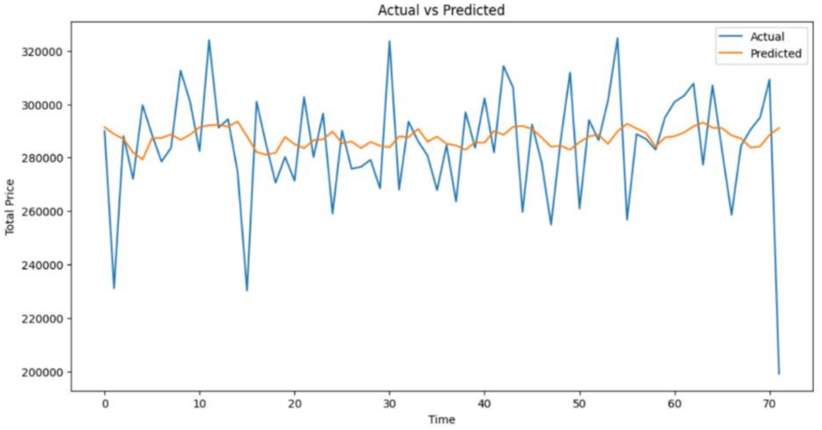

Figure 12: LSTM plot showing the Actual vs Predicted weekly sales

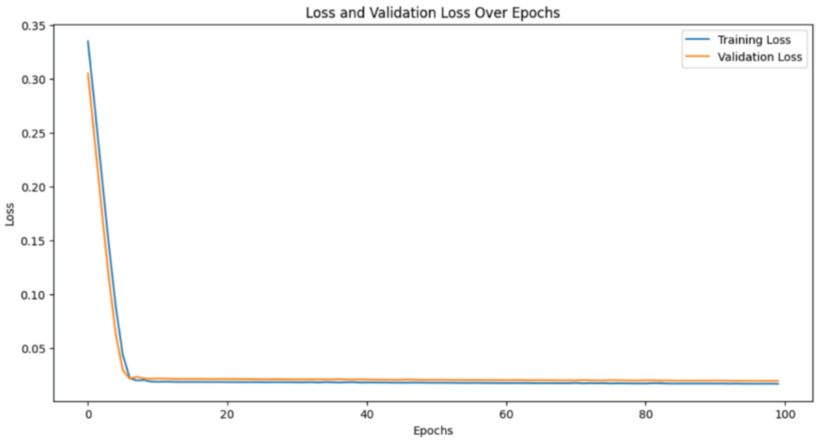

Figure 13: LSTM Plot showing the Training Loss vs Validation Loss

PSO-LSTM Model

The MinMaxScaler was used to normalize the data, scaling it to a range of 0 to 1. This is another important step as it helps the LSTM model work better, as each feature is scaled to the range defined by the MinMaxScaler. We arrange the data for the LSTM model as sequences of fixed length. Here, the input is the sales of the last three weeks, whereas the target is the sales of the following week. The purpose of this sequence length selection was to ensure providing enough historical context while keeping the input size to a tractable level. This sequence of length three was optimized using PSO. The next step involves the hyperparameter tuning of the LSTM model using the PSO implemented for the Regression problem discussed in the methodology section, where a real-world dataset was employed to build and train the final LSTM model. Once the optimized hyperparameters from the PSO algorithm were obtained, the final model was built and trained. The data was split into training and testing sets (80 percent training, 20 percent testing) to preserve the temporal order. The model performance was evaluated on a test set using MAE and

RMSE. The final model had MAE=15,282 and RMSE=21,459 for the test set.

To check robustness, we performed 5-fold rolling cross-validation, keeping the temporal order of data in every fold using the Time Series Split feature in Scikit-learn. In rolling cross-validation on a sliding window, the model is built, trained, and tested on each subset. Performance metrics (MAE and RMSE) are stored in a list, and their cross-validation performance is shown below:

Fold 1: MAE = 16939, RMSE = 21108

Fold 2: MAE = 16167, RMSE = 20737

Fold 3: MAE = 13684, RMSE = 17670

Fold 4: MAE = 17329, RMSE = 21613

Fold 5: MAE = 15332, RMSE = 21462.

This analysis shows that the LSTM model performs well in predicting sales for each week. PSO was used to select the optimal hyperparameters for the model, which was then validated using rolling cross-validation to assess its predictive ability. The model's learning capacity is evident in the visualization.

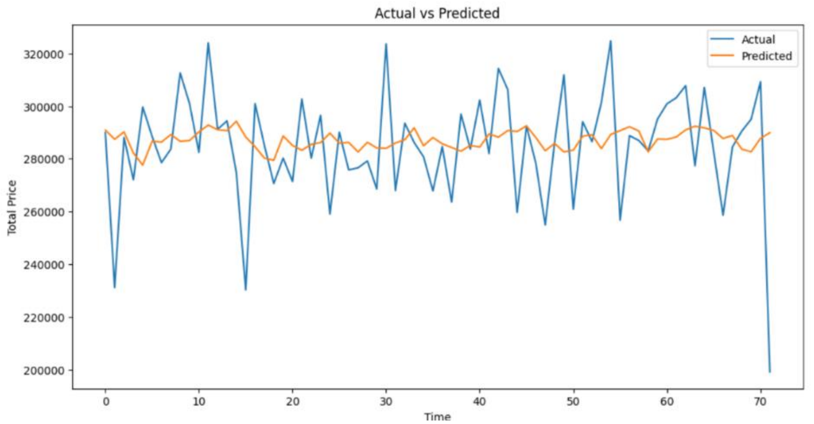

Figure 16: PSO-LSTM plot showing the Actual vs Predicted weekly sales

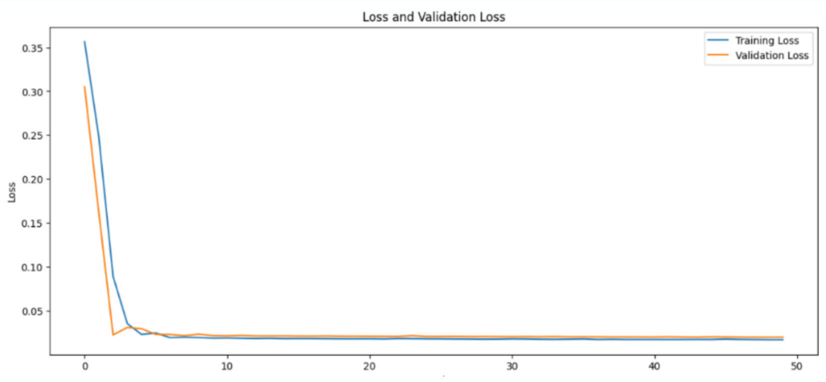

Figure 17: PSO-LSTM Training Loss vs Validation Loss

5 Discussion

Four predictive models are being compared based on their highly accurate performance, stability, complexity, training computation time, inventory management, and strategic planning. The aim is to analyze the models quantitatively for sales forecasting. We have ARIMA1, ARIMA2, LSTM, and PSO-LSTM, which is a hybrid model that combines Particle Swarm Optimisation (PSO) with LSTM.

i. ARIMA1 Model

The ARIMA1 Model obtained by fitting the auto-ARIMA function gave a Mean Absolute Error (MAE) of 15,460, and a Root Mean Square Error (RMSE) of 21,557. Cross-validation with five folds yields MAE = [13,367, 14,772, 16,218, 14,165, 14,381] and RMSE = [17,208, 18,849, 21,342, 18,944, 19,335]. This model can be given as a recommendation for basic tasks. It is a simple and easy-to-implement model that is also easy to interpret.

ii. ARIMA2 Model

The ARIMA2 model – a moving average model – with an MAE of 15,547 and an RMSE of 21,652 has also been considered, and the analyses can be seen as robust, with ACF and PACF plots and stationarity assessments performed via the Augmented Dickey-Fuller (ADF) test, whilst cross-validation shows good performance across multiple subsets of data. The model emphasizes simplicity and is suitable for moderately complex forecasting needs.

iii. LSTM Model

The LSTM model achieved an MAE of 15,441 and an RMSE of 21,792 due to the complete use of its capability with sequential data. The 5-fold rolling cross-validation helps to prove its capability to capture some complicated patterns in the historical sales data. The LSTM model incurs a greater computational cost and requires longer training times. It learns from longer sequences and is therefore capable of handling more complex forecasting tasks.

iv. PSO-LSTM Model

This model added PSO to the LSTM to obtain an MAE of 15,282 and an RMSE of 21,459. Cross-validation results show slightly better performance in statistics. In addition to the above-mentioned, the PSO-LSTM is a tool for sales forecasting. It is more accurate and robust when compared with LSTM. Nevertheless, it is more complex and time-consuming to train than an LSTM. More computational resources are required.

v. Comparative Analysis

The results of four different predictive models, ARIMA1, ARIMA2, LSTM, and PSO-LSTM, show that the last two models, PSO-LSTM and LSTM, are more accurate than ARIMA models. The easy adoption and fast implementation of ARIMA models make them suitable for daily operations and provide quick answers to managers' questions regarding total consumption. On the contrary, the LSTM and PSO-LSTM models are more complex and time-consuming to train, requiring more resources due to their deep training process.

Table 1 Paired Sample T-Test

Model	MAE	RMSE	Cross Validation MAE Range	Cross-Validation RMSE Range	Complexity	Training Time	Key Strength
ARIMA1	15,460	21,557	13,367 - 16,218	17,208 - 21,342	Low	Fast	Simplicity, interpretability
ARIMA2	15,547	21,652	13,367 - 16,197	17,192 -21,325	Low	Fast	Simplicity

LSTM	15,441	21,792	13,874 - 17,025	17,787 - 21,254	High	Moderate	Captures complex patterns
PSO-LSTM	15,2821	21,459	13,683 - 17,329	17,670 - 21,613	Very High	Slow	Optimized accuracy, robust

Despite the intricacy and resource consumption, the PSO-LSTM model outperformed the LSTM model, not only in terms of the flat top of the predicted data but also in its ability to capture more complex patterns within the data, such as a slight increase shortly after the decline. These two models are more suitable for strategic-level forecasting. The comparative analysis is shown in the Table 1.

5.1 Conclusion and Future Work

This study demonstrates the superior performance of more accurate forecasting models, particularly PSO-LSTM, over traditional ARIMA in predicting retail sales. Better prediction accuracy of PSO-LSTM can help store owners save on both oversupply and stockout costs, enabling better inventory management to optimize their capital usage and storage costs. Besides, the sales trend and customer information collected by the model will help stores leverage their targeting to attract customers more effectively and encourage them to make more purchases. Future work will examine the effect of other variables, such as economic factors, social trends, and competitive actions, on forecasting accuracy. To improve the PSO-LSTM model further or experiment with other hybrid models that mainly combine different machine learning algorithms to increase the forecasting performance. To determine if the analysis can be reproduced elsewhere in other retail sectors or different geographic contexts, experts need to apply it to other circumstances. To test out online forecasting models that can respond in real-time to changes in market dynamics and reflect dynamic market changes more rapidly than traditional models.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request. All relevant data generated or analyzed during this study have been included in the article.

Conflict of Interest Statement

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

6 References

Angeletos, G. M., Lian, C., & Wolf, C. K. (2024). Can deficits finance themselves? *Econometrica*, *92*(5), 1351–1390. https://doi.org/10.3982/ECTA21791

Apache Spark. (2025, May 19). *PySpark overview*. https://spark.apache.org/docs/latest/api/python/index.html

Bandara, K., Bergmeir, C., & Smyl, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert Systems with Applications*, *140*, Article e112896. https://doi.org/10.1016/j.eswa.2019.112896

Beaumont, C., Makridakis, S., Wheelwright, S. C., & McGee, V. E. (1984). Forecasting: Methods and applications. *The Journal of the Operational Research Society*, *35*(1), Article e79. https://doi.org/10.2307/2581936

Bilal, A. I., Bititci, U. S., & Fenta, T. G. (2024). Effective supply chain strategies in addressing demand and supply uncertainty: A case study of Ethiopian Pharmaceutical Supply Services. *Pharmacy*, *12*(5), Article e132. https://doi.org/10.3390/pharmacy12050132

Brau, R. I., Sanders, N. R., Aloysius, J., & Williams, D. (2024). Utilizing people, analytics, and AI for decision making in the digitalized retail supply chain. *Journal of Business Logistics*, *45*(1), Article e12355. https://doi.org/10.1111/jbl.12355

Calzone, O. (2022, February 21). *An intuitive explanation of LSTM.* Medium. https://medium.com/@ottaviocalzone/an-intuitive-explanation-of-lstm-a035eb6ab42c

Chen, H., Chiang, H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, *36*(4), 1165–1188. https://doi.org/10.2307/41703503

Cherepovitsyn, A., Mekerova, I., & Nevolin, A. (2025). Analysis of the Palladium Market: A Strategic Aspect of Sustainable Development. *Mining*, *5*(3), Article e39. https://doi.org/10.3390/mining5030039

Cui, R., Lu, Z., Sun, T., & Golden, J. M. (2024). Sooner or later? Promising delivery speed in online retail. *Manufacturing & Service Operations Management*, *26*(1), 233–251. https://doi.org/10.1287/msom.2021.0174

De Almeida, F. M., Martins, A. M., Nunes, M. A., & Bezerra, L. C. T. (2022, June 15–16). *Retail sales forecasting for a Brazilian supermarket chain: An empirical assessment* [Paper presentation]. Proceedings of IEEE 24th Conference on Business Informatics (CBI), Amsterdam, Netherlands. https://doi.org/10.1109/CBI54897.2022.00014

Eberhart, R., & Kennedy, J. (1995, October 4–6). *A new optimizer using particle swarm theory* [Paper presentation]. Proceedings of the Sixth International Symposium on Micro Machine and Human Science, Nagoya, Japan.

El-Adly, M. I., & Eid, R. (2016). An empirical study of the relationship between shopping environment, customer perceived value, satisfaction, and loyalty in the UAE malls context. *Journal of Retailing and Consumer Services*, *31*, 217–227. https://doi.org/10.1016/j.jretconser.2016.04.002

Elsayed, S., Thyssens, D., Rashed, A., Jomaa, H. S., & Schmidt-Thieme, L. (2021). *Do we really need deep learning models for time series forecasting?* Cornell University. https:/doi.org/10.48550/arXiv.2101.02118

Fildes, R., Ma, S., & Kolassa, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*, *38*(4), 1283–1318. https://doi.org/10.1016/j.ijforecast.2019.06.004

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054

George, E. P. B., & Gwilym, M. J. (1970). *Time series analysis: Forecasting and control*. Holden-Day.

Haselbeck, F., Killinger, J., Menrad, K., Hannus, T., & Grimm, D. G. (2022). Machine learning outperforms classical forecasting on horticultural sales predictions. *Machine Learning with Applications*, *7*, Article e100239. https://doi.org/10.1016/j.mlwa.2021.100239

He, Q.-Q., Wu, C., & Si, Y.-W. (2022). LSTM with particle Swam optimization for sales forecasting. *Electronic Commerce Research and Applications*, *51*, Article e101118. https://doi.org/10.1016/j.elerap.2022.101118

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Islam, M. M. (2023). *ECommerce data analysis*. https://www.kaggle.com/datasets/mmohaiminulislam/ecommerce-data-analysis?

Kholod, M., Celani, A., & Ciaramella, G. (2024). The analysis of customers' transactions based on POS and RFID data using big data analytics tools in the retail space of the future. *Applied Sciences*, *14*(24), Article e11567. https://doi.org/10.3390/app142411567

Kitchens, B., Dobolyi, D., Li, J., & Abbasi, A. (2018). Advanced customer analytics: Strategic value through integration of relationship-oriented big data. *Journal of Management Information Systems*, *35*(2), 540–574. https://doi.org/10.1080/07421222.2018.1451957

Kolassa, S. (2022). Commentary on the M5 forecasting competition. *International Journal of Forecasting*, *38*(4), 1562–1568. https://doi.org/10.1016/j.ijforecast.2021.08.006

Kumar, S., & Mahapatra, R. P. (2021). Design of multi-warehouse inventory model for an optimal replenishment policy using a rain optimization algorithm. *Knowledge-Based Systems*, *231*, Article e107406. https://doi.org/10.1016/j.knosys.2021.107406

Kumar, V., & Venkatesan, R. (2021). Transformation of metrics and analytics in retailing: The way forward. *Journal of Retailing*, *97*, 496–506. https://doi.org/10.1016/j.jretai.2021.11.004

Kunz, W. H., & Wirtz, J. (2024). Corporate digital responsibility (CDR) in the age of AI: implications for interactive marketing. *Journal of Research in Interactive Marketing*, *18*(1), 31–37. https://doi.org/10.1108/JRIM-06-2023-0176

Lens, V., Bushman-Copp, L., & Wimer, C. (2024). Fostering Financial and Family Well-Being: A Qualitative Study on How Parents Utilized the Expanded Child Tax Credit. *Journal of Poverty*. Advance online publication. https://doi.org/10.1080/10875549.2024.2379761

Liu, S., Liao, G., & Ding, Y. (2018, 31 May– 2 June). *Stock transaction prediction modeling and analysis based on LSTM* [Paper presentation]. Proceedings of 13th IEEE Conference on Industrial Electronics and Applications, Wuhan, China. https://doi.org/10.1109/ICIEA.2018.8398183

Macas, C. V. M., Aguirre, J. A. E., Arcentales-Carrión, R., & Peña, M. (2021, March 23–25). *Inventory management for retail companies: A literature review and current trends* [Paper presentation]. Proceedings of Second International Conference on Information Systems and Software Technologies, Quito, Ecuador.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLOS ONE*, *13*(3), Article e0194889. https://doi.org/10.1371/journal.pone.0194889

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). The M5 competition: Background, organization, and implementation. *International Journal of Forecasting*, *38*(4), 1325–1336. https://doi.org/10.1016/j.ijforecast.2021.07.007

Mustapha, O. O., & Sithole, T. (2025). Forecasting Retail Sales using Machine Learning Models. *American Journal of Statistics and Actuarial Sciences*, *6*(1), 35–67. http://dx.doi.org/10.47672/ajsas.2679

Nordin, F., & Ravald, A. (2023). The making of marketing decisions in modern marketing environments. *Journal of Business Research*, *162*, Article e113872. https://doi.org/10.1016/j.jbusres.2023.113872

Ramos, P., Santos, N., & Rebelo, R. (2015). Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and Computer-Integrated Manufacturing*, *34*, 151–163. https://doi.org/10.1016/j.rcim.2014.12.015

Rana, M. E., Yao, T. W., & Hameed, V. A. (2023). *Optimizing sales forecasting: Implementing a deep learning model for accurate and timely predictions* [Paper presentation]. Proceedings of 6th International Conference on Big Data and Artificial Intelligence, Jiaxing, China. https://doi.org/10.1109/BDAI59165.2023.10256821

Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning : A systematic literature review: 2005–2019. *Applied Soft Computing*, *90*, Article e106181. https://doi.org/10.1016/j.asoc.2020.106181

Shao, B., Li, M., Zhao, Y., & Bian, G. (2019). Nickel price forecast based on the LSTM neural network optimized by the improved PSO Algorithm. *Mathematical Problems in Engineering*, *2019*(1), Article e1934796. https://doi.org/10.1155/2019/1934796

Shultz, C., Hoek, J., Lee, L., Leong, W. Y., Srinivasan, R., Viswanathan, M., & Wertenbroch, K. (2022). JPP&M's global perspective and impact: An agenda for research on marketing and public policy. *Journal of Public Policy & Marketing*, *41*(1), 34–50. https://doi.org/10.1177/07439156211049216

Soori, M., Arezoo, B., & Dastres, R. (2023). Artificial intelligence, machine learning and deep learning in advanced robotics, a review. *Cognitive Robotics*, *3*, 54–70. https://doi.org/10.1016/j.cogr.2023.04.001

Stanelytė, G. (2021). *Inventory optimization in retail network by creating a demand prediction model* [Doctoral dissertation, Vilniaus Gedimino Technikos Universitetas]. eLABa. https://gs.elaba.lt/object/elaba:80509157/index.html

Wang, F., & Aviles, J. (2023). Enhancing operational efficiency: Integrating machine learning predictive capabilities in business intellgence for informed decision-making. *Frontiers in Business, Economics and Management*, *9*(1), 282–286.

Wang, J., Liu, G. Q., & Liu, L. (2019, March 15–18). *A selection of advanced technologies for demand forecasting in the retail industry* [Paper presentation]. Proceedings of 4th International Conference on Big Data Analytics, Suzhou, China. https://doi.org/10.1109/ICBDA.2019.8713196

Yan, K., & Deng, D. (2024). Research on supermarket replenishment volume prediction based on LSTM and PSO algorithm. *Journal of Education, Humanities and Social Sciences*, *25*, 129–137. https://doi.org/10.54097/e8w01s21

Yan, S., Archibald, T. W., Han, X., & Bian, Y. (2022). Whether to adopt "buy online and return to store" strategy in a competitive market? *European Journal of Operational Research*, *301*(3), 974–986. https://doi.org/10.1016/j.ejor.2021.11.040

Yu, H. (2023). *Pop-up retail strategies in an omnichannel context*.

Zeroual, A., Harrou, F., Dairi, A., & Sun, Y. (2020). Deep learning methods for forecasting COVID-19 time-Series data: A comparative study. *Chaos, Solitons & Fractals*, *140*, Article e110121. https://doi.org/10.1016/j.chaos.2020.110121