

عثمان عثمان العمودي

Salesforecasting revised .docx

 Midocean University

Document Details

Submission ID

trn:oid:::2945:337422736

Submission Date

Dec 15, 2025, 1:25 PM GMT+2

Download Date

Dec 15, 2025, 5:58 PM GMT+2

File Name

Salesforecasting revised .docx

File Size

1.8 MB

32 Pages

4,728 Words

30,659 Characters

58% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Detection Groups



37 AI-generated only 58%

Likely AI-generated text from a large-language model.



0 AI-generated text that was AI-paraphrased 0%

Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?

The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.





Sales Forecasting

Prepared By

Name: Othman Omar Alamoudi

ID No.: 243000199

Submitted in fulfillment of the requirements for the degree Master of
Science in Artificial Intelligence

At

Collage of Informatics

Midocean University

UAE

Supervisor Name: Wesam Ahmed

College of Informatics, Midocean University, Comoros

Email:.....

Acknowledgements

I want to express my deepest gratitude to **Dr. Wesam Ahmed**, whose guidance, encouragement, and expertise were instrumental throughout this research. Dr. Wesam's insightful feedback and unwavering support helped shape the direction of this thesis and inspired me to explore new dimensions in machine learning and sales forecasting.

Their mentorship not only enhanced the academic rigor of this work but also fostered a deeper appreciation for the value of data-driven decision-making in real-world business contexts. I am genuinely thankful for their patience, constructive criticism, and commitment to excellence.

Abstract

List of Publications

Table of Contents

List of Abbreviations	vi
List of Figures.....	vii
List of Tables	viii

Chapter 1: Introduction.....	8
1.1 Problem Definition	9
1.2 Thesis Objectives	9
1.3 Thesis Contribution	10
1.4 Thesis Organization.....	10
 Chapter 2: Overview.....	 11
2.1 background	12
2.2 Educational Data Mining	6
2.3 Applications of Educational Data Mining	7
Chapter 3: Related Work	31
3.1 Introduction	31
3.2 History	31
Chapter 4: Proposed Approach	78
4.1 Introduction	78
 Chapter 5: Experimental Results and Discussion	 86
 Chapter 6: Conclusion and Future Work	 127
7.1 Important Features	127
7.3 Future Work	129

References 130

List of Abbreviations

- **AI:** Accuracy
- **LSTM:** Artificial Neural Network
- **XGBOOST:** Analysis of Variance
- **ARIMA:** Receiver Operating Characteristic

List of Figures

2.1.1 Core AI technologies between ML and DL role in classification.....	13
2.3.1.1 XGBoost	8
2.3.3.1 Prophet flowchart.....	16
2.4.1.1 LSTM	21
2.5 Educational Process Mining (EPM): A Learning Analytics Data Set ...	23

List of Tables

3.1.1 Summary Table of ML.....	21
--------------------------------	----

Chapter 1 Introduction

Recently, machine learning has greatly contributed to the accuracy of sales forecasting models. Traditional statistical methods tend to suffer from non-linearity and the variability in market dynamics. Neural networks, especially such as Long Short-Term Memory (LSTM) models, however, have been robust in representing time scale dependencies. An LSTM-based model has been used to predict furniture sales in Central and Eastern Europe, recording low forecast error without seasonal assumptions, accurately monitoring demand variability and so on [1]. Sales prediction is an essential part of business forecasting and planning, especially in e-commerce where the order is cyclical and affected by different dynamic features. Receive accurate estimates improve visibility in the planning for inventory administration, stock planning, production planning, and cash flow, thus lowering human error and minimizing human intervention. Time series forecasting through traditional statistical methods such as the ARIMA model has been widely used, yet the model tends to fail in the face of nonlinear patterns or drastic changes in the markets. The development of AI is progressing faster than we can anticipate, and hybrid models that marry statistical and neural methods are now being developed. To take an example of this, Vavliakis et al. proposed a hybrid ARIMA-LSTM model by including linear time series analysis and nonlinear residual modeling through a LSTM network. Their approach also includes retail price as a characteristic to enhance the prediction model performance in price sensitive e-commerce settings [2]. Sales forecasting forms the framework of supply chain management and refers to the process that allows organizations to project demand, control stock levels, and lower expenses. Big data and machine learning are making more sophisticated algorithms capable of dealing with much larger data sets, replacing traditional forecasting methods. Forecasting sales is a significant key in supply chain management, allowing an organization to predict demand, manage inventory efficiently, and minimize operational costs. In this context, the data-driven forecasting methods that used to be performed over decades in the past have been replaced by algorithms that, based on data analysis and machine learning insights, can adapt to the complex and nonlinear patterns observed in consumer behavior as they evolve over time. One of the algorithms mentioned is Extreme Gradient Boosting (XGBoost), an ensemble learning method associated with speed and accuracy. It has been applied to different domains, such as predicting retail sales. It's not only great for analyzing big data, but also in finding patterns that are often overlooked. Akande et al. (2022) used XGBoost in forecasting sales through historical data from 45 Walmart stores.

Their research found that XGBoost predicted sales very well with a high accuracy, and achieved better performances than previous methods based on Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 criteria. Thus, it is a suitable guide for retail decision-makers [3], offering actionable intelligence for pricing strategies and inventory planning. A potential implication for predictive models, like XGBoost for instance, is that the prediction accuracy will be better than before.

1.1 Problem Definition

Sales forecasting plays a pivotal role in guiding inventory management, budgeting, and strategic decision-making across industries. However, traditional forecasting techniques—such as moving averages and linear regression—often fall short in the face of the complexity of modern sales environments. These methods struggle to account for nonlinear relationships, seasonal variations, promotional effects, and external disruptions, such as economic shifts or competitor actions. As businesses accumulate increasingly diverse and high-volume datasets, the limitations of conventional models become more pronounced. Inaccurate forecasts can lead to overstocking, missed revenue opportunities, and inefficient resource allocation. This is especially problematic in fast-paced sectors such as retail and e-commerce, where agility and precision are crucial. Machine learning and deep learning offer promising alternatives by enabling models to learn patterns directly from historical data, adapt to changing trends, and incorporate a wide range of variables that influence them. Techniques such as decision trees, ensemble methods (e.g., Random Forest, XGBoost), and support vector machines have shown strong potential in capturing complex dependencies. In particular, Long Short-Term Memory (LSTM) networks—a type of recurrent neural network—have demonstrated exceptional performance in modeling sequential data, making them well-suited for time series forecasting tasks where long-term dependencies and temporal dynamics are crucial.

1.2 Thesis Objectives

The primary objective of this thesis is to investigate and evaluate the effectiveness of machine learning and deep learning techniques in enhancing the accuracy and reliability of sales forecasting models. By leveraging historical sales data and relevant external variables, the study aims to develop predictive models that outperform traditional statistical approaches in dynamic and data-rich environments.

1.3 Thesis Contribution

1.4 Thesis Organization

This thesis contributes to the growing body of research on intelligent forecasting systems by demonstrating how machine learning techniques can enhance the accuracy, scalability, and interpretability of sales predictions. Unlike traditional statistical models, which often rely on rigid assumptions and limited feature sets, the proposed approach leverages data-driven algorithms to uncover complex patterns and nonlinear relationships within historical sales data.

Chapter 2 Overview

This thesis explores the application of machine learning and deep learning models to enhance the accuracy and reliability of sales forecasting. The study is motivated by recent research that demonstrates the superiority of intelligent forecasting systems over traditional statistical methods. For example, Li et al. (2020) developed an automatic LSTM-based forecasting system that effectively modeled sequential sales data [1]. Vavliakis et al. (2021) proposed a hybrid ARIMA-LSTM approach that improved prediction accuracy in e-commerce environments [2]. Akande et al. (2021) applied XGBoost to Walmart's historical sales data, achieving high performance and interpretability [3]. Building on these foundations, this research aims to design a robust forecasting framework that integrates data preprocessing, feature selection, and model tuning. The models will be evaluated using real-world datasets and performance metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 score. Ultimately, the goal is to provide actionable insights that support better inventory planning, budgeting, and strategic operations.

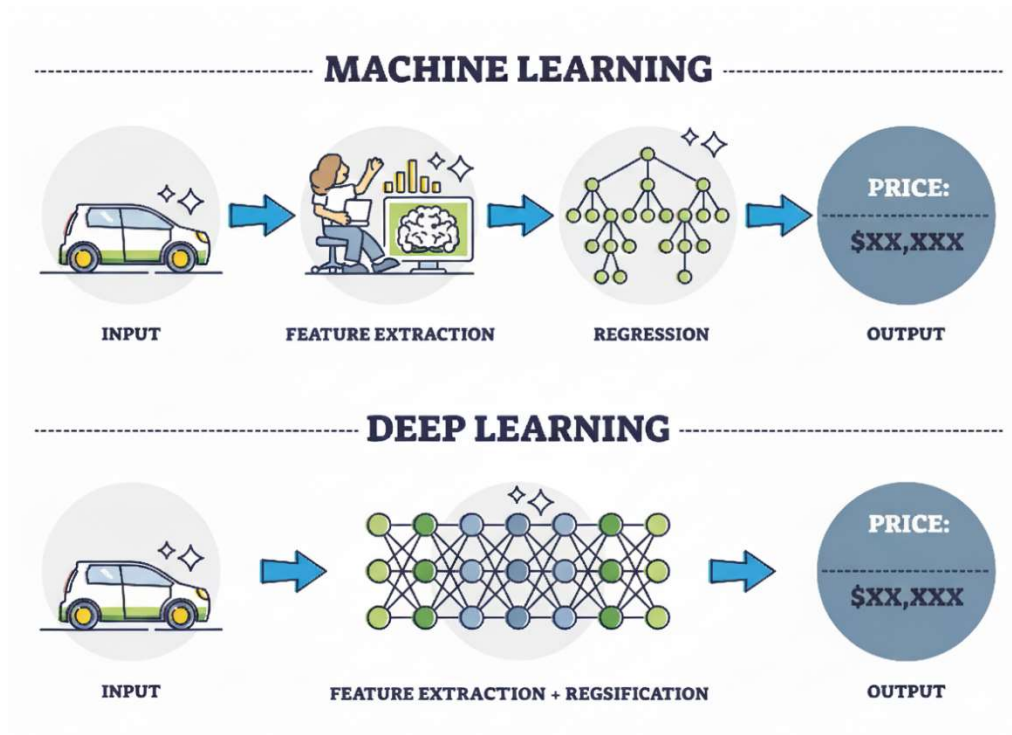


Figure 2.1.1: Core AI technologies between ML and DL role in classification.

2.1 Background

Sales forecasting is a foundational element of business operations, particularly in the retail and e-commerce sectors, where demand variability can significantly impact inventory control, financial planning, and strategic decision-making. Traditional forecasting models, such as moving averages and linear regression, have been widely used due to their simplicity and interpretability. However, these approaches often struggle to capture the nonlinear relationships, seasonal fluctuations, and external disruptions that characterize modern sales environments.

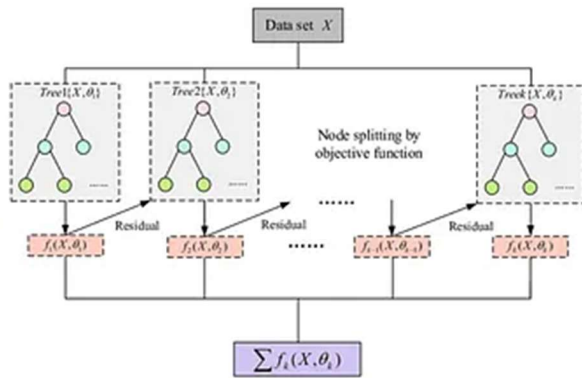
With the increasing availability of high-volume and diverse datasets, the limitations of conventional models have become more pronounced. Inaccurate forecasts can lead to overstocking, missed revenue opportunities, and inefficient resource allocation. To address these challenges, researchers have turned to machine learning and deep learning techniques, which offer greater flexibility and predictive power. Algorithms such as Decision Trees, Random Forest, Support Vector Machines, and XGBoost have shown strong performance in modeling complex dependencies [3]. Additionally, Long Short-Term Memory (LSTM) networks—a type of recurrent neural network—have demonstrated exceptional capabilities in capturing temporal dynamics and long-term dependencies in sequential data [1][2].

2.2 Machine Learning Techniques

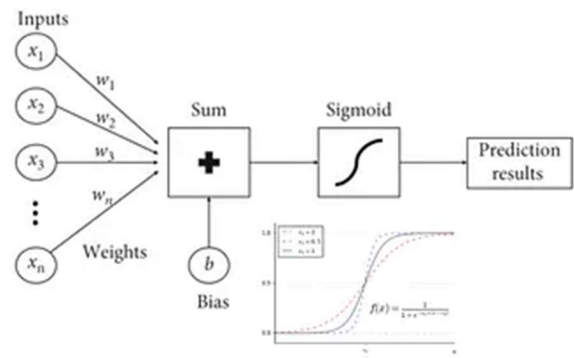
Machine Learning (ML) is a subfield of artificial intelligence that enables computer systems

2.2.1 *Extreme Gradient Boosting (XGBoost)*

Extreme Gradient Boosting (XGBoost) is a scalable and efficient ensemble learning algorithm that builds upon gradient boosting. It is particularly effective for structured data problems such as sales forecasting, where nonlinear relationships, seasonal effects, and external variables (e.g., promotions, holidays) influence demand. XGBoost builds an additive model by sequentially training decision trees, each one correcting the errors of its predecessor.



Flow chart of XGBoost



Flow chart of Logistic Regression

Figure 2.2.1.1: XGBoost.

The objective function in XGBoost combines a regularized loss function and a model complexity term:

$$Obj = \sum_{i=1}^n l(y_i, y_i^{(t-1)}) + \sum_{k=1}^K \Omega(f_k)$$

Where:

n = number of training samples

$l(y_i, y_i^{(t-1)})$: Loss function (e.g., Mean Squared Error)

$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2$: Regularization term

T : Number of leaves in the tree

w_j : Leaf weights

γ, λ : Regularization parameters

Each tree is built using the gradient and second-order derivative (Hessian) of the loss function to determine optimal splits. This second-order optimization improves convergence and stability.

XGBoost uses shrinkage (learning rate), column subsampling, and early stopping to prevent overfitting. Its ability to rank feature importance also makes it interpretable for business decisions.

In sales forecasting, XGBoost has shown strong performance across various retail datasets. For example, Massaro et al. (2021) applied XGBoost to large-scale retail data and achieved a **Mean Absolute Percentage Error (MAPE) below 10%**, outperforming traditional models and enabling more accurate inventory planning [4]

2.2.2 Prophet

Prophet is a robust time-series forecasting model developed by Meta (Facebook) that handles irregular seasonal patterns, holiday effects, and missing data. It is particularly effective for sales forecasting, where demand fluctuates due to promotions, holidays, and consumer behavior. Prophet decomposes time series into additive components: trend, seasonality, and holiday effects. [5]

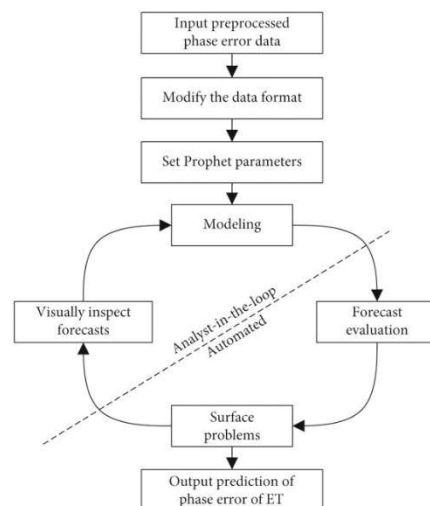


Figure 2.2.2.1: Prophet flowchart.

The forecasting function is defined as:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

Where:

- $g(t)$: Trend function (linear or logistic)
- $s(t)$: Seasonal component modeled via Fourier series
- $h(t)$: Holiday effects
- ϵ_t : Error term

Prophet supports both linear and logistic growth:

- **Linear:**

$$g(t) = kt + m$$

- **Logistic:**

$$g(t) = \frac{C}{1 + \exp(-k(t - m))}$$

Where:

- k : Growth rate
- m : Offset
- C : Carrying capacity

Seasonality is captured using a Fourier series:

$$s(t) = \sum_{n=1}^N \left[a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right]$$

Where:

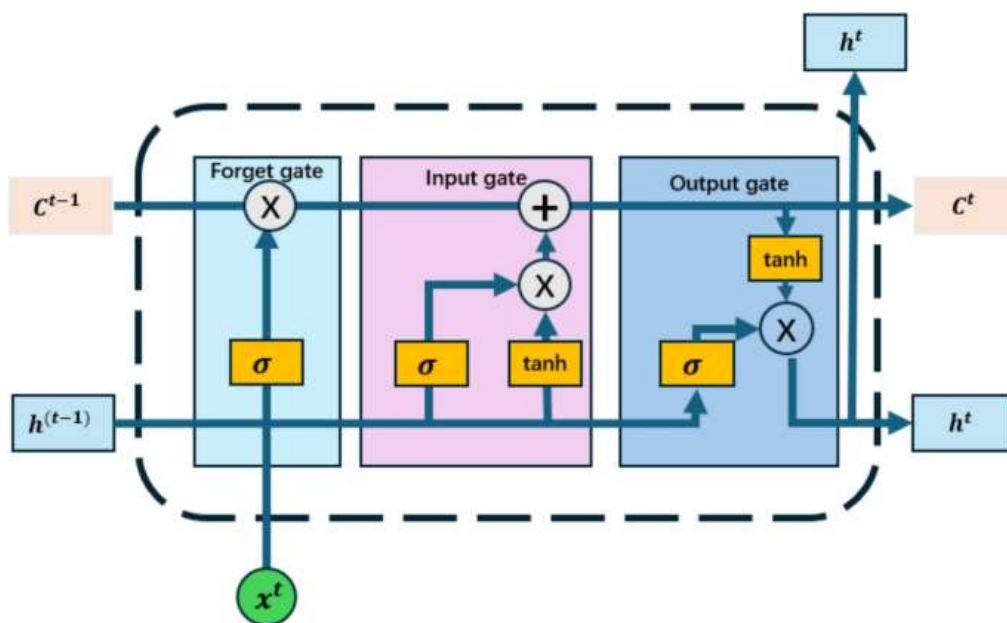
- P : Period (e.g., 365 for yearly)
- a_n, b_n : Coefficients learned from data.

2.3 Deep Learning Techniques

Deep Learning (DL) is a subfield of artificial intelligence that enables computer systems

2.3.1 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a specialized form of recurrent neural network (RNN) designed to capture long-term dependencies in sequential data. It is particularly effective for sales forecasting, where historical demand patterns, seasonality, and external factors influence future sales. Unlike traditional RNNs, LSTMs mitigate the vanishing gradient problem through gated memory cells, allowing them to learn from long sequences [5].



The LSTM cell updates its internal state using three gates: **input**, **forget**, and **output**, defined as:

$$x_t = \sigma(W_f \cdot [h_t - 1, x_t] + b_f)$$

$$i_t = \sigma(W_i \cdot [h_t - 1, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_t - 1, x_t] + b_C)$$

$$C_t = f_t \cdot C_t - 1 + i_t \cdot \tilde{C}_t$$

$$o_t = \sigma(W_o \cdot [h_t - 1, x_t] + b_o)$$

$$h_t = o_t \cdot \tanh(C_t)$$

Where:

- x_t : Input at time t
- h_t : Hidden state
- c_t : Cell state
- σ : Sigmoid activation
- $\tanh$: Hyperbolic tangent activation
- W, b : Weights and biases for each gate

2.5 Statistical Model Techniques

2.5.1 Auto Regressive Integrated Moving Average (ARIMA)

Auto Regressive Integrated Moving Average (ARIMA) is a classical statistical model used for time series forecasting. It combines autoregressive (AR), differencing (I), and moving-average (MA) components to model temporal dependencies and trends in the data. ARIMA is particularly effective for univariate forecasting tasks where the data exhibits non-seasonal patterns and can be made stationary through differencing.

The general ARIMA model is denoted as:

ARIMA(p,d,q)

Where:

- pp: Number of autoregressive terms
- dd: Degree of differencing
- qq: Number of moving average terms

2.6 Evaluation Metrics and Confusion Matrix

Evaluation Metrics:

Mean Absolute Error (MAE) : Average of the absolute differences between actual and predicted values.

$$MAE = \frac{1}{n} \sum_{i=0}^n |y_i - \hat{y}_i|$$

Where:

$n \rightarrow$ Number of observations (data points in the test set).

$y_i \rightarrow$ Actual value of sales at time i .

$\hat{y}_i \rightarrow$ Forecasted (predicted) value of sales at time i .

$|y_i - \hat{y}_i| \rightarrow$ Absolute error (difference between actual and predicted).

Root Mean Squared Error (RMSE): Square root of the average squared differences between actual and predicted values.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=0}^n (y_i - \hat{y}_i)^2}$$

Where:

$n \rightarrow$ Number of observations (data points in the test set).

$y_i \rightarrow$ Actual value of sales at time i .

$\hat{y}_i \rightarrow$ Forecasted (predicted) value of sales at time i .

$(y_i - \hat{y}_i)^2 \rightarrow$ Squared error (used in RMSE to penalize large deviations).

Mean Absolute Percentage Error (MAPE): Average absolute error expressed as a percentage of actual values

$$MAPE = \frac{100}{n} \sum_{i=0}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Where:

$n \rightarrow$ Number of observations (data points in the test set).

$y_i \rightarrow$ Actual value of sales at time i .

$\hat{y}_i \rightarrow$ Forecasted (predicted) value of sales at time i .

$\left| \frac{y_i - \hat{y}_i}{y_i} \right| \rightarrow$ Relative error (used in MAPE to express error as a percentage).[5]

Chapter 3 Related Work

3.1 Introduction

In this section, we examine recent studies that investigate the application of artificial intelligence to sales forecasting. Researchers have experimented with a range of models, from traditional machine learning (ML) to more advanced approaches, convolutional neural networks (CNN), long short-term memory (LSTM) networks. Some of these works go a step further, introducing hybrid models that blend different AI techniques or combine them with domain expertise to achieve greater accuracy. The following sub-sections break down these contributions, highlighting the progress made across each category of models.

3.2 Machine Learning Studies

Ensafi et al. (2022)[6] benchmarked forecasting models for furniture sales, comparing classical (SARIMA, Holt-Winters) and advanced methods (Prophet, LSTM, CNN). They found Stacked LSTM to be the most accurate, with Prophet and CNN also performing well. The study offers a practical model comparison but is limited to one category and does not explore hybrid combinations of the top-performing techniques.

Loukili et al. (2024) (2025)[7] conduct a direct "algorithm showdown" comparing various machine learning models for sales prediction. Their work identifies top-performing algorithms, such as tree-based ensembles or neural networks, offering a practical guide for model selection. However, the study's narrow focus on algorithm comparison may not address real-world complexities like promotional effects or data sparsity..

Hobor et al. (2025)[8] benchmark modern ML models, including Transformers and Temporal Fusion Transformers, for retail sales forecasting. They show these advanced architectures outperform conventional models by capturing complex patterns, though their implementation requires significant computational resources and expertise.

3.3 Sales Forecasting Using Deep Learning

De Castro Moraes et al. (2024)[9] pioneered a CNN-LSTM hybrid model for retail sales forecasting. Their approach, tested on real Brazilian retail data, successfully captures both short-term promotional spikes and long-term trends, outperforming traditional models. A key advantage is its automated feature learning, though they note its computational cost and data hunger as practical limitations.

Ahmadov and Helo (2023)[10] directly tackle the difficult problem of forecasting intermittent online sales using deep learning. They employ an LSTM-based model specifically designed to handle sporadic demand patterns typical in e-commerce. Their model demonstrates strong performance on a real multichannel retail dataset, effectively learning from sequences of intermittent data. This work highlights deep learning's potential for a notoriously challenging forecasting niche.

Boukrouh et al. (2025)[11] introduced a novel triple-hybrid model combining LSTM, SARIMA, and Holt-Winters for sales forecasting. Their model significantly outperformed each individual component, with MAPE improvements of up to 43%. The key advantage is its comprehensive capture of complex, seasonal, and recent trends in one framework. The main drawback is the increased complexity and integration effort required for the three-model ensemble.

3.4 Summary Related Work

Ref	Year	Dataset	Models	Results	Advantages	Limitations
Ensafi et al. [6]	2022	Public dataset of a retail store's sales history (Furniture category)	Classical: S ARIMA, Holt-Winters. Advanced: Prophet, LSTM, CNN, Stacked LSTM.	Stacked LSTM demonstrated superior performance. Prophet and CNN also showed strong results compared to classical time series models.	Provides a clear, empirical benchmark across a wide spectrum of methods—from classical to modern deep learning—for a specific retail category (furniture).	The analysis is confined to a single product category; uses an unspecified public dataset, limiting reproducibility; does not explore hybrid or ensemble techniques combining the best models.
Loukili et al. [7]	2024	Sales dataset, specifics not provided	Multiple machine learning algorithms (e.g.,	The study identifies the best-performing	Offers a direct "algorithm showdown," providing clear performance	Lack of dataset details limits reproducibility; likely does

			Random Forest, XGBoost, SVM, Neural Networks)	algorithm(s) among those tested, likely highlighting tree-based ensembles or neural networks as top contenders .	rankings to guide practitioners in model selection for sales prediction.	not address complex retail challenges like promotions, new products, or hybrid modeling approaches.
[8] Hobor et al.	2025	Retail sales data, specific source not named in reference	Modern ML Models (e.g., Transformer-based models, Temporal Fusion Transformers (TFT), N-BEATS, XGBoost, LSTM)	Demonstrated that state-of-the-art models like Transformers or TFT can achieve superior performance , especially on complex retail series with covariates.	Focuses on cutting-edge, modern ML architectures beyond classic models, highlighting performance gains from attention mechanisms and specialized design.	High computational requirements and complexity; may require extensive data preprocessing and hyperparameter tuning for practical deployment.
[9] de Castro Moraes	2024	Public data set from kaggle for retail sales from 2013-2017	CNN-LSTM(parallel / stacked)	Achieved (MAPE) of 13.43 For all and MAPE and 19.51 for 1 product	The study done upon three phases one month , three months and 1 year	
[10]Ahmadv & Helo	2023	Real-world multichannel retail dataset (intermittent online sales)	LSTM-based model	Demonstrated strong performance in forecasting sporadic, intermittent demand patterns common in	Effectively tailored for a notoriously difficult niche (intermittent demand), showing deep learning's applicability to sparse data sequences.	Focus is narrower (intermittence) ; may not explicitly model complex promotional shapes or multiscale features like a

				e-commerce.		hybrid architecture.
[11]Boukr ouh et al.	2025	Marketplace data - specific source not named in abstract	Hybrid Model: LSTM + SARIMA + Holt- Winters (Triple Exponenti al Smoothing)	Outperform ed individual component s: 9% better than LSTM, 39% better than SARIMA, and 43% better than Holt- Winters (based on MAPE).	Synthesizes strengths to capture complex patterns (LSTM), linear/seasonal trends (SARIMA), and recent trends (Holt-Winters) for highly accurate, reliable forecasts.	High complexity from integrating three distinct models; requires careful tuning and likely higher computational cost than any single model.

Chapter 4 Proposed Approach

4.1 Introduction

This chapter presents the proposed hybrid sales forecasting framework developed in this study. The approach integrates classical time-series modeling with deep learning techniques in order to exploit both linear trend modeling and nonlinear temporal pattern learning. The framework is implemented entirely in Python and is evaluated using real sales data, following the methodological objectives defined in Chapters 1–3. The main focus of this chapter is to describe the data preparation process, model architectures, training strategy, and the hybrid integration mechanism.

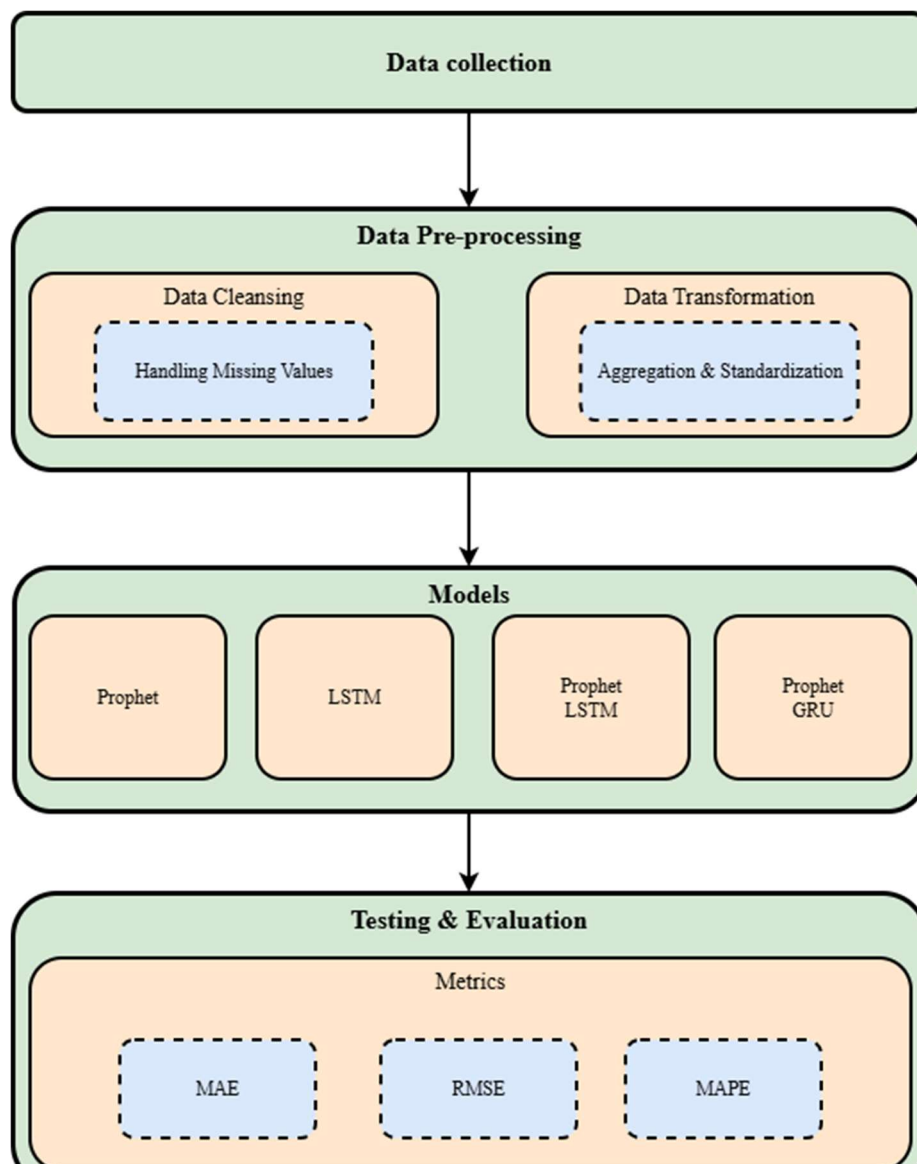


Fig. 4.1 Proposed Framework

4.2 Data Description and Preprocessing

The dataset used in this research contain 149116 and it consists of historical sales records indexed by time. Prior to model training, the data was cleaned to ensure consistency and completeness. Missing values in the date (ds) and sales (y) columns were checked and confirmed to be absent. The dataset was then split chronologically into training, validation, and test sets to preserve the temporal order and avoid information leakage. Feature engineering was applied to enrich the dataset with additional explanatory variables. These features included lagged sales values, rolling statistics, and calendar-based attributes. All numerical features were scaled to improve model convergence, particularly for deep learning models. For sequence-based models, the data was transformed into fixed-length sliding windows, with a sequence length of 30-time steps. This allowed the models to learn temporal dependencies over approximately one month of historical data.[12]

Table 4.1 Dataset description

column	Description	Datatype
transaction_id		numeric
transaction_date		date
transaction_time		time
transaction_qty		numierc
store_id		numierc
store_location		text
product_id		numeric
unit_price		numeric
product_category		text
product_type		text
product_detail		text

4.3 Baseline Forecasting Models

4.3.1 Prophet Model

Prophet was used as a strong baseline statistical model due to its ability to model trend and seasonality explicitly. The model was trained on the training portion of the data and then used to generate forecasts for the validation and test sets. Prophet outputs were aligned with the sequence-based deep learning inputs to ensure consistency during hybrid modeling. The Prophet model serves two roles in this study: as a standalone forecasting benchmark and as the linear component in the hybrid framework.

4.3.2 LSTM Model

A Long Short-Term Memory (LSTM) neural network was implemented to capture nonlinear temporal patterns in the sales data. The LSTM model was trained using the engineered sequential features. The architecture consists of an LSTM layer followed by a fully connected output layer. Mean Squared Error (MSE) was used as the loss function, and the Adam optimizer was employed for training. Early stopping based on validation loss was applied to prevent overfitting.

4.4 Proposed Hybrid Forecasting Framework

In addition to the LSTM-based hybrid model, this study also incorporates a GRU-based hybrid architecture to provide a comparative deep learning perspective. GRU was selected due to its simplified gating mechanism and lower computational complexity compared to LSTM, while still maintaining the ability to model nonlinear temporal dependencies. Including both architectures allow for a more comprehensive evaluation of recurrent neural networks within the proposed hybrid framework. The core contribution of this research is the hybrid forecasting approach that combines Prophet with deep learning models. The hybrid strategy is based on residual learning. First, Prophet is used to generate baseline forecasts. The residuals (i.e., the difference between actual sales and Prophet predictions) are then computed. These residuals represent nonlinear patterns not captured by the statistical model.

Two hybrid variants were implemented:

- **Prophet + LSTM:** An LSTM model is trained on the residual sequences to learn remaining nonlinear temporal dependencies. (parameters)

- **Prophet + GRU:** A Gated Recurrent Unit (GRU) model is trained on the residual sequences as an alternative recurrent architecture with fewer parameters.

In both cases, the final hybrid forecast is obtained by adding the deep learning–predicted residuals back to the original Prophet forecasts. This design allows the hybrid model to benefit from Prophet’s interpretability and robustness while enhancing accuracy through deep learning.

4.5 Model Training Configuration

All deep learning models were trained for a maximum of 50 epochs with early stopping enabled. Training was performed using GPU acceleration when available. The dataset sizes after preprocessing were 1,353 samples for training, 193 for validation, and 386 for testing. Performance was monitored using validation MSE to ensure generalization.

4.6 Evaluation Metrics

Model performance was evaluated using three standard forecasting metrics:

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error (MAPE)
- R²

These metrics provide complementary perspectives on forecast accuracy and are widely used in sales forecasting literature.

Chapter 5

5.1 Introduction

This chapter presents and analyzes the experimental results obtained from the proposed models. The performance of the baseline Prophet model is compared against the two hybrid approaches: Prophet+LSTM and Prophet+GRU. All results are reported on the held-out test set to ensure a fair evaluation.

5.2 Quantitative Results

Table 5.1 summarizes the performance comparison between the standalone Prophet model and the two proposed hybrid approaches.

Table 5.1: Performance Comparison of Forecasting Models

Model	MAE	RMSE	MAPE (%)
Prophet	184.12	231.04	13.84
Prophet + LSTM	181.12	228.38	13.47
Prophet + GRU	181.75	229.59	13.54

Overall, both hybrid models outperformed the standalone Prophet model across all evaluation metrics. Among the two hybrid variants, the Prophet+LSTM model achieved the best overall performance, with the lowest MAE, RMSE, and MAPE values.

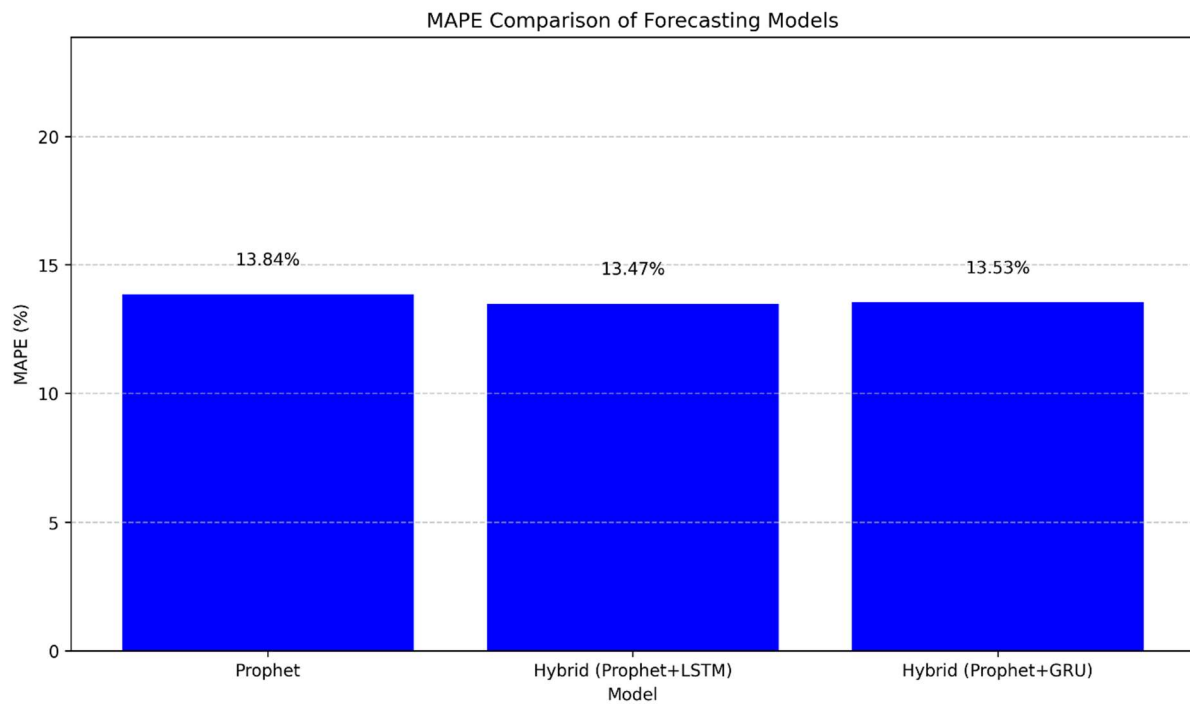


Fig 5.1 Performance Comparison of Forecasting Models

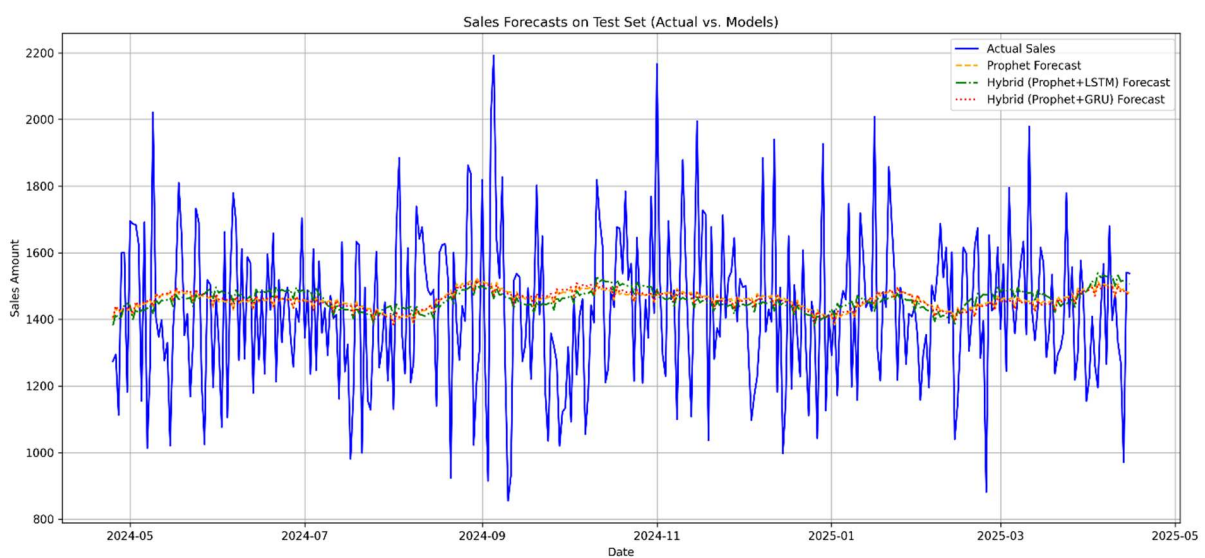


Fig 5.2 Sales Forecast on test set

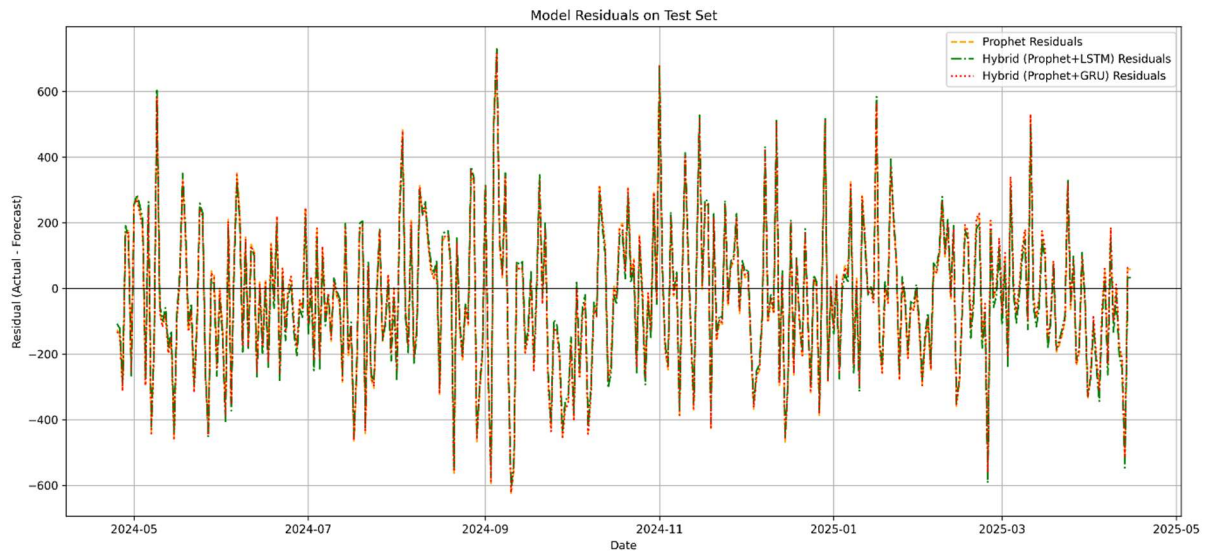


Fig5.3 Model residuals on test set

5.3 Comparative Analysis

The comparative results highlight the advantages of integrating deep learning models with a statistical baseline. Both hybrid models—Prophet+LSTM and Prophet+GRU—consistently outperformed the standalone Prophet model across all evaluation metrics. This confirms that residual-based learning is effective in capturing nonlinear patterns that are not fully modeled by traditional time-series techniques. When comparing the two hybrid approaches, the Prophet+LSTM model achieved the lowest MAE, RMSE, and MAPE values, indicating superior predictive accuracy. This performance advantage can be attributed to LSTM's more expressive memory cell structure, which is designed to capture long-term dependencies in sequential data. In contrast, the Prophet+GRU model demonstrated slightly higher error values but remained very close to the LSTM-based hybrid. GRU employs a simplified gating mechanism with fewer parameters, which often leads to faster training and reduced computational cost. Despite its simpler architecture, the GRU-based hybrid was able to learn meaningful residual patterns and significantly improve upon the Prophet baseline. Overall, the comparison suggests that while LSTM provides the highest accuracy, GRU offers a competitive and computationally efficient alternative. The choice between the two may therefore depend on the specific application requirements, such as available computational resources and real-time forecasting constraints.

5.4 Discussion

The experimental findings align with the objectives of this research and support the hypothesis that hybrid approaches can enhance sales forecasting accuracy. The improvements observed, although incremental, are meaningful in practical business contexts where small error reductions can translate into significant financial benefits. The results also confirm that combining statistical models with deep learning leads to more robust forecasting systems than relying on a single modeling paradigm.

5.5 Summary

This chapter demonstrated that hybrid forecasting models integrating statistical and deep learning approaches provide superior performance compared to standalone models. Both Prophet+LSTM and Prophet+GRU hybrids achieved lower forecasting errors than the Prophet baseline, confirming the effectiveness of residual-based hybrid learning. Among the evaluated models, Prophet+LSTM delivered the highest predictive accuracy, while Prophet+GRU offered a strong balance between accuracy and computational efficiency. These results validate the proposed hybrid framework and highlight its suitability for real-world sales forecasting applications. The insights gained from this chapter directly inform the conclusions and future research directions discussed in Chapter 6. This chapter demonstrated that the proposed hybrid Prophet-deep learning framework outperforms the standalone statistical model. The Prophet+LSTM hybrid achieved the best overall performance, validating the effectiveness of residual learning for sales forecasting. These findings provide a strong foundation for the conclusions and future research directions discussed in the next chapter.

Chapter 6 Conclusion and Future Work

References

1. Li, X., Du, J., Wang, Y., & Cao, Y. (2020, November). Automatic sales forecasting system based on LSTM network. In *2020 International Conference on Computer Science and Management Technology (ICCSMT)* (pp. 393-396). IEEE Computer Society.
2. Vavliakis, K. N., Siailis, A., & Symeonidis, A. L. (2021). Optimizing Sales Forecasting in e-Commerce with ARIMA and LSTM Models. In *WEBIST* (pp. 299-306).
3. Akande, Y. F., Idowu, J., Misra, A., Misra, S., Akande, O. N., & Ahuja, R. (2021, October). Application of XGBoost algorithm for sales forecasting using Walmart dataset. In *International Conference on Advances in Electrical and Computer Technologies* (pp. 147-159). Singapore: Springer Nature Singapore.
4. Massaro, A., Panarese, A., Giannone, D., & Galiano, A. (2021). Augmented data and xgboost improvement for sales forecasting in the large-scale retail sector. *Applied Sciences*, 11(17), 7793.
5. Brykin, D. (2024). Sales Forecasting Models: Comparison between ARIMA, LSTM and Prophet. *Journal of Computer Science*, 20(10), 1222-1230.
6. Ensafi, Y., Amin, S. H., Zhang, G., & Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning—A comparative analysis. *International Journal of Information Management Data Insights*, 2(1), 100058.
7. Loukili, M., Messaoudi, F., Ghazi, M. E., & Azirar, H. (2023, November). Predicting future sales: A machine learning algorithm showdown. In *The International Conference on Artificial Intelligence and Smart Environment* (pp. 26-31). Cham: Springer Nature Switzerland.
8. Hobor, L., Brcic, M., Polutnik, L., & Kapetanovic, A. (2025). Comparative Analysis of Modern Machine Learning Models for Retail Sales Forecasting. *arXiv preprint arXiv:2506.05941*.
9. de Castro Moraes, T., Yuan, X. M., & Chew, E. P. (2024). Hybrid convolutional long short-term memory models for sales forecasting in retail. *Journal of Forecasting*, 43(5), 1278-1293.
10. Ahmadov, Y., & Helo, P. (2023). Deep learning-based approach for forecasting intermittent online sales. *Discover Artificial Intelligence*, 3(1), 45.
11. Boukrouh, I., Idiri, S., Tayalati, F., Azmani, A., & Bouhsaien, L. (2025). A hybrid approach for sales forecasting: combining deep learning and time series analysis. *Int. J. Eng*, 38(4), 859-870.
12. <https://www.kaggle.com/datasets/ahmedabbas757/coffee-sales/discussion?sort=hotness> Last accessed: 14/12/2025

REF