

عثمان عثمان العمودي

البحث معدل.xcod.

- Final Thesis Submission after Reduce Similarity
- Thesis_(2)_AI901_Dr_Wessam_Ahmed_Third Trimester 2025 (Moodle PP)
- Midocean University

Document Details

Submission ID

trn:oid::1:3451808355

Submission Date

Dec 31, 2025, 1:35 PM GMT+4

Download Date

Dec 31, 2025, 3:40 PM GMT+4

File Name

xcod.3724495471_211621_عثمان_عثمان_العمودي_البحث_معدل_82842

File Size

1.7 MB

49 Pages

7,462 Words

46,244 Characters

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

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.

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 would like to extend my sincere appreciation to Dr. Wesam Ahmed for their support, encouragement, and expertise, which played a crucial role in this study. Dr. Wesam's thoughtful suggestions and encouragement paved the way for this thesis and further led me to investigate uncharted territories in machine learning and sales forecasting. Their supervision not only assisted in elevating the academic standards for this thesis but instilled in me a sense of importance with regards to data-informed decision-making in practical business settings.

List of Publications

Table of Contents

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

<u>Chapter 1: Introduction.....</u>	<u>4</u>
<u>1.1 Problem Definition</u>	<u>5</u>
<u>1.2 Thesis Objectives</u>	<u>6</u>
<u>1.3 Thesis Contribution</u>	<u>6</u>
<u>1.4 Thesis Organization.....</u>	<u>7</u>
<u>Chapter 2: Overview.....</u>	<u>9</u>
<u>2.1 background</u>	<u>10</u>
<u>2.2 Machine Learning Techniques.....</u>	<u>10</u>
<u>2.3 Deep Learning Techniques..</u>	<u>12</u>
<u>2.4 Statical Model Techniques..</u>	<u>13</u>
<u>2.5 Evaluation Metrics.....</u>	<u>15</u>
<u>Chapter 3: Related Work</u>	<u>18</u>
<u>3.1 Introduction</u>	<u>18</u>
<u>3.2 Machine Learning Studies</u>	<u>18</u>
<u>3.3 Deep Learning Studies</u>	<u>19</u>
<u>3.4 Summary Related Work.....</u>	<u>19</u>
<u>Chapter 4: Proposed Approach</u>	<u>24</u>
<u>4.1 Introduction.....</u>	<u>24</u>
<u>4.2 Dataset.....</u>	<u>25</u>
<u>4.3 Data Preprocessing</u>	<u>26</u>
<u>4.4 Baseline Forecasting Models.....</u>	<u>26</u>
<u>4.5 Proposed Hybrid Forecasting Framework</u>	<u>27</u>
<u>4.6 Model Training Configuration</u>	<u>28</u>

<u>4.7 Evaluation Metrics</u>	<u>29</u>
 <u>Chapter 5: Experimental Results and Discussion</u>	<u>32</u>
<u>5.1 Introduction</u>	<u>32</u>
<u>5.2 Quantitative Results</u>	<u>32</u>
<u>5.3 Comparative Analysis</u>	<u>34</u>
<u>5.4 Discussion</u>	<u>35</u>
<u>5.5 Summary</u>	<u>35</u>
 <u>Chapter 6: Conclusion and Future Work</u>	<u>37</u>
<u>6.1 Important Features</u>	<u>37</u>
<u>6.2 Future Work</u>	<u>37</u>
 <u>References</u>	<u>38</u>

List of Abbreviations

- **AI:** Artificial Intelligence
- **ML:** Machine Learning
- **DL:** Deep Learning
- **LSTM:** Long Short-Term Memory
- **XGBOOST:** Analysis of Variance
- **ARIMA:** Auto Regressive Integrated Moving Average
- **MAE:** Mean Absolute Error
- **RMSE:** Root Mean Square Error
- **MAPE:** Mean Absolute Percentage Error
- **R²:** Coefficient of Determination

List of Figures

2.1.1 Core AI technologies between ML and DL role in classification.....	9
2.2.1.1 XGBoost	11
2.3.1.1 LSTM	12
2.4.2.1 Prophet flowchart.....	14
4.1 Proposed Framework.....	24
5.1 Performance Comparison of Forecasting Models.....	32
5.2 R^2 Comparison Models	32
5.3 Model Forecasting Prediction comparison.....	33

List of Tables

3.4 Summary Related Work.....	19
4.1 Dataset description	25
4.6.1 Model architectures and hyperparameters	28
5.1 Performance Comparison of Forecasting Models.....	31

Abstract

Precision in forecasting time series data is critical in decision-making in financial, power, and supply chain applications. The manuscript evaluates the forecast accuracy of the basic Prophet model as well as the hybrid models Prophet+LSTM and Prophet+GRU on real data. The hybrid models combine the decomposition of Prophet model's trends and seasonal patterns with the strengths of recurrent neural network models in processing sequences. The results on the testing data set show that both hybrid models outperform the basic Prophet model on all considered measures, MAE, RMSE, and MAPE. Among the two hybrid models, Prophet+LSTM produces smaller error values (MAE = 119.41, RMSE = 148.32, MAPE = 8.58%, R2 = 0.58%), demonstrating its effectiveness in using the combined strengths of modeling paradigms in improving time series forecasts.

CHAPTER 1

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 ML algorithms continue to make it possible for more advanced algorithms to process larger amounts of data in a manner that has surpassed more traditional models of forecasting tools and techniques in business applications. One of the key applications of business and organizational management and operations involves sales forecasting. According to this aspect of business and organizational operations, sales business applications and tools used in predicting the future sales of their products and services have become essential in developing vehicles that can effectively predict sales and enable a company to plan appropriately based on the resulting outputs or data derived from previous sales and market trends of their products and services in the past. According to this case study and its applications in developing sales business tools and techniques in business and organizational operations and applications, more advanced algorithms that involve developing sales business tools and techniques that can

Page | 4

process larger amounts of data have emerged in light of the complex sales trends and patterns in financial and business applications and operations. One of the algorithms described in this case is XGBoost, an innovative algorithm used in many applications related to sales business tools and techniques that has the ability and capability to process larger amounts of business and financial data and applications. Its applications have spanned many areas in business and financial operations and applications related to sales business tools and techniques. XGBoost has stood out and emerged as an effective and powerful algorithm that has not only overcome limitations related to processing larger amounts of financial and business data and applications in predicting sales but has also recognized patterns that other algorithms and tools in sales business applications and operations have not seen and viewed in the past. According to a case study by Akande et al. in 2022 on predicting sales in Walmart stores based on their past sales data, the algorithm XGBoost emerged with remarkable accuracy and surpasses existing algorithms and their applications in predicting sales based on MAE, RMSE, and R^2 tests and assessments. According to this case study on the application and use of XGBoost in predicting sales in Walmart stores based on their past sales data and applications in financial and business operations and applications and tools in developing sales business applications and techniques that can process larger amounts of financial and business data and applications in.

1.1 Problem Definition

Sales forecasting is a basic element of inventory and budget planning. The existing approaches for forecasting sales, Moving Average and Linear Regression analysis among others, have been prone to the complexities associated with the dynamic and unpredictable natures of the sales environment. The limitation associated with the existing approaches to forecasting is more pronounced, especially for sectors that require precision and rapid decision implementation. One of the business sectors that have been severely affected due to the limitations of the traditional approach to forecasting is the retail business. Machine learning and deep learning offer an encouraging paradigm because they provide the ability to learn from past data and use that learning while considering a host of variables that may affect sales. While advances in artificial intelligence, machine learning, and deep learning techniques have brought forth innovative approaches towards optimizing decision-making processes in various fields such as sales and marketing, the daunting task of making precise forecasts and projections about future sales remains a pressing issue among various organizations today because conventional

methods and approaches towards making such forecasts and projections tend not to grasp the intricate patterns of sales data, which change with trending and behavioral patterns among consumers, among other variables. The lack of precise and trustworthy forecasting models makes it challenging in operational domains such as inventory management, the modeling of supply chains, and warehouse management because it may result in inventory buildup, shortages, wastage of inventory, increased operating costs, and poor customer satisfaction. Thus, there arises an immediate need for the development of more complex forecasting systems that can efficiently identify patterns from historical sales data and be dynamic enough to handle the changing marketplace. While there is hope that the use of machine learning and deep learning is beneficial in the development of patterns and nonlinear relationships, the utilization of these systems alone may not be an optimal way to make predictions using limited noisy data points.

1.2 Thesis Objectives

The main objective of the thesis is to analyze and evaluate the efficacy of machine learning and deep learning techniques in enhancing the accuracy and efficiency of sales forecasting models. Through the use of past sales data and relevant contextual variables, the research intends to create models that outperform conventional statistical models in volatile and complex settings.

1.3 Thesis Contribution

Chapter 2 outlines the theoretical background and conceptual ideas that are most closely tied to sales forecasting. This chapter discusses popular methodologies in machine learning, deep learning, and statistical modeling for time series data analysis, such as XGBoost, Prophet, LSTM, and ARIMA. This chapter forms the theoretical foundation for understanding the models discussed in this thesis.

Chapter 3 discusses and critically analyzes the related work done in the field of sales forecasting. Different studies that use machine learning, deep learning, and combinations of different methodologies are discussed. Through a comparison, research gaps are identified that create a necessity for the proposal of a new hybrid model.

Chapter 4 discusses in detail their proposed forecasting framework, as well as their experiment method. They include their data preparation processes, models, and their proposed combination

approaches of Prophet with DL models in this chapter. They also mention their training and evaluation criteria in this chapter.

Chapter 5 presents the results and discussion of the experiment. This chapter includes the quantitative comparison of the performance of statistical models, deep learning models, and hybrid models based on common assessment criteria. Chapter 5 also discusses the results and gives an interpretation of the efficiency of the hybrid method in the context of sales forecasting problems.

Concluding the thesis, chapter 6 will summarize the primary findings and contributions achieved. The limitations of the present study are defined as well as possible future studies on hybrid and data-driven sales forecasting methods.

1.4 Thesis Organization

This thesis contributes to the ever-evolving state of the art in intelligent forecasting tools by showing that ML algorithms can enhance the accuracy and explainability of sales forecasts while being scalable at the same time. By contrast, statistical models that have been prevalent in sales forecasts tend to be rigid in their assumptions and make use of a limited number of variables in their construction..

CHAPTER 2

Chapter 2 Overview

This paper discusses the utilization of machine learning and deep learning approaches to improve sales forecasting solutions. The rationale is based on empirical studies performed on brief time durations, which assess and verify the efficiency of intelligent sales forecasting tools against traditional sales forecasting approaches. For example, Li et al. 's empirical work published in 2020 demonstrated the development of a self-service sales forecasting system involving LSTM, which efficiently identified sales process intrinsic characteristics and patterns [1]. Similarly, based on their empirical work published in 2021, Vavliakis et al. introduced an integrated LSTM-ARIMA technique, achieving higher sales forecasting efficiency on an online platform compared to traditional approaches [2]. On the other hand, from their empirical work published in 2021, Akande et al. used XGBoost as the sales forecasting technique on Walmart, achieving accurate sales forecasting results while emphasizing interpretability [3]. Therefore, against the identified background, the proposed work clearly articulates its objective, which is the development of a capable and novel sales forecasting system involving preprocessing, feature development, and optimization.

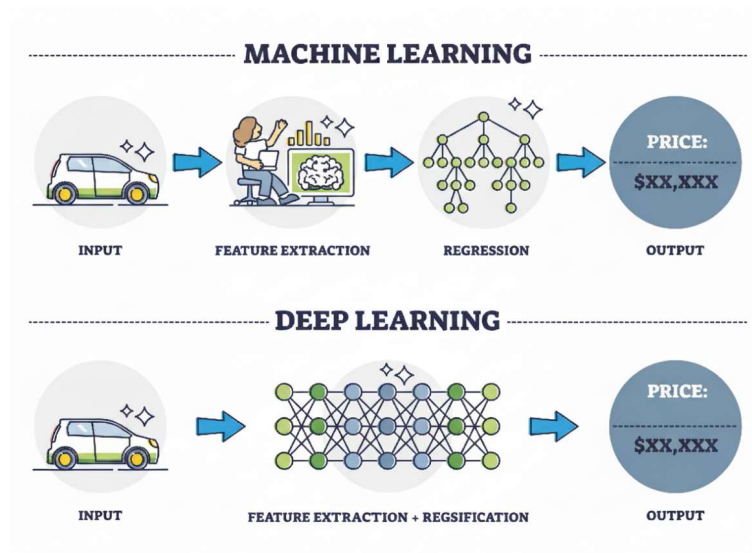


Figure 2.1.1: Core AI technologies between ML and DL role in classification.

2.1 Background

Sales forecasting is a vital aspect of business analysis, and more particularly of online businesses and retail, where the dynamics of sales and demand tend to significantly impact inventory and resource planning. Conventional approaches like moving average and linear regression have long been used, owing to their simplicity and interpretability of results. However, these approaches have tended to fare poorly and inadequately articulate the non-linearities and externalities of the modern-day environment of doing business. The emergence of big and diverse datasets representing vast areas of interest further underlines the limitations of conventional approaches. It is noted that errors of sales forecasting could result in inventory buildup, forgone revenues, and inefficient resource allocation. Concerned by these aspects, there is currently a gradually rising trend of research work underlining the need for utilizing machine learning and deep learning approaches, which are significantly more flexible and accurate. Techniques such as decision trees, random forest, support vector machines, and XGBoost have shown favorable results from the perspective of sales forecasting and articulation of complex relationships between various factors of doing business [3]. On the other hand, Recurrent Neural Network, particularly Long Short-Term Memory networks, have

shown promising results underlining their capabilities of sales forecasting of sequence data and understanding long-range dependencies therein [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)

is an advanced ensemble-based machine learning technique derived from the gradient boosting framework, designed to deliver high computational efficiency and scalability. The algorithm is well suited for structured and tabular datasets, including sales forecasting applications, where demand patterns are shaped by complex nonlinear interactions, seasonal variations, and exogenous factors such as promotional activities and calendar effects. XGBoost constructs its predictive model in a stage-wise manner by iteratively fitting decision trees, with each subsequent tree focusing on minimizing the residual errors produced by the preceding ensemble.

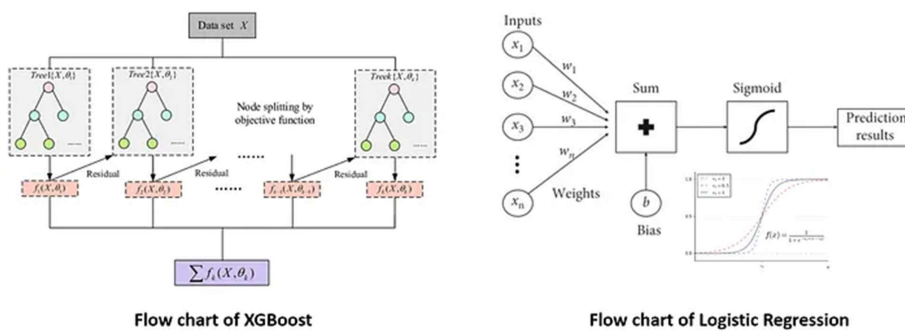


Figure 2.2.1.1: XGBoost.

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

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

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]

2.3 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) networks represent a specialized architecture within recurrent neural networks (RNNs), explicitly developed to model long-range temporal dependencies in sequential data. This capability makes LSTM particularly suitable for sales forecasting tasks, where future demand is affected by historical sales behavior, seasonal dynamics, and exogenous influences. In contrast to conventional RNN structures, LSTM employs gated memory mechanisms that regulate information flow, thereby alleviating the vanishing gradient problem and enabling effective learning across extended time horizons [5].

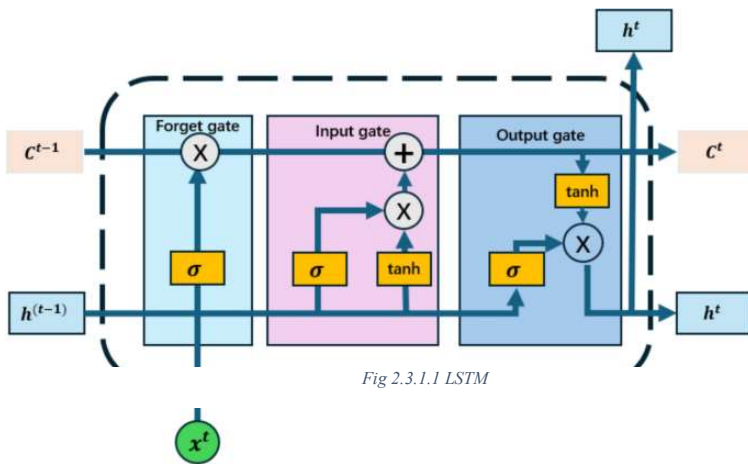


Fig 2.3.1.1 LSTM

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

$$x_t = \sigma(W_f \cdot [h_t - 1, x_t] + b_f) \quad (2)$$

$$i_t = \sigma(W_i \cdot [h_t - 1, x_t] + b_i) \quad (3)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_t - 1, x_t] + b_C) \quad (4)$$

$$C_t = f_t \cdot C_t - 1 + i_t \cdot \tilde{C}_t \quad (5)$$

$$o_t = \sigma(W_o \cdot [h_t - 1, x_t] + b_o) \quad (6)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (7)$$

Where:

x_t : Input at time t

h_t : Hidden state

$\tilde{C}_t$: Cell state

σ : Sigmoid activation

$\tanh$: Hyperbolic tangent activation

W, b : Weights and biases for each gate

2.4 Statistical Model Techniques

2.4.1 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

2.4.2 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]

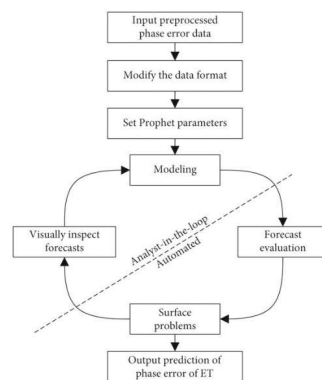


Figure 2.4.2.1: Prophet flowchart.

The forecasting function is defined as:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (8)$$

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:**

$$g(t) = kt + m \quad (9)$$

- **Logistic:**

$$g(t) = \frac{C}{1 + \exp(-k(t - m))} \quad (10)$$

Where:

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

Seasonality is captured using a Fourier series:

$$s(t) = \sum_{n=1}^N \left[a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right] \quad (11)$$

Where:

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

2.5 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.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=0}^n (y_i - \hat{y}_i)^2} \quad (12)$$

Where:

n → Number of observations (data points in the test set).

y_i → Actual value of sales at time i .

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

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

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

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (13)$$

Where:

n → Number of observations (data points in the test set).

y_i → Actual value of sales at time i .

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

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

Coefficient of Determination (R^2): commonly denoted as R^2 , is a statistical measure used to evaluate the **goodness of fit** of a predictive model. It represents the proportion of variance in the dependent variable that is explained by the independent variables in a regression or forecasting model. R^2 provides an indication of how well the model captures the underlying patterns in the data. Its value ranges between **0 and 1**, where higher values indicate better explanatory power. R^2 is mathematically defined as:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (14)$$

Where:

y_i → is the actual observed value

$\hat{y}_i$ → is the predicted value generated by the model.

$\bar{y}$ → is the mean of the observed values.

the numerator represents the residual sum of squares (RSS), while the denominator represents the total sum of squares (TSS).

Formatted: Font color: Text 1, Shadow

CHAPTER 3

Chapter 3 Related Work

3.1 Introduction

In this section, we will examine recent work undertaken to explore the application of artificial intelligence in the field of sales forecasting. Various studies have considered a range of approaches, from the use of machine learning to more recent approaches utilizing the power of convolutional and long short-term memory networks. Some studies even go ahead to propose improved solutions by combining various techniques of artificial intelligence in an efficient manner. The various subtopics ahead will highlight the advancements made in each case.

3.2 Machine Learning Studies

Ensafi et al. (2022)[6] Classical methods of furniture sales forecasting, such as SARIMA and Holt Winters, were compared to state-of-the-art models, including Prophet, LSTM, and CNN. Analysis of the findings revealed that the highest accuracy was achieved by the Stacked LSTM model, followed by the Prophet and CNN models. Even though it is a case study on the comparison of models, it neither focuses on a specific category of products nor pursues a combination of the best models.

Loukili et al. (2024) (2025)[7] The research embarks on a direct comparison, known as an “algorithm showdown,” of different machine learning models for sales forecasting. It highlights the most effective models, including tree-based ensemble models and neural networks, and acts as a guide in choosing the right models. However, the research scope in investigation might overlook the complexities of real-world data, especially promotional impact and data sparsity.-

Hobor et al. (2025)[8] They benchmark the effectiveness of modern machine learning techniques, such as Transformers and Temporal Fusion Transformers, to be applied to the forecasting of retail sales. The findings show that these modern approaches outperform traditional ones by incorporating complex patterns, but their execution is highly resource- and skill-intensive.

3.3 Sales Forecasting Using Deep Learning

De Castro Moraes et al. (2024)[9] A CNN-LSTM hybrid model was developed for the forecasting of retail sales. The proposed solution was tested using a real-world retail dataset in Brazil and proved the capability to learn short-term promotional events and long-term trends. Among the advantages is the capability to automate the learning of features, while the limitations lie in the computational complexity and data requirements.

Ahmadov and Helo (2023)[10] The research specifically tackles the difficult task of intermittent online sales forecasting and uses a deep learning method. The research uses a model based on LSTM to handle intermittent sales patterns typically found in online business environments. The proposed model performs very well on a real multichannel store dataset by learning intermittent sales examples. The research is a clear indication of the importance of deep learning in a forecast application area long proven to be very challenging.

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

						g the best models.
Loukili et al. [7]	2024	Sales dataset, specifics not provided	Multiple machine learning algorithms (e.g., Random Forest, XGBoost, SVM, Neural Networks)	The study identifies the best-performing algorithm(s) among those tested, likely highlighting tree-based ensembles or neural networks as top contenders.	Offers a direct "algorithm showdown," providing clear performance rankings to guide practitioners in model selection for sales prediction.	Lack of dataset details limits reproducibility; likely does not address complex retail challenges like promotions, new products, or hybrid modeling approaches.
Hobor et al. [8]	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.
de Castro Moraes [9]	2024	Public data set from kaggle for retail sales from 2013-2017	CNN-LSTM(parallel / staked)	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	
Ahmad ov & Helo [10]	2023	Real-world multichannel retail dataset (intermittent online sales)	LSTM-based model	Demonstrated strong performance in forecasting sporadic,	Effectively tailored for a notoriously difficult niche (intermittent demand), showing deep learning's	Focus is narrower (intermittence); may not explicitl

				intermittent demand patterns common in e-commerce.	applicability to sparse sequences.	y model complex promotional shapes or multiscale features like a hybrid architecture.
Boukro uh et al. [11]	2025	Marketplace data - specific source not named in abstract	Hybrid Model: LSTM + SARIMA + Holt-Winters (Triple Exponential Smoothing)	Outperformed individual components: 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.
Ahmed et al.[12]	2024	Retail store sales	Hybrid statistical + LSTM	Hybrid model improved MAPE by 10%	Combines interpretability and accuracy	Complex integration
Bandar a et al.[13]	2021	Multiple retail datasets	Random Forest, Gradient Boosting	Ensemble models reduced MAE by up to 15%	Robust to noise, handles nonlinearities	Requires extensive feature engineering
Borovykh et al.[14].	2022	Daily sales time series	XGBoost regression	XGBoost achieved lowest forecasting error	High accuracy, interpretable features	Not sequence-aware
Hewamalage et al.[15]	2023	Real-world retail datasets	LightGBM, ARIMA	ML models outperformed classical methods	Efficient training, scalable	Weak long-term dependency learning

Karim et al.[16]	2022	Multivariate sales data	LSTM, GRU	GRU achieved comparable accuracy to LSTM	Fewer parameters, faster training	Slightly lower accuracy
Lim et al.[17]	2021	Retail and energy datasets	Temporal Fusion Transformer	Achieved state-of-the-art accuracy	Handles multivariate time series	Complex architecture
Makridakis et al.[18]	2020	M4 competition datasets	ML regression models	ML competitive with statistical methods	Benchmark validation	No deep temporal modeling
Sezer et al.[19]	2020	Sales and demand datasets	LSTM, CNN	Deep models outperformed classical models	Strong nonlinear learning	High computational cost
Wen et al.[20]	2023	E-commerce transaction data	Deep RNN models	Reduced RMSE by 18%	Captures seasonality and trends	Requires large datasets
Zhang et al.[21]	2020	Retail chain sales data	Linear Regression, SVR	SVR outperformed linear regression in RMSE	Simple implementation, strong baseline	Limited nonlinear temporal modeling

CHAPTER 4

Chapter 4 Proposed Approach

4.1 Introduction

This chapter outlines the proposed hybrid sales forecasting model developed in this research. This model is designed to incorporate both linear and non-linear pattern identification through classic and deep learning paradigms. This model is implemented entirely in Python code and is tested on actual sales data, as is consistent with the aims of this research set out in Chapters 1–3. The salient aspect of this chapter is to describe how data processing steps, models, and training processes are implemented in order to achieve this hybrid integration.

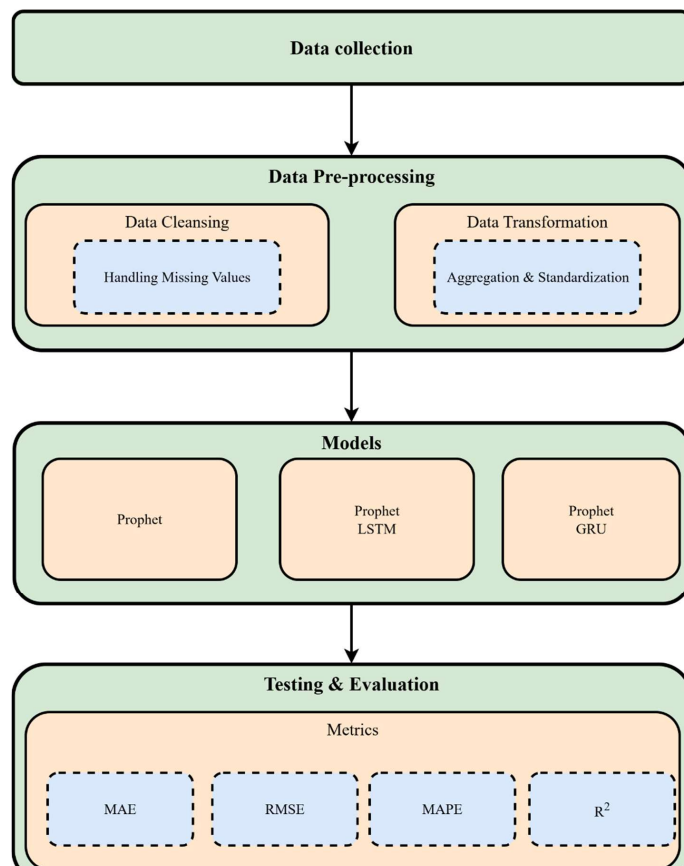


Fig. 4.1 Proposed Framework

4.2 Dataset Description

For this research, a real-world transactional data with the title “Barista Coffee Shop Sales Data” featuring 100,001 sales transactions is used. Each sales transaction is a distinct record that involves a transaction at a coffee shop with specific details relating to the time of transaction, product details, quantity sold, pricing, and geographical location of the store. This data records sales with a refined time resolution by including time and date variables that can be aggregated at a daily and higher level of time granularity for sales forecasting purposes. For this research work, the target variable is sales amount that denotes total sales derived per transaction. Owing to the transactional characteristics of the data, aggregation is required to convert the raw data into a time series format congenial with sales forecasting algorithms and techniques. This data has realistic sales patterns based on customer demand cycles and product types at different stores, making this data apt for assessment of statistical-based, deep learning-based, and hybrid sales forecasting algorithms and techniques.[22]

Table 4.1 Dataset description

column	Description	Datatype
order_date	Transaction date	Date
order_time	Transaction time	Time
product_category	Category of the product (e.g., Coffee, Tea)	Categorical
qty	Sales qty	Intege
unit_price	unitprice	Float
store_id	Store Id	text
store_location	The location of the store	text
customer_location	Describe the location of the customer	text
customer_id	Customer Id	text
Customer_age	The age of the customer	Integer
Customer_gender	Category of the customer (e.g., Male, Female,Other)	Categorical
loyalty_member	Category of the Loyalty (e.g., True, False)	Categorical
is_repeat_customer	Category of the Repeated customer (e.g., True, False)	Categorical
customer_discovery_source	Describe how did the customer find out about the shop	text
payment_method	Describes the payment method	text
discount_applied	Discount Amount	Float
total_amount	Amount of product	Float
promo_applied	Describes if the promo applied	Categorical
promo_type	Describes the promo type	Categorical

4.3 Data Preprocessing

Before training these models, several preprocessing steps had been performed to increase the quality and compatibility of these datasets with the forecasting models. First, these datasets had been checked for missing values on both the date and sales columns. Since no significant number of missing values was found, it became possible to continue with the analysis without conducting any imputation. Then, some data transformation methods had been used to optimize these models. In deep learning models, to increase convergence speed, a sales normalization technique had been used on these sales values using a suitable scaling approach. Another approach, which utilized lag values and rolling statistics, had been used to enable these machine learning models to access historical context information for better predictions. For sequence models like LSTM/GRU networks, these sales values had been represented as fixed-length input sequences with a sliding window approach. A fixed length of 30 had been considered for these input sequences to enable these networks to capture temporal dependencies for one month of historical sales information. These preprocessing steps had been uniformly applied on all training, validation, as well as test datasets.

4.4 Baseline Forecasting Models

4.4.1 Prophet Model

Prophet was used as a strong baseline statistic model for its deliberate ability to handle trend and seasonality. The model was then trained on the training part of the data and used for making predictions on the validation and test data. The output of the Prophet model was then matched to the sequence-based input format of the deep learning architecture to ensure consistency across the hybrid model framework. In this research, the Prophet model plays a dual role: it serves both as a forecast model on its own and as the linear model part of the hybrid model framework.

4.4.2 LSTM Model

For addressing nonlinear patterns in the sales data, an LSTM (Long Short-Term Memory) neural network model has been used. For training this model, engineered features with a sequence aspect were utilized. The network architecture consists of an LSTM layer in combination with an output layer that is fully connected. Mean squared error functions as the

loss function for this network along with an Adam optimizer. For preventing overfitting, early stopping on the basis of validation loss is applied.

4.5 Proposed Hybrid Forecasting Framework

Aside from the use of the hybrid model based on LSTM, the study also involves the use of another hybrid model based on GRU in order to obtain a comparative insight in the use of deep learning. GRU was chosen over LSTM due to its simpler gate control and less computational complexity in order to capture the same capacity in modeling nonlinear temporal relations. The use of two separate models adds towards an in-depth study of the use of recurrent neural networks in the proposed hybrid model. The main contribution described in the study involves the use of the hybrid forecasting model that combines Prophet and deep models. In the hybrid model, the use of residual learning is considered, in which Prophet models were first used for baseline forecasting, and the residual defined by the difference between observed sales and the Prophet forecast was computed.

Two forms of hybridization were used:

- Prophet + LSTM: For the remaining nonlinear patterns, the model uses an LSTM on the residual sequences.
- Prophet + GRU A Gated Recurrent Unit (GRU) network is also developed to learn from the residuals, utilizing an architecture with fewer model parameters.

In both versions, the final combined forecast is derived by summing the residuals predicted by deep learning to the original forecasts in the Prophet model. This also makes possible the use of interpretability and robustness in the combined model by incorporating those qualities of Prophet.

4.6 Model Configuration

Table 4.6.1 Model architectures and hyperparameters

Category	Model	Parameter	Value
Baseline	Prophet	Yearly seasonality	Enabled
		Weekly seasonality	Enabled
		Daily seasonality	Disabled
		External regressors	promo, discount, avg_age
		Training data	70% of time series
		Prediction horizon	Test set (15%)
		Residual window length	7 days
Hybrid A	Prophet + LSTM	Input features	1 (Prophet residuals)
		LSTM units	50
		Activation	ReLU
		Output layer	Dense(1)
		Optimizer	Adam
		Loss function	Mean Squared Error
		Batch size	16
		Max epochs	100
		Early stopping	Patience = 5
		Residual window length	7 days
		Input features	1 (Prophet residuals)
		Hidden units	128
Hybrid B	Prophet + GRU	GRU layers	2
		Dropout	0.3
		Optimizer	Adam (lr = 0.001)
		Loss function	Mean Squared Error
		Batch size	16
		Max epochs	50
		Early stopping	Patience = 8
		Residual window length	7 days
		Input features	1 (Prophet residuals)
		Hidden units	128

4.7 Evaluation Metrics

In order to make an objective assessment and comparison of the predictive ability of the proposed models, three popular error measurement criteria, namely, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), were used in this research work. These criteria are most widely used in sales forecasting problems as they provide complementary information about model accuracy and validity at different levels. The MAE values represent the average absolute difference between actual and predicted values, and it does not provide any scale-free information about model accuracy as MAE is measured in the same scale as that of the actual phenomenon under investigation. This makes MAE particularly useful in making an assessment about the actual impact of inaccuracy in predictions on inventory management decisions in organizations. MAE does not give any special emphasis on different errors, and it is least affected by high values. The RMSE value measure is based on the square root of the average of the squares of the prediction errors, and by taking square values, RMSE favors higher absolute values. This makes RMSE particularly useful in making overall model validity assessments in a sales forecasting problem where high values of inaccuracy in predictions can result in high financial losses in businesses, especially due to missed sales opportunities, and RMSE is most commonly used in determining overall model validity during peak seasons. The MAPE values provide scale-free quantitative information about model accuracy, as MAPE measures model accuracy in absolute percentage values, and MAPE is most useful in making inter-model comparisons about model accuracy, particularly in inter-company comparisons, and MAPE is most preferred by organizational management in making model assessment decisions. Besides the error-based measures, the Coefficient of Determination (R^2) was also employed to evaluate the explanatory roles of the forecasting models. This measure indicates the percentage of variation in actual sales that the forecasting model is able to explain. The values of this measure always range from 0 to 1, and high values always indicate an accurate relationship between actual and forecasted values. The closer the value to 1, the better the model's ability to explain the underlying behavior in the sales series. The values closer to zero indicate poor ability to explain. The R^2 measure has been widely used in sales forecasting to investigate the ability of the forecasting models to explain the overall demand behavior without necessarily relying on the accuracy of each individual observation. In being able to integrate R^2 in addition to MAE, RMSE, and MAPE in the evaluation, this paper offers more than just a full comparison between models. Though MAE,

Page | 30

RMSE, and MAPE can measure the accuracy and extent of errors in prediction, R^2 can measure, in essence, how well the model explains the variability in sales. With all four metrics in play, it would be possible to make an apt comparison between statistical models, deep models, and hybrid models.

Formatted: Normal (Web), Justified, Space After: 0 pt,
Line spacing: 1.5 lines

|

CHAPTER 5

Chapter 5 Experimental Results and Discussion

5.1 Introduction

This chapter entails the analysis of experimental outcome based on the proposed models. For comparison purposes, the performance of the original Prophet model is analyzed alongside two hybrid models: Prophet+LSTM and Prophet+GRU. All these are based on their performance on the test dataset.

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 (%)	R ²
Prophet	131.38	160.21	9.72 %	0.51
Prophet + LSTM	119.41	148.32	8.58	0.58
Prophet + GRU	119.73	148.37	8.71	0.58

In general, both hybrid models performed better than the standalone Prophet model on all evaluation criteria. Of the two hybrid models, the best performer was the Prophet+LSTM model, as reflected in terms of minimum MAE, RMSE, MAPE and R².

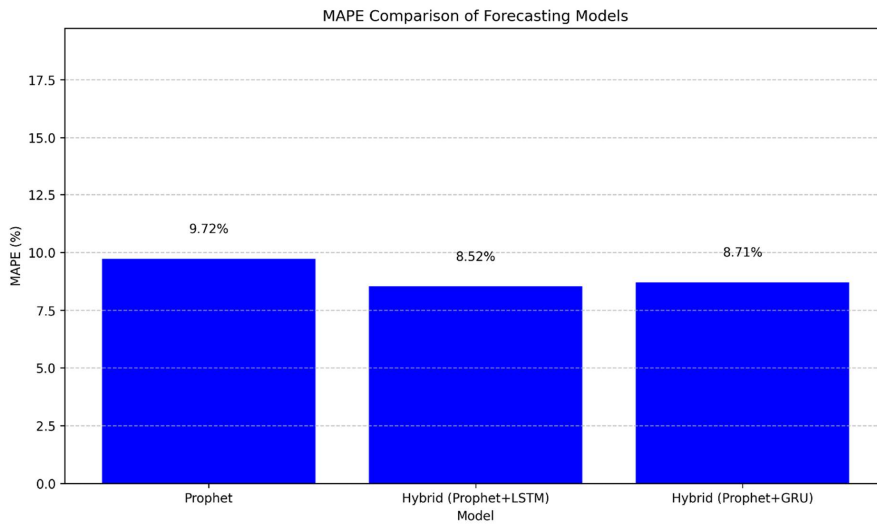


Fig 5.1 Performance Comparison of Forecasting Models

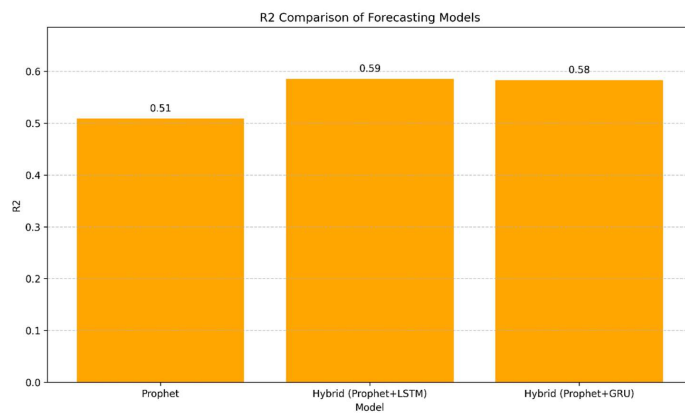


Fig 5.2 R² Comparison Of Forecasting Models

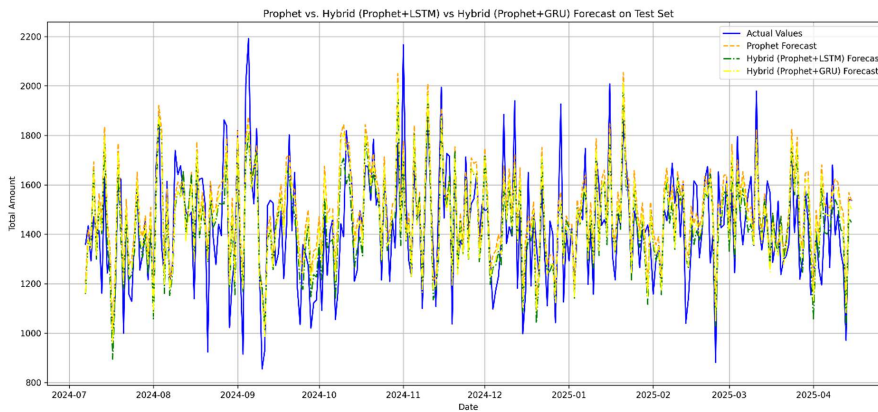


Fig5.3 Model Comparison

5.3 Comparative Analysis

The results of these comparisons highlight the need and importance of combining a deep learning technique with a statistical approach. Both of these combined setups, Prophet+LSTM and Prophet+GRU, perform better than the separate Prophet model on all fronts of comparison. This is consistent with and supports the importance of learning from residuals to overcome and compensate for deficiencies in prevailing time series solutions. On comparing these combined approaches to handling time series problems, the best results in terms of MAE, RMSE, and MAPE are obtained by Prophet+LSTM. The better performance of this model may be attributed to its complex memory cell structure design, which is efficient for dealing with long-term dependencies within sequential information. The relatively high errors shown by the combined Prophet+GRU model, although closer to those of its corresponding LSTM-based version, may be due to its more efficient gate handling and lower number of parameters. Despite such a difference in results, however, this combined GRU model is able to extract valuable features from residuals and outperform significantly its separate Prophet version. As such, although the best accuracy is from its LSTM-based version, another efficient alternative is proposed with comparable performance characteristics. The choice of either GRUs or LSTMs would be a function of application requirements.

5.4 Discussion

The empirical evidence confirms the goals of this work, as well as the hypothesis that hybrid methods have the ability to enhance the accuracy of sales forecasts. The improvements, although marginal, are of great worth when considering that even small percentages of error differences can result in significant gains for businesses. Finally, the results indicate that the combination of statistical methods with deep learning is able to produce more accurate models than the others.

5.5 Summary

This chapter has revealed that hybrid forecasting models based on combining statistical approaches and deep learning methods demonstrate better performance compared to single models. Firstly, both of the hybrid models, namely Prophet+LSTM and Prophet+GRU, have been able to produce better forecasting performance compared to the baseline Prophet model, thus affirming that the application of residual-based hybrid learning is beneficial and effective. Among all models tested within this chapter, the Prophet+LSTM model has demonstrated the highest forecasting accuracy, while the other hybrid model, namely Prophet+GRU, has offered a good trade-off between accuracy and efficiency. The conclusions and future work mentioned within Chapter 6 have been directly supported by this chapter and its findings, and this chapter has offered strong confirmation that the hybrid model based on combining deep learning and Prophet outperforms the standalone statistical model, and that the model with LSTM is better and thus supports residual-based hybrid learning for sales forecasting effectively.

Chapter 6

Chapter 6 Conclusion and Future Work

6.1 Conclusion

From the comparative analysis, the impact of combining the Prophet model with recurrent neural networks is that there is a significant improvement in the prediction capability compared to the use of the Prophet model alone. The two combined models, namely the Prophet+LSTM model and the Prophet+GRU model, result in a decrease in the error values for all metrics. This result tends to show that the two combined models are capable of detecting the nonlinear relationships in the data that the individual Prophet model is incapable of. Between the two combined models, the Prophet+LSTM model performs better. This result tends to show that the LSTM model performs well when combined with the trend and seasonality components in the Prophet model.

6.2 Future Work

Future research directions could be to further extend the hybrid forecasting paradigms using attention mechanisms, transformers, or other state-of-the-art sequence modeling approaches. Research directions related to ensemble techniques to combine different hybrid patterns can improve robustness and help to reduce uncertainty within the models. It can further be researched how hybrid paradigms can be applied on multivariate time series, high-frequency series, or irregular patterns to validate their practical performance. The hyperparameter optimization, along with computational complexity, will play a crucial role regarding the application of hybrid forecasting paradigms.

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. Ahmed, R. S., Hasnain, M., Mahmood, M. H., & Mehmood, M. A. (2024). Comparison of deep learning algorithms for retail sales forecasting. *IECE Transactions on Intelligent Systematics*, 1(3), 101-115.
13. Bandara, K., Bergmeir, C., & Smyl, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert systems with applications*, 140, 112896.
14. Borovykh, A., Bohte, S., & Oosterlee, C. W. (2017). Conditional time series forecasting with convolutional neural networks. *arXiv preprint arXiv:1703.04691*.
15. Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting*, 37(1), 388-427.
16. Karim, F., Majumdar, S., Darabi, H., & Chen, S. (2017). LSTM fully convolutional networks for time series classification. *IEEE access*, 6, 1662-1669.

17. Lim, B., Arik, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International journal of forecasting*, 37(4), 1748-1764.
18. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of forecasting*, 34(4), 802-808.
19. Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied soft computing*, 90, 106181.
20. Wen, Q., Sun, L., Yang, F., Song, X., Gao, J., Wang, X., & Xu, H. (2020). Time series data augmentation for deep learning: A survey. *arXiv preprint arXiv:2002.12478*.
21. Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, 14(1), 35-62.
22. <https://www.kaggle.com/datasets/yashparab/barista-coffee-sales-data-for-eda-csv/data>
Last accessed:20/12/2025