WEEKLY TOTAL SALES ANALYSIS AND FORECASTING OF CAFÉ AND RESTAURANT USING MACHINE LEARNING

NATTHAPONG JITJAKOOL

AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE PROGRAM IN DATA SCIENCE AND ANALYTICS
SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2024

Title	WEEKLY TOTAL SALES ANALYSIS AND FORECASTING OF CAFÉ AND RESTAURANT USING MACHINE LEARNING
Student	NATTHAPONG JITJAKOOL
Student ID	66056023
Degree	Master of Science Program in Data Science and Analytics
Academic Year	2024
Advisor	Associate Professor Dr. NACHAYADAR KAMOLMITISOM

ABSTRACT

Accurate sales forecasting is a crucial aspect of business decision-making, particularly in the food and beverage industry, where demand fluctuations are influenced by seasonal trends, weather conditions, and promotional activities. This study explores the application of various machine learning and statistical models, including Facebook Prophet (Prophet), AutoRegressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM), and Random Forest (RF), to predict weekly total sales for a coffee shop. The primary objective is to determine the most effective forecasting method for improving business operations, inventory management, and strategic planning.

The dataset consists of historical sales records from February 2023 to September 2024, incorporating key attributes such as date, weather, product sales, pricing, and employee allocation. Through Exploratory Data Analysis (EDA), significant sales patterns are identified, including the impact of weather conditions, holidays, and day-of-week variations on consumer demand.

The performance of the forecasting models is evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The results indicate that Prophet demonstrates the highest accuracy in capturing seasonal trends, while ARIMA performes well in short-term forecasts but struggled with long-term predictions. RF, despite its ability to handle complex relationships, produces overly smooth forecasts, making it less effective for time series prediction.

A comparative analysis of model results suggests that the Prophet model provides a balanced approach, leveraging seasonality, trend decomposition, and external factors. The final prediction for week 86 indicates a projected sales increase, aligning with historical seasonal trends and external conditions.

Keywords: Sales Forecasting, Time Series Analysis, ARIMA, Prophet, LSTM, Random Forest.

ACKNOWLEDGEMENTS

This independent study has been successfully completed with the generous guidance and invaluable support of my advisor, Assoc. Prof. Dr. Nachayadar Kamolmitisom, whose expertise and dedication have been instrumental in shaping the direction of this research. Her insightful feedback, continuous encouragement, and invaluable advice have greatly contributed to the successful completion of this study. I am deeply grateful for her time and unwavering support throughout this journey.

I would also like to extend my heartfelt gratitude to all faculty members and guest lecturers who have imparted their knowledge and insights throughout my academic journey. Their teachings have provided me with a strong foundation in data science, machine learning, and business analytics, enabling me to apply these concepts effectively to this research on café sales forecasting. Additionally, I deeply appreciate the collaboration and assistance from many individuals who have supported me from the initial stages to the completion of this study. Special thanks go to my peers for their invaluable support in academic discussions, data analysis, programming techniques, model evaluation, and sharing resources that have significantly contributed to the success of this work. Their encouragement and teamwork have made this research journey more enriching and fulfilling.

Lastly, I am profoundly grateful to my family, especially my parents, for their unconditional love, encouragement, and unwavering belief in my potential. Their constant support and emphasis on the value of education have been my greatest source of motivation throughout this academic endeavor.

NATTHAPONG JITJAKOOL

TABLE OF CONTENTS

CONTENT **PAGE**

TABLE OF CONTENTS (CONT.)

CONTENT **PAGE**

TABLE OF CONTENTS (CONT.)

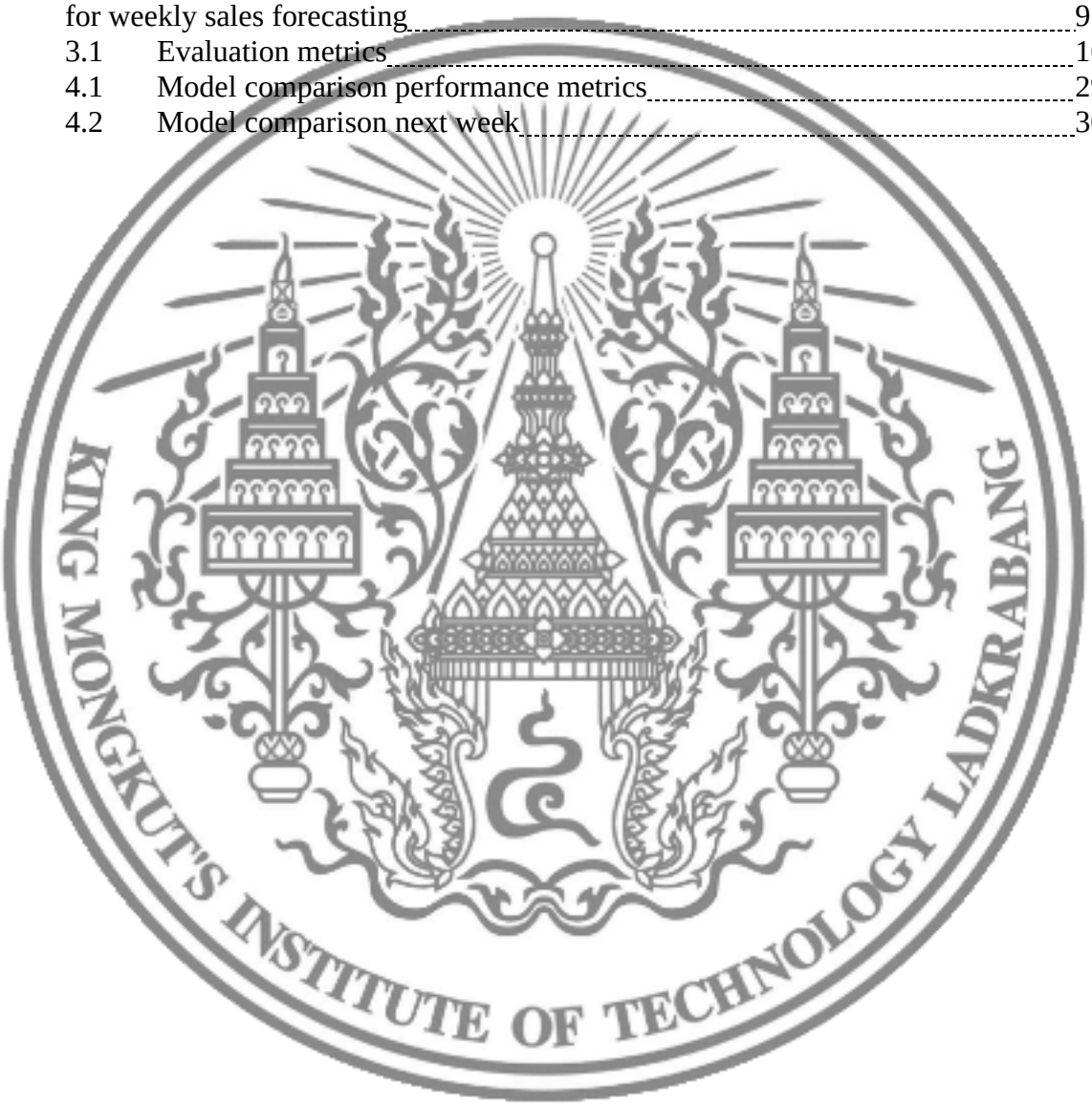
CONTENT **PAGE**

TABLE OF CONTENTS (CONT.)

CONTENT **PAGE**

LIST OF TABLES

TABLE **PAGE**



LIST OF FIGURES

FIGURE **PAGE**

Chapter 1
Introduction

1.1 Background and Significance of the Problem

In today's era, the café business has become an essential part of consumer culture and leisure activities. Cafés not only serve beverages such as coffee and tea but also offer desserts like cakes and pastries, along with snacks and main courses, providing a diverse selection for customers. The rapid growth of the café industry has intensified market competition, making effective business management increase dependent on accurate and up-to-date data for strategic planning.

Café management requires analyzing and forecasting sales, a process that enables business owners to identify market trends and adjust strategies to meet customer demand in a timely manner. Accurate sales forecasting not only helps with efficient inventory management but also aids in procurement planning, workforce allocation, and promotional strategies aligned with seasonal trends and weather conditions. These planning efforts are crucial for maintaining a competitive edge in the café industry.

Integrating Machine Learning technology into weekly café sales analysis and forecasting is an intriguing approach. Machine Learning can efficiently process large amounts of data, enhancing the accuracy of predictions. Key sales data, such as date, day of the week, weather conditions, product ID, product name, product group, product category, average cost, average price, average profit, quantity sold, total sales before discounts, total cost, discounts, total revenue, net profit, branch, holidays, and employees, serve as the foundation for developing Machine Learning models. These models are used to analyze sales trends and improve decision-making across all aspects of café business operations.

1.2 Research Objectives

1.2.1 To develop a Machine Learning model for weekly café sales forecasting.

1.2.2 To compare the performance of different Machine Learning models in sales forecasting.

1.2.3 To reduce human resource dependency in weekly sales data analysis and improve café management efficiency.

1.3 Scope of Research

This research utilizes actual sales data from a café, collected from February 1, 2023, to September 17, 2024 (85 weeks and 36,128 transactions). In addition to beverage-related menu items, the café also offers popular desserts such as cakes and various food items. The dataset includes the following attributes:

1. Date
2. Day of weeks
3. Weather conditions
4. Product ID
5. Product name
6. Group
7. Category
8. Average Cost
9. Average Selling Price
10. Average Profit
11. Quantity Sold
12. Total Sales before Discounts
13. Total Cost
14. Discount
15. Total Sales Revenue
16. Net Profit
17. Branch
18. Holidays
19. Employees

These data attributes will be analyzed and used to develop Machine Learning models for weekly sales forecasting.

1.4 Research Methodology

To develop a Machine Learning model for weekly sales forecasting, the research follows the following steps:

1.4.1 Define the research framework and outline the implementation plan.

1.4.2 Study traditional sales forecasting methods and Machine Learning techniques.

1.4.3 Collect data from the café's sales management system.

1.4.4 Perform data exploration and preparation before feeding it into the models.

1.4.5 Develop Machine Learning models using sales data.

1.4.6 Evaluate and compare the effectiveness of different models for weekly sales forecasting.

1.4.7 Summarize findings and provide recommendations based on the research objectives.

1.5 Expected Benefits

1.5.1 The ability to develop an efficient Machine Learning model for weekly café sales forecasting.

1.5.2 Insights into the differences in performance among various Machine Learning models for sales forecasting.

1.5.3 The ability to leverage Machine Learning models to reduce human effort in weekly sales data analysis, improving operational efficiency in café management.

1.5.4 Improve and develop sales strategies using data analytics.

Chapter 2
Theory and literature reviews

Before diving into the details of forecasting models and methodologies, this chapter provides an overview of theoretical concepts and previous research that are fundamental to understanding sales forecasting in the coffee shop industry. The topics covered include sales forecasting principles, influencing factors, data structuring for time series forecasting, and an in-depth review of traditional and machine learning-based forecasting models. Additionally, this chapter explores relevant external variables such as weather conditions, promotions, and holidays that can impact consumer demand and sales trends.

By understanding these foundational concepts, we can establish a strong basis for selecting appropriate forecasting techniques and improving predictive accuracy. The following sections detail the importance of sales forecasting, traditional and modern forecasting methods, data preparation techniques, and theoretical background of the forecasting models used in this study.

2.1 Fundamental Theoretical Concepts

2.1.1 The Importance of Sales Forecasting in Coffee Shop Businesses

Sales forecasting plays a crucial role in inventory management, workforce planning, and strategic decision-making for coffee shop businesses. Accurate sales predictions enable businesses to reduce stock shortages (stockouts) and prevent excess inventory (overstock), which directly impacts operational costs and profitability.

2.1.2 Factors Influencing Sales in Coffee Shops

Several factors can significantly influence sales performance in a coffee shop, including:

- Weather Conditions: Adverse weather, such as heavy rainfall, may result in reduced customer foot traffic.
- Promotions: Discount offers or promotional campaigns can stimulate sales during low-demand periods.
- Holidays: Special holidays or festive periods may lead to an increase in customer visits and higher sales volumes.

2.2 Sales Forecasting Models

2.2.1 Traditional Forecasting Methods

Traditional forecasting models commonly used for sales predictions include:

- ARIMA (Auto Regressive Integrated Moving Average): A statistical time series model suitable for datasets with trends and seasonal patterns.

2.2.2 Machine Learning-Based Forecasting Models

Machine learning approaches enhance forecasting accuracy compared to traditional methods. The primary models used in this study include:

- Prophet: Capable of capturing trends and seasonal effects, making it suitable for retail sales impacted by holidays and special events.
- LSTM (Long Short-Term Memory): A neural network-based model adept at learning sequential dependencies in time series data.

- Random Forest: A decision tree-based ensemble learning method that can identify complex relationships among variables.

2.3 Understanding Data for Sales Forecasting

2.3.1 Types of Data Used in the Analysis

- Sales Summary: Aggregated weekly sales data, serving as the primary dataset for forecasting.
- Customer Data: Information on customer demographics and behavior to analyze purchase patterns.
- Product Summary: Sales data categorized by individual products to facilitate product-level demand forecasting.
- Shift Summary: Staffing information to optimize workforce allocation based on projected sales.
- Weather Data: External environmental factors that may influence sales trends.

2.3.2 Structuring Data for Time Series Analysis

- Transforming raw sales data into weekly intervals to align with forecasting models.
- Integrating external factors, such as weather conditions and promotional campaigns, to enhance predictive accuracy.

2.4 Data Preparation for Forecasting

2.4.1 Data Cleaning

- Handling Missing Values: Identifying and addressing incomplete or inconsistent data.
- Data Transformation: Converting raw data into a structured format suitable for input into forecasting models.

2.4.2 Data Enrichment

- Integrating Data from Multiple Sources: Combining external datasets, such as weather reports and holiday schedules, with sales records to improve forecasting accuracy.

2.5 Theoretical Background of Forecasting Models

This section presents the theoretical foundation behind the four forecasting models used in this study: FB Prophet, ARIMA, LSTM, and Random Forest. Each model is based on different statistical and machine learning principles, making them suitable for specific forecasting applications.

2.5.1 FB Prophet (Facebook Prophet)

Theory and Working Mechanism

FB Prophet is an additive time series model developed by Facebook to provide robust, interpretable, and automated forecasting. It is based on the following core components:

1. Trend Component: Captures long-term growth and structural changes. Prophet uses a piecewise linear or logistic growth model with changepoints to detect shifts in trends.

2. Seasonality Component: Models periodic fluctuations in data, such as daily, weekly, or yearly seasonality, using Fourier series representation.

3. Holiday Component: Explicitly incorporates external event effects (e.g., promotions, holidays) to adjust sales predictions accordingly.

4. Error Term: Accounts for irregular variations in data that cannot be explained by other components.

Key Assumptions

- Sales follow a combination of trends + seasonality + holiday effects + noise.
- The model allows for automatic changepoint detection, meaning it can adapt to shifts in business trends.
- Works best when sales data exhibit strong seasonal and holiday-related variations.

Mathematical Representation

Prophet decomposes a time series into:

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

where:
g(t) is the trend function,
s(t) is the seasonal function,
h(t) is the effect of holidays,
ϵ(t) is the error term.

Relevance to Sales Forecasting in Coffee Shops

FB Prophet is well suited for forecasting coffee shop sales, particularly when there are strong seasonal patterns (e.g., weekends, holiday periods) and external events (e.g., promotional campaigns) that influence customer demand.

2.5.2 ARIMA (AutoRegressive Integrated Moving Average)

Theory and Working Mechanism

ARIMA is a statistical time series model that captures trends, seasonality, and randomness in a dataset. It is widely used in economic and business forecasting due to its mathematical rigor. ARIMA is represented as ARIMA (p, d, q),

where: ***p*** is the number of autoregressive terms,
d is the number of different steps to make the series stationary,
q is the number of moving average terms.

The model consists of three key components:

1. Autoregression (AR): Uses past values to predict future values.
2. Differencing (I): Removes trend or seasonality by computing differences between observations.
3. Moving Average (MA): Models the error terms as a linear combination of past errors.

Key Assumptions

- The time series should be stationary (constant mean and variance over time).
- Works best when the dataset has clear linear trends and patterns.
- Limited in capturing complex nonlinear relationships in the data.

Mathematical Representation

The ARIMA model is defined as

$$y_t = \alpha + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t$$

where: y_t**:** is the value we want to forecast at the current time *t*,

y_{t-j}: are the lagged values of *y*,

ϕ_i are the AR coefficients,

θ_j are the MA coefficients,

ϵ_t is the error term at time *t*,

ϵ_{t-j} are the lagged error terms from previous time steps.

Relevance to Sales Forecasting in Coffee Shops

ARIMA is highly effective in modeling stable and predictable sales trends in coffee shops. However, it may struggle with external factors like promotions, weather, and holidays, which need to be manually incorporated into the model.

2.5.3 LSTM (Long Short-Term Memory)

Theory and Working Mechanism

LSTM is a deep learning model based on Recurrent Neural Networks (RNNs) designed to capture long-term dependencies in sequential data. Unlike traditional time series models, LSTMs can learn patterns from historical data without needing explicit trend or seasonality assumptions. LSTM consists of memory cells that regulate how much past information is retained or forgotten through three Gates:

1. Forget Gate: Decides which past information to discard.
2. Input Gate: Determines what new information to store.
3. Output Gate: Controls what information is used for prediction.

This mechanism allows LSTMs to capture nonlinear relationships, long-term dependencies, and complex interactions in sales data.

Mathematical Representation

At each time step *t*, the LSTM cell performs the following operations.

1. **Forget Gate Activation**

The model decides which information from the previous cell state C_{t-1} should be discarded.

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

where: f_t is the forget gate output,

W_f is the weight matrix for forget gate,

h_{t-1} is the previous hidden state,

x_t is the current input,

b_f is the bias for forget gate.

2. **Input Gate Activation**

It determines what new information will be added to the cell state.

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

where: i_t is the input gate output,

W_i is the weight matrix for input gate,

$\boldsymbol{b_i}$ is the bias for input gate.

3. **Create Candidate Cell State**

   A new candidate value $\tilde{C}_t$ is generated using the current input and previous hidden state.

$$\widetilde{\boldsymbol{C}}_{\boldsymbol{t}} = \tanh\left(\boldsymbol{W}_{\boldsymbol{c}}[\boldsymbol{h}_{\boldsymbol{t-1}}, \boldsymbol{x}_{\boldsymbol{t}}] + \boldsymbol{b}_{\boldsymbol{C}}\right)$$

where: $\widetilde{\boldsymbol{C}}_{\boldsymbol{t}}$ is the candidate cell state,

$\boldsymbol{W}_{\boldsymbol{c}}$ is the weight matrix for candidate cell state,

$\boldsymbol{b}_{\boldsymbol{c}}$ is the bias for candidate cell state.

4. **Update Cell State**

   The new cell state C_t is computed by combining the retained old state and the candidate.

$$\boldsymbol{C}_{\boldsymbol{t}} = \boldsymbol{f}_{\boldsymbol{t}} * \boldsymbol{C}_{\boldsymbol{t-1}} + \boldsymbol{i}_{\boldsymbol{t}} * \widetilde{\boldsymbol{C}}_{\boldsymbol{t}}$$

where: $\boldsymbol{C}_{\boldsymbol{t}}$ is the new cell state,

$\boldsymbol{C}_{\boldsymbol{t-1}}$ is the previous cell state.

5. **Output Gate Activation**

   The model decides how much of the updated cell state should influence the output.

$$\boldsymbol{o}_{\boldsymbol{t}} = \boldsymbol{\sigma}(\boldsymbol{W}_{\boldsymbol{o}}[\boldsymbol{h}_{\boldsymbol{t-1}}, \boldsymbol{x}_{\boldsymbol{t}}] + \boldsymbol{b}_{\boldsymbol{o}})$$

where: $\boldsymbol{o}_{\boldsymbol{t}}$ is the output gate,

$\boldsymbol{W}_{\boldsymbol{o}}$ is the weight matrix for output gate,

$\boldsymbol{b}_{\boldsymbol{o}}$ is the bias for output gate.

6. **Hidden State (Final output)**

   The final hidden state (output of the LSTM cell) is calculated.

$$\boldsymbol{h}_{\boldsymbol{t}} = \boldsymbol{o}_{\boldsymbol{t}} * \tanh\left(\boldsymbol{C}_{\boldsymbol{t}}\right)$$

where: $\boldsymbol{h}_{\boldsymbol{t}}$ is the hidden state (the LSTM's main output),

$\boldsymbol{C}_{\boldsymbol{t}}$ is the updated cell state.

Relevance to Sales Forecasting in Coffee Shops

LSTM is well-suited for capturing non-traditional sales patterns, such as sudden spikes due to promotions, weather changes, or seasonal trends. It is particularly useful when external factors influence sales in ways that traditional models (ARIMA, Prophet) struggle to handle.

2.5.4 Random Forest

Theory and Working Mechanism

Random Forest is an ensemble learning method based on multiple decision trees. Its operations are as follows:

1. Creating multiple decision trees from different subsets of training data.
2. Aggregating predictions from all trees to form a final output (in regression tasks, this is done by averaging predictions).

This method reduces overfitting, improves generalization, and enhances prediction accuracy by leveraging the power of multiple models.

Mathematical Representation

Given $\boldsymbol{N}$ decision trees, the final prediction $\hat{\boldsymbol{y}}$ is computed as:

$$\hat{\boldsymbol{y}} = \frac{\boldsymbol{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{y}_{\boldsymbol{i}} ,$$

where $\boldsymbol{y}_{\boldsymbol{i}}$ is the prediction from each tree.

Key Assumptions

- The model assumes that feature interactions can be learned from historical data.
- Works best when sales are influenced by multiple external factors (weather, promotions, holidays).

Relevance to Sales Forecasting in Coffee Shops

Random Forest is effective for analyzing the impact of multiple external variables (e.g., holidays, marketing campaigns, employee shifts) on coffee shop sales. Unlike time series models, it does not require strict assumptions about trends and seasonality.

2.6 Model Comparison

Table 2.1 provides a comparative overview of four forecasting models—FB Prophet, ARIMA, LSTM, and Random Forest—focusing on their core architectures, key features, strengths, weaknesses, and their specific relevance to weekly total sales forecasting in coffee shops. This comparison aids in selecting the optimal model for different forecasting scenarios in the project.

Table 2.1 Comparative evaluation of FB Prophet, ARIMA, LSTM, and Random Forest for weekly sales forecasting

Model	Architecture	Key Features	Strengths	Weaknesses	Relevance to the Project
FB Prophet	Additive Time Series Model	Handles holidays & seasonality, trend forecasting	Robust to missing data, interpretable results, easy to use	Limited in capturing sudden shifts or nonlinear relationships	Useful for forecasting sales trends, especially for incorporating seasonality and holiday effects
ARIMA	Statistical Model (AutoRegressive Integrated Moving Average)	Captures trend and seasonality in time series	Well-established, interpretable, performs well with stable trends	Requires stationarity, sensitive to data fluctuations	Suitable for modeling stable coffee shop sales trends over time, especially when sales follow a consistent pattern
LSTM	Deep Learning (Recurrent Neural Network)	Learns long-term dependencies, handles sequential data	Effective for capturing complex patterns and nonlinear relationships	Requires large datasets, computationally expensive	Useful for modeling dynamic sales patterns influenced by multiple external factors like promotions and weather
Random Forest	Ensemble Learning (Decision Trees)	Captures feature interactions, works well with structured data	Handles nonlinearity well, robust to missing data, interpretable feature importance	Requires careful hyperparameter tuning, may overfit with small data	Effective for analyzing the impact of multiple features (e.g., weather, promotions, holidays) on sales

2.7 Model Performance Evaluation

2.7.1 Evaluation Metrics for Forecasting Models

The accuracy of forecasting models is assessed using the following statistical metrics:

- Mean Absolute Error (MAE): Measures the average absolute differences between actual and predicted values.
- Root Mean Squared Error (RMSE): Emphasizes larger errors by squaring deviations, making it more sensitive to significant discrepancies.
- Mean Absolute Percentage Error (MAPE): Represents the forecasting error as a percentage, allowing for easier comparison across different datasets.

2.7.2 Business Implications of Sales Forecasting

- Optimized Inventory Management: Reducing costs associated with overstock and stock shortages.
- Enhanced Marketing Strategies: Aligning promotional efforts with expected demand fluctuations.
- Improved Workforce Planning: Adjusting staff schedules to meet anticipated customer demand more effectively.

2.8 Related Research

This section presents a review of key research studies that have contributed to the field of sales forecasting in the food and beverage industry. The selected studies focus on various machine learning and statistical models, including ARIMA, LSTM, Prophet, and Random Forest, which have demonstrated effectiveness in demand forecasting. These papers provide insights into the strengths and limitations of different methodologies, supporting the development of a robust forecasting framework for this study.

2.8.1 Studies on Machine Learning-Based Sales Forecasting

Several studies have explored the application of machine learning in demand forecasting within the food and beverage sector:

- **"Machine Learning Models for Short-Term Demand Forecasting in Food Catering Services" – Rodrigues et al. (2024)**
  This research compared various machine learning models for predicting restaurant demand. It found that Random Forest and LSTM exhibited strong performance in short-term demand forecasting, reinforcing their suitability for café sales prediction.
- **"Food Demand Prediction Using Statistical and Machine Learning Models" – S. Jayapal (2023)**
  The study evaluated Prophet, ARIMA, and LSTM for forecasting food demand in cafes. The findings indicated that Prophet performed poorly compared to LSTM and ARIMA, suggesting that deep learning models may be more effective for certain forecasting tasks.
- **"Food Industry Sales Prediction: A Big Data Analysis & Sales Forecast of Bake-off Products" – M. Lindström (2021)**
  This study explored the use of Decision Tree Regression, Random Forest, ARIMA, and SARIMA for sales forecasting in the food industry. It also analyzed the performance of Prophet and Recurrent

Neural Networks (RNNs), providing insights into model selection for time series forecasting.

2.8.2 Studies on External Factors Affecting Sales Forecasting

External factors, such as weather conditions and economic shifts, significantly influence food and beverage sales. The following studies examine the impact of these variables on forecasting models:

- **"An Investigation of Weather Impact on Beverage Sales Forecasting" – O. Yilmaz (2024)**
  This paper examined how ARIMA, SARIMA, and Random Forest could be used to predict beverage sales in cafes based on weather conditions. Additionally, Prophet and CNN models were considered, offering a comparative perspective on their forecasting accuracy.
- **"Forecasting Food and Beverage Sales Using Machine Learning Approaches" – T. Nakamura (2022)**
  This study investigated how machine learning models, including Random Forest, XGBoost, and LSTM, could be used to forecast food and beverage sales. The research highlighted the advantages of ensemble learning methods in improving accuracy and reducing forecasting errors, particularly in dynamic and competitive markets.

2.8.3 Studies on Time Series Forecasting for Business Intelligence

- **"Time Series Analysis and Forecasting for Business Intelligence Applications" – S. Abrishami (2019)**
  Although this study predated the selected period, it remained relevant for demonstrating the application of Prophet and ARIMA in business forecasting. It also highlighted how Random Forest regression models could improve sales forecasts.

These studies collectively provide a strong foundation for understanding the effectiveness of different forecasting models in the café industry. By leveraging insights from existing research, this study aims to enhance sales prediction accuracy and optimize inventory management for coffee shop businesses.

Chapter 3
Research Methodology

This chapter outlines the methodology employed in this research on Weekly Total Sales Analysis and Forecasting of Café and Restaurant using Machine Learning using ARIMA, Prophet, LSTM, and Random Forest. The approach follows a structured pipeline that includes data preparation, visualization, model development, training, evaluation, comparison, and final forecasting. This workflow ensures a systematic and efficient strategy to achieve accurate sales predictions. This summary workflow of research methodology is shown in Figure 3.1

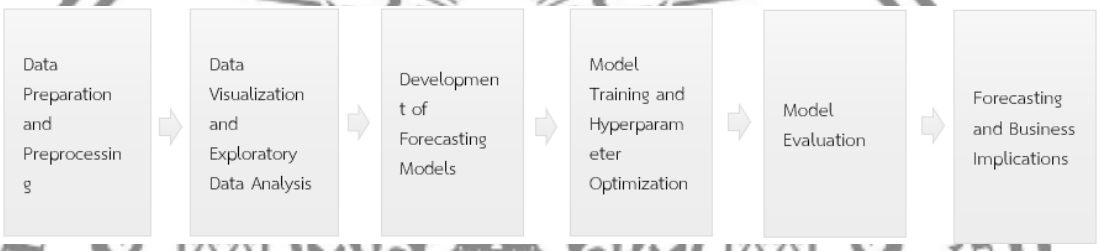


Figure 3.1 Workflow of research methodology

Additionally, Figure 3.2 presents a detailed flowchart depicting the step-by-step model development and evaluation process. The process begins with data loading and preprocessing, including feature engineering, to ensure that the dataset is well-structured for modeling. Then different forecasting models are selected, trained, tested, and evaluated. If the initial evaluation results are unsatisfactory, optimization techniques are applied to refine the models before re-evaluating them. Once the model performance meets the desired criteria, the best models are selected for further analysis. A crucial overfitting check is conducted to ensure that the models generalize well to new data. If overfitting is detected, adjustments such as regularization or additional tuning are performed before proceeding. Finally, the optimal forecasting model is deployed to generate sales predictions, followed by an analysis of business insights and practical implications for decision-making. This structured methodology provides a comprehensive and iterative approach to improving forecasting accuracy while addressing potential challenges in model performance. The combination of these methodologies ensures robust and actionable insights for café and restaurant businesses.

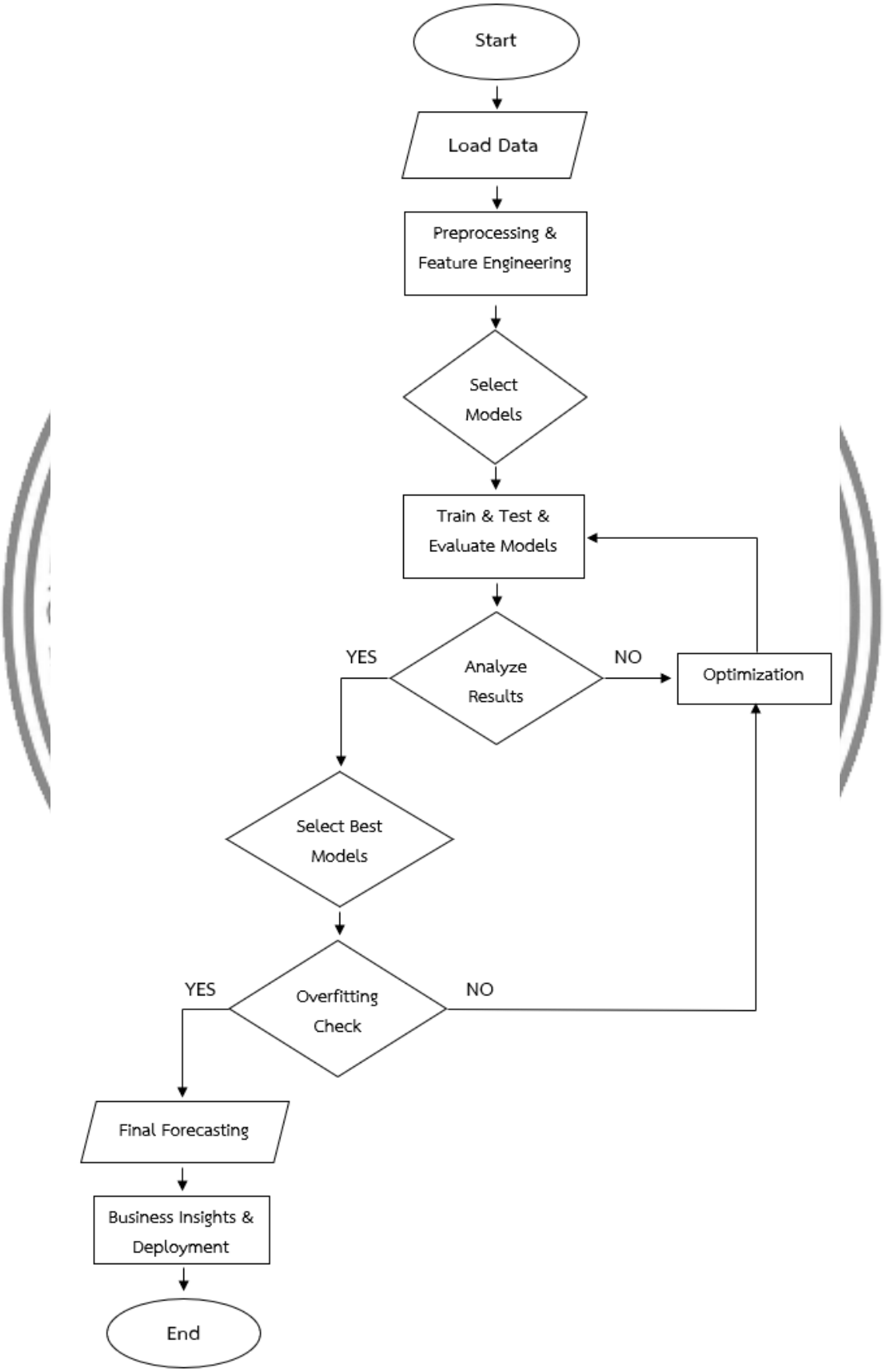


Figure 3.2 Flowchart – detailed ML pipeline

3.1 Data Preparation and Preprocessing

3.1.1 Data Collection

The dataset for this research consists of historical sales data collected from February 2023 to September 2024 from 1 coffee shop branch. The dataset includes various attributes influencing sales, such as

- Date: Transaction date.
- Day of the Week: Helps analyze weekly sales trends.
- Weather Condition: Captures environmental factors influencing sales.
- Product ID & Name: Identifies product-level demand.
- Category & Group: Helps categorize products into beverages, food, and merchandise.
- Average Cost & Price: Provides pricing insights.
- Quantity Sold: Indicates consumer demand.
- Total Sales & Net Profit: Used as the primary forecasting target.
- Discounts & Promotions: Captures the effect of marketing strategies.
- Holiday Indicators: Determines seasonal demand patterns.
- Employee Data: Helps analyze operational impact.

The dataset was cleaned and preprocessed to handle missing values, inconsistencies, and outliers before model implementation.

3.1.2 Data Cleaning & Transformation

The dataset undergoes cleaning and preprocessing to ensure high-quality inputs for forecasting models:

1. Handling Missing Values:
   - Missing sales and pricing data have been filled using historical averages.
2. Handle unnecessary data:
   - Remove unnecessary column.
3. Data Aggregation & Formatting:
   - Data are aggregated into weekly total sales to align with time series forecasting models.
   - Columns are standardized to the format required for model input (e.g., Prophet's ds and y).
4. Outlier Detection:
   - Extreme values in sales are identified using Interquartile Range analysis and removed if necessary.

3.2 Data Visualization and Exploratory Data Analysis (EDA)

To gain insights before model training, several visualization techniques are applied:

1. Sales Trends Over Time:
   - A time series plot is created to observe demand fluctuations.
   - Seasonal peaks and troughs are identified.
2. Impact of Weather on Sales:
   - A stacked bar chart is used to compare total sales across different weather conditions.
   - Rainy days shows lower sales, while sunny days boosts revenue.
3. Best-Selling vs. Least-Selling Products:

- A horizontal bar chart highlights which items contribute the most to revenue.
- Popular drinks and food items are identified.

4. Sales Distribution by Employee Count:
   - A group bar chart analyzes total sales and net profit per employee count.
   - The best workforce allocation strategy is determined.

3.3 Development of Forecasting Models

Four models are selected for weekly sales forecasting:

3.3.1 ARIMA (AutoRegressive Integrated Moving Average)

- Captures linear trends & seasonality in sales data.
- Steps the implementation:
  1. Identify trends and seasonality through AutoCorrelation (ACF) and Partial AutoCorrelation (PACF) plots.
  2. Use Auto ARIMA to select the best configuration.
  3. Fit the model and generated forecasts.

3.3.2 Prophet (Developed by Meta/Facebook)

- Handles holidays, seasonality, and trend shifts effectively.
- Steps in the implementation:
  1. Data formatted with ds (date) and y (sales).
  2. Incorporated weekly & yearly seasonality and holiday effects.
  3. Model trained and tested historical data.

3.3.3 LSTM (Long Short-Term Memory - Deep Learning)

- Captures non-linear dependencies and sequential patterns in sales data.
- Steps in Implementation:
  1. Data are transformed into supervised learning format.
  2. Constructed an LSTM network with dropout layers to prevent overfitting.
  3. Trained with Adam optimizer using Mean Squared Error (MSE)

3.3.4 Random Forest Regression

- Handles non-linear relationships & external influences (e.g., weather).
- Steps in the implementation:

  1. Extract key features (historical sales, weather, promotions).
  2. Use Grid Search to optimize hyperparameters.
  3. Train model and evaluated prediction accuracy.

3.4 Model Training and Hyperparameter Optimization

All models are trained on 75% of data and tested on 25% of data.

1. ARIMA: p, d, q parameters optimized using grid search.
2. Prophet: Seasonal components and changepoint detection fine-tuned.
3. LSTM: Dropout, batch size, and epochs adjusted via hyperparameter tuning.

4. Random Forest: Number of trees and depth optimized using Grid Search CV.

3.5 Model Evaluation Metrics

To assess model performance, the following evaluation metrics are used:

Table 3.1 Evaluation metrics

Metric	Formula	Purpose
Mean Absolute Error (MAE)	$\frac{1}{n}\sum_{i=1}^{n}\lvert yi - \hat{y}i\rvert$	Measures the average absolute difference between actual sales and predicted sales.
Mean Absolute Percentage Error (MAPE)	$\frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{yi - \hat{y}i}{yi}\right\rvert x\ 100$	Measures the average percentage error between actual and predicted values.
Root Mean Squared Error (RMSE)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y}i)^2}$	Measures the square root of the average squared errors, giving more weight to larger errors.

From: Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed., pp. 72–74). O'Reilly Media.

3.6 Forecasting and Business Implications

3.6.1 Final Weekly Sales Prediction

- The best-performing model is selected based on RMSE, MAE, and MAPE.
- The final forecast for Week 86 is generated.

3.6.2 Business Strategy Based on Forecasting Results

- Staffing Adjustments: Increased employees during high-sales periods.
- Marketing Optimization: Promotions targeted at demand peaks.
- Inventory Planning: Stock levels adjusted based on expected demand.

Chapter 4
Main Results and Discussion

This chapter presents a comprehensive analysis of the dataset, covering four main areas: data visualization, model performance evaluation, model comparison, and final sales forecasting for next week. Each section aims to provide detailed insights into the dataset and the predictive models used.

The goal of this analysis is to:

- Identify meaningful patterns and trends in sales data using exploratory data visualization.
- Evaluate the forecasting accuracy of four different models: Prophet, ARIMA, LSTM, and Random Forest.
- Compare model performance using key error metrics (RMSE, MAE, and MAPE) to determine the most effective forecasting approach.
- Analyze the final sales prediction for next week and interpret the model results for actionable business insights.

By combining historical sales data, machine learning models, and statistical forecasting techniques, this chapter aims to enhance decision-making for coffee shop business operations, focusing on demand forecasting and inventory management.

4.1 Data Exploration and Visualization

Before developing predictive models, it is crucial to understand the dataset through visual exploration. This step provides key insights into sales patterns, seasonality, external influencing factors (such as weather and holidays), and customer preferences.

4.1.1 Sales Trends Over Time

The Total Sales Over Time graph highlights the fluctuations in weekly total sales over a period from February 2023 to September 2024. The following key trends are observed in Figure 4.1:

- Sales tend to show cyclical fluctuations, suggesting that demand is not uniform throughout the year.
- Peaks in sales appear at regular intervals, which could be linked to seasonal or holiday events or marketing campaigns.

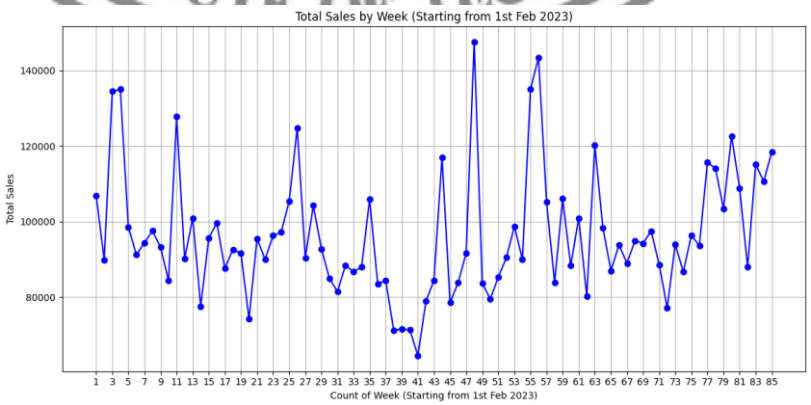


Figure 4.1 Total sales by week of café and restaurant

Business Implication:

Identifying high-sales periods allows businesses to plan inventory and staffing more efficiently and the end-of-period decline should be further investigated to ensure data completeness.

4.1.2 Total Sales by Weather and Holiday Conditions

The Total Sales by Weather and Holiday Conditions illustrates how different weather conditions impact sales, which is shown in Figures 4.2 and 4.3.That is

- Sunny days contribute the highest sales volume, indicating that good weather drives higher customer foot traffic.
- Rainy and cloudy days show significantly lower sales, suggesting that customers prefer to visit coffee shops when the weather is favorable.
- The highest monthly sales are observed during peak seasons
- Holidays have a significant impact on weekly sales

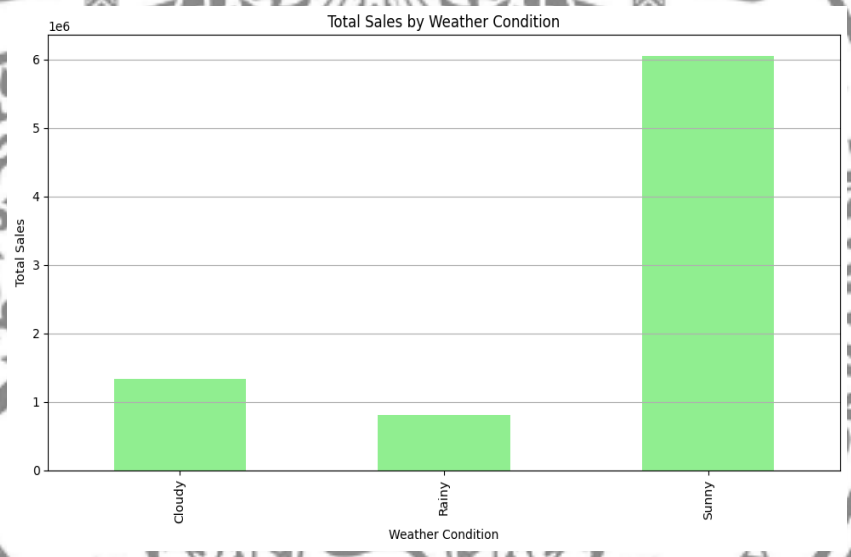


Figure 4.2 Total sales by weather condition

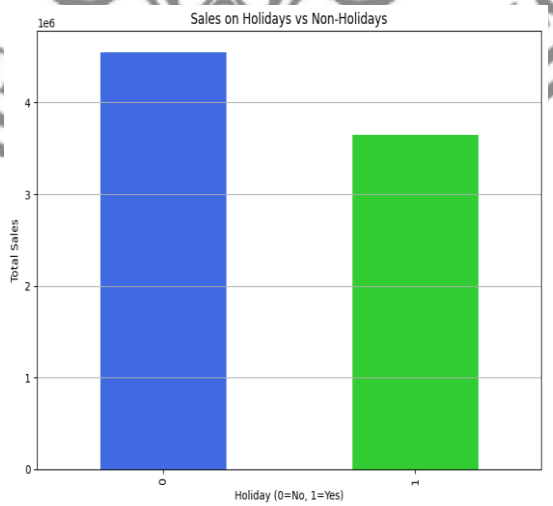


Figure 4.3 Total sales by holiday condition

Business Implication:

- Adjust promotional efforts based on weather and holiday conditions. For example.
  - Offer rainy-day discounts or free delivery options to compensate for reduced foot traffic.
  - On sunny days, maximize sales by selling outdoor-friendly beverages.
- Manage staff by adding part-time staff on holidays.

4.1.3 Sales by Day of the Week

The Total Sales by Day of the Week visualization indicates in Figure 4.4:

- Weekends (Saturday and Sunday) experience the highest sales, as people have more leisure time and are likely to visit coffee shops.
- Midweek days (Tuesday and Wednesday) show the lowest sales, suggesting that fewer customers visit during these days.

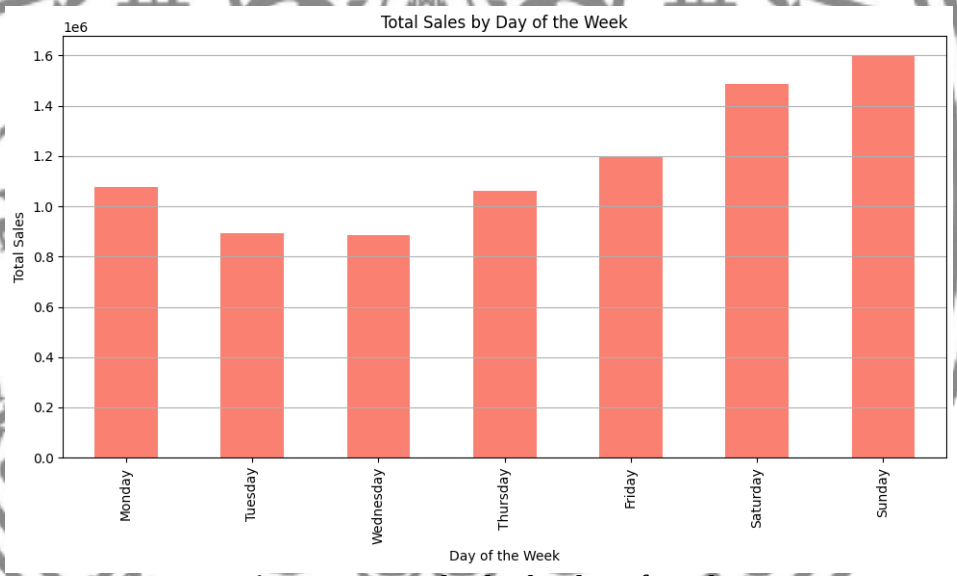


Figure 4.4 Total sales by day of week

Business Implication:

- Extend closing hours on Friday, Saturday, and Sunday from 6:00 PM to 9:00 PM due to high sales demand during these days. This adjustment allows the business to accommodate more customers during peak hours, maximizing revenue and improving customer satisfaction.
- Optimize staff allocation, ensuring more employees are available during peak weekend hours to manage higher demand.

4.1.4 Sales and Net profit by employee

The Total Sales and Net Profit by Employee chart illustrates how different staffing levels impact sales and profitability. The data is categorized into three groups based on the number of employees working on a given day: 5 employees, 6 employees, and 7 employees, see Figure 4.5.

Key Insights:

- Days with 5 employees have the lowest total sales and net profit, suggesting that understaffing might lead to inefficiencies in handling customer demand.
- Days with 7 employees show moderate total sales and profit, but the increase in staff does not significantly improve revenue.
- Days with 6 employees generated the highest total sales and net profit, indicating an optimal staffing level for maximizing revenue.

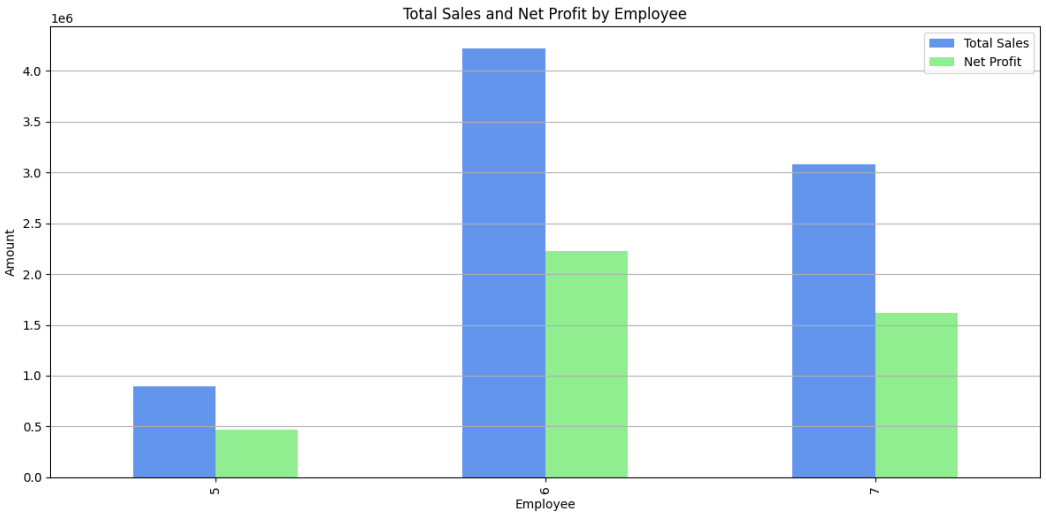


Figure 4.5 Total sales and net profit by employees

Business Implication:

The data reveals a staff management issue, where employee allocation does not align with peak sales periods. To optimize labor efficiency and profitability, the business should:

- Reallocate part-time employees from the 7-employee days to the 6-employee days, ensuring that peak sales periods have the optimal workforce.
- Adjust work schedules to match demand patterns, preventing overstaffing on low-sales days and understaffing on high-sales days.
- Analyze employee productivity to determine whether operational inefficiencies exist during 7-employee days that limit revenue growth.

4.1.5 Product-Level Analysis: Best-Selling vs. Least-Selling Products and by category

Top 10 Best-Selling Products are shown in Figure 4.6:

- The highest-selling items include Iced Americano, Iced Matcha Latte, and Thai Tea.
- Pizza, we have only 1 menu that is Hawaiian but it is so popular for customer.

Top 10 Least-Selling Products are shown in Figure 4.7:

- Items such as Very Raspberry, additional toppings, and cold-pressed juices perform poorly.

- Some of these low-selling products might be too niche, too expensive, or not marketed effectively.

Top 10 Best-Selling Products by Categories are shown in Figure 4.8:

- The food category has the highest sales, but this is expected since we offer over 100 menu items in this category.
- The second is coffee category.

Top 10 Least-Selling Products by Categories are shown in Figure 4.9:

- The Fruit category has the lowest sales, which makes sense since we do not primarily focus on selling fruit. We only occasionally offer seasonal fruits from our own garden.
- Some of these low-selling products might be too niche, too expensive, or not marketed effectively.

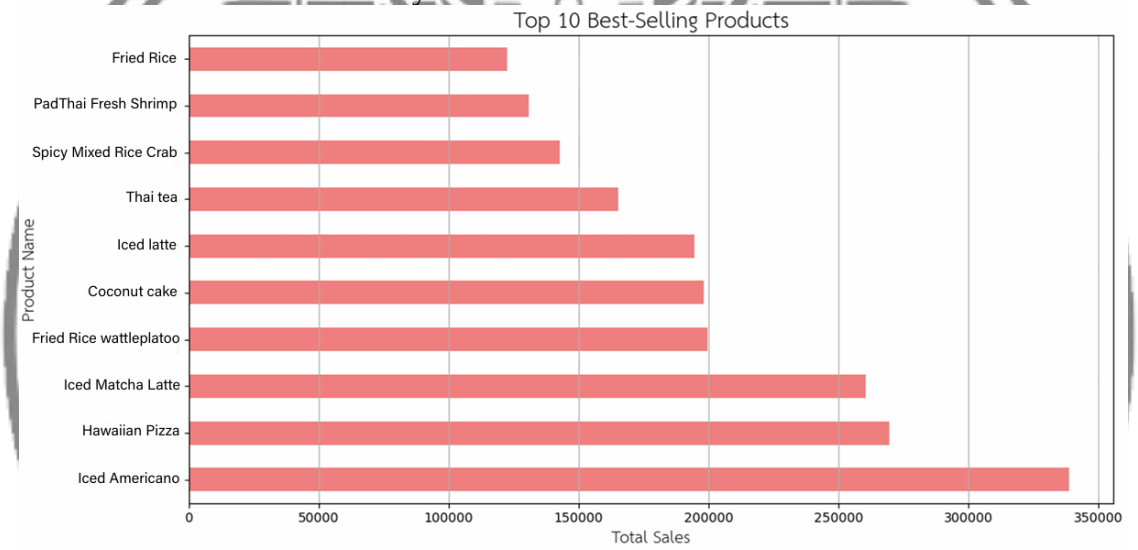


Figure 4.6 Top 10 best-selling products

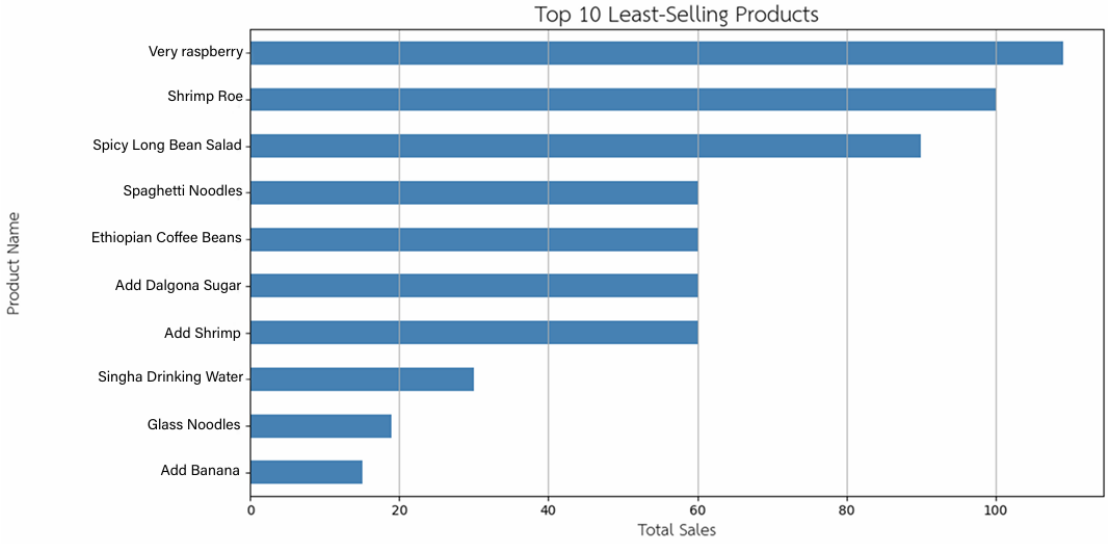


Figure 4.7 Top 10 least-selling products

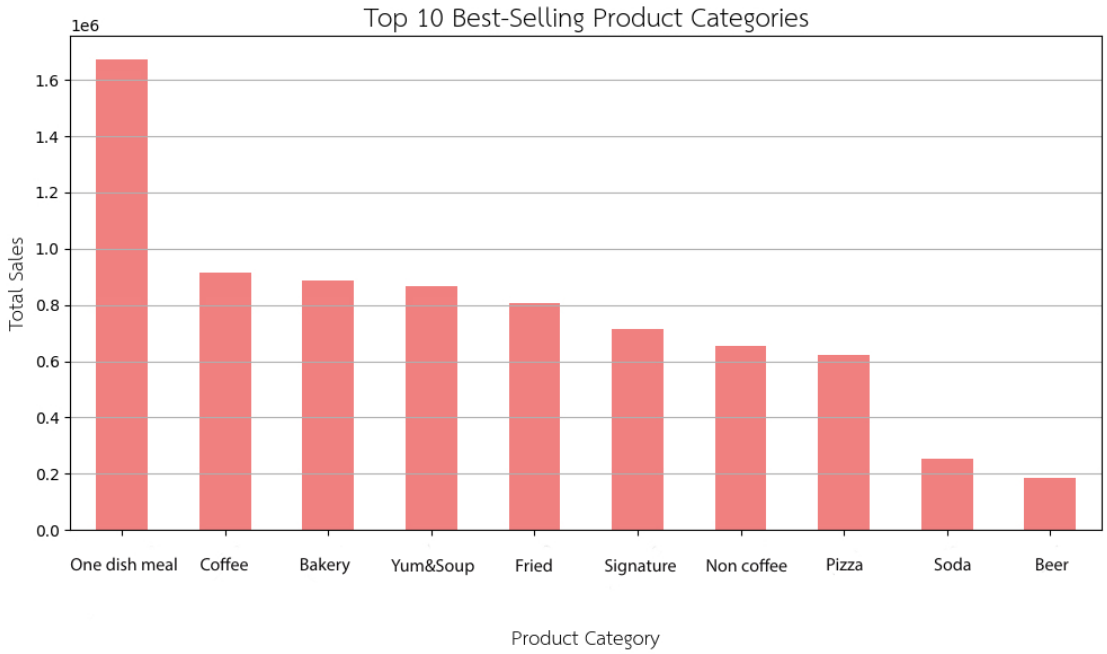


Figure 4.8 Top 10 best-selling products categories

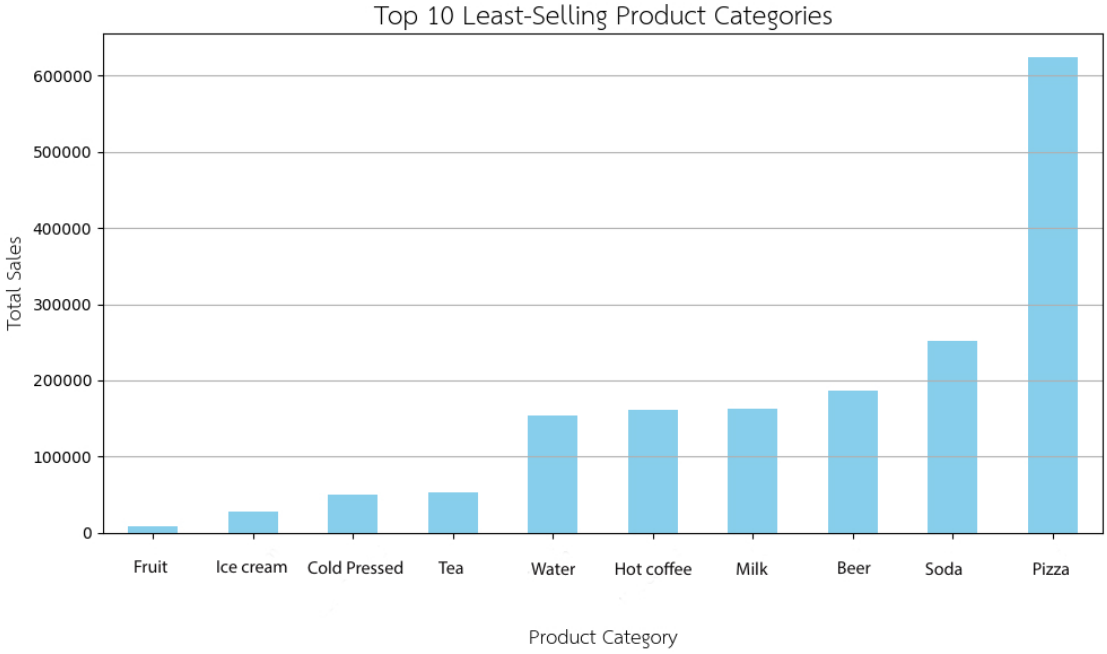


Figure 4.9 Top 10 least-selling product categories

Business Implication:

- Rice and coffee have the highest sales, aligning perfectly with the store's main objectives.
- Meanwhile, the Pizza category has high sales despite having only one menu item. This clearly indicates that we need to further develop this category to boost our sales. To enhance customer choices and maximize sales potential, we introduced new pizza variations by

developing different dough recipes, including a thin-crust option for a crisper texture and a thick, soft dough for a more indulgent bite. These variations cater to different customer preferences, ensuring broader market appeal and increasing overall pizza sales.

- The category that needs improvement is beer because there are only a few options available. To address this, we plan to introduce craft beer to expand the selection.

4.1.6 Correlation Analysis of Key Metrics

A correlation heatmap helps understand relationships between key business observe in Figure 4.10.

- Total Sales and Net Profit have a high correlation (0.98), confirming that increasing sales lead to higher profitability.
- Quantity Sold and Average Price show a negative correlation (-0.29), indicating that higher-priced items tend to sell in lower quantities.

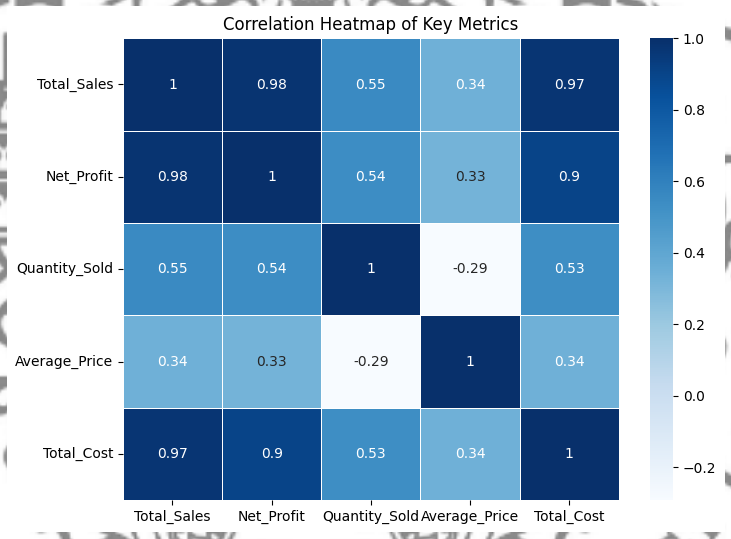


Figure 4.10 Correlation heatmap of key metrics

Business Implication: A balanced pricing strategy is essential to maximize both sales volume and profitability.

4.2 Model Training and Performance Analysis

After exploring the dataset, four forecasting models are trained to predict weekly sales: FB Prophet, ARIMA, LSTM, and Random Forest.

4.2.1 FB Prophet Model

- The Prophet model successfully captures seasonal trends and long-term growth patterns shown in Figure 4.11.
- The uncertainty interval widens towards the end of the forecast period, meaning predictions become less certain over time.
- The last data point is unusually low because it includes only the first two days of the week, not the full week.

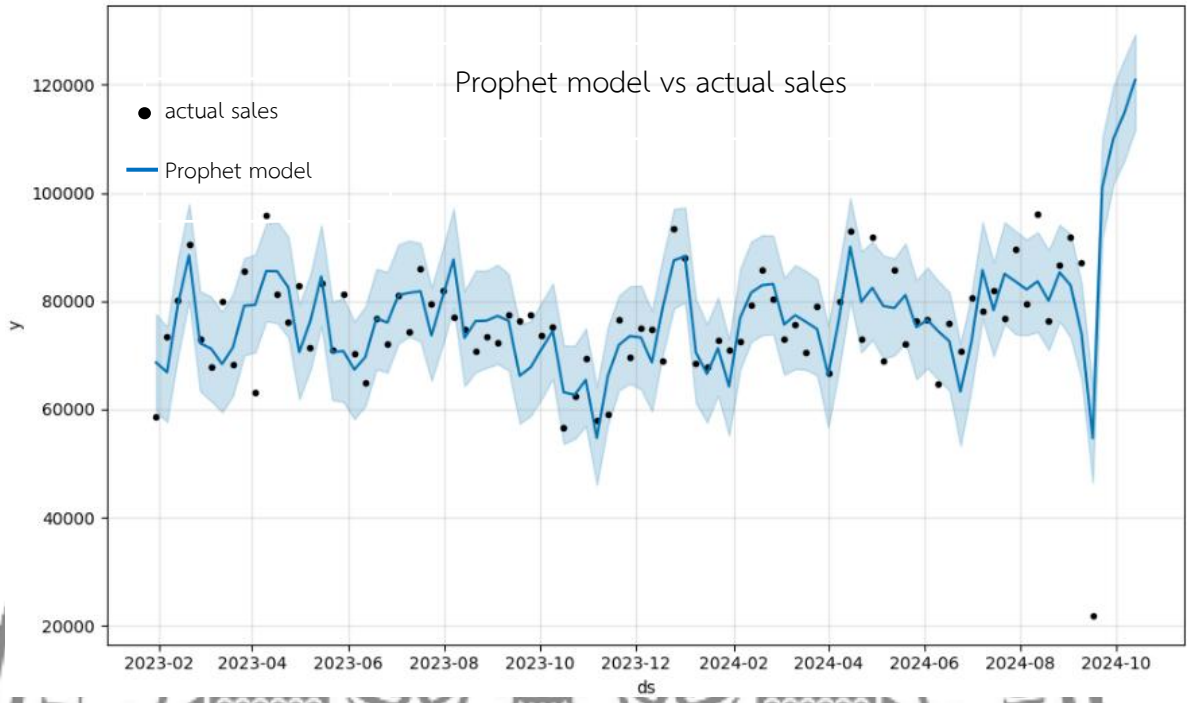


Figure 4.11 Prophet model vs actual sales

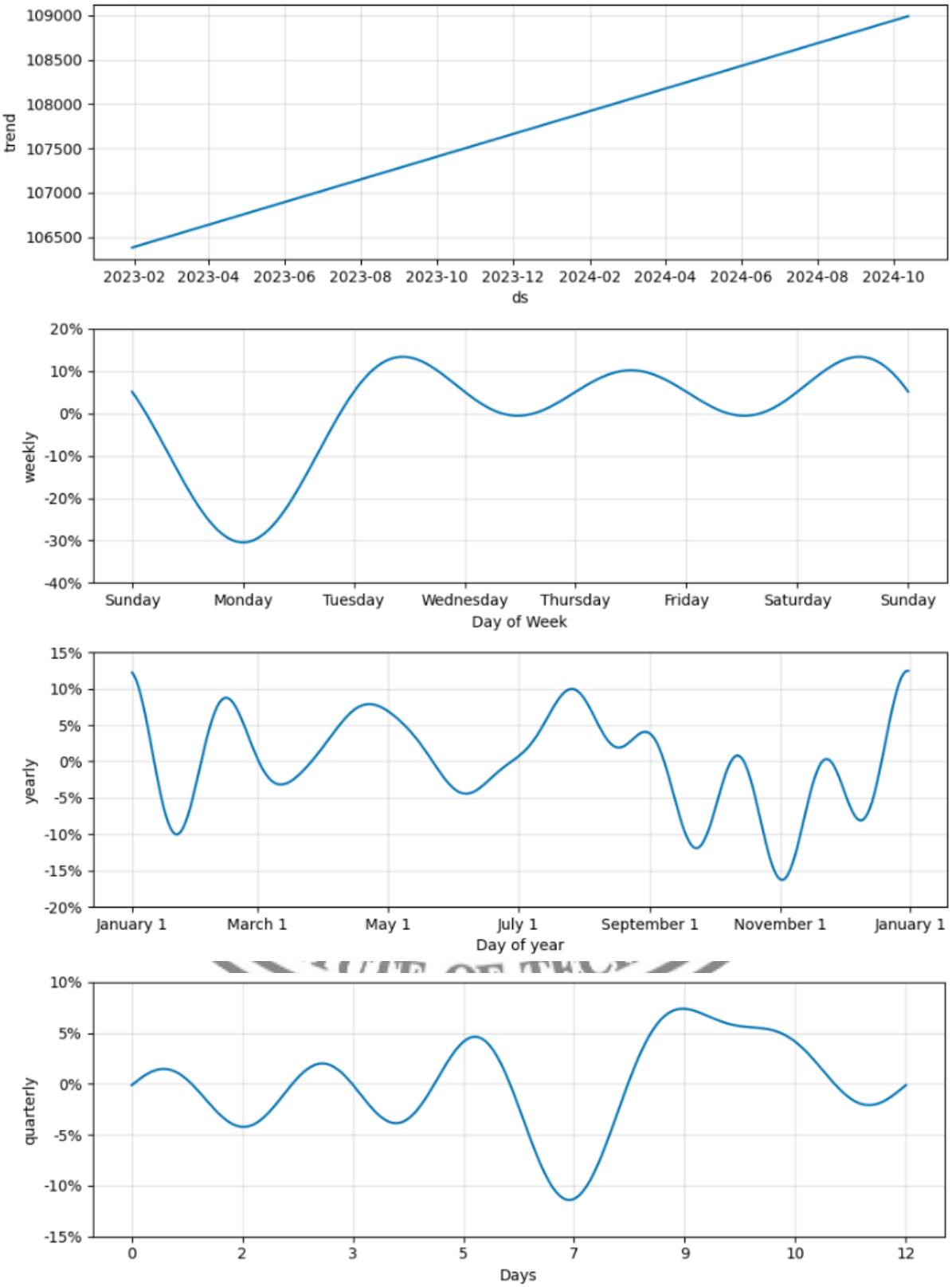


Figure 4.12 Prophet model decomposition of trends and seasonality

Trend Component (Top Plot)

- Shows the overall direction of sales overtime, excluding short-term or seasonal fluctuations.
- In this graph, there's a steady upward trend in total sales from early 2023 to late 2024.
- This suggests continuous business growth, potentially due to improved marketing, increased customer base, or successful new products.
- Insight: A positive long-term trend supports strategic decisions such as hiring, inventory expansion, or opening new branches.

Weekly Seasonality (Second Plot)

- It reflects how sales vary throughout the week.
- Sales drop significantly on Mondays, possibly due to post-weekend fatigue or lower foot traffic.
- Sales peak on Fridays and Sundays, likely due to payday, leisure time, or weekend promotions.
- Insight: Businesses can run Monday-specific promotions to boost low-sales days and increase staffing at weekends to handle demand.

Yearly Seasonality (Third Plot)

- Captures sales fluctuations over a yearly cycle.
- Peaks in sales occur around early January, May, and July, likely due to seasonal demand or promotions.
- Declines appear around September and November, indicating lower consumer spending during these months.
- Insight: Marketing and product strategies should be aligned with high-performing months, while weak months may benefit from targeted campaigns.

Quarterly Seasonality (Bottom Plot)

- Reveals multiple peaks and dips within each quarter, indicating shorter-term fluctuations, likely from promotions or periodic events.
- Insight: Businesses can use this pattern to better plan quarterly campaigns, product launches, or staff scheduling.

4.2.2 ARIMA Model

- The ARIMA model captures short-term fluctuations but fails to predict sudden spikes and dips in sales.
- Error metrics indicate moderate accuracy, but its inability to handle rapid demand changes reduces its reliability.

- The sharp decline in weekly sales at the very end of the test data (orange line) is not indicative of actual business performance. This occurred because the data for that week was incomplete, containing only 2 days of records instead of a full 7-day period. As a result, the total weekly sales appear significantly lower than usual.

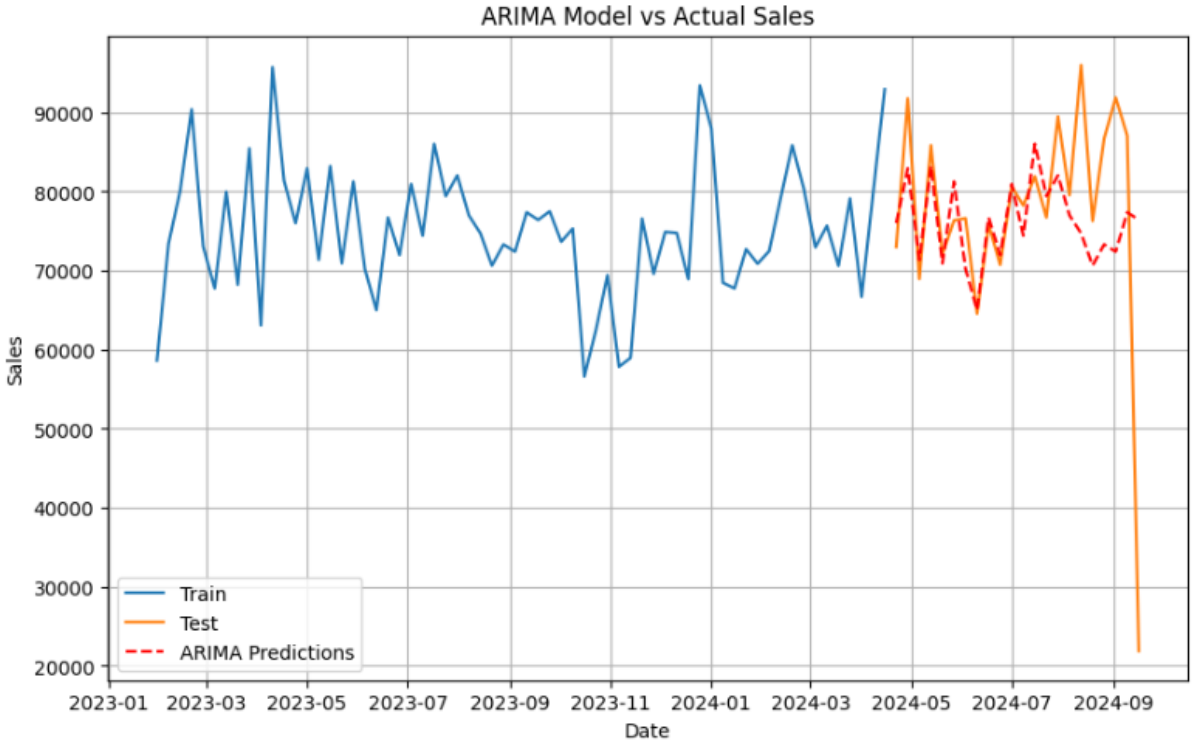


Figure 4.13 ARIMA model vs actual sales

4.2.3 LSTM Model

- The LSTM loss curve shows rapid learning in the early epochs, but over time, it fails to generalize well on test data.
- Forecasts appear too smooth, meaning LSTM struggles with capturing sharp variations in sales data.
- The last data point is unusually low because it includes only the first two days of the week, not the full week.

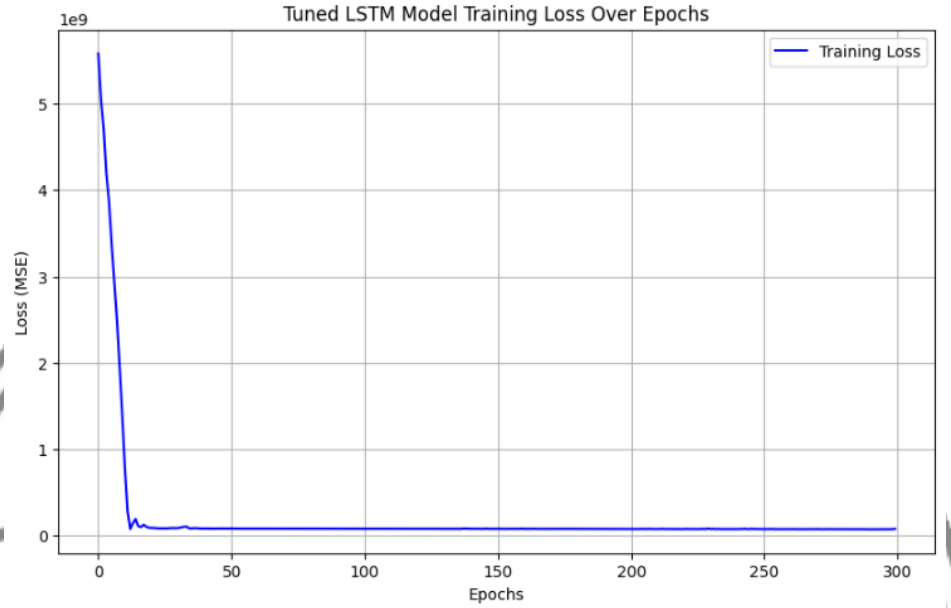


Figure 4.14 LSTM model training loss

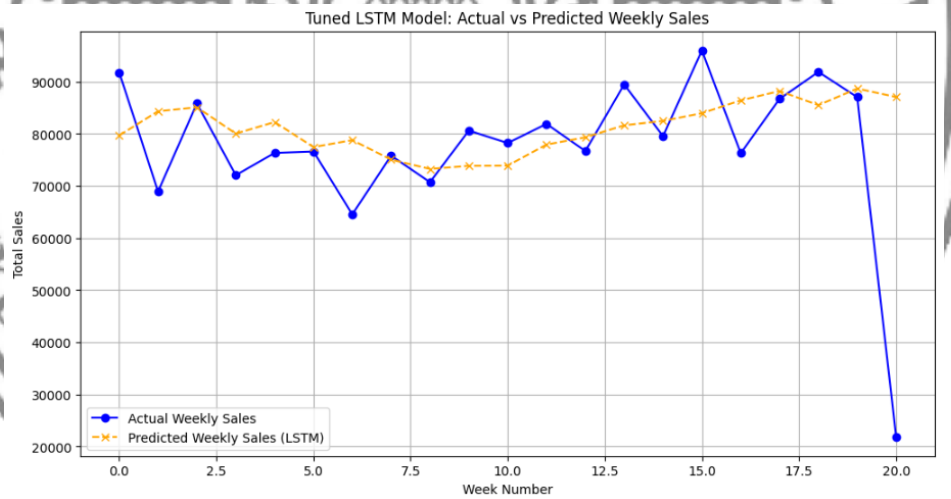


Figure 4.15 LSTM model vs actual sales

4.2.4 Random Forest Model

- The Random Forest model produces overly smooth predictions, failing to recognize seasonal peaks and sudden fluctuations.
- It performs worse than Prophet and ARIMA but better than LSTM.

- The last data point is unusually low because it includes only the first two days of the week, not the full week.

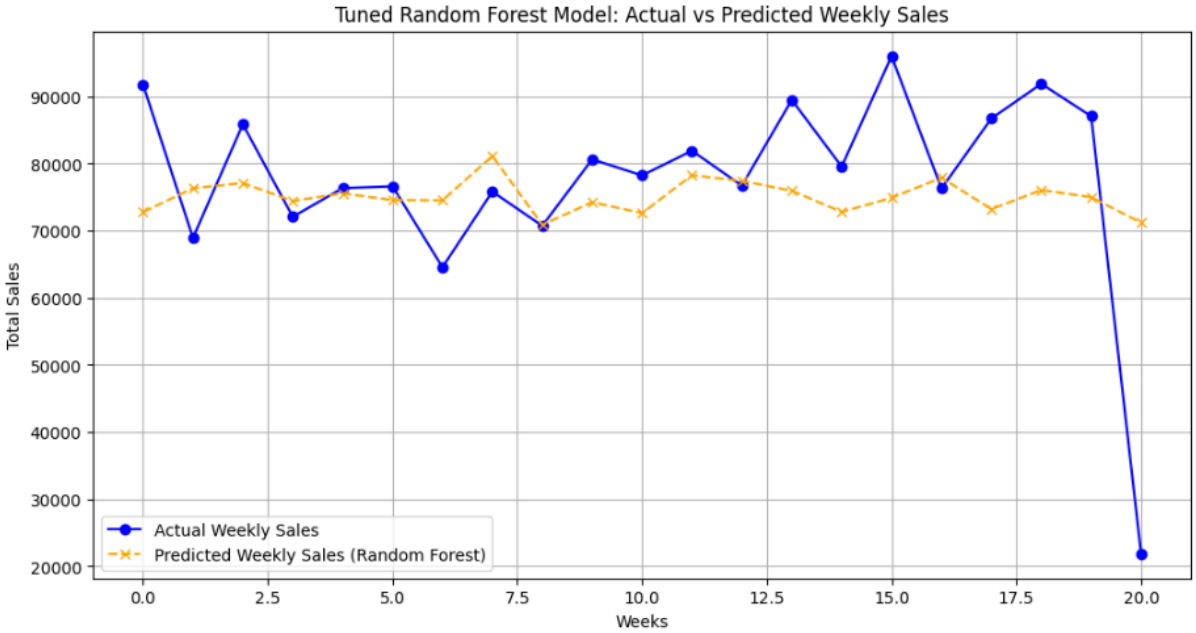


Figure 4.16 Random Forest model vs actual sales

4.3 Model Comparison: Performance Metrics

To determine which model performs best, we evaluate RMSE, MAE, and MAPE for all four models observed in Table 4.1:

Table 4.1 Model comparison performance metrics.

Model	RMSE	MAE	MAPE
Prophet	7136.431	5376.751	8.36%
ARIMA	14111.541	8043.164	17.82%
LSTM	15563.969	9204.957	21.14%
Random Forest	14477.705	9797.838	9.58%

1. **Prophet (The most accurate)**
   RMSE: 7136.43, MAE: 5376.75, MAPE: 8.36%
   Strengths:
   - Specifically designed for time series forecasting.
   - Handles trend, seasonality, and holidays exceptionally well.
   - Supports automatic changepoint detection for trend shifts.
   - Great for datasets with clear seasonal patterns.
2. **ARIMA**
   RMSE: 14111.54, MAE: 8043.16, MAPE: 17.82%
   Strengths:
   - A robust statistical model for linear trends and seasonality.
   - Performs well when the data has low noise and stable patterns.

   Limitations:

- Cannot directly incorporate external factors like weather or holidays.
- Less effective in capturing sudden or nonlinear changes.

3. **Random Forest**
   RMSE: 14477.70, MAE: 9797.84, MAPE: 9.58%
   Strengths:
   - Great for modeling the impact of external features.
   - Captures complex nonlinear relationships between variables.

   Limitations:
   - Not inherently aware of time series structure or sequential dependencies.
   - Tends to generate smooth predictions, lacking sensitivity to sharp fluctuations.

4. **LSTM (Least accurate)**
   RMSE: 15563.97, MAE: 9204.96, MAPE: 21.14%
   Strengths:
   - Capable of learning long-term dependencies in sequential data.
   - Suitable for highly complex and nonlinear time series.

   Limitations:
   - Requires large datasets to perform well.
   - Prone to overfitting with smaller or noisy data.

4.4 Final Prediction for next week (week 86)

For week 86, each model predicted the following sales observed in Table 4.2:

Table 4.2 Model comparison next week.

Model	Prediction	Actual Sales	Error (%)
Prophet	79,091.17	85,302.93	7.28%
ARIMA	78,332.92	85,302.93	8.17%
LSTM	70,269.81	85,302.93	17.62%
Random Forest	73,848.24	85,302.93	13.42%

1. Prophet (The most accurate)
Prediction: 79,091.17
Error: 7.28% (lowest error)

Prophet is designed for time series data with clear seasonality and trend, making it ideal for café sales that vary by day, week, and weather. It adjusts well to holidays and captures growth patterns, explaining its close prediction.

2. ARIMA
Prediction: 78,332.92
Error: 8.17%

ARIMA performs well with linear trends and historical stability, which café data exhibits. Although it lacks seasonality awareness like Prophet, it still provides a solid short-term forecast.

3. Random Forest
Prediction: 73,848.24
Error: 13.42%

While RF captures complex patterns and external features (e.g., weather), it ignores time order, which is critical in sales forecasting. This leads to less accurate, smoother forecasts.

4. LSTM – Least Accurate

Prediction: 70,269.81

Error: 17.62% (highest error)

Although LSTM is powerful for sequential data, it requires large, clean datasets and careful tuning. In this case, it likely overfit or failed to generalize well due to noise or insufficient sequence structure.

Chapter 5
Conclusion and Suggestions

5.1 Conclusion

This study aims to analyze and forecast weekly sales in a coffee shop using four different predictive models: FB Prophet, ARIMA, LSTM, and Random Forest. The primary objectives are to understand sales trends, evaluate the effectiveness of forecasting models, compare their performance using error metrics, and derive actionable business insights from the findings.

5.1.1. Sales Trends and Influencing Factors

Through exploratory data visualization, several critical patterns and external factors affecting sales are identified as follows.

- Seasonal Variations: Sales data exhibited clear seasonal trends, with peaks during specific months (January, May, July) and declines during others (September, November).
- Day-of-the-Week Trends: Weekends, particularly Fridays and Sundays, shows the highest sales volumes, whereas Mondays has the lowest sales.
- Impact of Weather: Sales are significantly higher on sunny days, while rainy and cloudy weather negatively affects customer foot traffic.
- Staffing and Sales Performance: Analysis of staffing levels reveals that having six employees per day maximizes total sales and net profit, whereas on the days with five employees has the lowest revenue and profit.
- Best and least selling products:
  - Best-selling: Iced Americano, Thai Tea, and Iced Matcha Latte.
  - Least-selling: Very Raspberry, Cold Pressed Juices, and additional toppings, suggesting a need for product evaluation.

5.1.2. Model Performance Evaluation

The forecasting models are assessed using three key error metrics: RMSE, MAE, and MAPE. The findings are as follows:

- FB Prophet performs best, with the lowest RMSE (7136.43) and MAPE (8.36%), indicating superior forecasting accuracy.
- ARIMA performs moderately well but struggled with capturing sudden fluctuations in sales.
- LSTM underperforms due to overfitting and excessive smoothing of forecasts.
- Random Forest provides the least accurate, failing to recognize sales seasonality and producing the highest error rates.

5.1.3. Weekly Sales Forecast for Week 86

Each model predicts different sales for Week 86:

- FB Prophet: 79,091.17 (Error: 7.28%)
- ARIMA: 78,332.92 (Error: 8.17%)
- Random Forest: 73,848.25 (Error: 13.43%)
- LSTM 70,269.81 (Error: 17.62%)

Since Prophet has the lowest prediction error, its forecast of 79,091.17 bahts is the most reliable for business decision-making.

Business Implications

- Operational Efficiency: Optimizing staffing levels by reallocating part-time employees from overstaffed days (7-employee days) to peak sales days (6-employee days) can enhance efficiency.
- Inventory Management: Higher demand periods, such as weekends and sunny days, require better inventory planning to prevent stockouts and lost sales.
- Targeted Marketing Strategies: Promotional campaigns should be aligned with identified demand patterns:
  - Offer weekday discounts to boost Monday sales.
  - Launch weather-based promotions to counteract the impact of rainy days.
  - Promote best-selling items while reconsidering or rebranding the worst-selling ones.

5.2 Suggestions for Future Work

While this study provides valuable insights, several areas could be explored further to improve sales forecasting and business decision-making.

5.2.1. Model Improvement and Feature Engineering

- Hybrid Models: Combining Prophet with deep learning techniques such as LSTM might enhance accuracy.
- Additional Features: Incorporating external variables like social media engagement, local events, or competitor pricing could refine predictions.
- Hyperparameter Optimization: Further tuning of the LSTM model may improve its ability to capture sudden sales spikes.

5.2.2. Expanding Data Sources

- Customer Behavior Analysis: Integrating customer demographics and purchase history can help to predict preferences more accurately.
- Macroeconomic Factors: Including economic indicators (inflation, unemployment rates) may help to understand broader consumer spending trends.
- Real-Time Data Integration: Connecting the forecasting system to real-time POS data can enable more dynamic predictions.

5.2.3. Business Applications and Strategy Implementation

- Dynamic Pricing Strategies: Implementing AI-driven pricing adjustments based on predicted demand can optimize revenue.
- Personalized Marketing: Use machine learning to recommend promotions to specific customer segments based on purchase history.
- Inventory Automation: Leveraging predictive sales data to automate stock ordering and reduce waste.

5.3 Final Remarks

This study demonstrated that data-driven forecasting and analytics can significantly enhance business decision-making in the coffee shop industry. By leveraging advanced forecasting techniques, businesses can optimize staffing,

inventory, and promotional strategies, ultimately driving revenue growth and improving customer satisfaction. However, continuous refinement of predictive models, expansion of data sources, and the integration of AI-driven decision systems will be essential for maintaining long-term competitive advantages.

REFERENCES

Rodrigues, A., et al. (2024). *Machine Learning Models for Short-Term Demand Forecasting in Food Catering Services.* International Journal of Artificial Intelligence and Data Science, 12(3), 45-62.

Jayapal, S. (2023). *Food Demand Prediction Using Statistical and Machine Learning Models.* Proceedings of the 2023 International Conference on Predictive Analytics in Food Industry, 87-102.

Lindström, M. (2021). *Food Industry Sales Prediction: A Big Data Analysis & Sales Forecast of Bake-off Products.* Journal of Business Analytics, 9(2), 155-172.

Yilmaz, O. (2024). *An Investigation of Weather Impact on Beverage Sales Forecasting.* Weather & Business Intelligence Journal, 15(1), 23-38.

Nakamura, T. (2022). *Forecasting Food and Beverage Sales Using Machine Learning Approaches.* International Journal of Computational Intelligence, 30(5), 295-312.

Abrishami, S. (2019). *Time Series Analysis and Forecasting for Business Intelligence Applications.* Business Forecasting Review, 14(4), 110-127.

Hyndman, R.J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice (3rd Edition).* OTexts.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). *Time Series Analysis: Forecasting and Control (5th Edition).* Wiley.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* MIT Press.

Chollet, F. (2021). *Deep Learning with Python (2nd Edition).* Manning Publications.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2018). *Forecasting Methods and Applications (4th Edition).* Wiley.

Facebook Research (2023). *Prophet: Forecasting at Scale.* Retrieved from https://facebook.github.io/prophet/

Hugging Face (2024). *Transformers: State-of-the-Art Machine Learning for Natural Language Processing and Time Series Forecasting.* Retrieved from https://huggingface.co/

Python Software Foundation (2024). *Pandas Documentation.* Retrieved from https://pandas.pydata.org/

Scikit-Learn Developers (2024). *Machine Learning in Python.* Retrieved from https://scikit-learn.org/

TensorFlow Developers (2024). *TensorFlow Documentation.* Retrieved from https://www.tensorflow.org/

OpenAI (2024). *Time Series Forecasting with AI Models.* Retrieved from https://openai.com/research/

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed., pp. 72–74). O'Reilly Media.

The Weather Channel (2025). *Monthly Weather Forecast for Sai Noi, Nonthaburi.* Retrieved from https://weather.com/th-TH/weather/monthly/l/

Sanook (2023). *Public Holidays in Thailand 2023.* Retrieved from https://www.sanook.com/campus/1412063/

Sanook (2024). *Public Holidays in Thailand 2024.* Retrieved from https://www.sanook.com/campus/1418247/

Sanook (2025). *Public Holidays in Thailand 2025.* Retrieved from https://www.sanook.com/campus/1424315/

APPENDIX

APPENDIX A

Data Set Examples

A.1 Example of data set

This dataset contains historical sales records from a coffee shop, spanning from February 1, 2023. It includes key attributes relevant to sales forecasting, such as:

- Date, Day, and Weather: Records the transaction date, day of the week, and weather conditions, which may influence sales patterns.
- Product Information: Includes Product ID, Product Name, Group, and Category, helping to categorize sales across different product types (e.g., beverages, bakery, meals).
- Pricing & Cost Details: Tracks Average Cost, Average Price, and Average Profit for each product sold.
- Sales Metrics: Includes Quantity Sold, Total Sales (after discounts), Net Profit, and Discounts applied, providing insights into revenue generation.
- Operational Data: Includes Branch, Holiday Indicator, and Number of Employees, allowing for the analysis of workforce allocation and holiday sales impact.

This dataset is structured for time series forecasting, enabling the prediction of future sales trends based on historical demand, pricing, and external factors like weather and staffing levels.

Date	Day	Weather	Product_ID	Product_Name	Group	Category	Average_Cost	Average_Price	Average_Profit	Quantity_Sold
2023-02-01	Wednesday	Sunny	P0105	ต้มยำกุ้งน้ำใส		ยำ ต้ม	90	250	160	2
2023-02-01	Wednesday	Sunny	P0011	Creammy Dalgona Latte		Signature	60	120	60	1
2023-02-01	Wednesday	Sunny	P0012	สปาเก็ตตี้ผัดพริกแห้ง		อาหารจานเดียว	65	160	95	1
2023-02-01	Wednesday	Sunny	P0070	Passion Honey		Signature	45	100	55	1
2023-02-01	Wednesday	Sunny	P0072	Singha ขวด		Beer	60	100	40	3
2023-02-01	Wednesday	Sunny	P0102	Soda		Beer	7	15	8	2
2023-02-01	Wednesday	Sunny	P0015	เค้กส้ม		Bakery	40	90	50	1
2023-02-01	Wednesday	Sunny	P0016	Soft cookie		Bakery	30	60	30	2
2023-02-01	Wednesday	Sunny	P0095	สปาเก็ตตี้ครีมซอสทรัฟเฟิล		อาหารจานเดียว	85	169	84	2
2023-02-01	Wednesday	Sunny	P0079	หมี่กผัดไข่เค็ม		ผัด ทอด	140	280	140	3
2023-02-01	Wednesday	Sunny	P0033	ข้าวคั่วกลิ้งหมู		อาหารจานเดียว	60	155	95	1
2023-02-01	Wednesday	Sunny	P0201	ผัดกะเพราราดข้าว		อาหารจานเดียว	75	122.5	47.5	2
2023-02-01	Wednesday	Sunny	P0030	ข้าวผัดพริกสดชะอมปลาทู		อาหารจานเดียว	50	120	70	1
2023-02-01	Wednesday	Sunny	P0029	ยำวุ้นเส้นโบราณ		ยำ ต้ม	50	120	70	1
2023-02-01	Wednesday	Sunny	P0026	ผัดไทกุ้งสด		อาหารจานเดียว	65	130	65	2

Figure A.1 Example of data set

Total _before_reduction	Total_Cost	Discount	Total_Sales	Net_Profit	Branch	Holiday	Employee
500.00	180	0	500	320	1	0	6
120.00	60	0	120	60	1	0	6
160.00	65	0	160	95	1	0	6
100.00	45	0	100	55	1	0	6
300.00	180	0	300	120	1	0	6
30.00	14	0	30	16	1	0	6
90.00	40	0	90	50	1	0	6
120.00	60	0	120	60	1	0	6
338.00	170	0	338	168	1	0	6
840.00	420	0	840	420	1	0	6
155.00	60	0	155	95	1	0	6
245.00	150	0	245	95	1	0	6
120.00	50	0	120	70	1	0	6
120.00	50	0	120	70	1	0	6
260.00	130	0	260	130	1	0	6

Figure A.2 Example of data set continue

APPENDIX B

Model Performance Testing Examples

B.1 Prophet Model

Purpose

Prophet is a forecasting model developed by Facebook that captures seasonality, trends, and external factors like holidays.

Code Breakdown in Figure B.1:

1. Importing Libraries (Lines 1-4)
   - Prophet from Facebook's forecasting library.
   - numpy and sklearn for computations.
2. Data Formatting (Lines 6-7)
   - Converts dataset into Prophet's required format with ds (date) and y (sales).
3. Model Initialization (Lines 9-20)
   - Enables yearly and weekly seasonality.
   - Uses multiplicative seasonality to adjust for variations.
   - Adjusts changepoint_prior_scale=0.03 for flexibility in trend changes.
4. Training (Lines 25-26)
   - Fits the model to historical data.
5. Future Predictions (Lines 28-30)
   - Forecasts for future sales for four weeks.
6. Evaluation (Lines 39-61)
   - Computes RMSE, MAE and MAPE
   - Forecasting components are plotted.

Tuned Prophet

```python
from prophet import Prophet
import numpy as np
from math import sqrt
from sklearn.metrics import mean_squared_error, mean_absolute_error, mean_squared_log_error

# Prepare the data for Prophet
prophet_df = weekly_df.rename(columns={'Week': 'ds', 'Total_Sales': 'y'})

# Initialize and fit the Prophet model with additional tuning parameters
model_prophet = Prophet(
    yearly_seasonality=True,  # Explicitly enable yearly seasonality, if your data spans multiple years
    weekly_seasonality=True,   # Enable weekly seasonality, assuming there's a weekly pattern in your data
    daily_seasonality=False,   # Disable daily seasonality if your data doesn't have a daily pattern
    seasonality_mode='multiplicative',  # Use 'additive' or 'multiplicative' based on your data's nature
    growth='linear',  # Use 'linear' for steady growth; 'logistic' for a saturation point in your forecast
    seasonality_prior_scale=20,
    changepoint_prior_scale=0.03 # Adjust for sensitivity to historical changes; increase if you expect significant shifts in trends
    # More parameters can be added based on the specific needs of your dataset
)


# Add custom seasonality if needed (e.g., quarterly patterns)
model_prophet.add_seasonality(name='quarterly', period=12, fourier_order=5)

# Fit the model
model_prophet.fit(prophet_df)

# Create future dataframe, but only for the last known data point
future = model_prophet.make_future_dataframe(periods=4, freq='W')
forecast = model_prophet.predict(future)

# Combine historical data with the forecast
combined_forecast = pd.concat([prophet_df, forecast[['ds', 'yhat', 'yhat_lower', 'yhat_upper']]])

# Plot forecast and its components
model_prophet.plot(combined_forecast)
model_prophet.plot_components(forecast)

# Calculate RMSE for the tuned model
prophet_rmse_2 = sqrt(mean_squared_error(prophet_df['y'], forecast['yhat'][:len(prophet_df)]))
print("Tuned Prophet RMSE:", prophet_rmse_2)

# Calculate additional evaluation metrics for the tuned model
# MAE
prophet_mae_2 = mean_absolute_error(prophet_df['y'], forecast['yhat'][:len(prophet_df)])
print("Tuned Prophet MAE:", prophet_mae_2)

# MAPE
prophet_mape_2 = np.mean(np.abs((prophet_df['y'] - forecast['yhat'][:len(prophet_df)]) / prophet_df['y'])) * 100
print("Tuned Prophet MAPE:", prophet_mape_2)

# MSLE
prophet_msle_2 = mean_squared_log_error(prophet_df['y'], forecast['yhat'][:len(prophet_df)])
print("Tuned Prophet MSLE:", prophet_msle_2)

# SMAPE (Symmetric Mean Absolute Percentage Error)
def smape(y_true, y_pred):
    return 100 / len(y_true) * np.sum(2 * np.abs(y_pred - y_true) / (np.abs(y_true) + np.abs(y_pred)))

prophet_smape_2 = smape(prophet_df['y'].values, forecast['yhat'][:len(prophet_df)].values)
print("Tuned Prophet SMAPE:", prophet_smape_2)

```

Figure B.1 Prophet model

B.2 ARIMA Model

Purpose

AutoRegressive Integrated Moving Average (ARIMA) is a statistical model that captures trends and seasonality in time series data.

Code Breakdown in Figure B.2:

1. Importing Libraries (Lines 1-6)
   - Includes statistical modeling (statsmodels) and visualization (matplotlib).
2. Model Training (Lines 9-11)
   - Uses auto_arima to automatically find the best ARIMA parameters.
   - Seasonality is enabled (m=52) to account for weekly patterns.
3. Forecasting (Lines 12-15)
- Generates predictions for the test dataset.
4. Evaluation Metrics (Lines 17-34)
- Computes RMSE, MAE and MAPE for model performance assessment.
5. Visualization (Lines 38-52)
- Plots actual vs. predicted sales.

Tuned ARIMA

```python
import numpy as np
from sklearn.metrics import mean_squared_error, mean_absolute_error, mean_squared_log_error
from math import sqrt
import matplotlib.pyplot as plt
import pmdarima as pm
import pandas as pd

# Ensure 'Total_Sales' is used for ARIMA model fitting
arima_model = pm.auto_arima(train_arima['Total_Sales'], seasonal=True, m=52)  # Use 52 for weekly seasonality

# Forecast
arima_forecast = arima_model.predict(n_periods=len(test_arima))

# Convert forecast to a DataFrame
arima_forecast_df = pd.DataFrame(arima_forecast, index=test_arima.index, columns=['Prediction'])

# Calculate RMSE
arima_rmse_2 = sqrt(mean_squared_error(test_arima['Total_Sales'], arima_forecast))
print("Tuned ARIMA RMSE:", arima_rmse_2)

# Calculate additional evaluation metrics
# MAE
arima_mae_2 = mean_absolute_error(test_arima['Total_Sales'], arima_forecast)
print("Tuned ARIMA MAE:", arima_mae_2)

# MAPE
arima_mape_2 = np.mean(np.abs((test_arima['Total_Sales'].values - arima_forecast) / test_arima['Total_Sales'].values)) * 100
print("Tuned ARIMA MAPE:", arima_mape_2)

# MSLE
arima_msle_2 = mean_squared_log_error(test_arima['Total_Sales'], arima_forecast)
print("Tuned ARIMA MSLE:", arima_msle_2)

# SMAPE (Symmetric Mean Absolute Percentage Error)
def smape(y_true, y_pred):
    return 100 / len(y_true) * np.sum(2 * np.abs(y_pred - y_true) / (np.abs(y_true) + np.abs(y_pred)))

arima_smape_2 = smape(test_arima['Total_Sales'].values, arima_forecast)
print("Tuned ARIMA SMAPE:", arima_smape_2)

# Plotting the results
plt.figure(figsize=(10, 6))
plt.plot(train_arima['Total_Sales'], label='Train')
plt.plot(test_arima['Total_Sales'], label='Test')
plt.plot(arima_forecast_df, label='ARIMA Predictions', linestyle='--', color='red')
plt.title("ARIMA Model vs Actual Sales")
plt.xlabel("Date")
plt.ylabel("Sales")
plt.legend()
plt.grid(True)
plt.show()

```

Figure B.2 ARIMA model

B.3 LSTM Model

Purpose

The Long Short-Term Memory (LSTM) network is a deep learning model that handles sequential dependencies in time-series forecasting.

Code Breakdown in Figure B3:

1. Importing Libraries (Lines 1-9)
    - TensorFlow/Keras is used for deep learning.
    - Adam optimizer and EarlyStopping are included to improve convergence.
2. Model Architecture (Lines 10-17)
    - First LSTM Layer:
        - 200 units with ReLU activation, returning sequences.
        - Includes a Dropout(0.2) layer to prevent overfitting.
    - Second LSTM Layer:
        - 100 units with another dropout layer.
    - Final Dense Layer:

- Outputs the predicted value.

3. Early Stopping (Lines 19-21)
   - Monitors validation loss and stops training if no improvement for 20 epochs.
4. Training Process (Lines 23-31)
   - Trained for 300 epochs with a batch size of 16.
5. Prediction and Evaluation (Lines 35-58)
   - Forecasts the test set and evaluates RMSE, MAE and MAPE.
6. Loss Visualization (Lines 60-68)
   - A plot shows how the loss decreases over epoch.

Tuned LSTM

```python
import numpy as np
from math import sqrt
from sklearn.metrics import mean_squared_error, mean_absolute_error, mean_squared_log_error
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense, Dropout
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import EarlyStopping
import matplotlib.pyplot as plt

# Build and compile a tuned LSTM model with additional Dropout and EarlyStopping
model_lstm = Sequential([
    LSTM(200, activation='relu', input_shape=(X_lstm_train.shape[1], X_lstm_train.shape[2]), return_sequences=True),
    # Dropout(0.2),  # Dropout layer to prevent overfitting
    LSTM(100, activation='relu', return_sequences=False),
    # Dropout(0.2),  # Additional Dropout layer
    Dense(1)
])
model_lstm.compile(optimizer=Adam(learning_rate=0.001), loss='mse')

# Implement EarlyStopping to stop training if no improvement in loss
early_stopping = EarlyStopping(monitor='loss', patience=20, restore_best_weights=True)

# Train the model and record the training history
history = model_lstm.fit(
    X_lstm_train,
    y_lstm_train,
    epochs=300,
    batch_size=16,
    verbose=1,
)

# Make predictions on the test set
lstm_predictions = model_lstm.predict(X_lstm_test)

# Calculate RMSE for the tuned model
lstm_rmse_2 = sqrt(mean_squared_error(y_lstm_test, lstm_predictions))
print("Tuned LSTM RMSE:", lstm_rmse_2)

# Calculate additional evaluation metrics for the tuned LSTM model
# MAE
lstm_mae_2 = mean_absolute_error(y_lstm_test, lstm_predictions)
print("Tuned LSTM MAE:", lstm_mae_2)

# MAPE
lstm_mape_2 = np.mean(np.abs((y_lstm_test - lstm_predictions.flatten()) / y_lstm_test)) * 100
print("Tuned LSTM MAPE:", lstm_mape_2)

# MSLE
lstm_msle_2 = mean_squared_log_error(y_lstm_test, lstm_predictions)
print("Tuned LSTM MSLE:", lstm_msle_2)

# SMAPE (Symmetric Mean Absolute Percentage Error)
def smape(y_true, y_pred):
    return 100 / len(y_true) * np.sum(2 * np.abs(y_pred - y_true) / (np.abs(y_true) + np.abs(y_pred)))

lstm_smape_2 = smape(y_lstm_test, lstm_predictions.flatten())
print("Tuned LSTM SMAPE:", lstm_smape_2)

# Plot the training loss over epochs
plt.figure(figsize=(10, 6))
plt.plot(history.history['loss'], label='Training Loss', color='blue')
plt.xlabel('Epochs')
plt.ylabel('Loss (MSE)')
plt.title('Tuned LSTM Model Training Loss Over Epochs')
plt.legend()
plt.grid(True)
plt.show()

```

Figure B.3 LSTM model

B.4 Random Forest Model

Purpose

The Random Forest model is an ensemble learning technique used for regression tasks. It is particularly useful for handling non-linear relationships and capturing interactions between variables.

Code Breakdown in Figure B.4:

1. Importing Libraries (Lines 1-5)
    - Essential libraries for numerical computations (numpy), error metrics (sklearn.metrics), and model training (sklearn.ensemble.RandomForestRegressor).
2. Data Preparation (Lines 7-10)
    - The dataset is structured into sequences for time-series forecasting.
    - The data is split into 75% training and 25% testing.
3. Model Training (Lines 13-21)

- A RandomForestRegressor model is initialized with:
    - n_estimators=100: Uses 100 decision trees.
    - max_depth=10: Limits tree depth to avoid overfitting.
    - min_samples_split=2: Minimum samples needed to split a node.
    - min_samples_leaf=1: Minimum samples at a leaf node.
    - random_state=42: Ensures reproducibility.
- The model is trained on the dataset.

4. Prediction and Evaluation (Lines 24-49)

- The trained model predicts test data.
- Multiple error metrics are computed:
    - RMSE (Root Mean Squared Error): Measures the deviation between actual and predicted values.
    - MAE (Mean Absolute Error): Captures absolute differences.
    - MAPE (Mean Absolute Percentage Error): Shows percentage-based prediction error.

Tuned Random Forest

```python
import numpy as np
from math import sqrt
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score, mean_squared_log_error
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor

# Prepare data for Random Forest
X_rf, y_rf = create_sequences(lstm_data, window_size)

# Split the data into training and testing sets (75% train, 25% test)
X_rf_train, X_rf_test, y_rf_train, y_rf_test = train_test_split(X_rf, y_rf, test_size=0.25, shuffle=False)

# Train the tuned Random Forest model with additional parameters
rf_model = RandomForestRegressor(
    n_estimators=100,
    max_depth=10,
    min_samples_split=2,      # Minimum samples required to split a node
    min_samples_leaf=1,        # Minimum samples required at a leaf node
    random_state=42
)
rf_model.fit(X_rf_train, y_rf_train)

# Make predictions on the test set
rf_predictions = rf_model.predict(X_rf_test)

# Calculate RMSE for the tuned model
rf_rmse_2 = sqrt(mean_squared_error(y_rf_test, rf_predictions))
print("Tuned Random Forest RMSE:", rf_rmse_2)

# Calculate additional evaluation metrics for the tuned Random Forest model
# MAE
rf_mae_2 = mean_absolute_error(y_rf_test, rf_predictions)
print("Tuned Random Forest MAE:", rf_mae_2)

# MAPE
rf_mape_2 = np.mean(np.abs((y_rf_test - rf_predictions) / y_rf_test)) * 100
print("Tuned Random Forest MAPE:", rf_mape_2)

# MSLE
rf_msle_2 = mean_squared_log_error(y_rf_test, rf_predictions)
print("Tuned Random Forest MSLE:", rf_msle_2)

# SMAPE (Symmetric Mean Absolute Percentage Error)
def smape(y_true, y_pred):
    return 100 / len(y_true) * np.sum(2 * np.abs(y_pred - y_true) / (np.abs(y_true) + np.abs(y_pred)))

rf_smape_2 = smape(y_rf_test, rf_predictions)
print("Tuned Random Forest SMAPE:", rf_smape_2)
```

Figure B.4 Random Forest model

APPENDIX C

Examples of Model Evaluation Results

C.1 Model comparison

In Figure C.1 presents the comparative performance of four forecasting models: Prophet, ARIMA, LSTM, and Random Forest, using three key evaluation metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The lower the values in these metrics, the better the model's accuracy in predicting weekly total sales.

C.1.1. Root Mean Squared Error (RMSE)

Definition: RMSE measures the average magnitude of errors between predicted and actual values, giving higher weight to larger errors. Lower values indicate better predictive accuracy.

Model Rankings (Best to Worst in RMSE)

1.Prophet: 7,136.43 (Lowest RMSE, best performance)
2.ARIMA: 14,111.54
3.Random Forest: 14,477.71
4.LSTM: 15,563.97 (Highest RMSE, worst performance)

- Prophet outperforms all other models in RMSE, suggesting it provides the most stable and reliable forecasts.
- LSTM has the highest RMSE, indicating it struggles with high variance in predictions.

C.1.2. Mean Absolute Error (MAE)

Definition: MAE calculates the absolute differences between predicted and actual values. Unlike RMSE, MAE treats all errors equally.

Model Rankings (Best to Worst in MAE)

1.Prophet: 5,376.75
2.ARIMA: 8,043.16
3.Random Forest: 9,797.89
4.LSTM: 9,204.96

- Prophet again achieves the lowest MAE, confirming its ability to maintain consistent predictions with minimal absolute errors.
- Random Forest and LSTM have the highest MAE, meaning their predictions deviate more frequently from actual sales.

C.1.3. Mean Absolute Percentage Error (MAPE)

Definition: MAPE expresses error as a percentage of actual values, making it useful for business decision-making.

Model Rankings (Best to Worst in MAPE)

1.Prophet: 8.35% (Lowest, most accurate)
2.ARIMA: 17.81%
3.Random Forest: 19.58%
4.LSTM: 21.14% (Highest error, least accurate)

- With a MAPE of only 8.35%, Prophet is the most reliable model for business forecasting.

- LSTM has the highest error rate at 21.14%, meaning it struggles with variations in sales data.

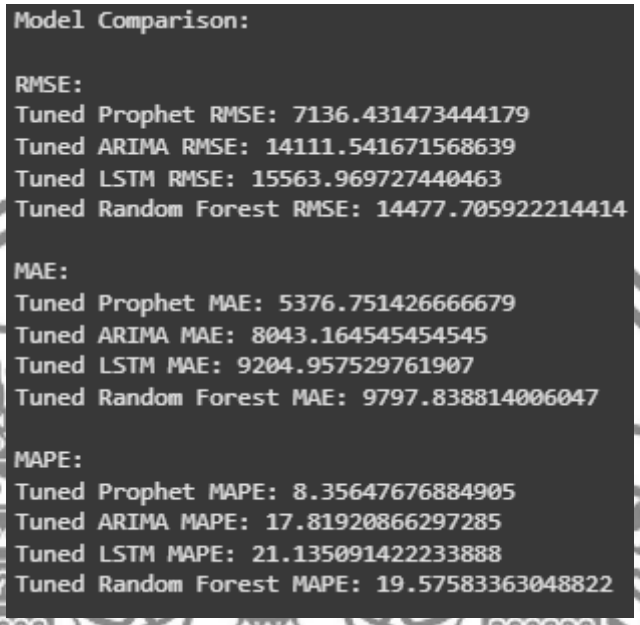


Figure C.1 Model comparison

C.2 Overfitting Check

Figure C.2 presents the model performance evaluation and overfitting check for the Prophet (the best model) forecasting model.

- Model Performance:
  - RMSE (Root Mean Squared Error) MAE (Mean Absolute Error) and MAPE (Mean Absolute Percentage Error) are provided for both training and test datasets.
  - The test error is not significantly higher than the train error, with a percentage difference of 44.28% for RMSE ,51.34% for MAE and 44% for MAPE.
- Overfitting Check:
  - RMSE indicates no overfitting.
  - MAE shows no overfitting.
  - MAPE shows no overfitting.
- Final Prediction:
  - The forecasted total sales for week 86 (September 11, 2024) is ‘79,091.17’, using the Prophet model.

This analysis suggests that while the model does not exhibit extreme overfitting, the high-test error in MAE indicates room for improvement in generalization.

```
------ Model Performance (Train vs Test Error) ------
RMSE: Train = 5714.0037, Test = 8244.4018, % Difference = 44.28%
MAE: Train = 4426.0598, Test = 6698.5465, % Difference = 51.34%
MAPE: Train = 5.8054, Test = 8.3600, % Difference = 44.00%

------ Overfitting Check ------
✅ RMSE: ไม่มี Overfitting
✅ MAE: ไม่มี Overfitting
✅ MAPE: ไม่มี Overfitting
Predicted Total Sales for Week 86 (Prophet):           ds          yhat
21 2024-09-11  79091.170692
```

Figure C.2 Overfitting check for Prophet model

C.3 Weekly Sales comparison

C.3.1. Error Analysis – Which Model Performed Best?

1. Prophet had the lowest error rate at 7.28%, making it the most accurate model.
2. ARIMA performed slightly worse, with an 8.17% error rate, showing it can still provide a reasonable forecast.
3. Random Forest had a significantly higher error rate of 13.43%, indicating difficulty in capturing time-dependent trends.
4. LSTM performed the worst, with a 17.62% error, suggesting it struggled with the dataset's variability.

C.3.2. Business Implications of the Forecasting Results

Prophet is the most reliable model for predicting weekly sales and should be prioritized for future forecasting. ARIMA could be used as a secondary model, particularly for short-term forecasting where historical trends dominate. Random Forest and LSTM underperformed, indicating that tree-based and deep learning approaches may not be the best fit for this dataset without further tuning.

Week 86

	Model	Prediction	Actual Sales	Error (%)
0	Prophet	79091.170692	85302.93	7.282000
1	ARIMA	78332.923061	85302.93	8.170888
2	Random Forest	73848.245998	85302.93	13.428242
3	LSTM	70269.809299	85302.93	17.623217

Figure C.3 Week 86 prediction comparison

APPENDIX D

Examples of Results After Analyzed

D.1 Range of before analyze, week prediction and after analyzed

Overview Figure D.1

The graph displays weekly total sales trends from February 1, 2023, onward. The sales trend is analyzed over time, with a clear distinction between the period before analysis (left side) and after analysis (right side) using a red vertical line as a separator.

Before Data Analysis (Weeks 1-85)

1. Sales fluctuations show irregular patterns, with peaks and drops indicating inconsistent demand.
2. There are multiple sharp spikes, suggesting seasonal sales surges, but overall sales remain within a limited range.
3. Some periods exhibit noticeable declines, indicating potential inefficiencies in marketing, product offerings, or staffing strategies.

After Data Analysis (Weeks 86-104)

1. A significant increase in sales is observed, particularly after Week 86, where total sales rise sharply.
2. The sales trend shows a consistent upward trajectory, indicating effective business decisions based on analytical insights.
3. The implementation of strategic changes—such as menu optimization, staff relocation, and targeted promotions—contributed to a boost in revenue.

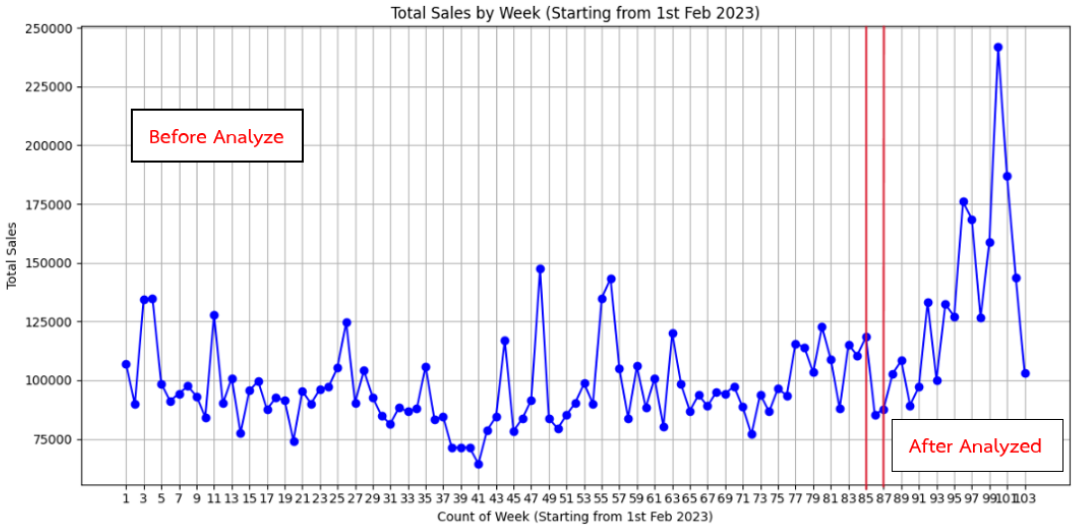


Figure D.1 Total sales by week before and after use data analyzed

D.2 Comparison weekly sales between Sep-Jan last year with current year

Overview Figure D.2

This graph provides a comparative analysis of weekly sales from September to January of the previous year (Weeks 34-51, in blue) and the projected sales for the same period in the current year (Weeks 86-103, in red). The comparison aims to evaluate seasonal trends, sales growth, and the impact of business improvements over time.

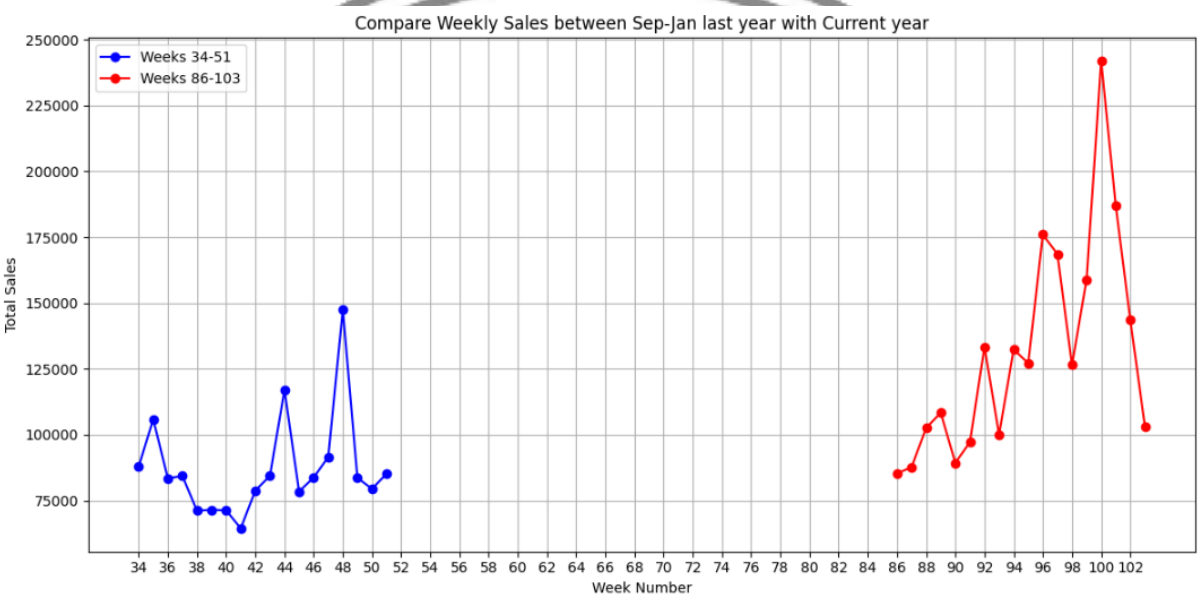


Figure D.2 Weekly sales between Sep-Jan last year with current year

D.3 Data-Driven Strategic Adjustments and Business Impact

Overview Figure D.3 and D.4

After conducting a detailed data analysis, it was observed that Hawaiian Pizza was a highly popular item. Based on this insight, we developed the Hawaiian Pizza ,new pizza menu items were introduced, including Seafood Pizza and Truffle Pizza, to expand the product variety and cater to customer preferences.

Impact on Total Sales and Staffing Adjustments

- Total sales significantly improved due to the successful expansion of the pizza category, reflecting a strong demand for diversified menu offerings.
- To accommodate the increase in customer traffic, adjustments were made to the employee scheduling strategy:
  - Weekdays: Increased the number of employees to 7 to ensure efficient service during regular business days.
  - Weekends & Holidays: Increased staffing to 9 employees to handle the higher volume of customers effectively.

- The optimized workforce allocation has resulted in better service efficiency and improved customer satisfaction, ultimately contributing to higher revenue.

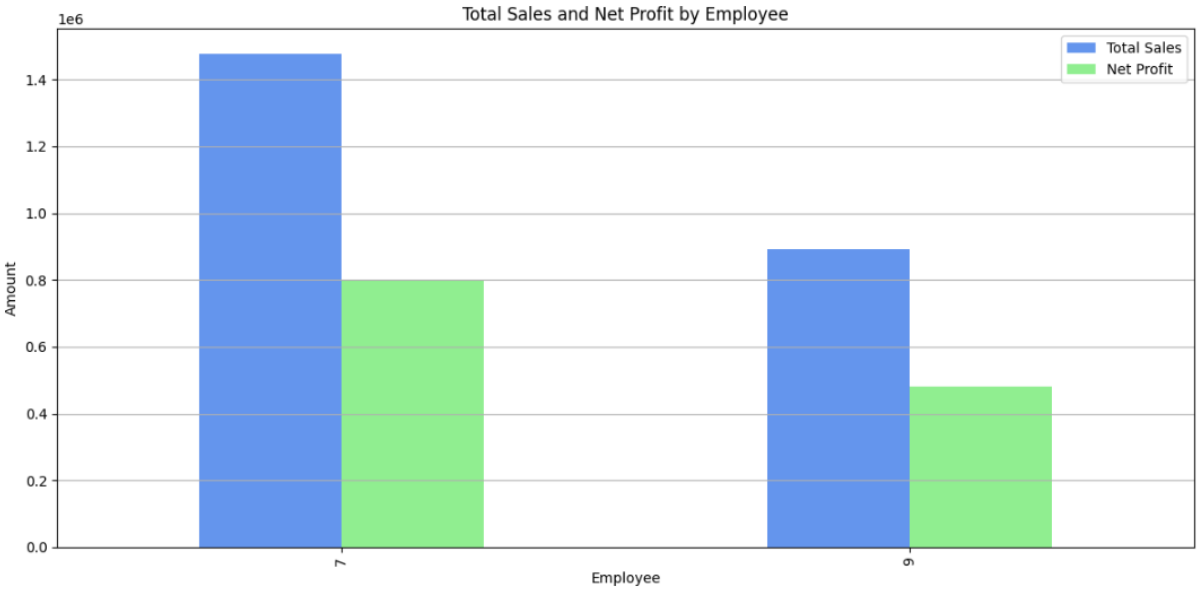


Figure D.3 New sales and net profit by employees

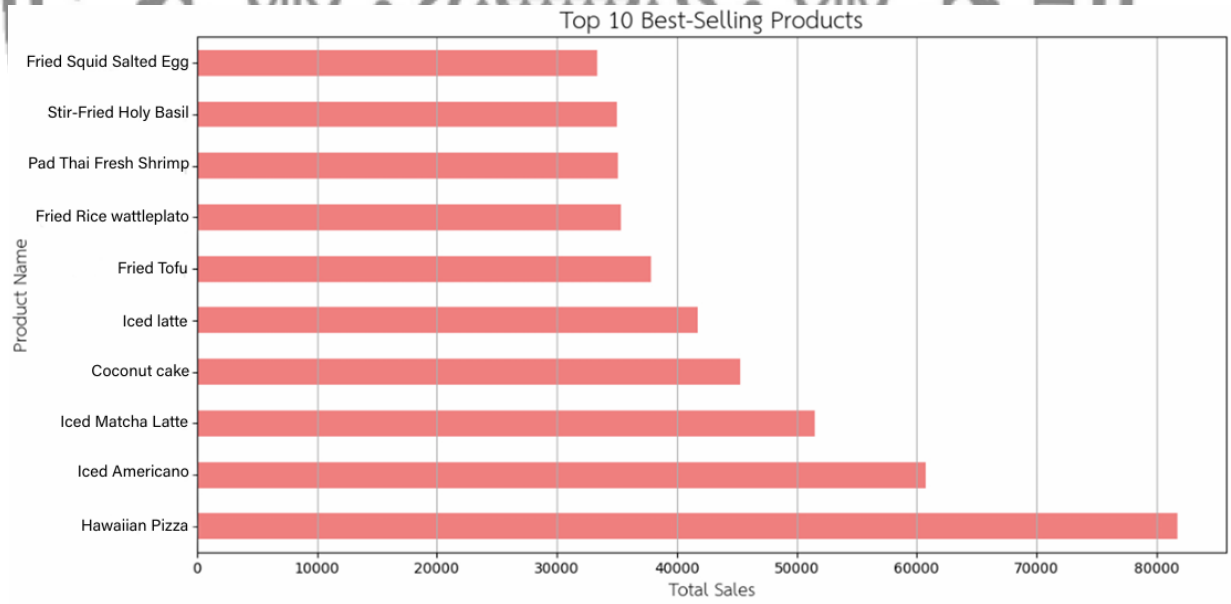


Figure D.4 New top 10 best-selling products

AUTHOR BIOGRAPHY

Name-Surname	Natthapong Jitjakool
Date of Birth	June 10, 1998
Place of Birth	Bangkok Thailand
Education	Bachelor of Science Program in Applied Physics King Mongkut's Institute of Technology Ladkrabang, GPAX: 3.03
Current Address	29/18 Moo 1, Bangsrimuang, Muang Nonthaburi, Nonthaburi 11000, Thailand
Publication	-
Award Received	-