Walmart Sales forecasting using Equilibrium optimized Deep LSTM

MSc Research Project
MSc in Data Analytics

Arul Selvam Perumal
Student ID:x21128847

School of Computing
National College of Ireland

Supervisor: Prof. Athanasios Staikopoulos

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Arul Selvam Perumal
Student ID:	x21128847
Programme:	MSc in Data Analytics
Year:	2022-2023
Module:	Research Project
Supervisor:	Prof. Athanasios Staikopoulos
Submission Due Date:	14th August 2023
Project Title:	Walmart Sales forecasting using Equilibrium optimized Deep LSTM
Word Count:	7303
Page Count:	27

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Arul Selvam Perumal
Date:	14th August 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Walmart Sales forecasting using Equilibrium optimized Deep LSTM

Arul Selvam Perumal
x21128847

Abstract

For sales prediction, timely prediction presents accurate information for companies when business trends are constantly developing in order to attain a strong balance between demand and supply. Sales forecasting is formulated as a time series forecasting issue that aspires to forecast the volume of future sales for diverse products with the observation of several significant parameters for instance, season, discount, brand, and so on., corresponding to historical sales records. Moreover, in order to carry out accurate sales forecasting, this research work intends to perform Walmart sales prediction using Equilibrium Optimizer (EO)-deep Long Short Term Memory (deep LSTM), called the EO-deep LSTM model. Initially, the input data is pre-processed using missing data imputation and data augmentation is done using min-max normalization. Then, technical indicators, namely Simple Moving Average (SMA), Volume Adjusted Moving Average (VAMA), Average Directional Movement Index (ADX), Weighted Moving Average (WMA), Trend Detection Index (TDI), and Exponential Moving Average (EMA) are extracted. With these indicators as features, the prediction is done using the Deep LSTM model, which is trained using an EO algorithm. The experimentation is carried out using Walmart sales forecasting dataset and the performance of the proposed approach is evaluated by parameters, like Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

Keywords: sales prediction, Walmart, pre-processing, data augmentation, deep LSTM

1 Introduction

In the last decade, online shopping has received remarkable fame worldwide. Shopping apps for mobile phones, a surge of minimal expensive smartphones, simple access to intent, a raise in youth awareness in smaller cities and rural areas, and also a rise in wealth are important factors in the increasing market share of online retail. Consumers favor online shopping for expediency, home delivery, and time-saving. Currently, numerous businesses experience the requirement for an online existence to raise their coverage and evade the costs related to setting up physical retail stores. E-commerce websites, such as Taobao, Amazon, eBay, Flipkart, and so on, get pleasure from a huge consumer base. Nonetheless, this domain is success more and more unpredictable and competitive, and consumer prospects concerning quality and price have risen severely. In manufacturing, a business that can no longer on cost benefits is also required to concentrate on having

ML	Machine Learning
ANN	Artificial Neural Network
DAGNN	Directed Acyclic Graph Neural Network
RNN	recurrent neural network
MSE	Mean Squared Error
LSTM	long short-term memory
LR	Linear regression
GRU	Gated Recurrent Unit
TCN	Temporal Convolutional Network
DT	Decision Tree
DNN	Deep Neural Network
RF	Random Forest
DL	Deep Learning
SVR	Support Vector Regression

effectual supply chain management with an enhanced comprehension of customer behavior and demands. Precise sales prediction is a significant segment of effectual supply chain management, as it assists to evade the underproduction as well as overproduction of products. By gaining a deeper understanding of customer demand, companies can mitigate risks like the Bullwhip effect, while greatly enhancing the effectiveness and efficacy of their supply chain.

In the 21st century, the continuous adoption and advancement of technology have led to a remarkable explosion in data generation. Retail giants, such as Walmart perceive this data as their most valuable asset, as it assists them to forecast future sales, understanding customer behavior, strategizing profit generation, and maintaining competitiveness over other firms. Walmart is an American multinational retail corporation, which operates a vast network of approximately 11,000 stores across more than 27 countries, using over 2.2 million associates. It is the world's largest company regarding revenue and offers a diverse range of products including home furnishings, electronics, groceries, body care products, clothing, and so on. By presenting customers the commitment to everyday low prices, Walmart's wide range of products generates annual revenue of nearly 500 billion dollars. This emphasizes the maximum importance for the company to employ comprehensive approaches in order to predict future sales and following profits. The data generated by Walmart plays an important role in predicting customer buying patterns, promotional strategies, future sales, and developing modern in-store technologies. The utilization of up-to-date technology techniques is essential for an organization to endure in the current cutting-edge global market and make services and products, which differentiate them from its competitors. Throughout the year, Walmart runs various promotional markdown sales following significant holidays in the United States. Analyzing the impact of these promotions on weekly sales is vital for Walmart to allocate resources effectively toward these strategic initiatives. Additionally, understanding user requirements and buying patterns is crucial for Walmart to enhance customer retention and increase demand, ultimately leading to higher profits.

Nowadays, a real-time decision is made by the customers utilizing the information present online in the e-commerce surroundings, and it is currently probable to forecast product sales and customer demands by evaluating the data available online. Maintaining a balance between demand and supply is important for retailers for several retail businesses and the precise sales prediction volume is attractive and crucial for commercial achievement. Over-estimated sales can affect extreme inventory, harmful cash flow, and insolvency, whereas underestimated sales might guide to displeased orders, and minimized business status, and income. Actually, sales prediction is devised as a time series forecast issue that aspires to forecast the volume of future sales on the basis of seen multivariate time series data that comprises a volume of historical sales as well as significant factors (for instance discount, season, brand, and so on). Therefore, logical modeling of the influential factors and historical sales volume must be carried out to effectively forecast sales volumeChen et al. (2020).

Sales forecasting is utilized to forecast the accessibility of several commodities presented at several stores in several cities within a Big Mart firm. Moreover, sales forecasting permits the e-commerce platform, in order to have a reliable prediction and enhanced accuracy that will assist them with competitive prices, inventory planning, and timely promotional schemes. Further, the Walmart sales prediction permits us to comprehend the Walmart platform lifecycle as its sales, stability, and development and also states how the sales can be affected by the short-range product objectives, like pricing, promotion, ranking online, and season. As the product volume and outlets expand significantly, manually predicting them becomes progressively challenging. Regarding time, geography, money, and time, forecasting the appropriate attention for an object is an enormous sensation for dealers. As quickly as feasible, due to the capacity and cash restraints, merchants might be restricted in time or there is an urgent requirement to sell their merchandise. Thus, the product appeal is ascertained by a variety of measures, like popularity, time, price, outlet position, outlet type, etc Thivakaran and Ramesh (2022). Even more challenging, promotional prediction is a complicated task that is compounded by the absence of historical sales data in the cold start scenario, where immediate action is crucial. Various researchers demonstrate that sales of conventional products are predicted by utilizing statistical approaches. The most important confront of new products is the restricted historical data availability that restricts the use of conventional forecasting approaches. The leading prediction techniques for new products are thus market research, managerial opinions, and sales force input. Even though the aforesaid approaches might be feasible to predict new product sales, quantitative approaches can considerably do better than these approaches based on judgmentElalem et al. (2022). To tackle the challenge of sales prediction, several methodologies and approaches can be applied. Traditional statistical models like time series forecasting, namely exponential smoothing or ARIMA, have been extensively employed to capture trends, seasonality, and patterns in sales data.

In order to overcome the challenges present in sales prediction, ML algorithms have been utilized, particularly in fields, like energy prediction, recommender systems, and online retailers Chen et al. (2020) Aher et al. (2021). The rise of ML systems has led to an increase in automated decisions that presents certain applications where it becomes challenging in trusting the decisions. In order to characterize a technique, prediction alone and measures computed in the predictions are not adequate Haselbeck et al. (2022). As expected, an applicant in the supply chain of a retailer may feel uncomfortable demanding a new product when there is limited information available regarding it. This circumstance becomes more complex when a decision regarding a number of units is made by an automated ML system that operates as black-box techniques without presenting any supporting feedback. To address this issue of mistrust, one potential solution is to integrate additional information that sheds how the decision was made. Ideally, a few levels of cooperation must exist among the ML system and humans wherein the user can control and enhance the outcomes.

Recently, time series prediction approaches are extensively used in numerous fields, like recommender systems, financial market prediction, and medical research. Besides these approaches, the detection of trending events or recurring patterns on the basis of clues from historical observation has gained interest in several applications, like solar intensity prediction, traffic modeling, and argument detection. Indeed, the identification of repeating trends plays an important role in sales prediction. This was done by aligning relative contextual information learned from the significant parameters, and this is called trend alignment. Nonetheless, for the trend alignment in sales forecast, both conventional auto-regressive-based techniques and current trend mining approaches are inefficient. While traditional statistical models have long been used for sales prediction, recent developments in data analytics and ML approaches have opened up new avenues for more accurate and sophisticated prediction techniques. In addition to well-established methods, several alternative approaches have emerged, offering unique perspectives and potential improvements in sales prediction. These techniques rely on the assumption that trends in time series data repeat periodic manner with a fixed time period. Therefore, for every application area domain knowledge is required and cautiously selected parameters on the basis of the data. Therefore, conventional approaches face limitations in aligning the same trends in sales time series, in which the sales patterns are irregular and delicate because of the complex real-world circumstance. This issue becomes even more complex when there are a huge number of diverse products.

The emergence of development in sales time series has precise contexts that can be designed by analyzing the interactions among several significant factors. Extensive research and application have focused on RNN techniques for learning vector representations from sequential inputs in the context of raw time series data. The RNN exhibit its benefits in flexible and discriminative modeling of a non-linear relationship unlike existing methods, such as kernel techniques and Gaussian process that are restricted by their predefined non-linear form. In addition, the performance of RNNs has been considerably developed by two variations, like GRU and LSTM. These variants have exhibited remarkable improvements in tasks associated with image captioning and neural machine translation. Within these applications, the encoder-decoder RNN framework utilizes two separate RNNs in order to encode the sequential inputs to latent contextual vectors. Then, decode these encoded contexts and generates the required interpretations. Subsequent to exhibiting its advantages in current time series modeling tasks, it is logical to consider encoder-decoder RNNs for sales forecasts by utilizing its ability to effectively capture the non-linear association amid the sales volume and influential factors. Nonetheless, even with the advancements in existing encoder-decoder RNN approaches, sales forecast remains a serious issue. This complexity arises from the interaction of multiple influential factors, which have diverse influences on diverse products. Figure 1 depicts the Walmart sales forecasting carried out by utilizing ML/DL approach. Here, the major processes, like data understanding, data preparation, modeling, evaluation, and deployment of the model in the business are performed.

1.1 Problem Statement

The main limitation in previous studies is the absence of sales history data for analysis. Also, there are a limited number of Walmart store details for analysis. Due to its limited

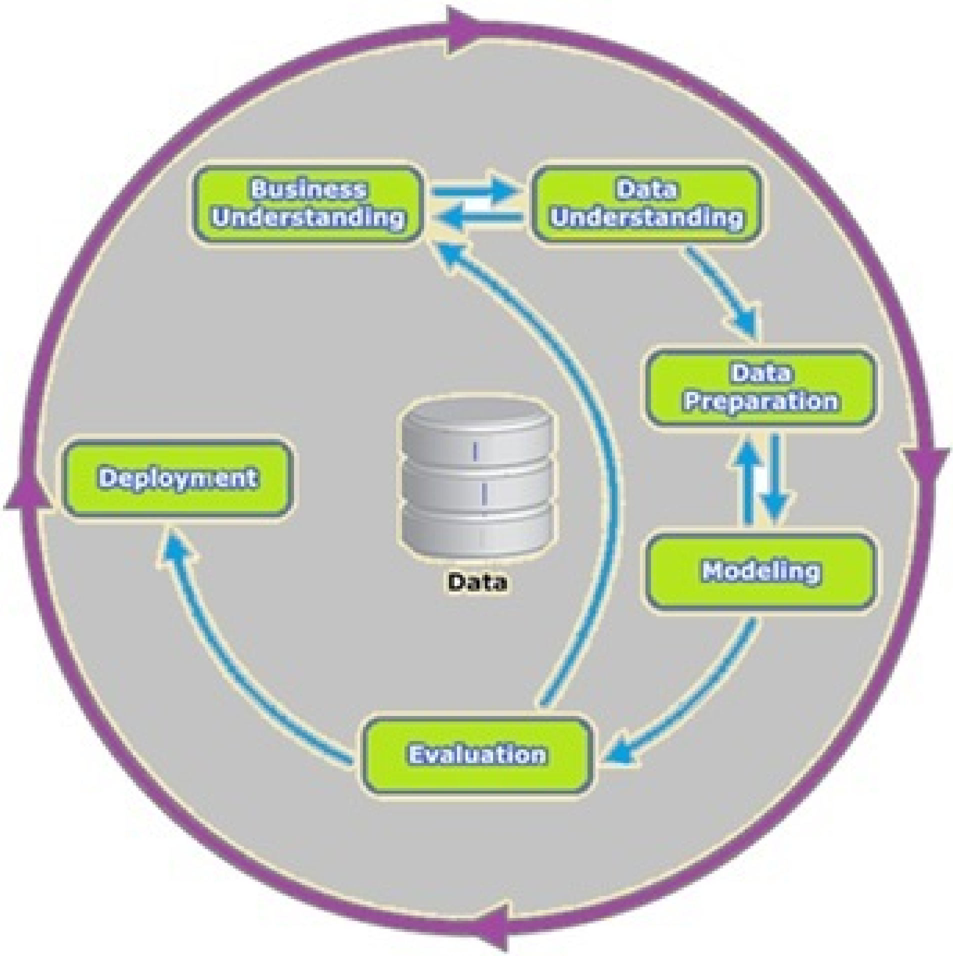

Figure 1: Flow model of Walmart sales prediction

past history data details, the techniques cannot be trained efficiently to provide accurate predictions and outcomes. Due to the absence of data availability, it is very difficult for training and tuning the approaches as an over-constrained approach may minimize the accuracy of the approach. A suitable number of training is needed to effectively train the approach and illustrate helpful insights. Moreover, approaches created have been designed on the basis of particular preset business conditions and assumptions. In addition, it is very difficult to forecast how customer purchasing behavior changes over the last decade or how the policies established by the management may influence the revenue of the company Loureiro et al. (2018). These parameters have a direct effect on Walmart's sales and it is required to constantly examine the developments in the market and evaluate them with conventional performance to develop policies and approaches to enhance the profitability.

1.2 Contribution

The major aim of the proposed work of sales prediction is stated below:

- To develop a novel deep learning approach, called Equilibrium Optimized-Deep LSTM (EO-Deep LSTM), that uses an optimization algorithm, Equilibrium Optimizer, in order to train the Deep LSTM such that the parameters are tuned optimally and thus, improve the performance of sales prediction.

- To employ various technical indicators to the proposed deep learning-based prediction model so that it can accurately predict Walmart sales.

1.3 Research Questions

- How can DL approaches be effectively used to reduce prediction errors and enhance the reliability of Walmart sales prediction models?

- What are the advantages of using DL methods over conventional methods for Walmart sales prediction?

- Does training the DL approach using an optimization algorithm improve the performance of the prediction?

- What are performance evaluation measures exploited to determine the effectiveness of the prediction approach?

1.4 Chapter Overview

The organization of this work is stated as follows:
Chapter 1: Introduction:- This chapter presents an overview of sales prediction and also includes the importance of ML and DL methods employed for sales prediction. Additionally, this chapter highlights the research's contribution regarding Walmart sales prediction and also outlines the research questions.
Chapter 2: Literature Survey:- This chapter presents a comprehensive review of the conventional sales prediction methods, which are categorized based on the methods employed to forecast the sales demand.
Chapter 3: Proposed method for Walmart sales forecasting:- In this chapter,

the proposed EO-Deep LSTM algorithm for Walmart sales prediction is clearly explained. Each section of this chapter focuses on different phases of the methodology and provides a comprehensive understanding of the workflow involved in the Walmart sales prediction process.

Chapter 4: Result Analysis This chapter delineates and explains the outcomes of the proposed model. Here, several performance evaluation metrics based on error are utilized to assess the effectiveness of the model. Also, this chapter includes essential graphs and analyses to illustrate the findings and provide insights into the performance of the proposed model.

Chapter 5: Conclusion and future studies: In this section, a summary of the study is provided and also emphasizes the key findings and contributions. The chapter also outlines the scope for future research and also suggests the potential areas of research that can further enhance the field of Walmart sales prediction.

2 Chapter

2.1 Introduction

Sales prediction is a critical business activity, and its accuracy has a significant influence on the market performance of any firm. Sales prediction is a crucial function in any retail business, and Walmart, as a retail giant, can leverage DL and ML techniques to improve sales forecasting accuracy and, finally, the company's bottom line. The purpose of this chapter was to search the existing approaches applied to predict Walmarts sales.

The retail industry has faced increasing pressure to provide accurate sales forecasts as they influence many significant business decisions. Accurate predictions are necessary for retail businesses since sales forecasts determine inventory management, staffing, and pricing decisions. The trend of utilizing advanced methods, such as DL and ML, for sales prediction in various industries, has been rising. DL and ML are approaches extensively utilized in sales forecasting as these approaches can provide robust and accurate forecasts that lead to better decision-making. Previous studies have shown the efficiency of ML and DL approaches in predicting sales in industries, such as fast food, fashion, and online apparel. The advantages of using DL and ML approaches include their capability to handle complex and nonlinear relationships between variables and analyze different types of data, including structured and unstructured data. In addition, these approaches present highly accurate forecasts than conventional linear models. The previous studies exhibited that a neural network with 10 hidden nodes was the most accurate model's optimal number, as it generated the utmost accuracy rate in sales prediction, in comparison with other approaches. This finding implies that by utilizing DL approaches, decision-makers in retail, namely Walmart, can choose a good sales prediction model according to the optimal number of hidden nodes for increased prediction accuracy. The primary advantage of using DL is the ability to handle complex and non-linear relationships between variables. DL designs the data with layers of abstraction that allow for complex mathematical transformations, thus providing superior forecasting accuracy. In addition, DL networks are flexible in modeling different data types, from structured to unstructured, making it easier to incorporate data from diverse sources.

Previous studies have shown the effectiveness of DL and ML in sales forecasting, and their benefits include their capability to handle complex, non-linear relationships between variables, flexibility in modeling diverse data types, combining enormous amounts of data from multiple sources, and the possibility of producing superior forecasting accuracy. By deploying DL and ML techniques, Walmart can enhance the accuracy of sales prediction and make more informed business decisions. In conclusion, using DL or ML approaches for sales prediction presents various advantages in capturing complex relationships and analyzing different types of data.

2.2 Specification of existing studies in sales prediction

Over the past few years, machine learning algorithms have been applied to variety of industries for predictive analytics including finance, healthcare, marketing, and sales. The use of machine learning in sales prediction has been quite popular in the business world, mainly due to its ability to identify patterns and make predictions based on large data sets. The existing studies have focused on exploring the potential of machine learning techniques to improve sales prediction accuracy. Studies have shown that these algorithms provide better results than traditional statistical models when used to predict sales. Sales prediction using machine learning has been an area of interest in recent years. The studies mentioned above have highlighted the potential of machine learning techniques for improving sales prediction accuracy. The use of these algorithms can provide better insights into customer behavior, preferences, and buying trends to predict future sales and inform business decisions.

Sales prediction is critical for forecasting the availability of various commodities available in Big Mart Company across different locations. As the number of products and outlets multiply exponentially, predicting sales manually becomes progressively challenging. Correctly predicting consumer interest in a product remains a huge challenge for merchants due to constraints in geography, time, and finances. Due to time constraints and capacity and budgetary limitations, merchants may feel compelled to sell their merchandise quickly. Hence, a product's appeal is influenced by numerous factors, including price, popularity, time, type of outlet, location, and others. Existing studies have explored various techniques, including traditional statistical methods and DL. Most have investigated the use of historical sales data to predict future sales, while few have examined the impact of economic factors or other external variables.

2.3 Related works

This section describes the conventional methods used for sales prediction in existing studies. Here, the methods reviewed are classified as a convolutional network-based model Yin and Tao (2021) Huang et al. (2022), Gradient boosting model Singh et al. (2020) Haselbeck et al. (2022) Aguilar-Palacios et al. (2020), LSTM-based model Yin et al. (2020) Sajawal et al. (2022), ANN-based model Thivakaran and Ramesh (2022) Sharma et al. (2019), and RF-based model Jeswani (2021) Aher et al. (2021).

a) Convolutional network-based prediction models:

For predicting time series, the CNN is an appropriate model since it presents widened convolutions, in that filters can be utilized to calculate dilations among cells. The space range among every cell permits the neural network to comprehend advance the associations among the diverse clarifications in the time-series. Figure 2 demonstrates the architecture of convolutional network model.

The study by Yin and Tao (2021), developed a sales prediction model for online products by considering various influencing factors and leveraging the benefits of deep

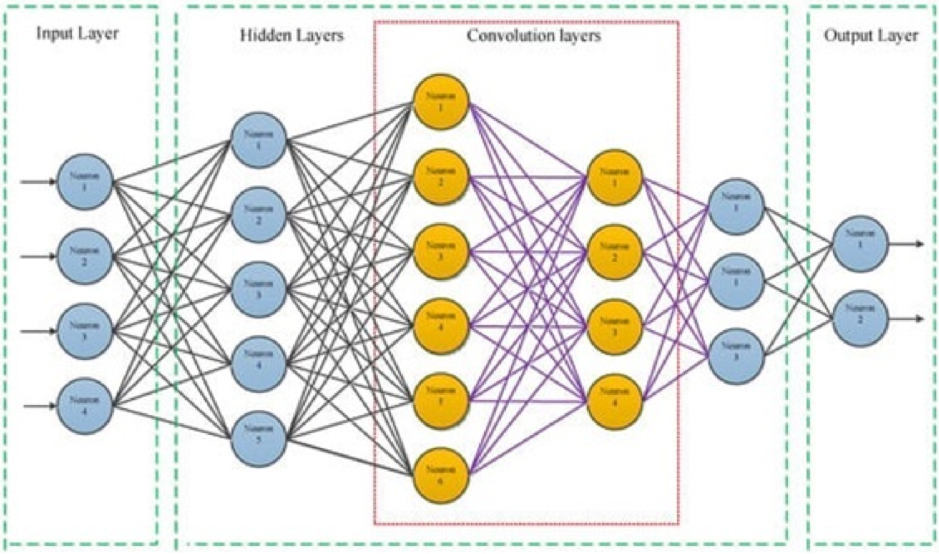

Figure 2: Convolutional network model (source used from: Abdolrasol et al. (2021))

learning algorithms. The study evaluated the model's adaptability across different types of online products. The researchers compared the performance of a fully connected model with a CNN model and found that the latter had superior accuracy and generalization ability. They also demonstrated that the CNN model outperformed non-deep learning models across different product categories. Furthermore, the study determined that an unsupervised pre-trained CNN model was the most effective and adaptable approach for sales forecasting. Huang et al. (2022) developed a robust e-commerce sales forecasting technique that considered multiple correlated features. The model operates in three layers: In the initial layer, the TCN extracts profound temporal patterns from historical sales data, preserving the integrity of temporal information. Moving to the second layer, a reinforcement learning-based feature selection technique is employed to filter the relevant correlation feature set. These selected features are then combined with the processed temporal characteristics, enhancing the informative input while avoiding high-dimensional feature spaces. In the third layer, the reformer model comprehensively learns all features and assigns varying degrees of attention to different importance levels, resulting in a reliable and consistent sales forecast. The model's performance was evaluated through experiments, comparing it with advanced sales forecasting models. The results unequivocally demonstrated that the model exhibits remarkable accuracy and stability.

b) Gradient boosting-based prediction models

The gradient-boosted trees approach is an ensemble learning model, which integrates a huge number of decision trees to generate the ultimate prediction. The "tree" part of gradient-boosted trees indicates the detail that the final approach is a forest of decision trees. Figure 3 exhibits the gradient boosting model.

A few of the gradient boosting-based approaches presented in the conventional studies regarding sales prediction are stated below:

Singh et al. (2020) aimed to research building machine learning algorithms for forecasting sales on an e-commerce platform. In conducting their study, they examined the literature reviews related to similar studies and systems in order to identify good machine-learning models for their project. The aim of the work was to understand which machine

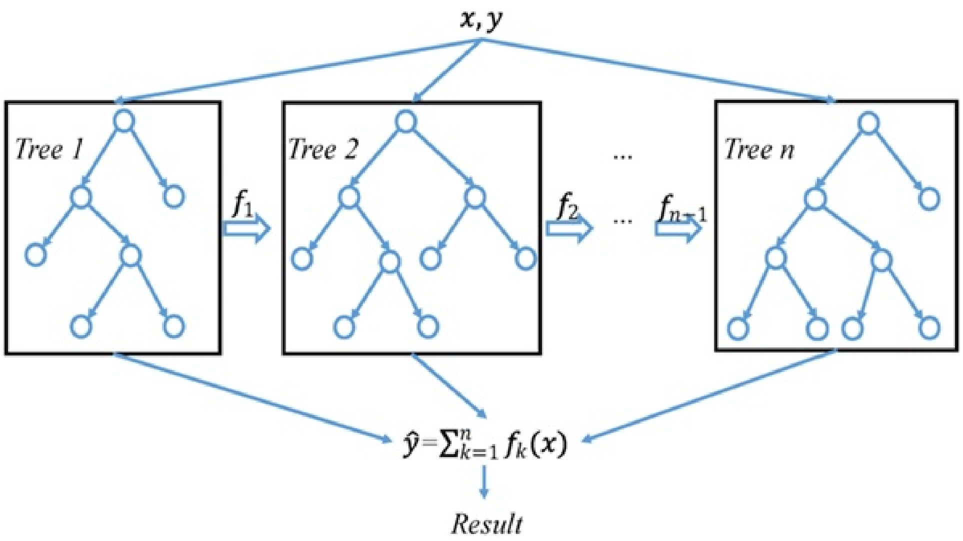

Figure 3: Gradient boosting model (source used from: Wang et al. (2019))

learning models were utilized in previous studies, which would aid in selecting the best models for the current project. Once the models were selected, the researchers built them and tested their accuracy, performance, and error. Finally, they compared the accuracy and errors of all the models to determine the best one that had a high level of accuracy and low error for forecasting sales. The model that fulfilled their criteria was integrated into the system that they had built, enabling it to present a view of the forecasted and current sales. The study by Florian Haselbeck et al., Haselbeck et al. (2022) evaluated the performance of nine latest machine learning and three traditional forecasting algorithms in predicting sales of horticulture. The authors found that machine learning techniques outperformed classical forecasting approaches in all research, with the gradient-boosted ensemble learner XGBoost emerging as the top-performing model in 14 out of 15 assessments. The authors also discovered that the advantage of machine learning over classical forecasting methods improved in datasets with multiple seasons. Moreover, the inclusion of supplementary external factors like weather and holiday information, along with meta-features, resulted in enhanced predictive performance. Furthermore, the study evaluated the algorithms' capability to predict the abrupt surge in demand for horticultural products during the SARS-CoV-2 pandemic in 2020. Among the algorithms considered, XGBoost exhibited superior performance in this specific scenario.

Aguilar-Palacios et al. (2020) developed a method for demand forecasting that integrated the predictive power of Gradient Boosted DT regression with interpretability of different clarifications. The method created implicit contrastive explanations by shaping the data, which leads to cold-start predictions relative to observed promotional sales of related products (i.e., "neighbors"). The selection of these neighbors is determined by their proximity to the predicted promotion that was determined through the calculation of variable significance during the training of the regressor. By leveraging this data, a proficient reviewer can modify the cold-start forecast by regulating the neighbors' contribution. To verify the outcomes, the methodology was tested on both surrogate and actual market data. The proxy model testing revealed that this technique precisely detects the attributes that impact sales and hand-picks the nearest neighbors to generate a comparative explanation. The outcomes from actual market data also suggest that this strategy was carried out comparably to popular methods, like NGBoost or AutoGluon, and CatBoost, while additionally delivering an advantage of interpretability.

c) LSTM-based prediction models

LSTM-based sales prediction models are a type of RNN to predict future sales. These approaches have gained popularity because of their ability to capture and model sequential patterns in time series data. LSTMs can be utilized to design univariate time series prediction issues. These are issues that consist of a single series of observations and an approach is needed to learn from the series of precedent observations to forecast the subsequent value in the series. Figure 4 exhibits the structure of the LSTM model.

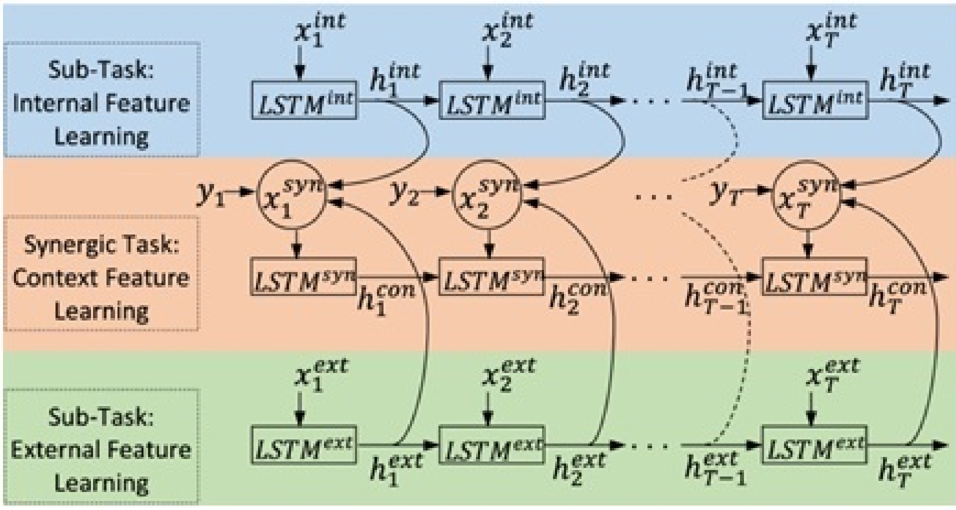

Figure 4: LSTM model (source used from: Chen et al. (2020))

A few of the LSTM-based approaches that are presented in the conventional studies regarding sales prediction are stated as below:

Yin et al. (2020) addressed the issue of low accuracy in forecasting latest product demand arising from the lack of historical data and incomplete contemplation of influencing factors. They presented a hybrid model for the latest product demand forecasting that combined clustering with deep learning. The researchers employed fuzzy clustering-rough set method to determine the weight of similarity in product attributes, according to the measurement of similarity in a product. This provided a foundation for obtaining and organizing historical sales information about similar products, as well as the identification of product similarity. The prediction error of the Bass model was modified using an LSTM neural network model that considers similarity, in which influencing factors, such as seasonality, sales time, and product differentiation were embedded to enhance demand forecasting accuracy. In their study, Sajawal et al. (2022) applied regression (LR, Gradient Boosting Regression RF Regression,) and time series (ARIMA LSTM) models to forecast sales and conducted extensive predictive study and estimation. The dataset utilized was attained from Citadel POS (Point Of Sale), a cloud-based application that facilitates retail stores in carrying out transactions, tender data locally, managing sales, and viewing reports, vendors, customers, and managing inventories, spanning from 2013 to 2018.

d) Random forest-based prediction models

Random forest-based prediction models are widely used in sales prediction tasks due to their effectiveness in handling complex, high-dimensional datasets. Generally, the

Random Forest model is called an ensemble learning model, which forms a set of decision trees and integrates them to make a further precise and constant prediction. Also, it is referred to as an influential approach for regression tasks and classification, and also sales forecasting. Figure 5 demonstrates the architectural model of the random forest.

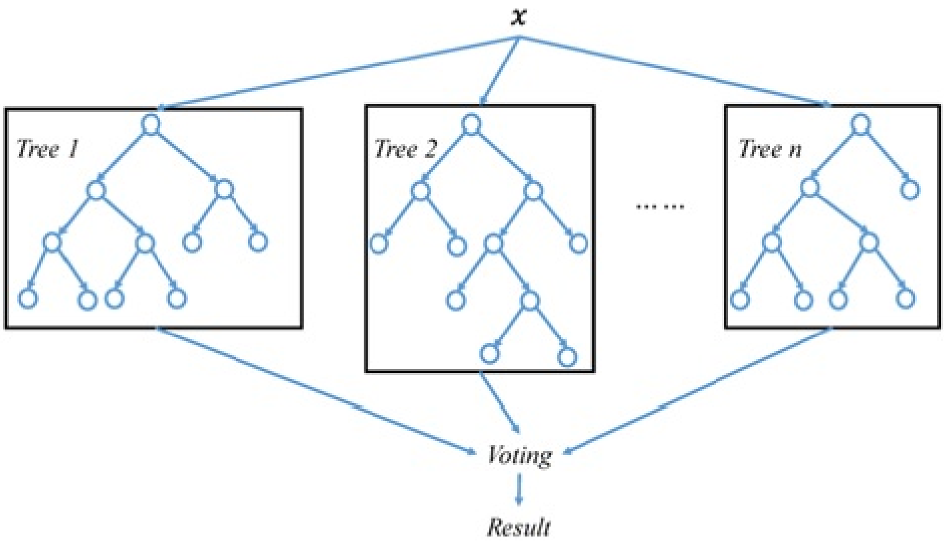

Figure 5: Random forest model (source used from: Wang et al. (2019))

Some of the random forest-based approaches that are presented in the state-of-the-art studies regarding sales prediction are stated below:

In the study, Jeswani (2021) evaluated the performance of Walmart store subsets and projected the future weekly sales of these stores using various models, including random forest, gradient boosting, and linear and lasso regression. A comprehensive exploratory data analysis was performed on the dataset to assess the influence of various factors, including temperature, holidays, and fuel prices, on the weekly sales of Walmart. This analysis aimed to gain insights into the relationships and potential impact of these factors on sales patterns. Moreover, Jeswani has developed a dashboard in Power BI, highlighting the projected sales for each store and department, providing a summary of the whole predicted sales. Aher et al. (2021) analyzed data to establish the basis for providing discounts on various product items. To analyze and predict sales, the authors utilized three models on the Black Friday Sales Dataset, which is presented on Kaggle. The models used in the study included RF Regressor, ridge regression, DT Regressor, lasso regression, and LR. The performance of the model was evaluated using MSE. The outcomes demonstrated that the Random Forest Regressor outperformed the other models and obtained the minimal MSE score.

e) ANN-based prediction models

Generally, ANN is an approach that is utilized for prediction owing to the abilities of machine learning Abdolrasol et al. (2021). Generally, the ANN approach is employed to forecast sales revenue. ANN has found rising consideration in prediction theory that leads to prosperous applications in explanatory and time series sales prediction. The architecture model of ANN is depicted in Figure 6.

Some of the ANN-based approaches that existed in the literature regarding sales prediction are stated as follows:

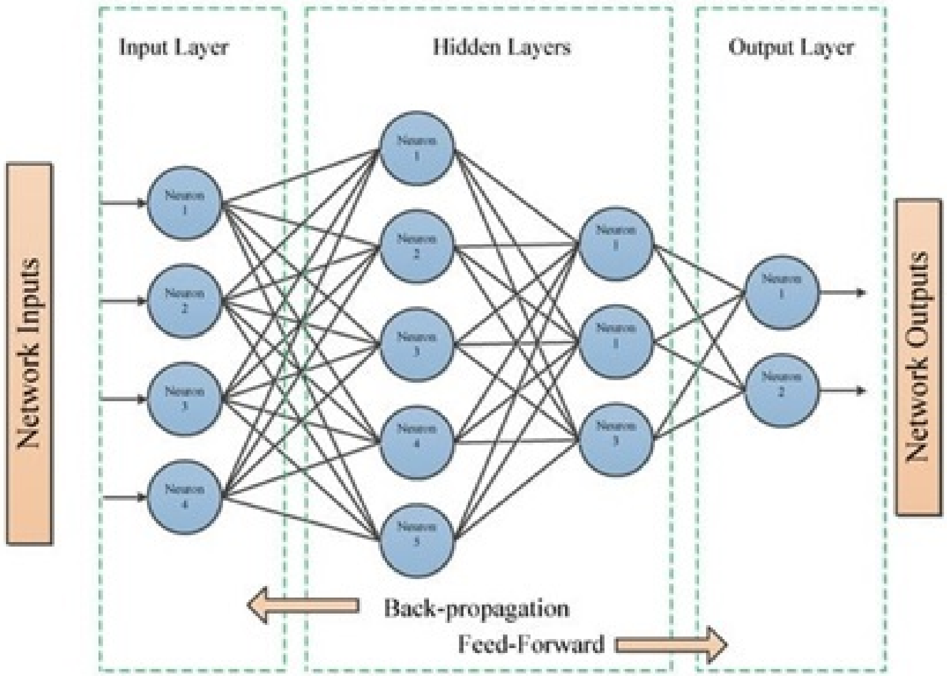

Figure 6: ANN model (source used from: Abdolrasol et al. (2021))

Thivakaran and Ramesh (2022) introduced a new method for demand forecasting. The model was designed for the business model of Big Mart, which involved multiple stores across the country selling identical items simultaneously through a marketplace model. The authors used both supervised and ANN approaches to generate dependable forecasts as opposed to alternative learning methods. Key aspects of the method include data exploration, feature selection, and data transformation. The model was tested on Big Mart Sales data, where relevant data was presented and analyzed to generate accurate predictions for future outcomes. Sharma et al. (2019) investigated the effectiveness of several modeling techniques such as for predicting book sales on Amazon.in. The study tested an ANN, regression analysis, and decision-tree analysis, utilizing relevant factors and their interactions as predictor variables. The researchers analyzed each independent predictor variable's importance, for instance, review sentiment and discount rate and determined the top significant predictors that marketers can control to influence sales. The artificial neural network model outperformed the decision-tree-based model in accuracy, while regression analysis, with and without interaction and sentiment factors, produced comparable outcomes.

f) Other approaches for sales prediction:

This section demonstrates the other approaches used for the sales prediction model. The study by Elalem et al. (2022) presented a structure that employed advanced techniques to predict sales of newly launched, short-lived products similar to preceding ones while there was restricted accessibility of historical sales details about the latest product. To generate sufficient sales data, the researchers utilized historical data through data augmentation and time-series clustering and considered two quantitative cluster assignment approaches. They applied a conventional statistical method (ARIMAX) and three DNN-based machine learning methods - gated recurrent units to the data, CNNs, and LSTM evaluated their performance. Utilizing two extensive datasets, the research investigated the forecasting methods' relative efficacy and found clustering usually resulted in considerably lesser prediction errors. The study's main discovery was that the straightforward ARIMAX approach outperformed advanced DNNs, exhibiting MAEs that are reduced by approximately 21% to 24%, for the larger dataset. Chen et al. (2020) presented a framework, named TADA, which utilized multi-task RNNs and dual-attention for sales prediction through trend alignment. Furthermore, the frequency of sales data acquisition was high and regular, making it complex for the model to sustain the accuracy of prediction with the latest data constantly being added. To address this, the authors extended the TADA framework to TADA+, which included an online learning module coupled with a new comparable reservoir. Unlike conventional random sampling-based reservoirs, this reservoir selected data samples that were challenging for the model to identify distinct dynamic patterns.

Loureiro et al. (2018) investigated the employment of DL in sales forecasting in the fashion industry, specifically for sales prediction of the latest personality products in future seasons. The objective of the research was to assist the purchasing operations of fashion Retail companies, and thus, the researchers used a real dataset presented by the company for analysis. The techniques constructed considered a comprehensive and different set of variables, including physical characteristics of products and domain experts' opinions. The study compared sales predictions produced by deep learning with several limited techniques, namely ANN, LR, RF, Decision Trees, and SVR. Although the DL model exhibited good performance in forecasting sales in the fashion retail market, in some evaluation metrics, the performance of the approach was not notably superior to certain basic techniques, particularly RF. Petroșanu et al. (2022) introduced an e-commerce sales forecasting technique that utilized a DAGNN for DL architecture. The method generated daily sales revenue predictions tailored to specific product categories, with long-term forecasting capabilities. The approach offered e-commerce store owners a highly accurate forecasting tool that predicts sales revenue for each product category up to three months in advance. This technique boasts remarkable scalability and generalization capabilities, thanks to the dynamic and incremental process of obtaining DAGNN's building blocks. Furthermore, the method combined multiple layers, resulting in optimal data usage that achieved excellent processing times and performance metrics. The technique for forecasting sales has the possibility of being extensively applied to anticipate sales up to three months in advance for other e-commerce retailers, even for larger enterprises.

Aguilar-Palacios et al. (2019) have developed an interpretable ML method, specifically designed to automatically predict promotional sales in real-world environments. The focus of their approach is on accurately forecasting sales during promotional periods. The developed approach leverages an automated weighted k-nearest neighbor's algorithm, which employs a feature selection process to measure the similarity of promotional sales. One of the significant benefits of this method is that it is capable of learning online, bypassing the requirement for repetitive model retraining and redeployment. To ensure the model's robustness, detailed surrogate models were used to infer the mechanisms underlying sales. The model's effectiveness was tested using real market data from a global retailer, encompassing diverse categories and countries, and accounting for various types of stores. The performance of this algorithm was benchmarked against an ensemble of RT and the retailer's own forecast, and it outperformed both, as measured by a merit figure that accounted not only for MAE but also for error deviations specific to retail businesses. In various categories and geographical locations, the method substantially enhanced the forecast's accuracy, resulting in practical benefits for supply chains. Furthermore, the team provides some insight into how this method can be deployed as a RESTful service in a production environment in the appendix.

2.4 Conclusion

Predicting sales is a difficult job since it involves using past data to anticipate future outcomes. Various ML models, algorithms, and features have been studied to aid in this task. Significant findings were made after analyzing these models. Firstly, the review volume emerged as the most crucial predictor of book sales on Amazon.in, as per all three models employed. Secondly, the average rating, discount amount, and discount rate had minor or negligible impacts on sales predictions. Thirdly, the regression and decision-tree models showed that positive and negative sentiments are individually significant predictors, but the neural network model contradicted prior research by indicating that negative sentiment has no more substantial impact than positive sentiment on sales predictions. Finally, all three models validated that the review volume's interaction with negative and positive sentiments served as significant predictors. Therefore, overall, the review volume, negative and positive sentiments, and their interactions were found to be the most significant drivers of book sales across all models. These results can help online sellers adjust these significant factors and manage their supply chain effectively by accurately predicting sales volume.

3 Chapter Proposed Optimized-Deep LSTM for Walmart sales Forecasting

3.1 Introduction:

Sales prediction is a significant task performed by Walmart and the prediction will be performed to present a fundamental effect on the business decision-making procedure. By performing the sales prediction, financial status can be understood easily to deal with the workforce and also to improve their supply chain management. This chapter explains the proposed EO-Deep LSTM algorithm developed for sales prediction in the following subsections.

a) Workflow:

The main aim of this thesis is to present a sales prediction approach for Walmart sales by exploiting a deep learning model. Here, the primary process is the pre-processing phase, which is performed using the missing data imputation. Then, the data augmentation is performed by utilizing the oversampling model. Subsequently, technical indicators, namely SMA, WMA, VAMA, ADX, TDI, and EMA are extracted. With these indicators as features, by utilizing the EO-Deep LSTM algorithm the prediction is performed. Here, EO-Deep LSTM is proposed by employing the EO to tune the parameters of the Deep LSTM for effective prediction. Figure 7 demonstrates the architecture model of Walmart sales prediction using the EO-Deep LSTM model.

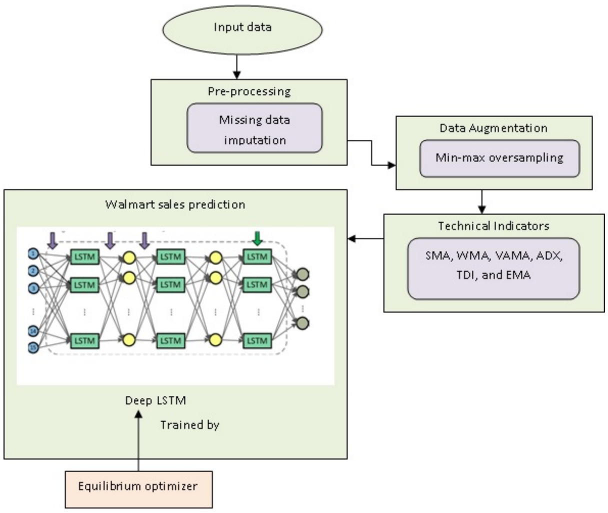

Figure 7: Architecture model of walmart sales prediction using EO-Deep LSTM

3.2 Pre-processing

Pre-processing is a procedure that defines the selection of data and how it be cleaned from all the outliers or noise. It refers to cleaning the data containing an enormous number of unnecessary information that may have numerous additional meanings but are not significant for the analysis. For instance, the product review will be present in the dataset, but the product review is not necessary for the sales prediction, thus here the product review is considered as the noise, which should be removed. In this research, the pre-processing phase is carried out by the missing data imputation Bertsimas et al. (2017) Haliduola et al. (2022) Furthermore, for the sales and price values, if the dataset has missing values, the missing values should be handled properly by replacing them with the average values (i.e.) missing data imputation to keep the data consistent Singh et al. (2020).

3.3 Data Augmentation

In the field of the DL approach, data augmentation is a renowned model, which is used to overcome the issue associated with overfitting during the training stage Sajawal et al. (2022). Here, the overfitting factor can be minimized and the learning can be enhanced by using the artificial transformation of the training data.

Oversampling In order to generate new data instances, oversampling is performed by using the transformation to conventional data instances to adjust the class imbalance. In this work, the oversampling is performed by using the min-max model. The data will be augmented based on the minimal and the maximal values of the features in each column. Then, the augmented data is normalized to make the further process easy. The min-max approach Patro and Sahu (2015) is defined as a linear transformation of the original data into the particular interval (new_{min}, new_{max}) as formulated below,

$$Z_m = new_{min} + (new_{max} - new_{min}) \times (\frac{Z_{max} - Z_{min}}{Z_{min} - Z_{max}}) \tag{1}$$

This approach scales data from (Z_{min}, Z_{max}) to (new_{min}, new_{max}) in proportion. The most important benefit of this model is that it conserves all associations of the data values accurately. It does not set up any possible bias in the data. Also, min-max normalization is the general manner to normalize the data. The minimal value of the feature gets transferred into "0", and the maximal value of the features gets transferred into "1". Additionally, all other values are transferred into a decimal between 1 and 0.

3.4 Technical Indicator Extraction

Technical indicators are the data points, which is used to analyze the asses traded on the basis of their history and price. Here, the technical indicators, SMA, WMA, VAMA, ADX, TDI, and EMA are extracted. The moving average is one of the well-known indicators of technical analysis. By smoothing the data, these measures are utilized to eradicate noise and detect trends Vaiz and Ramaswami (2016). It is extensively utilized due to its ease and its probability to integrate various moving averages together. The diverse kinds of moving averages used in this work are stated as follows: SMA: It presents all the days' equivalent weights, which are computed as follows:

SMA: It presents all the days' equivalent weights, which are computed as follows:

$$SMA = \frac{\sum_{1}^{N} P_r}{N} \tag{2}$$

where, N indicates a number of total durations, and P_r represents price.

EMA: Not like SMA, the EMA presents greater priority to the actual data. The current value obtains greater significance in the computation of EMA evaluated with the further ones.

$$EMA = EMA_{-1} + M \times (input - EMA_{(-1)}) \tag{3}$$

where, $M = \frac{2}{(N+1)}$

WMA: It presents greater significance to actual days and lesser significance to the further days typically. However, the trader has to take a choice on which day is supposed to be of higher or lesser importance.

$$WMA = \frac{P_r \times N \times P_{r(-1)} \times (N-1) + P_{r(-2)} \times (N-2) + ... + P_{r(-1(N-1)} \times 1}{N!} \tag{4}$$

VAMA: It is a type of moving average that utilizes volume as a weighting factor. It refers that days with greater volume carry more weight in the computation. Not like cumulative moving averages, VWMA only considers data within the specified time period for its computation.

$$VAMA = \frac{\sum_{1}^{N} P_r \times volume}{\sum_{1}^{N} volume} \tag{5}$$

In the case of all moving averages, if the closing price is superior to the moving average, it suggests a buy signal. Conversely, if the closing price is lesser than the moving average, it indicates a sell signal. Since the price and moving average values change frequently, the moving average curve can be followed, if it is increasing, it is referred as a purchase signal or it is a retail signal.

ADX: The ADX is employed to assess the weakness or strength of a trend, rather than determining the actual direction. Here, the directional movement is explained as $-E_i$ and $+E_i$. and . The Directional Indicator $(+E_i)$ and Directional Indicator $-E_i$ balance the ADX by providing additional information on trend direction. The true range percentage that is upward in direction is called $+E_i$, whereas $-E_i$ is the true percentage that is downward in direction. Utilized jointly, chartists can ascertain both the strength and direction of the development. The scheme is to buy while $+E_i$ is higher than $-E_i$ and the scheme is to sell while $-E_i$ is higher. In order to form a complete trading system refers the crosses of these directional indicators can be integrated with ADX. While ADX is above 25, the trend is said to be strong, and when ADX is below 20, no trend is present. Between the values 20 and 25, there exists a range to be a gray zone. However, the low value of ADX refers to a weak trend, and the high value refers to a strong trend.

$$ADX = \frac{ADX_{-1} \times (N-1) + DX}{N} DX = \frac{(+EI) - (-EI)}{(+EI) + (-EI)} \tag{6}$$

TDI: It is utilized to recognize while a trend has started and while it has come to an ending. It can also be utilized as a standalone indicator or integrated with others; it will carry out well in recognizing the starting of trends. Using a positive direction indicator value, an uptrend is signaled but while using a negative value, a downtrend is signaled. If the director indicator is negative and TDI is positive, it refers to a sell signal and if both the director indicator and TDI are positive it is a buy signal.

$$20 - dayTDI = (ab20) - \left\{ (\sum aM40) - (\sum aM20) \right\} \tag{7}$$

whereas, $ab20$ = absolute value of the summation of 20-day momenta of the final 20 days, $\sum am20$ refers to sum of 20 day absolute momenta of the last 20 days, and $\sum am40$ refers the sum of 20 day absolute momenta of the last 40 days.

3.5 Sales forecasting using EO-Deep LSTM

This section describes sales forecasting using the EO-Deep LSTM approach in a detailed manner. The advantage of the deep LSTM model is that it has the ability to capture and remember long-term dependencies in sequential data. Moreover, EO algorithm has the advantages of strong robustness and fast convergence speed. Thus, in this work, the deep LSTM model is used and it is trained by the EO algorithm for the prediction.

a) Deep LSTM

The deep LSTM model comprises numerous hidden layers, which involve both the fully connected layers and LSTM layers Devi and Arunachalam (2023). In temporal space $(\rho = 1, 2, ..., t)$, the Deep LSTM network pairs output and input sequences together. For the implementation of a deep LSTM network trained with multiple datasets, both output and input sequences should be arranged as 3-D arrays. The first dimension consists of entries with samples, the second dimension consists of time steps, and the third dimension consists of input or output channels/features.

The architecture of deep LSTM is displayed in Figure 8; where each LSTM cell is astonishingly the same as the neural node in existing neural networks and comprises independent sets of bias and weights shared over complete temporal space within the layer. The four interconnected segments that form an LSTM are the input and forget gate, internal cell, and output gate. From the previous step, an internal cell can remember its state because of self recurrent link. The flow of input activation is regulated by the input gate into the internal state of the cell, while the output activation flow is regulated by the output gate into the LSTM cell output. Adaptively, it has the potential to forget or rest its memories as the internal cell state is scaled by the forget gate. The input manages the cell state, as well as output gates in each cell of the LSTM, which can be utilized to indicate long short-term temporal dependence of the dynamical system.

Let us consider, $(t = 1, ..., N)$, wherein N represents the overall integers of time steps among L^{th} LSTM network layer, the forget gate is indicated as F_t^L , the input gate is indicated as I_t^L , the input state LSTM is defined as X_t^L , the output of the hidden state is represented as H_t^L and the output gate is indicated as O_t^L , C_{t-1}^L indicates the cell state memory at preceding time step $t-1$ and H_{t-1}^L refers the hidden state output. The below formulations are utilized to describe the connection amid these defined variables.

$$F_t^{(L)} = \lambda(w_{XF}^L X_t + w_{HF}^L H_{t-1} + B_F^{(L)}) \tag{8}$$

$$I_t^{(L)} = \lambda(w_{XI}^L X_t + w_{HI}^L H_{t-1} + B_I^{(L)}) \tag{9}$$

$$O_t^{(L)} = \lambda(w_{XO}^L X_t + w_{XO}^L H_{t-1} + B_0^{(L)}) \tag{10}$$

$$\tilde{C}_t^L = tanh(w_{OI}^{(L)})X_t + w_{OI}^{(L)} H_{t-1} + B_o^{(L)}) \tag{11}$$

$$C_t^{(L)} = F_t^L \Theta C_{t-1}^L + I_t^(L)\Theta \tilde{C}_t^L \tag{12}$$

$$H_t^L = O^(L)_t tanh \Theta \tilde{C}_t^{(L)} \tag{13}$$

whereas, $W_{\beta\delta}^L$ with $\beta = \{X, H\}$ and $\delta = \{F, I, C, O\}$ indicates the weight matrices equivalent to several inputs (for instance H_t^L or X_t^L) within several gates (for instance, tanh layer, forget gate, input or output gate), in which B_δ^L indicates related bias vectors. For example, W_{XF}^L and W_{HF}^L correspond to the input vector X_t and H_t correspondingly within forget gate. Moreover, represents the hyperbolic tangent function. Θ represents the product, C_r^L represents the tanh layer produced vector of center candidate numbers, and λ indicates the logistics sigmoid function. In sequence modeling, the deep LSTM network is successful as each LSTM cell has a complicated connection method. Temporal feature map translation to suitable output space can be performed with the assistance of fully connected layers.

b) Equilibrium Optimizer

Equilibrium Optimizer, called as EO, is a population-based, nature-inspired meta-heuristics that belongs to the class of Physics-based optimization approaches, inspired by dynamic source and sink approaches with a physics foundation. EO is an intelligent approach that was enthused by the mass balance formulation in physics. The mass balance formulation reflects the physical procedures of mass entering, producing, and leaving in the control volume. In the optimizer, each particle concentration is updated arbitrarily till it attains equilibrium. EO approach designs three mathematical approaches Faramarzi et al. (2020):

Step1: Initialization phase

As same as conventional meta-heuristic approaches, by initializing the population, the EO starts the optimization process Also, by randomly initializing the particles the

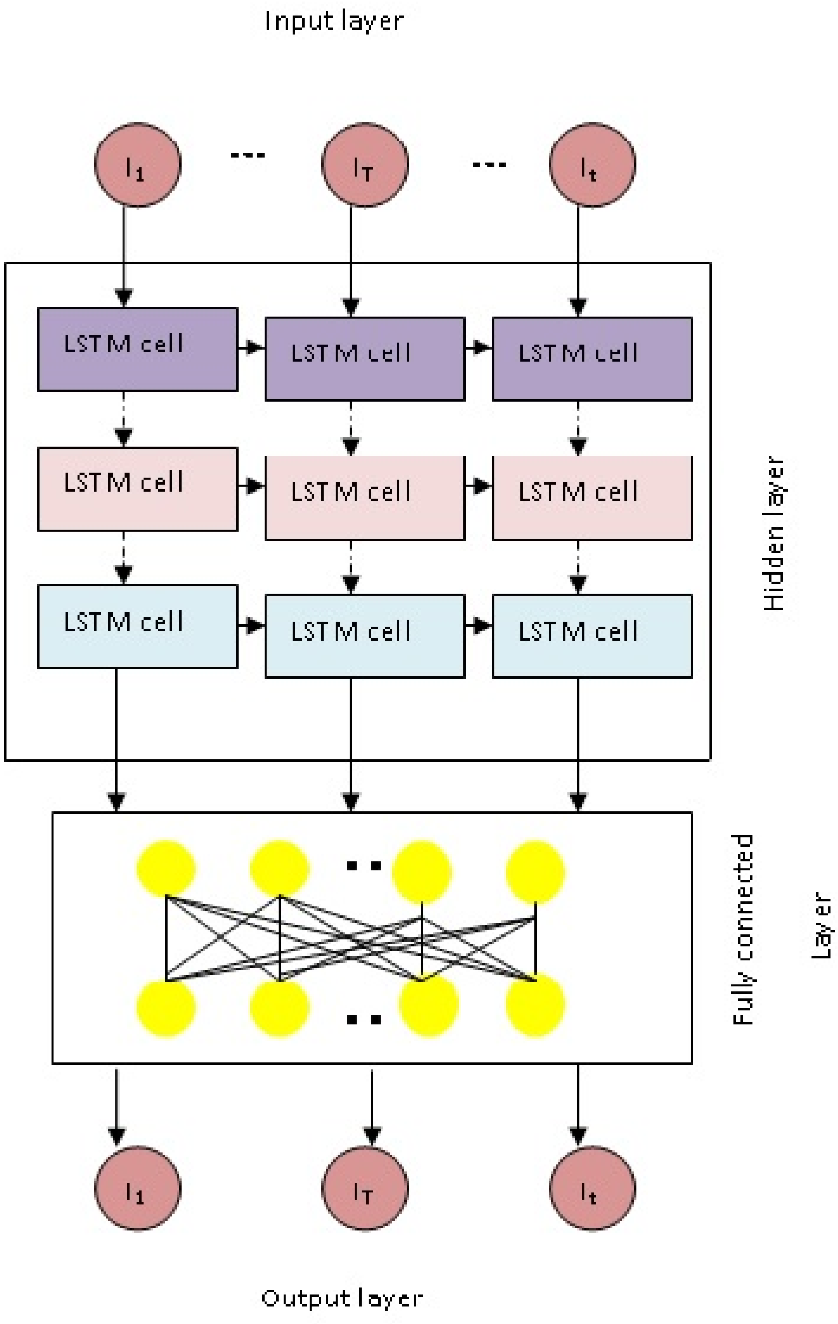

Figure 8: Architectural model of Deep LSTM

initial concentration designed in the D_m represents the dimensional search space. For each particle, the initial concentration is stated as follows:

$$Y_a^{initial} = lb + rn_a(ub - lb)a = 1, 2, .., n \tag{14}$$

wherein $y_a^{initial}$ signifies the primary concentration of the a_{th} particle, Ib and ub denote the minimum and maximum values of particles in the search space, n indicates the number of particles, and rn_a denotes an arbitrary vector in the range of [0,1].

Step 2: Equilibrium candidates and pool ($\vec{Y}eq_avg$)
The main aim of the equilibrium candidates and the pool is to enhance the global search capability of the approach and avoid falling into the local best solution of minimal quality. Subsequent to finishing the initialization stage, the generated particle concentration is calculated. The four particles with the maximum fitness values are chosen to prepare for the equilibrium pool formation. During the algorithm optimization procedure, the equilibrium pool is utilized to present the candidate solutions. It comprises four particles with optimal fitness values and one average particle created at the time of the initialization stage. The mathematical formulation is stated as follows:

$$\vec{y}eq_avg = \frac{\vec{y}_{eq1} + \vec{y}_{eq2} + \vec{y}_{eq3} + \vec{y}_{eq4}}{4} \tag{15}$$

$$\vec{y}eq_pool = \vec{y}_{eq1} + \vec{y}_{eq2} + \vec{y}_{eq3} + \vec{y}_{eq4}, \vec{eq}_{avg} \tag{16}$$

Among them, $\vec{Y}eq_avg$ signifies the average particle, $\vec{Y}eq1$ to $\vec{Y}eq4$ signifies four particles with the maximum concentration chosen subsequent to the algorithm initialization, and$\vec{Y}eq_pool$ signifies the equilibrium pool. Particularly, the four particles with the maximum concentration contribute to the algorithm exploration in the equilibrium pool, whilst the average particle plays a considerable role in the exploitation stage. In the iterative process of the approach, each particle is chosen from the five candidate particles in the equilibrium pool with a similar probability that contributes to the global optimal solution generation.

Step 3. Updating the concentration: The $\vec{f}$ refers to the exponential term, which is a significant indicator to balance the exploration and exploitation ability of the EO approach. The $\vec{f}$ computation is stated as below:

$$\vec{f} = d_1 sgn\vec{R} - 0.5).[e^{-\vec{ot}} - 1] \tag{17}$$

wherein, $sgnR \vec{-} 0.5$ refers to the direction of exploration and exploitation capability of the approach, d_1 indicates the constant, which controls the exploration capability of the approach, t denotes coefficient update with a number of iterations, σ denotes vectors within an interval of [0,1] that can be computed as below:

$$t = \left(1 - \frac{itr}{max - itr}\right)^{\left(d_2 - \frac{iter}{max - iter}\right)} \tag{18}$$

wherein, $max - itr$ refers maximum iteration of the approach, itr denotes the current iteration of the approach, and d_1 refers to the constant, which controls the exploitation capability of the approach. An equivalently significant indicator $\vec{g}$, is created to develop the exploitation ability of EO model which is stated below:

$$\vec{g} = g\vec{c}p(\vec{y}_{eq} - \vec{\sigma}\vec{y}).\vec{f} \tag{19}$$

$$g\vec{c}p = \left\{ 0.5 R_1 0 R_2 \geq gp R_2 < gp \right\} \tag{20}$$

The above formulation, $\vec{y}_{eq}$ indicates a concentration of the present particle, $g\vec{c}p$ indicates the control parameter vector of generation rate, and gp represents a constant value of 0.5. For each particle, the updated formulation is stated below:

$$\vec{y} = \vec{y}_{eq} + (\vec{y} - e\vec{q}).\vec{f} + \frac{\vec{g}}{\vec{\sigma} v} \tag{21}$$

wherein, v represented as a unit.

4 Chapter

4.1 Results and Discussion

The simulation of the proposed scheme was done in the PYTHON tool and the dataset employed was "Walmart Sales forecasting" for conducting tests. The performance analysis of the proposed scheme was performed using evaluation metrics, like RMSE, MAE, and MSE, which were evaluated with that of conventional approaches.

4.2 Dataset Description:

Walmart sales forecasting dataset taken from https://data.world/mayagupta/walmart-sales forecasting (2023): This dataset is used for experimentation. In this dataset, historical sales data is accessible for 45 Walmart stores that are situated in diverse regions. On each day certain holidays and events affect sales. Because of unexpected demands, the business is facing a challenge and runs out of stock sometimes, owing to an unsuitable machine learning approach. Walmart would like to forecast the sales and demand precisely.

4.3 Evaluation Measures

(a) MSE: This metric evaluates the mean or average of the square of difference among the actual and estimated values.

$$MSE = \frac{1}{L} \sum_{m=1}^{L} (Z_m - \breve{Z}_m)^2 \tag{22}$$

where, Z_m denotes observed values, L refers number of data points, and $\breve{Z}_m$ indicates the predicted values.

(b) RMSE: This measure is employed to evaluate the average deviation between values predicted by a model and the actual values.

(c) MAE: MAE is a metric of errors between paired observations that expresses a similar occurrence.

4.4 Performance analysis

Figure 9 delineates the performance analysis of the EO-Deep LSTM model by varying the delay. The MSE value of the EO-Deep LSTM mechanism was displayed in Figure 9 (a). With respect to the delay 50000, the MSE value of the EO-Deep LSTM mechanism was 0.423 for iteration 40, however, the value decreased to 0.223 at iteration 100. For delay 100000, the MSE value of the EO-Deep LSTM mechanism was 0.426 for iteration 20, whereas the MSE value was reduced to 0.214 for iteration 100. With respect to the delay of 150000, the MSE value of the EO-Deep LSTM mechanism was 0.415 for iteration 20, which was reduced to 0.198 at iteration 100. Figure 9 (b) demonstrates the RMSE value of the EO-Deep LSTM approach. Regarding delay 100000, the RMSE value of the EO-Deep LSTM mechanism was 0.653 for iteration 20, whereas it decreased to 0.463 at iteration 100. Regarding delay 150000, the RMSE value of the mechanism was 0.645 for iteration 20, but the value decreased to 0.449 at iteration 100. With respect to the delay of 200000, the RMSE value of the mechanism was 0.593 for iteration 20, which decreased to 0.419 at iteration 100. Figure 9 (c) depicts the MAE value of the EO-Deep LSTM approach. Regarding the delay 50000, the MAE value of the EO-Deep LSTM approach was 0.583 for iteration 20, whereas it decreased to 0.365 at iteration 100. With respect to delay 100000, the MAE value of the EO-Deep LSTM approach was 0.578 for iteration 20, whereas it decreased to 0.335 at iteration 100. For the delay of 200000, the MAE value of the mechanism was 0.454 for iteration 20, but the MAE value decreased to 0.247 at iteration 100.

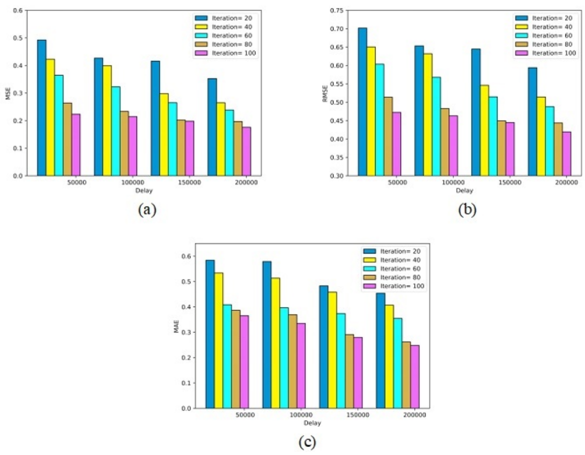

Figure 9: Performance analysis of the mechanism (a) MSE (b) RMSE and (c) MAE

4.5 Comparative analysis

This section elaborates on the comparative investigation of the proposed EO-Deep LSTM scheme by comparing the existing techniques, such as CNN Yin and Tao (2021), DNN Loureiro et al. (2018), and RNN Chen et al. (2020) mechanisms.

Figure 10 depicts the comparative investigation of the EO-Deep LSTM mechanism over the existing schemes. Figure 10 (a) delineates the comparative evaluation of the EO-Deep LSTM approach regarding the MSE. With respect to the delay 100000, the EO-Deep LSTM obtained the least MSE (i.e., 0.214), whereas, the CNN attained a high

MSE (i.e., 0.404), DNN attained the MSE value of 0.366, and the RNN attained the MSE value 0.242. The comparative evaluation of the EO-Deep LSTM approach regarding the RMSE is illustrated in Figure 10 (b). With respect to the delay 150000, the EO-Deep LSTM achieved the least RMSE value of 0.445, but, the CNN acquired a high RMSE value of 0.621, DNN attained a high RMSE value of 0.580, and the RNN also acquired a high RMSE value of 0.445. Figure 10 (c) demonstrated the comparative evaluation of the EO-Deep LSTM approach to the MAE. With regard to the delay 100000, the EO-Deep LSTM achieved the least MSE value of 0.335, whereas, the CNN attained a high MSE value of 0.565, DNN attained a high MSE value of 0.437, and the RNN also attained a high MSE value of 0.365.

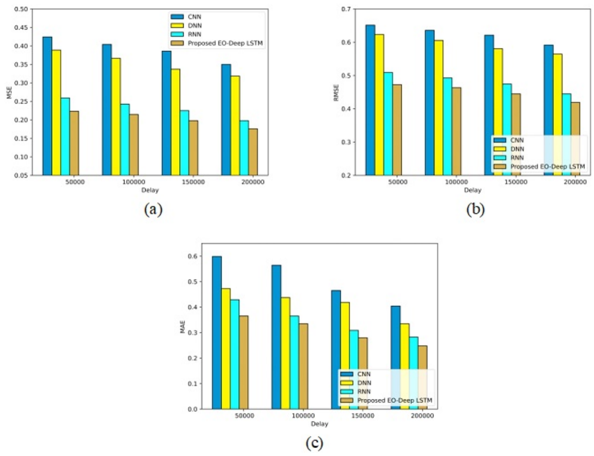

Figure 10: Comparative analysis of the mechanism (a) MSE (b) RMSE and (c) MAE

Table 1: Comparative Discussion

Methods	CNN	DNN	RNN	Proposed EO-Deep LSTM
MSE	0.350	0.318	0.198	0.176
RMSE	0.621	0.564	0.445	0.419
MAE	0.404	0.335	0.283	0.248

Table 1 depicts the comparative discussion of the EO-Deep LSTM mechanism over the existing mechanisms. In terms of MSE, the EO-Deep LSTM was 98.8% better than the CNN, 81% better than the DNN, and 13% better than RNN. Moreover, the EO-Deep LSTM was 48% better than the CNN, 34% better than the DNN, and 7% better than RNN regarding the RMSE. With respect to MAE, the EO-Deep LSTM was 62% better than the CNN, 35% better than the DNN, and 14% better than RNN. Thus, it is obvious that the proposed approach has the least error values and superior performance than the comparative techniques; thereby proving its effectiveness in sales prediction.

5 Chapter Conclusion and Future Work

Conclusion:

Sales prediction is an important up till now unsolved issue because of the subtle significant patterns amid diverse measures and the uneven sales trends activated by intricate real-life circumstances. This work aspires to predict Walmart sales by employing the EO-Deep LSTM technique. The research workflow involves several steps, such as pre-processing, data augmentation, technical indicator extraction, and prediction. Initially, the input data was subjected to pre-processing, using the missing data imputation. To enhance the data quality, data augmentation is done by oversampling. Next, a set of technical indicators, such as SMA, WMA, VAMA, ADX, TDI, and EMA, were extracted from the data. These indicators serve as features for the prediction mechanism. The prediction task was carried out using Deep LSTM which was a kind of RNN. It was trained using an equilibrium optimizer to fine-tune the parameters. The implementation of the research work was conducted in Python, using the Walmart sales forecasting dataset for carrying out tests. The performance evaluation of the proposed mechanism was conducted using parameters, namely MSE, MAE, and RMSE. These metrics provide insights into the efficacy of the sales prediction mechanism. The performance of the proposed EO-Deep LSTM achieved an MSE value of 0.176, an RMSE value of 0.419, and the MAE value of 0.248.

Future Work: Walmart can concentrate on the e-commerce features of the business with the rising technology and growing demand of consumers. Also, it makes it a lot simpler to recognize customer buying patterns. A significant feature of this study in the future is also to attempt and comprehend consumer buying behavior on the basis of department sales. One more feature that would be value investigating with this research is recognizing its trends with sales for each of the stores. Additionally, forecasting future trends on the basis of the available sales data. In the future, Walmart's sales prediction can be enhanced by integrating external factors such as economic indicators, weather conditions, and social media trends. This can provide a complete understanding of the factors that influence sales and enhances the accuracy of predictions. Also, in the future hybrid deep learning mechanisms can be developed to optimize the performance of the mechanism. Moreover, fine-tuning parameters and a hybrid optimizing process can be included in the training process. By performing this, the reliability of the sales forecasting mechanism can be enhanced.

References

Abdolrasol, M. G., Hussain, S. S., Ustun, T. S., Sarker, M. R., Hannan, M. A., Mohamed, R., Ali, J. A., Mekhilef, S. and Milad, A. (2021). Artificial neural networks based optimization techniques: A review, *Electronics* **10**(21): 2689.

Aguilar-Palacios, C., Muñoz-Romero, S. and luis Rojo-Álvarez, J. (2020). Cold-start promotional sales forecasting through gradient boosted-based contrastive explanations, *IEEE Access* **8**: 137574–137586.

Aguilar-Palacios, C., Muñoz-Romero, S. and Rojo-Álvarez, J. L. (2019). Forecasting promotional sales within the neighbourhood, *Ieee Access* **7**: 74759–74775.

Aher, A., Rajeswari, K. and Vispute, S. (2021). Data analysis and price prediction of black friday sales using machine learning technique, *IJERT* .

Bertsimas, D., Pawlowski, C. and Zhuo, Y. D. (2017). From predictive methods to missing data imputation: an optimization approach., *J. Mach. Learn. Res.* **18**(1): 7133–7171.

Chen, T., Yin, H., Chen, H., Wang, H., Zhou, X. and Li, X. (2020). Online sales prediction via trend alignment-based multitask recurrent neural networks, *Knowledge and Information Systems* **62**(6): 2139–2167.

dataset taken from https://data.world/mayagupta/walmart-sales forecasting (2023).

Devi, R. A. and Arunachalam, A. (2023). Enhancement of iot device security using an improved elliptic curve cryptography algorithm and malware detection utilizing deep lstm, *High-Confidence Computing* **3**(2): 100117.

Elalem, Y. K., Maier, S. and Seifert, R. W. (2022). A machine learning-based framework for forecasting sales of new products with short life cycles using deep neural networks, *International Journal of Forecasting* .

Faramarzi, A., Heidarinejad, M., Stephens, B. and Mirjalili, S. (2020). Equilibrium optimizer: A novel optimization algorithm, *Knowledge-Based Systems* **191**: 105190.

Haliduola, H. N., Bretz, F. and Mansmann, U. (2022). Missing data imputation using utility-based regression and sampling approaches, *Computer Methods and Programs in Biomedicine* **226**: 107172.

Haselbeck, F., Killinger, J., Menrad, K., Hannus, T. and Grimm, D. G. (2022). Machine learning outperforms classical forecasting on horticultural sales predictions, *Machine Learning with Applications* **7**: 100239.

Huang, J., Chen, Q. and Yu, C. (2022). A new feature based deep attention sales forecasting model for enterprise sustainable development, *Sustainability* **14**(19): 12224.

Jeswani, R. (2021). Predicting walmart sales, exploratory data analysis, and walmart sales dashboard.

Loureiro, A. L., Miguéis, V. L. and Da Silva, L. F. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail, *Decision Support Systems* **114**: 81–93.

Patro, S. and Sahu, K. K. (2015). Normalization: A preprocessing stage, *arXiv preprint arXiv:1503.06462* .

Petroșanu, D.-M., Pîrjan, A., Căruţaşu, G., Tăbușcă, A., Zirra, D.-L. and Perju-Mitran, A. (2022). E-commerce sales revenues forecasting by means of dynamically designing, developing and validating a directed acyclic graph (dag) network for deep learning, *Electronics* **11**(18): 2940.

Sajawal, M., Usman, S., Alshaikh, H. S., Hayat, A. and Ashraf, M. U. (2022). A predictive analysis of retail sales forecasting using machine learning techniques, *Lahore Garrison University Research Journal of Computer Science and Information Technology* **6**(04): 33–45.

Sharma, S. K., Chakraborti, S. and Jha, T. (2019). Analysis of book sales prediction at amazon marketplace in india: a machine learning approach, *Information Systems and e-Business Management* **17**: 261–284.

Singh, K., Booma, P. and Eaganathan, U. (2020). E-commerce system for sale prediction using machine learning technique, **1712**(1): 012042.

Thivakaran, T. and Ramesh, M. (2022). Exploratory data analysis and sales forecasting of bigmart dataset using supervised and ann algorithms, *Measurement: Sensors* **23**: 100388.

Vaiz, J. and Ramaswami, M. (2016). Forecasting stock trend using technical indicators with r, *Int. J. Comput. Intell. Inform* **6**(3).

Wang, Y., Pan, Z., Zheng, J., Qian, L. and Li, M. (2019). A hybrid ensemble method for pulsar candidate classification, *Astrophysics and Space Science* **364**: 1–13.

Yin, P., Dou, G., Lin, X. and Liu, L. (2020). A hybrid method for forecasting new product sales based on fuzzy clustering and deep learning, *Kybernetes* **49**(12): 3099–3118.

Yin, X. and Tao, X. (2021). Prediction of merchandise sales on e-commerce platforms based on data mining and deep learning, *Scientific Programming* **2021**: 1–9.