FORECASTING AIRASIA’S PROFITABILITY BASED ON FUEL PRICE TRENDS USING ARIMA AND XGBOOST

MIRA EDORA BINTI YUNOS

UNIVERSITI TEKNOLOGI MALAYSIA

PSZ 19:16 (Pind. 1/23)

UNIVERSITI TEKNOLOGI MALAYSIA
DECLARATION OF Choose an item.

Author's full name	:	Mira Edora Binti Yunos			
Student's Matric No.	:	MCS241057	Academic Session	:	I
Date of Birth	:	24 JULY 2002	UTM Email	:	miraedora@graduate.utm.my
Title	:	FORECASTING AIRASIA'S PROFITABILITY BASED ON FUEL PRICE TRENDS USING ARIMA AND XGBOOST			

I declare that this Choose an item. is classified as:

☒	**OPEN ACCESS**	I agree that my report to be published as a hard copy or made available through online open access.
☐	**RESTRICTED**	Contains restricted information as specified by the organization/institution where research was done. *(The library will block access for up to three (3) years)*
☐	**CONFIDENTIAL**	Contains confidential information as specified in the Official Secret Act 1972)

(If none of the options are selected, the first option will be chosen by default)

I acknowledged the intellectual property in the Choose an item. belongs to Universiti Teknologi Malaysia, and I agree to allow this to be placed in the library under the following terms :

1. This is the property of Universiti Teknologi Malaysia
2. The Library of Universiti Teknologi Malaysia has the right to make copies for the purpose of research only.
3. The Library of Universiti Teknologi Malaysia is allowed to make copies of this Choose an item. for academic exchange.

Signature of Student:

Signature :

Full Name: Mira Edora Binti Yunos
Date : 30 June 2025

Approved by Supervisor(s)

Signature of Supervisor I:	Signature of Supervisor II
Full Name of Supervisor I	Full Name of Supervisor II
Date :	Date :

NOTES : If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction

Date:

Librarian

Jabatan Perpustakaan UTM,

Universiti Teknologi Malaysia,

Johor Bahru, Johor

Sir,

CLASSIFICATION OF THESIS AS RESTRICTED/CONFIDENTIAL

TITLE: FORECASTING AIRASIA'S PROFITABILITY BASED ON FUEL PRICE TRENDS USING ARIMA AND XGBOOST

AUTHOR'S FULL NAME:MIRA EDORA BINTI YUNOS

Please be informed that the above-mentioned thesis titled __________should be classified as RESTRICTED/CONFIDENTIAL for a period of three (3) years from the date of this letter. The reasons for this classification are

(i)

(ii)

(iii)

Thank you.

Yours sincerely,

SIGNATURE:

NAME: Prof. Madya. Ts. Dr. Mohd Shahizan bin Othman

ADDRESS OF SUPERVISOR:

"I hereby declare that I have read this thesis and in my
opinion this thesis is sufficient in term of scope and quality for the
award of the degree of Master in Data Science"

Signature	:	___________________________
Name of Supervisor I	:	PROF. MADYA. TS. DR. MOHD SHAHIZAN BIN OTHMAN
Date	:	30 JUNE 2025

FORECASTING AIRASIA’S PROFITABILITY BASED ON FUEL PRICE TRENDS USING ARIMA AND XGBOOST

MIRA EDORA BINTI YUNOS

A project report submitted in partial fulfilment of the
requirements for the award of the degree of
Master in Data Science

School of Education
Faculty of Social Sciences and Humanities
Universiti Teknologi Malaysia

JUNE 2025

DECLARATION

I declare that this thesis entitled *"Forecasting Airasia's Profitability Based on Fuel Price Trends Using Arima and Xgboost"* is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature :

Name : MIRA EDORA BINTI YUNOS

Date : 30 JUNE 2025

ACKNOWLEDGEMENT

Throughout the course of this thesis, I have had the privilege of interacting with numerous scholars, practitioners, and researchers whose guidance and insights have greatly enriched my academic journey. I would like to express my sincere gratitude to Prof. Madya. Ts. Dr. Mohd Shahizan bin Othman , my lecturer, for his unwavering encouragement, expert guidance, and constructive feedback throughout the writing process.

I am deeply grateful to my mother, who has been my greatest source of inspiration and motivation. Her unconditional love and support have given me the strength to pursue my Master's degree at Universiti Teknologi Malaysia (UTM). Her belief in me has been a constant driving force in my life.

Lastly, I also extend my heartfelt thanks to my fellow postgraduate students, whose friendship, collaboration, and assistance were invaluable during this challenging yet rewarding journey. I sincerely appreciate the support from my colleagues and all those who offered their time, opinions, and advice at various stages of this research. Though I cannot name each one individually, I am truly grateful for their contributions.

ABSTRACT

This thesis focuses on forecasting AirAsia’s profitability in response to fuel price trends using a hybrid approach that combines ARIMA for univariate time series analysis and XGBoost for multivariate machine learning modeling, with the aim of improving forecasting accuracy and providing actionable insights for financial planning and strategic decision-making in the airline industry. The dataset's used spans a minimum of 3–5 years, contains between 500,000 and 1,000000 records, and includes at least 5–7 key variables such as fuel prices, revenue, operating costs, and passenger demand, ensuring it is comprehensive yet manageable in size and complexity. Data preprocessing steps such as handling missing values, outlier detection, normalization, and feature engineering including lagged features, rolling averages, and seasonal indicators are applied to enrich the dataset's and enhance model performance. The study also explores the relationship between fuel price volatility and profitability, highlighting the importance of integrating time series decomposition and exploratory data analysis (EDA) to uncover meaningful patterns and trends. Ultimately, the research contributes to advancing the application of machine learning and time series analysis in aviation finance, offering a framework that supports cost management, risk mitigation, and strategic planning in an economically volatile environment.

Table of Contents

Chapter 1: Introduction

1.0 Introduction 1
1.1 Problem Background 2
1.2 Problem Statement 3
1.3 Research Questions 4
1.4 Objectives of the Research 5
1.5 Scope of the Study 6
1.6 Significance of the Research 7
1.7 Structure of the Thesis 8
1.8 Summary 9

Chapter 2: Literature Review

2.0 Introduction 10
2.1 Overview of AirAsia and the Volatile Airline Sector 11
 2.1.1 Global and Regional Airline Industry Landscape 11
 2.1.2 AirAsia's Business Model and Sensitivity to Fuel Prices 12
2.2 Fuel Price Volatility and Its Impact on Airline Profitability 13
 2.2.1 Fuel Pricing Trends in Malaysia and Global Drivers 13
 2.2.2 Fuel as a Primary Operational Cost 14
2.3 Forecasting Models ARIMA and XGBoost 15
 2.3.1 ARIMA in Fuel Price Forecasting 15
 2.3.2 Limitations of ARIMA for Nonlinear Events 16
 2.3.3 Applications of XGBoost in Energy and Aviation Forecasting 17
2.4 Overview of Hybrid Models 18
2.5 Summary 19

Chapter 3: Research Methodology

3.0 Introduction 20

3.1 Research Framework 21
3.2 Problem Formulation 22
3.3 Data Collection 23
3.4 Data Pre-processing 24
3.5 Feature Engineering and Time Series Decomposition 25
3.5.1 Lagged Features and Rolling Averages 25
3.5.2 Seasonal Features and Derived Metrics 26
3.5.1.1 Decomposition of Time Series 26
3.5.1.2 Visualization for EDA 27
3.5 Model Development and Evaluation 28
3.6 Summary 29

Chapter 4: initial Results
4.0 Introduction 30
4.1 Exploratory Data Analysis (EDA) 31
4.2 Data Preparation and Cleaning 32
4.2.1 Overview of the Dataset 32
4.2.2 Correlation and Profit Trends Analysis 33
4.2.3 Outlier and Missing Value Treatment 34
4.3 Initial Model Results 35
4.3.1 XGBoost Model Results 35
4.3.2 ARIMA Model Results 36
4.4 Summary 37

Chapter 5: Discussion, Conclusion and Future Work
5.0 Introduction 38
5.1 Summary of Findings 39
5.2 Conclusions 40
5.3 Future Work 41

CHAPTER 1

INTRODUCTION

1.0 Introduction

AirAsia, a leading low-cost carrier in Asia, operates within one of the most volatile and economically sensitive sectors: the airline industry. With fuel accounting for approximately 30% of total airline expenditures, fuel price fluctuations play a significant role in determining overall profitability (Chen et al., 2024). Despite AirAsia's aggressive digital transformation initiatives—ranging from AI-driven customer service platforms to predictive analytics for operational efficiency—its exposure to external economic shocks, particularly in fuel markets, remains substantial (Wu, 2024). Given the narrow profit margins characteristic of low-cost carriers, even a modest rise in fuel prices can cause disproportionate impacts on financial performance, with estimates showing that a 10% fuel price increase could reduce profits by up to 15% for carriers like AirAsia (Cai et al., 2025).

In light of these vulnerabilities, accurate fuel price forecasting has become essential for risk mitigation, financial planning, and long-term strategic decision-making. In Malaysia, fuel prices particularly Ron97, which is directly linked to international crude oil markets without subsidies are influenced by multiple factors such as global oil supply, foreign exchange volatility, and domestic tax policies (Sokkalingam et al., 2021). Traditional statistical methods such as ARIMA have been successfully used to model such time series due to their strength in capturing linear temporal dependencies. However, the rise of machine learning (ML) techniques like XGBoost offers complementary benefits by modelling nonlinear relationships and incorporating multiple external features, thereby improving predictive accuracy (Pin Li & Zhang, 2018).

This research project aims to forecast AirAsia's profitability by modelling historical fuel price trends as the core external cost driver. By integrating ARIMA and XGBoost into a hybrid predictive framework, the study seeks to assess the effectiveness of combining time series analysis and structured data modelling to capture both trend-based and nonlinear impacts of fuel prices on profitability. Unlike past approaches that examined financial performance in isolation or relied solely on statistical or Machine learning models, this study bridges the gap by adopting a hybrid, data-driven methodology. Through this approach, the project not only enhances predictive capability but also provides actionable insights for AirAsia's financial resilience and operational planning in a fuel-volatile environment.

1.2 Problem Background

AirAsia, a major player in the low-cost airline sector, operates within a highly volatile and cost-sensitive industry where fuel expenses can account for up to 50% of operating costs (Cai et al., 2025). In recent years, escalating fuel price fluctuations—driven by global oil supply shocks, currency volatility, and evolving government subsidy policies—have posed significant challenges to the airline's profitability and financial planning (Sokkalingam et al., 2021). Although AirAsia has taken strategic steps to digitalize its operations and improve cost efficiency through predictive maintenance, mobile platforms, and AI-driven customer engagement (Wu, 2024), fuel volatility continues to represent a critical risk factor.

Traditional statistical forecasting models like ARIMA have proven effective for analyzing linear trends in time series data. However, these models often fall short when addressing complex, nonlinear patterns that dominate fuel price movements. As a result, hybrid models that combine ARIMA with machine learning algorithms like XGBoost are gaining traction for their ability to enhance predictive performance by capturing both temporal dependencies and multifactorial interactions (Baumann et al., 2021).

Despite their success in sectors like energy forecasting, the application of such hybrid models in aviation-specific profitability forecasting—particularly within the context of Malaysian fuel pricing structures—remains underexplored. In Malaysia, retail fuel prices such as those under the Managed Float System fluctuate weekly based on global benchmarks and domestic policies, yet limited access to proprietary data like MOPS hampers accurate prediction (Sokkalingam et al., 2021). This unpredictability has raised concerns among aviation stakeholders, particularly low-cost carriers like AirAsia, which operate on tight profit margins and have limited hedging capacities (Cai et al., 2025).

Accordingly, this project aims to fill a critical research and practical gap by leveraging a hybrid ARIMA and XGBoost framework to forecast AirAsia's profitability based on historical fuel price trends. By integrating time series and machine learning approaches, the study seeks to deliver a more robust and adaptive forecasting model that supports informed decision-making, minimizes financial exposure, and promotes long-term business sustainability in a turbulent operating environment.

1.3 Problem Statement

At present, there is no integrated forecasting framework specifically designed to measure and anticipate the financial impact of fuel price volatility on the profitability of low-cost airlines such as AirAsia. While traditional statistical models have been used in isolation for financial forecasting, they often fall short in capturing the nonlinear patterns and complex external influences that affect aviation fuel pricing and operational costs. This limitation becomes especially critical in an industry where fuel expenses represent a significant portion of total operating costs and where profit margins are already thin.

This project aims to address this gap by developing a hybrid predictive model that combines ARIMA for time-dependent trend analysis with XGBoost for advanced machine learning-based prediction. The model will analyse historical data to forecast fuel price trends and their projected influence on AirAsia's profitability, offering a more comprehensive and adaptable tool for financial planning and risk management. Beyond improving forecast accuracy, the project also seeks to uncover patterns in fuel cost fluctuations, economic conditions, and operational sensitivities that influence airline revenue and expenditures. The insights gained can be used to enhance strategic decision-making, optimize cost-control measures, and support the development of more resilient financial models. Ultimately, this project aspires to empower AirAsia with a data-driven approach to proactively navigate market uncertainty and maintain financial sustainability in an increasingly volatile aviation environment.

1.4 Research Questions

1. How accurately can ARIMA and XGBoost forecast AirAsia's profitability based on historical fuel price trends?

2. What are the key patterns and influencing factors behind fluctuations in fuel prices that affect airline profitability?
3. How can a hybrid forecasting model improve financial planning and risk mitigation for low-cost carriers like AirAsia?

1.5 Objectives of the Research

1. Develop and apply an ARIMA model to forecast fuel price trends over time.
2. Utilize XGBoost to model the nonlinear relationships between fuel prices and AirAsia's profitability.
3. Integrate both models into a hybrid framework to improve the accuracy of financial forecasting.

1.6 Scope of the Study

1. The project focuses on forecasting the profitability of AirAsia in relation to fuel price trends in Malaysia.
2. Historical data on fuel prices and AirAsia's financial performance will be utilized, primarily from open data platforms or official reports.
3. The study applies ARIMA for time series analysis and XGBoost for structured data modeling, combining them into a hybrid forecasting model.
4. The analysis is limited to economic and operational factors that influence profitability, with fuel price being the central external variable considered.

1.7 Significance of the Research

The significance of this research lies in its potential to provide a more accurate and data-driven approach to forecasting profitability in the airline industry, particularly for low-cost carriers like AirAsia. By developing a hybrid model that integrates ARIMA and XGBoost, this study aims to improve predictive accuracy in assessing the financial impact of fluctuating fuel prices—one of the most critical cost components in aviation operations.

The findings of this project can support more informed decision-making in areas such as cost management, risk mitigation, and strategic planning. In addition, the proposed forecasting framework may serve as a valuable tool for financial analysts, airline executives, and policymakers seeking to enhance the financial resilience and sustainability of aviation businesses in volatile economic conditions. Ultimately, the study contributes to advancing the application of machine learning and time series analysis in the field of aviation finance.

1.8 Structure of the Thesis

This thesis is organized into five comprehensive chapters, each addressing different aspects of the research in a systematic manner. The structure is designed to guide the reader from the research context through to the findings and implications of the study.

Chapter 1: Introduction

This chapter introduces the core research problem—forecasting AirAsia's profitability in the context of volatile fuel price trends. It outlines the background of the aviation industry, emphasizing the sensitivity of low-cost carriers like AirAsia to fluctuations in fuel costs. The chapter also presents the problem statement, research questions, objectives, scope, significance, and the overall structure of the thesis. The aim is to establish a clear understanding of why this research is relevant and necessary in today's dynamic airline and energy environment.

Chapter 2: Literature Review

This chapter provides a review of existing literature related to airline profitability, fuel price forecasting, and the application of time series and machine learning models such as ARIMA and XGBoost. It explores previous studies on cost drivers in the aviation industry, the role of predictive analytics, and hybrid forecasting techniques. Gaps in current research are identified, particularly the limited adoption of hybrid models in aviation financial forecasting within the Malaysian context. This review sets the foundation for the proposed methodological framework used in this study.

Chapter 3: Research Methodology

The methodology chapter describes the research design, data sources, tools, and techniques used to conduct the study. It details the process of collecting historical fuel price and financial performance data, data preprocessing steps, and the implementation of ARIMA and XGBoost models. The rationale for using a hybrid approach is explained, alongside the evaluation metrics used to assess model performance. This chapter ensures transparency and reproducibility by outlining every step taken during the forecasting process.

Chapter 4: Data Analysis and Results

This chapter presents the outcomes of the exploratory data analysis and model implementation. Descriptive statistics, trend visualizations, and diagnostics are provided to understand the structure of the data. The performance of both the ARIMA and XGBoost models is evaluated individually, followed by the integration and analysis of the hybrid model. Forecasting results are presented, interpreted, and discussed in relation to the research objectives. Insights into how fuel price movements influence AirAsia's profitability are highlighted.

Chapter 5: Discussion, Conclusion, and Future Work

In the final chapter, the findings of the study are discussed in depth. The research outcomes are connected back to the initial objectives and are compared with findings from previous studies to highlight their implications. Strategic recommendations are proposed for improving financial forecasting and risk management in the airline sector. Limitations of the study are acknowledged, and potential areas for future research are outlined, such as integrating additional economic indicators or applying the model to other regional carriers.

1.9 Summary

This chapter has introduced the research project focused on forecasting AirAsia's profitability based on historical fuel price trends using a hybrid ARIMA and XGBoost model. The problem statement identified the lack of an integrated and accurate predictive framework tailored to the Malaysian aviation context, highlighting the need for advanced forecasting techniques that combine time series analysis with machine learning.

The chapter also presented the research questions and objectives, which aim to develop a robust forecasting model capable of capturing both linear and nonlinear patterns in fuel pricing data and their relationship to AirAsia’s financial performance. The scope and significance of the study were defined, emphasizing its practical value in supporting strategic planning, cost management, and risk mitigation in the airline sector.

Finally, the chapter outlined the structure of the thesis, providing an overview of the five chapters and how they collectively contribute to addressing the research problem. In essence, Chapter 1 has laid the groundwork for a data-driven investigation into financial forecasting in aviation, setting the stage for a deeper exploration of relevant literature, methodology, and analytical results in the chapters that follow.

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

The environment of the airline industry is a highly turbulent one since the demand is broadly cyclical, there is high operating leverage, and the industry has great exposure to financial shocks from adverse economic conditions and rising fuel prices (Chen et al., 2024). Though low-cost carriers like AirAsia have models that look cost-efficient, they are extremely vulnerable because of narrow profit margins and heavy dependence on fuel where it accounts for up to 30% of expenditure. This history proves that AirAsia can be very financially resilient it shifted from being a highly indebted company to becoming one of Asia's major low-cost airlines. However, with more rules on global carbon emissions and volatility in fuel, this will be an even bigger challenge for them to stay profitable. Sustainable aviation fuel (SAF) has been highlighted as a long-term decarbonization strategy, yet its limited commercial adoption, high production cost, and minimal penetration into global fuel markets which is less than 1% as of 2019 continue to pressure airlines financially (Chen et al., 2024). As fuel prices remain a dominant and unstable cost factor, accurately forecasting their impact on profitability becomes vital for financial planning and risk mitigation in the aviation sector.

In recent years, artificial intelligence and machine learning techniques have become very popular in big data environments as well as in the oil and gas industries because, through AI models, predictive efficiency and operational accuracy have been improved. Applications of this approach include production forecasting, optimization of drilling processes, and reservoir analysis; in all these applications, AI demonstrated strong ability to learn from nonlinear high-dimensional data. From this base, therefore, the same predictive analytical frameworks such as ARIMA for time series forecasting and XGBoost for structured data modelling, are being increasingly adopted for airline financial forecasting wherein fuel cost trends are critical. The fusion of time-series analysis with advanced ML algorithms like XGBoost offers a powerful modelling approach to capture both temporal dependencies and complex feature interactions. As such, this study adopts ARIMA and XGBoost to forecast AirAsia's profitability by examining historical fuel price patterns, addressing a critical gap in financial forecasting practices tailored to aviation dynamics.

2.1 Overview of AirAsia and the Volatile Airline Sector

This section provides an analysis of the dynamic nature of the global airline industry, highlighting its susceptibility to economic fluctuations, regulatory challenges, and environmental concerns. It discusses the essential role of sustainable aviation fuel (SAF) as a real solution to lower carbon emissions and meet long-term goals of reducing carbon impact. Also looked at are the obstacles to using SAF, like being very expensive and not able to grow easily, while stressing the importance of rules to help and new technology. This part also highlights how important it is to include better forecasting models like XGBoost to deal with business and money risks in the field leading toward greener and stronger airline services.

2.1.1 Global and Regional Airline Industry Landscape

The global airline industry is vulnerable to several factors that create a significant operational and financial environment for the business. These factors include shocks on oil prices, fluctuations in economic activities, climate risks, and media-driven investor sentiment. Regarding what pertains to the shocks of oil price volatility particularly supply and demand shocks, there is a direct relationship with profitability since fuel accounts for 20 until 50% of operating costs (Cai et al., 2025). Economic activity shocks like cyclical changes in demand are more fundamental drivers of airline stock returns while the volatility is disproportionately led by oil supply and consumption demand shocks. For example, sensitivity to oil supply shock is more pronounced in airlines such as Southwest, American Airlines while the volatility consequent upon consumption demand fluctuation is more pronounced among low-cost carriers like Ryanair. Climate risks are also a key factor, as severe weather events increase the need for route adjustments and raise maintenance expenses. This, in turn, heightens fluctuations in profitability and makes stock prices more sensitive—especially during busy travel periods (Montero et al., 2024). Climate risks further amplify operational disruptions and stock price volatility, particularly during peak travel seasons, as extreme weather events increase costs such as rerouting, maintenance in which attract media attention, led to exacerbates investor uncertainty (Jin & Cairang, 2025).

These factors vary in the different regions. For instance, Asian and European carriers like China Southern and Air France are exposed to extreme climate-related variability because of seasonal demand high plus regulatory pressure, while the North American counterparts, such as Delta and United are more susceptible to inventory demand shocks of oil. The extremes of climate risks reported in the media disproportionally affect the stock performance in markets that are more densely populated (Jin & Cairang, 2025). After the pandemic, this increased further as economic activity and stability in the oil market increased reliance on the industry; therefore, strategies need to change fuel hedging adaptive strategies with safe adoption to mitigate risks (Cai et al., 2025). These strategies underscore the need for region-specific risk management frameworks to address interconnected financial, operational and even environmental pressures.

2.1.2 AirAsia's Business Model and Sensitivity to Fuel Prices

AirAsia, a pioneer in the Asian low-cost carrier (LCC) market, adopts a business model which is highly efficient and low-cost with aggressive fleet expansion characterized by high aircraft utilization and ancillary revenue streams. This allows AirAsia to use secondary airports, optimize its route network, and minimize operational frills scaling the company to more than 60 destinations in 16 countries with 84 aircraft by 2009 (Chen et al., 2024). For AirAsia, the dominance of fuel as a cost driver is made worse by its short-haul high-frequency flights which augment fuel consumption relative to revenue generation.

As a result of external economic fluctuations and fuel price volatility, AirAsia's profitability risks are heightened significantly. LCCs are adversely affected by oil price shocks because their profit margins are thin and hedging capacity is limited (Cai et al., 2025). It is estimated that a 10% increase in fuel prices can result in a loss of up to 15% of an LCC's net profit, owing to their inability to quickly pass on costs to price-sensitive customers (Cai et al., 2025). Further, AirAsia's financial performance is dependent upon macroeconomic conditions during recessions, as seen in its 2001 debt restructuring in the context of rising oil prices (Chen et al., 2024). Therefore, this underscored the airline's vulnerability to exogenous shocks, necessitating adaptive strategies like fuel hedging and fleet modernization to mitigate risks.

2.2 Fuel Price Volatility and Its Impact on Airline Profitability

This section explores how fuel price fluctuations significantly affect airline profitability. It begins by examining fuel price trends in Malaysia and the key factors that drive volatility, such as taxes, global oil supply, and geopolitical events. It also explains how hybrid forecasting models like ARIMA and XGBoost to help improve prediction accuracy. This section also highlights how jet fuel is the largest operating cost for airlines and how rising prices can reduce profit margins, even with strategies like fuel hedging and fleet upgrades. Finally, it looks at how climate-related disruptions increase costs through rerouting and maintenance, adding more financial uncertainty.

2.2.1 Fuel Pricing Trends in Malaysia and Global Drivers

Several factors influence fuel pricing trends, including taxation policies, global oil supply and demand dynamics, and geopolitical risk factors. Prices are volatile due to these factors, which are critical for industries that rely on fuel, such as aviation. It has been shown by Pin Li and Jin-Suo Zhang (2018) that advanced hybrid forecasting models such as ARIMA and XGBoost are highly accurate in predicting the security of energy supplies and fuel prices in China, thereby providing valuable insight into how similar methodologies could be applied to analysis of Ron97 pricing trends in Malaysia. Several factors influence retail fuel prices, including supply-demand imbalances, geopolitical tensions, and taxation, making it essential for stakeholders to utilize robust predictive tools to anticipate and mitigate risks.

The interconnected nature of global energy markets means that disruptions in one region can ripple across the globe, influencing fuel costs and availability. For instance, geopolitical events such as trade disputes or sanctions often lead to sudden spikes in oil prices, affecting airlines' operational budgets and profitability. In their study, Pin Li and Jin-Suo Zhang (2018) highlight the importance of integrating multiple indicators—such as energy dependence, production diversity, and clean energy adoption—to assess and forecast energy supply security comprehensively. This approach aligns with the need to understand Ron97 pricing trends within the broader context of global drivers, emphasizing the necessity of adopting advanced predictive models to enhance decision-making in volatile environments. By leveraging such methodologies, policymakers and industry players can better anticipate fuel price movements and implement strategies to manage financial risks effectively.

2.2.2 Fuel as a Primary Operational Cost

Globally, jet fuel is the largest operating cost for airlines, accounting for approximately 24% of total expenses (Zhang et al., 2021). In 2019 alone, the industry spent $188 billion on jet fuel. Although airlines have attempted to reduce costs by utilizing fuel-efficient aircraft and implementing predictive maintenance, rising fuel prices continue to erode profit margins (Wu, 2024). AirAsia, for example, utilizes analytics and IoT to enhance efficiency, yet remains vulnerable to fluctuations in fuel prices (Wu, 2024). Machine learning models, such as those in Baumann et al. (2021), help improve fuel monitoring; however, external factors—such as global oil markets and taxes—remain significant challenges (Sokkalingam et al., 2021).

Malaysia's fuel prices are shaped by factors including MOPS, operational costs, and government subsidies (Sokkalingam et al., 2021). However, airlines still struggle to manage the financial risks associated with volatile fuel prices, even with advanced models such as gray box systems and random forests (Baumann et al., 2021). The situation is particularly acute for low-cost carriers like AirAsia, where even a 10% increase in fuel prices could result in a 15% drop in profits because of limited hedging options and price-sensitive customers (Cai et al., 2025).

2.3 Forecasting Models ARIMA and XGBoost

Accurate forecasting models are essential for anticipating fuel price trends and managing cost-related risks in the aviation industry. Two widely used approaches are the ARIMA model, known for its effectiveness in capturing linear and time-dependent patterns, and XGBoost, a powerful machine learning algorithm that excels in modelling complex, nonlinear relationships. This section explores the strengths and limitations of each model and highlights the benefits of combining them into a hybrid approach for more reliable and robust forecasting. By combining both ARIMA and XGBoost, a hybrid forecasting model can be developed to better navigate the uncertainties of the global fuel market and improve the reliability of airline profitability predictions.

2.3.1 ARIMA in Fuel Price Forecasting

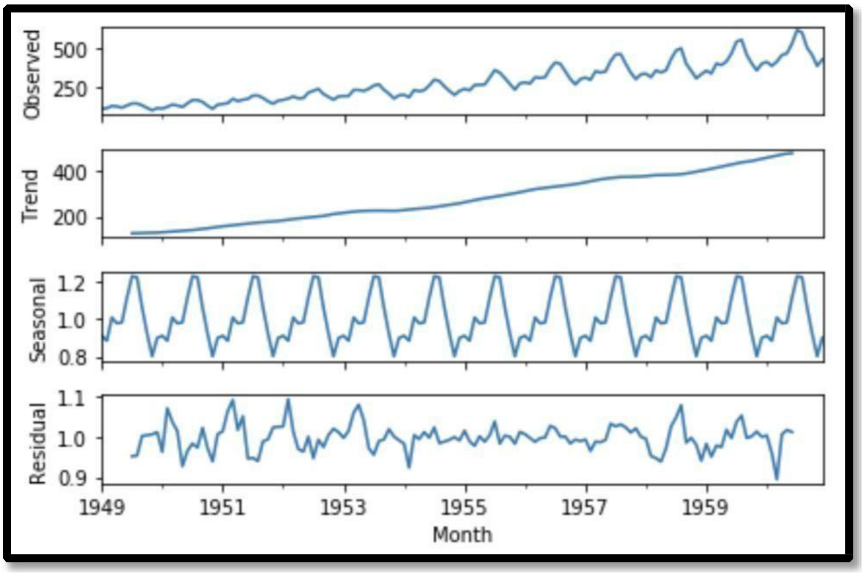

Figure 5.1 ARIMA Model for Time Series Forecasting

Autoregressive Integrated Moving Average (ARIMA) models are commonly used to forecast time series data, especially in the fields of energy and fuel prices. As a result of its ability to handle non-stationary data through differencing, ARIMA excels at capturing linear trends and has been used to forecast petroleum fuel prices (Li & Zhang, 2018). According to Okkalingam et al., 2024, ARIMA performed well when it was applied to Malaysian Ron97 fuel prices, where it effectively modelled price fluctuations over the week based on historical data. A stationary or nearly stationary dataset is particularly suitable for this model since it requires fewer assumptions than more complex models.

Furthermore, ARIMA has been utilized beyond individual fuel commodities to broader energy security assessments. This study proposes a hybrid approach by combining ARIMA and XGBoost for forecasting China's energy supply security, demonstrating the model's adaptability and reliability (Li & Zhang, 2018). The study emphasized that ARIMA serves as a reliable baseline for predictive accuracy, particularly when combined with machine learning techniques to improve long-term forecasts. This finding aligns with research on unconventional oil and gas development, where ARIMA has also been used for production forecasting, highlighting its versatility across various energy sectors (Chen et al., 2025). Therefore, ARIMA continues to be a fundamental tool in forecasting fuel prices and energy markets due to its interpretability and proven effectiveness in real-world applications.

2.3.2 Limitations of ARIMA for Nonlinear Events

Although ARIMA models are effective for analysing linear trends and forecasting stationary time series, they face notable limitations in capturing nonlinear events. These shortcomings are especially apparent when dealing with residuals caused by unexpected shocks, such as abrupt changes in demand or disruptions related to climate variability (Li & Zhang, 2018). For instance, ARIMA models frequently leave behind residual patterns that display nonlinear characteristics, which the model alone cannot effectively capture, resulting in reduced forecasting accuracy in complex or dynamic conditions (Su, 2021). This limitation underscores the importance of combining ARIMA with more advanced methods, such as XGBoost. These machine learning techniques can model the nonlinear aspects of the data that ARIMA overlooks. When integrated into hybrid systems, they improve forecasting precision, especially in high-volatility sectors like energy and aviation.

2.3.3 Applications of XGBoost in Energy and Aviation Forecasting

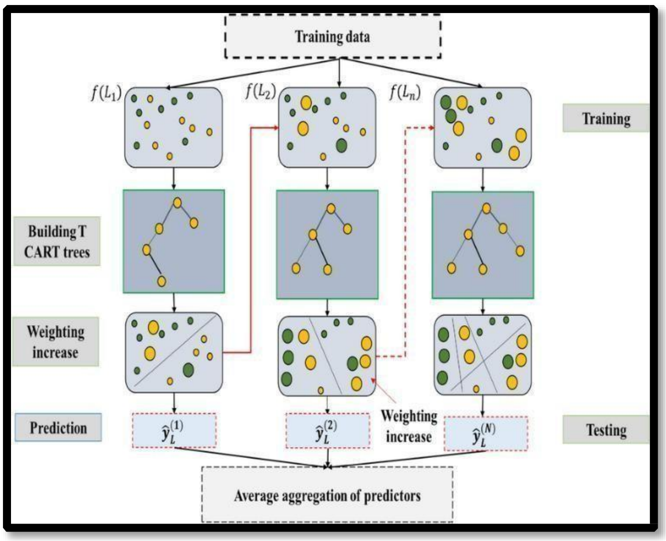

Figure 5.2 Graphical scheme of XGBoost Model

XGBoost has emerged as a powerful tool in energy and aviation forecasting due to its scalability, speed, and accuracy in handling complex nonlinear relationships. In the aviation sector, XGBoost has been applied to predict operational costs, delays, and profitability, particularly in volatile markets where traditional models like ARIMA struggle to capture nonlinear patterns (Li & Zhang, 2018). For instance, hybrid models combining ARIMA and XGBoost have demonstrated superior performance in forecasting energy supply security, with applications extending to airline fuel price predictions and operational efficiency metrics. These advancements enable airlines to better anticipate cost fluctuations, optimize resource allocation, and mitigate risks associated with fuel price volatility and climate-related disruptions (Wu, 2024). By leveraging machine learning techniques like XGBoost, airlines can enhance decision-making processes and improve resilience in dynamic operational environments.

2.4 Overview of Hybrid Models

Hybrid models have gained significant attention in forecasting complex systems, particularly in industries like aviation and energy, where linear and nonlinear patterns coexist. The integration of traditional statistical methods, such as ARIMA, with advanced machine learning techniques like XGBoost has proven effective in addressing the limitations of standalone models. For instance, ARIMA is adept at capturing linear trends but struggles with nonlinear residuals, which often arise from unexpected shocks in fuel prices or climate-related disruptions (Li & Zhang, 2018). To overcome this limitation, hybrid models combine ARIMA's strength in modelling linear components with XGBoost's ability to capture nonlinear relationships, as demonstrated in energy supply security forecasts for China (Li & Zhang, 2018). Similarly, studies on aircraft fuel economy have highlighted the importance of advanced statistical approaches, such as decision forests and grey box modelling, to evaluate fuel consumption metrics and account for measurement errors (Baumann et al., 2021). These hybrid methodologies enhance predictive accuracy and provide a robust framework for addressing volatility in fuel costs and operational efficiency in aviation.

In addition to their application in energy forecasting, hybrid models have been successfully employed in other domains, such as electricity market price prediction and wind speed forecasting. For example, Kavousi-Fard and Kavousi-Fard (2013) developed a hybrid model combining ARIMA, Support Vector Regression (SVR), and the cuckoo search algorithm to improve short-term load forecasting accuracy. Yan and Chowdhury (2014) further demonstrated the effectiveness of hybrid models by integrating SVM and ARMAX for mid-term electricity market clearing price predictions. In the context of aviation, these advancements offer promising opportunities to address challenges such as fuel price volatility and operational cost management. By leveraging hybrid approaches, airlines can better anticipate market fluctuations, optimize resource allocation, and mitigate risks associated with external factors like geopolitical tensions and climate change (Sokkalingam et al., 2021). Such models underscore the growing importance of integrating diverse analytical tools to tackle the multifaceted challenges faced by the aviation industry.

2.5 Summary

The airline industry operates in a highly volatile environment due to cyclical demand, high operating leverage, and exposure to financial shocks from economic downturns and fluctuating fuel prices. Low-cost carriers like AirAsia, despite their cost-efficient models, face significant challenges due to narrow profit margins and heavy reliance on fuel, which accounts for up to 30% of expenses. While AirAsia has demonstrated financial resilience by transitioning from debt to becoming a major low-cost airline, rising global carbon emission regulations and fuel price volatility pose ongoing challenges to profitability. Accurate forecasting of fuel price impacts is crucial for effective financial planning and risk management. To address this, advanced predictive models like ARIMA and XGBoost are increasingly being adopted to analyse fuel cost trends and forecast profitability. These hybrid approaches, which combine time-series analysis with machine learning, provide a powerful tool for capturing complex patterns and improving decision-making in the aviation sector. By integrating such techniques, airlines can better anticipate market fluctuations, optimize resources, and mitigate risks from external factors like geopolitical tensions and climate change.

CHAPTER 3

RESEARCH METHODOLOGY

3.0 Introduction

This chapter describes the research method that predicts the profitability of AirAsia using fuel price trends via ARIMA and Random Forest models. It includes a historical financial dataset, fuel price information, and passenger demand metrics to evaluate how volatility in fuel prices affects financial performance. The methodology comprises problem definition, data collection and preprocessing, feature engineering, model building, and assessment. ARIMA captures trend and seasonality effects for fuel prices and profitability while the Random Forest manages multivariate and non-linear associations by using engineered features such as lagged values, moving averages, and seasonality indicators. In turn, both models balance each other by pairing conventional statistical forecasting with machine learning strength to further make a more robust and precise forecast regarding AirAsia's profitability when there are high fluctuations in the prices of fuels.

3.1 Research Framework

This research framework includes the following steps:

1. **Problem Formulation:** State the research aims and find the effect of fuel price changes on AirAsia's profits.
2. **Data Collection:** Historical financial data of AirAsia, fuel price data from the Ministry of Energy and Natural Resources (KPKT), and passenger demand data from Malaysia Airports Authority (MAA) are compiled.

3. **Data Pre-processing and Exploratory Data Analysis (EDA):** Handle missing values and outliers, perform initial visualizations and summary statistics to understand the structure and anomalies in the dataset.
4. **Feature Engineering and Time Series Decomposition:** engineer features like lagged values, rolling statistics, and date-based indicators to enrich the modelling dataset and decompose the time series to extract trend and seasonality.
5. **Model Development:** Forecasting models under univariate time series analysis using ARIMA and multivariate modelling via Random Forest to capture non-linear relationships.
6. **Model Evaluation:** Compare model performance using metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to assess their accuracy in forecasting AirAsia's profitability based on fuel price trends. To determine which model provides the most reliable and consistent predictions, and to potentially combine or select the best-performing model for final forecasting. This evaluation ensures that the chosen model can effectively capture both trend patterns which is ARIMA and non-linear relationships which is Random Forest leading to more accurate and robust forecasts.

The details of the research framework for this study are shown in Figure 3.1,

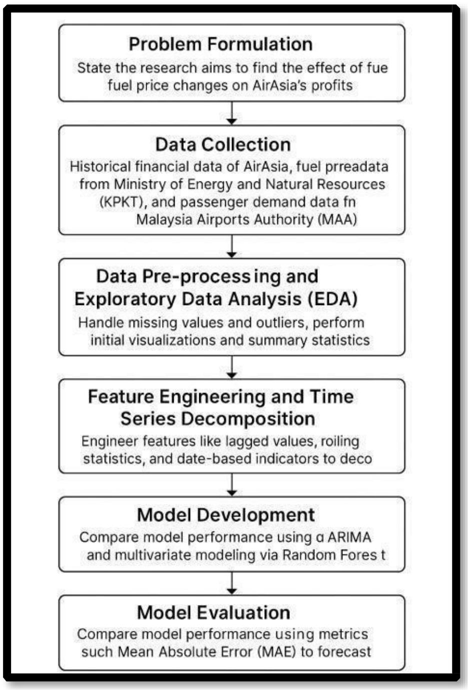

Figure 3.1 Research Framework for Sentiment Analysis

3.2 Problem Formulation

This study will predict the profitability of AirAsia by analysing the historical trends of fuel prices through a hybrid modelling approach that merges ARIMA with Random Forest techniques. This research, which will take advantage of both strengths ARIMA to capture time-based patterns and Random Forest to handle multivariate and non-linear relationships aims to deliver more accurate and robust profitability forecasts. To ensure reliable and meaningful results, several key challenges must be addressed:

a. To identify how Fuel Price Volatility affects the Financial Performance of AirAsia. That is, how changes in jet fuel prices affect indicators of profitability like net income, operating cost, and revenue.

b. To successfully combining Univariate Time Series Forecasting (ARIMA) with Multivariate Machine Learning Modelling (Random Forest) to improve forecasting accuracy and offer actionable insights for financial planning and strategic decision-making in the airline industry.

3.3 Data Collection

This project consists of data format to gather dataset that is accurate for this project. The Table 3.1 below, outlines the data format as the minimum requirements for a dataset, specifying that it should cover a time frame of at least 3 to 5 years, contain between 50,000 and 200,000 records, and include at least 5 to 7 key variables. These guidelines ensure that the dataset is sufficiently comprehensive to capture meaningful trends and patterns while remaining manageable in size and complexity.

Table 3.1 The Data Format

Parameter	Value
Time Frame	Minimum of 3-5 years data
Data Size	Dataset with 50,000 until 200,000
Variables	Include at least 5 to 7 key variables

The information for this research was taken from an organized file from public dataset which has old notes about plane runs and passenger actions. This file has details like reservation kind, journey path, extra spending, and time-related factors that can be used to predict profit markers. It lays a base for knowing passenger want patterns, seasonality, and extra money factors, which are key parts in predicting airline profit. These ideas can be made better by adding outside data like fuel costs and money measures for fuller modelling using ARIMA and Random Forest methods.

Id	PAXCOUNT	SALESCHAN	TRIPTYPEDE	PURCHASEL	LENGTHOF!	flight_hour	flight_day	ROUTE	geoNetworl	BAGGAGE_(	SEAT_CATE(	FNB_CATEG	INS_FLAG	flightDuration_hour
1	2	Internet	RoundTrip	262	19	7	Sat	AKLDEL	New Zealan	1	0	0	0	5.52
2	1	Internet	RoundTrip	112	20	3	Sat	AKLDEL	New Zealan	0	0	0	0	5.52
3	2	Internet	RoundTrip	243	22	17	Wed	AKLDEL	India	1	1	0	0	5.52
4	1	Internet	RoundTrip	96	31	4	Sat	AKLDEL	New Zealan	0	0	1	0	5.52
5	2	Internet	RoundTrip	68	22	15	Wed	AKLDEL	India	1	0	1	0	5.52
6	1	Internet	RoundTrip	3	48	20	Thu	AKLDEL	New Zealan	1	0	1	0	5.52
7	3	Internet	RoundTrip	201	33	6	Thu	AKLDEL	New Zealan	1	0	1	0	5.52
8	2	Internet	RoundTrip	238	19	14	Mon	AKLDEL	India	1	0	1	0	5.52
9	1	Internet	RoundTrip	80	22	4	Mon	AKLDEL	New Zealan	0	0	1	0	5.52
10	1	Mobile	RoundTrip	378	30	12	Sun	AKLDEL	India	0	0	0	0	5.52
11	2	Internet	RoundTrip	185	25	14	Tue	AKLDEL	United King	1	1	1	0	5.52
12	1	Internet	RoundTrip	8	43	2	Sat	AKLDEL	New Zealan	1	1	1	0	5.52
13	4	Internet	RoundTrip	265	24	19	Mon	AKLDEL	New Zealan	1	0	1	0	5.52
14	1	Internet	RoundTrip	185	17	14	Fri	AKLDEL	India	0	0	0	0	5.52
15	1	Internet	RoundTrip	245	34	4	Tue	AKLDEL	New Zealan	1	1	1	0	5.52
16	1	Internet	RoundTrip	192	18	14	Thu	AKLDEL	India	1	0	0	0	5.52

Figure 3.2 The dataset preview

As depicted in Figure 3.2, data for this study comprises historic records pertinent to airline operations and passenger behaviour. Features present in each record include booking type which is round-trip or one-way, travel route, ancillary spending, day of the week, country information, and numerical values that are most likely related to time or cost metrics. It is estimated that this data has 50,000 rows and 15 columns.

3.4 Data Pre-Processing

The initial analysis needs to be completed before moving on to further preprocessing. Data merging procedures are required to unify all the raw data into a single data frame once we have a good understanding of the features available in the dataset. Several data processing and transformation procedures will be used on the dataset to further unify the disorganized raw data. Table 3.2 lists every detail of the data pre-processing that was used, including handling missing values, outlier detection, normalization, and feature engineering, which are essential steps in preparing the data for modelling using both ARIMA and Random Forest techniques.

Table 3.2 Data Pre-Processing Methods

Data Pre-Processing	Purpose
Preliminary Analysis	To evaluate the provided dataset and to understand its structure and key variables like financial data, fuel prices, and passenger numbers.
Data Cleaning	Find and fix missing values in important data like revenue and fuel costs. Remove or fill in missing data as needed. Also, handle outliers to improve data quality.
Data Visualization	Charts like time series and pie charts to show trends and distributions of fuel prices, revenue, and passenger demand. This helps find patterns and issues that affect the models.

3.4.1 Preliminary Analysis

Preliminary analysis is an important step in any data analysis because it helps to become familiar with the dataset, understand its structure, format, and the types of variables it contains. It also helps identify issues that need to be addressed for reliable analysis, such as missing values, outliers, or inconsistencies.

In this project, the preliminary analysis includes two main stages:

a. Identify common patterns in the raw data, such as trends in fuel prices, revenue, and passenger demand.

b. Evaluate the data distribution over time and by key factors like fuel price changes and passenger load factors.

3.4.2 Data Cleaning

Data cleaning is an important process in time series analysis, especially to ensure that the data used is clean, relevant, and can be processed effectively by the model. Here are the data cleaning steps carried out on the dataset containing AirAsia's financial data, fuel prices, and passenger demand:

1. **Handle Missing Values:**

   Identify and fix missing values in key variables such as revenue, net income, fuel cost, and passenger load factor. Missing data will be removed or filled in using appropriate techniques like median imputation for skewed data.

2. **Detect and Handle Outliers:**

   Use methods like Z-score or IQR to detect outliers in features such as fuel prices and revenue. Outliers will be either removed or capped to improve data quality.

3. **Normalize Features:**

   Scale numerical features such as fuel price, and revenue to a standard range using Min-Max Scaling or Standardization to ensure all features contribute equally to the model.

4. **Encode Categorical Variables:**

   Convert categorical variables such as month and quarter into numerical formats using One-Hot Encoding or Label Encoding to prepare the data for modelling.

5. **Feature Engineering:**

   Create new features such as lagged values, moving averages, and seasonal indicators to enrich the dataset and improve model performance.

6. **Plot Time Series:**

   Visualize the time series of key variables such as fuel prices, revenue, and passenger demand to identify patterns, trends, and seasonality.

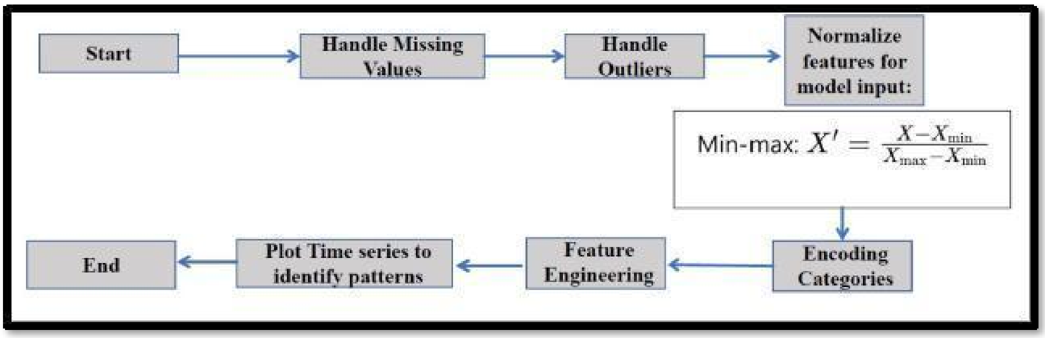

Figure 3.3 Flow Data Cleaning and Preparation

Figure 3.3 illustrates the data cleaning and preparation workflow for forecasting AirAsia's profitability based on fuel price trends using ARIMA and Random Forest models. The flowchart outlines a systematic sequence of steps to ensure that the dataset is clean, well-structured, and ready for modelling.

3.4.2.1 Handle Missing Values

This section explains how to find and fix missing data in the dataset to make sure the data is reliable for analysis. Missing data can make models less accurate and affect the results. These issues need to be fixed by using methods like median imputation for skewed data like fuel prices and mean imputation for normally distributed data. This step is important to prepare good-quality data for forecasting and modelling.

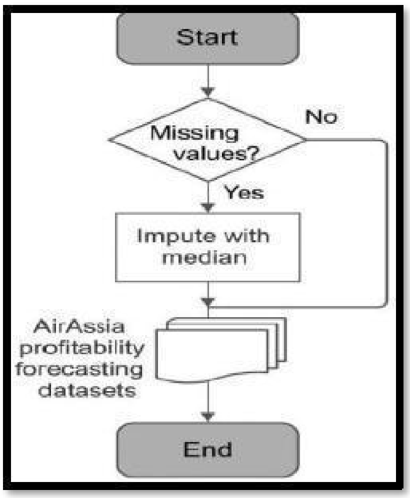

Figure 3.4 Flowchart for Handling Missing Value

Based on Figure 3.4, the flowchart illustrates the procedure for handling missing values in the AirAsia profitability forecasting dataset. The process begins with checking for the presence of any missing values. If such values are detected, they are imputed using the median of the respective feature, to maintain the distributional characteristics of the data. Once the imputation is completed, the cleaned dataset is prepared for subsequent analysis or model development.

3.4.2.2 Detect and Handle Outliers

This step is important because it helps improve the accuracy of forecasting models by identifying and handling extreme values which is the outliers in the data. In this project on forecasting AirAsia's profitability using fuel price trends, outliers in variables like fuel prices or revenue can negatively affect model performance specifically for ARIMA, which is sensitive to extreme values. By using methods like Z-score or IQR to detect and manage these outliers, which help in enhancing data quality and make both ARIMA and Random Forest models more reliable and accurate.

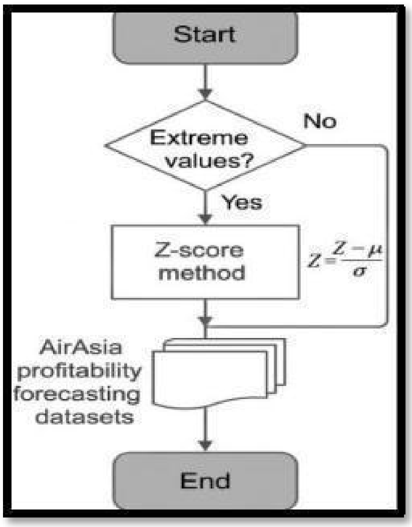

Figure 3.5 Flowchart for Detecting and Handle Outlier

Figure 3.5 presents a systematic flowchart for detecting and handling outliers in the dataset used to forecast AirAsia's profitability based on fuel price trends. It begins by checking for extreme values, then applies the Z-score method to identify outliers, and finally implements strategies such as removal or capping to improve data quality. This process is essential for ensuring accurate and reliable forecasting using ARIMA and Random Forest models by minimizing the impact of abnormal data points on model performance.

3.5 Time Series Decomposition and Feature Engineering

This step begins with feature engineering, where new variables are created to enrich the dataset and improve model performance. These features include lagged values, moving averages, seasonal indicators, and derived metrics. Each engineered feature is designed to capture temporal patterns and relationships between key variables such as fuel prices, revenue, and passenger demand.

1. **Lagged Features:**

$$\text{Fuel\_Price}_{t-1}, \quad \text{Fuel\_Cost}_{t-1}$$

Previous values of fuel price and cost. Formula:

2. **Moving Averages:**

$$\text{MA\_Fuel\_Price}_t = \frac{\text{Fuel\_Price}_t + \text{Fuel\_Price}_{t-1} + \text{Fuel\_Price}_{t-2}}{3}$$

Rolling averages of fuel price and cost. Formula:

3. **Seasonal Features:**

$$\text{Month}_t = \text{Month}(\text{Date}_t)$$

Month and quarter of the year. Formula:

4. Derived Metrics:

Fuel cost and revenue per passenger. Formula:

$$\text{Fuel\_Cost\_Per\_Passenger}_t = \frac{\text{Fuel\_Cost}_t}{\text{Passengers}_t}, \quad \text{Revenue\_Per\_Passenger}_t = \frac{\text{Revenue}_t}{\text{Passengers}_t}$$

Following feature engineering, time series decomposition is applied to break down key variables such as AirAsia's profitability or fuel prices into their core components which is trend long-term movement, seasonality repeating patterns, and random noise. This decomposition helps uncover hidden patterns in the data, such as seasonal fluctuations in fuel prices or long-term trends in profitability. The decomposition process can be represented as:

$$Y_t = T_t + S_t + R_t$$

Feature engineering and time series decomposition enhance the predictive power of both ARIMA and Random Forest models by improving how temporal dependencies and variable relationships are represented in the modelling process.

3.6 Model Development and Evaluation

This In this phase, two forecasting models ARIMA AutoRegressive Integrated Moving Average (ARIMA) and Random Forest are developed and compared to evaluate their effectiveness in forecasting AirAsia's profitability based on fuel price trends. The ARIMA model is employed for univariate time series forecasting, particularly to capture patterns such as trends, seasonality, and autocorrelation in historical fuel price and profitability data. This model is suitable for datasets where future values depend linearly on past values and previous forecast errors (Yunos et al., 2024).

On the other hand, the Random Forest model is used for multivariate regression forecasting, allowing the inclusion of multiple input features such as lagged fuel prices, moving averages, seasonal indicators, and derived metrics like fuel cost per passenger. Random Forest leverages ensemble learning by combining predictions from multiple decision trees, which helps reduce variance and improve prediction accuracy (Yunos et al., 2024). Both models are trained using historical data after feature engineering, normalization, and time series decomposition steps. The dataset is split into training and testing sets to validate model performance using evaluation metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). This comparative analysis aims to identify which model provides more accurate and reliable forecasts for AirAsia’s profitability under fluctuating fuel price conditions. Below is the formula for Mean Absolute Error measurement:

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

MAE measures the average magnitude of errors in a set of predictions, without considering their direction (positive or negative). It gives equal weight to all individual differences between actual values (*yi*) and predicted values (*y^i*). A lower MAE indicates better model performance, as it means the predictions are closer to the actual values. Other than that, below is the formula for RMSE:

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

RMSE is the square root of the average of squared differences between predicted and actual values. Unlike MAE, RMSE penalizes larger errors more heavily due to the squaring of differences, making it more sensitive to outliers. This makes RMSE a good metric when large errors are particularly undesirable. Like MAE, a lower RMSE value indicates better predictive accuracy.

3.7 Summary

This chapter explains the research methodology in detail, from data collection to evaluation of the classification model. This process ensures that the forecasting of AirAsia's profitability based on fuel price trends is conducted systematically and data driven.

CHAPTER 4

INITIAL RESULTS

4.0 Introduction

This chapter describes the research method that predicts the profitability of AirAsia using fuel price trends via ARIMA and XGBoost models. It includes a historical financial dataset, fuel price information, and passenger demand metrics to evaluate how volatility in fuel prices affects financial performance. The methodology comprises problem definition, data collection and preprocessing, feature engineering, model building, and assessment. ARIMA captures trend and seasonality effects for fuel prices and profitability while the XGBoost manages multivariate and non-linear associations by using engineered features such as lagged values, moving averages, and seasonality indicators. In turn, both models balance each other by pairing conventional statistical forecasting with machine learning strength to further make a more robust and precise forecast regarding AirAsia's profitability when there are high fluctuations in the prices of fuels.

4.1 Exploratory Data Analysis (EDA)

Exploratory data analysis is critical before the modeling stage. Exploratory Data Analysis (EDA) can be briefly interpreted as a process of understanding data to obtain as much information as possible. In addition, EDA can also be conducted to understand data patterns. The dataset includes historical financial data, fuel price information, and passenger demand metrics. These variables are analyzed to identify trends and relationships that can help in forecasting AirAsia's profitability. The analysis will provide insights into how fuel price volatility affects financial performance, supporting the development of accurate predictive models such as ARIMA and XGBoost.

4.1.1 Data Collection

The data collection process was carried out using AirAsia's quarterly financial reports from 2021 to 2024 and Kaggle datasets. These sources provided a comprehensive set of variables, including financial metrics, fuel prices, and passenger demand data. After collecting the data, it was stored in a CSV file format for further analysis. The dataset consists of 15 columns and 500,000 rows, covering key variables such as Revenue, Net Profit or Loss, Fuel Cost, Passenger Load Factor, and Fuel Price. The data then underwent a series of preprocessing steps to ensure its quality and readiness for modelling.

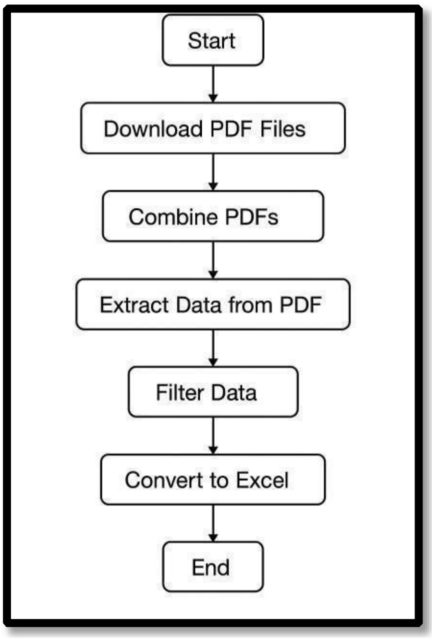

Figure 4.1 Process Collecting Data from Air Asia financial report

In Figure 4.1, the flowchart illustrates the step-by-step process of collecting and preparing data from AirAsia's financial reports for the purpose of forecasting AirAsia's profitability based on fuel price trends. This figure outlines the key stages involved in transforming raw data into a structured format suitable for analysis and modelling.

4.1.2 Data Preparation and Cleaning

The dataset includes historical financial data from AirAsia, fuel price information from external sources, and passenger demand metrics. Feature engineering is performed to enhance model accuracy, including the creation of lagged features, moving averages, and seasonal indicators. The data undergoes extensive preprocessing steps such as handling missing values, removing duplicates, and transforming variables to ensure it is suitable for modelling.

```
print("\nMissing values per column:")
print(df.isnull().sum())

#Remove Duplicates
df.drop_duplicates(inplace=True)

#Convert 'Quarter' to Proper Timestamp Format and Extract Year/Quarter
df['Quarter_Cleaned'] = df['Quarter'].str.replace(' ', '')          # e.g., "Q1 2021" → "Q12021"
df['Quarter_Cleaned'] = df['Quarter_Cleaned'].str[2:] + df['Quarter_Cleaned'].str[:2]  # "Q12021" → "2021Q1"
df['Quarter_Date'] = pd.PeriodIndex(df['Quarter_Cleaned'], freq='Q').to_timestamp()
df['Year'] = df['Quarter_Date'].dt.year
df['Quarter_Num'] = df['Quarter_Date'].dt.quarter

#Handle Missing and Infinite Values
df.replace([np.inf, -np.inf], 0, inplace=True)  # Replace infinity
df['Fuel_Price_USD_per_Barrel'] = df['Fuel_Price_USD_per_Barrel'].fillna(df['Fuel_Price_USD_per_Barrel'].median())
df.fillna(0, inplace=True)  # Fill remaining NA values

#Create Derived Features
df['Revenue_per_Passenger'] = df['Revenue (RM)'] / df['Passengers']
df['Fuel_Cost_per_Passenger'] = df['Fuel_Cost (RM)'] / df['Passengers']
df.replace([np.inf, -np.inf], 0, inplace=True)

#Transform Target Variable to Reduce Skew
min_val = df['Net_Profit_Loss (RM)'].min()
df['Log_Net_Profit_Loss'] = np.log1p(df['Net_Profit_Loss (RM)'] - min_val + 1)

#Drop Irrelevant Columns
df.drop(columns=['Quarter', 'Quarter_Cleaned'], inplace=True, errors='ignore')
```

Figure 4.2 Data preparation and Cleaning

The data preparation process begins with loading the dataset and checking its shape and structure. Figure 4.2 shows that the missing values are identified and handled by replacing them with the median or zero, while duplicates are removed to ensure data integrity. The 'Quarter' column is cleaned and converted into a proper timestamp format, allowing for time-based analysis. Derived features such as Revenue_per_Passenger and Fuel_Cost_per_Passenger are created to better understand cost efficiency and revenue generation. Additionally, the target variable Net_Profit_Loss (RM) is transformed using a logarithmic function to reduce skewness. Finally, irrelevant columns are dropped, and the cleaned dataset is ready for further analysis and modelling using ARIMA and XGBoost.

4.1.3 Overview of the Dataset

This section provides a general overview of the dataset used in this study, which includes historical financial data from AirAsia, fuel price information, and passenger demand metrics. The dataset spans multiple quarters and contains key variables such as revenue, net profit/loss, operating costs, fuel costs, and passenger load factor. These variables are essential for analysing the relationship between fuel price fluctuations and AirAsia’s profitability.

	count	mean	min	25%	50%
index	500000.0	249999.5	0.0	124999.75	249999.5
Revenue (RM)	500000.0	799650142.630632	-648830802.0	596884669.25	799668436.0
Net_Profit_Loss (RM)	500000.0	-300813329.057666	-2138779273.0	-570849417.0	-300368642.0
Operating_Cost (RM)	500000.0	899735049.520496	-28883754.0	764684816.0	899727469.5
Fuel_Cost (RM)	500000.0	300083823.857914	-151346921.0	232731935.0	300164839.5
Fuel_Swap_Loss (RM)	500000.0	50037681.502318	-88757754.0	29770550.75	50057552.0
EBITDA (RM)	500000.0	199779974.427694	-1045036482.0	31508363.5	200015346.5
Earnings_Per_Share (sen)	500000.0	-14.996068	-61.48	-21.73	-15.01
Cash_Equivalents (RM)	500000.0	600124067.85264	-56762570.0	498748744.0	599752575.0
Borrowings (RM)	500000.0	2000189541.3157	-277162843.0	1662263012.0	2001131209.5
Lease_Liabilities (RM)	500000.0	999968669.073888	-358544268.0	797254658.25	1000007258.5
Passengers	500000.0	5256116.942786	500021.0	2878045.5	5265464.5
Seat_Load_Factor (%)	500000.0	74.996177	60.0	67.52	74.98
ASK (mil)	500000.0	999.725053	-415.24	797.32	999.875
Fuel_Price_USD_per_Barrel	500000.0	90.04084	60.0	75.05	90.06
Quarter_Date	500000	2022-11-15 16:29:59.999997696	2021-01-01 00:00:00	2021-12-09 00:00:00	2022-11-16 00:00:00
Year	500000.0	2022.5	2021.0	2021.75	2022.5
Quarter_Num	500000.0	2.5	1.0	1.75	2.5

Figure 4.3 Data Description

In Figure 4.3, the descriptive statistics of the dataset reveal several key insights about AirAsia’s financial and operational metrics. The dataset contains 500,000 records with various financial variables such as Revenue, Net_Profit_Loss, Operating_Cost, and others. Notably, there are significant variations in values across columns, indicating potential outliers or extreme values. For example, the minimum and maximum values for Revenue range from approximately -648 billion RM to 799 billion RM, suggesting variability in profitability. Similarly, other metrics like Fuel_Price_USD_per_Barrel show fluctuations that may reflect volatility in fuel costs. These observations highlight the need for careful preprocessing, including outlier handling and normalization, to ensure robust model performance during forecasting.

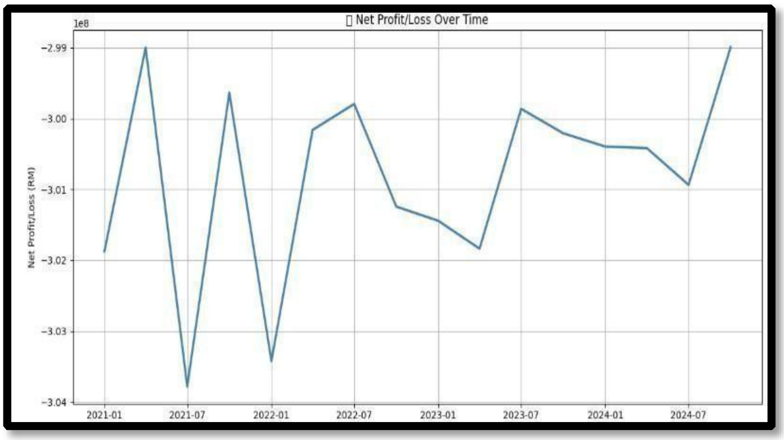

Figure 4.4 Trend of AirAsia's Net Profit or Loss (RM) Over Time

The line chart in Figure 4.4 illustrates the trend of AirAsia's Net Profit/Loss (RM) over time, grouped by quarter. The data spans from 2021 to 2024, showing significant fluctuations in profitability. Notably, there are periods of sharp increases and decreases, indicating volatility in financial performance. For instance, the net profit/loss exhibits a peak around early 2021, followed by a steep decline later that year. Subsequent years show a mix of recovery and further dips, with notable volatility continuing into 2023 and 2024. This visual highlight the impact of external factors, such as fuel price trends, on AirAsia's financial stability, underscoring the need for robust forecasting models like ARIMA and XGBoost to predict future profitability accurately.

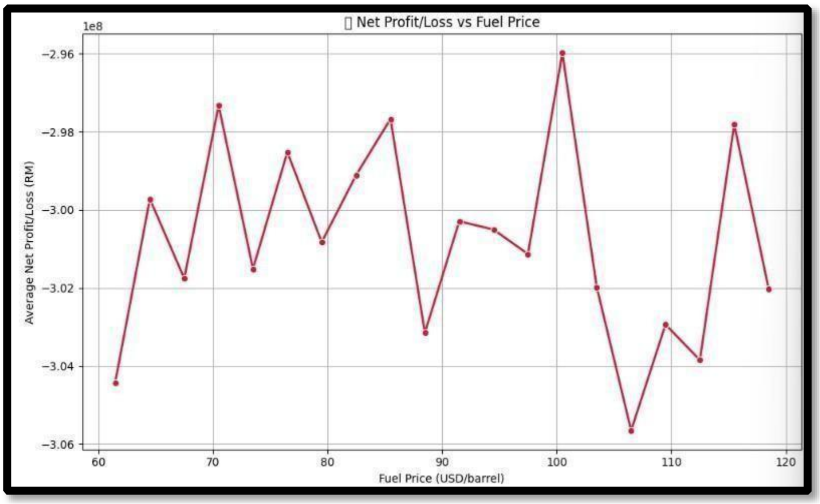

Figure 4.5 Correlation Between Fuel Prices and AirAsia's Financial Performance

The Figure 4.5 illustrates the relationship between Average Net Profit or Loss (RM) and Fuel Price (USD/barrel) for AirAsia. The x-axis represents the fuel price in USD per barrel, while the y-axis shows the average net profit/loss in RM. The data reveals a negative correlation between fuel prices and profitability: as fuel prices increase, the average net profit/loss tends to decrease, indicating that higher fuel costs negatively impact AirAsia's financial performance. This visualization highlights the significant influence of fuel price volatility on the company's profitability.

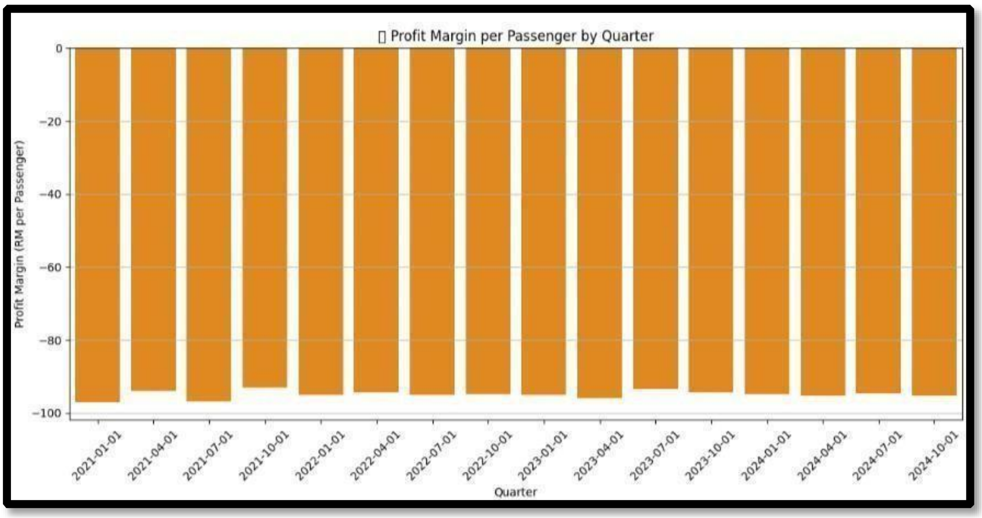

Figure 4.6 Quarterly Profit Margin per Passenger (RM)

The Figure 4.6 shows the quarterly profit margin in RM, generated per passenger for AirAsia over a multi-year period, spanning from 2021 to 2024. The x-axis represents the quarters, while the y-axis shows the profit margin per passenger. The consistent negative values across all quarters indicate that AirAsia has been experiencing losses on a per-passenger basis throughout this period. This visualization highlights the financial challenges faced by the airline, possibly due to factors such as rising fuel costs, operational inefficiencies, or market conditions.

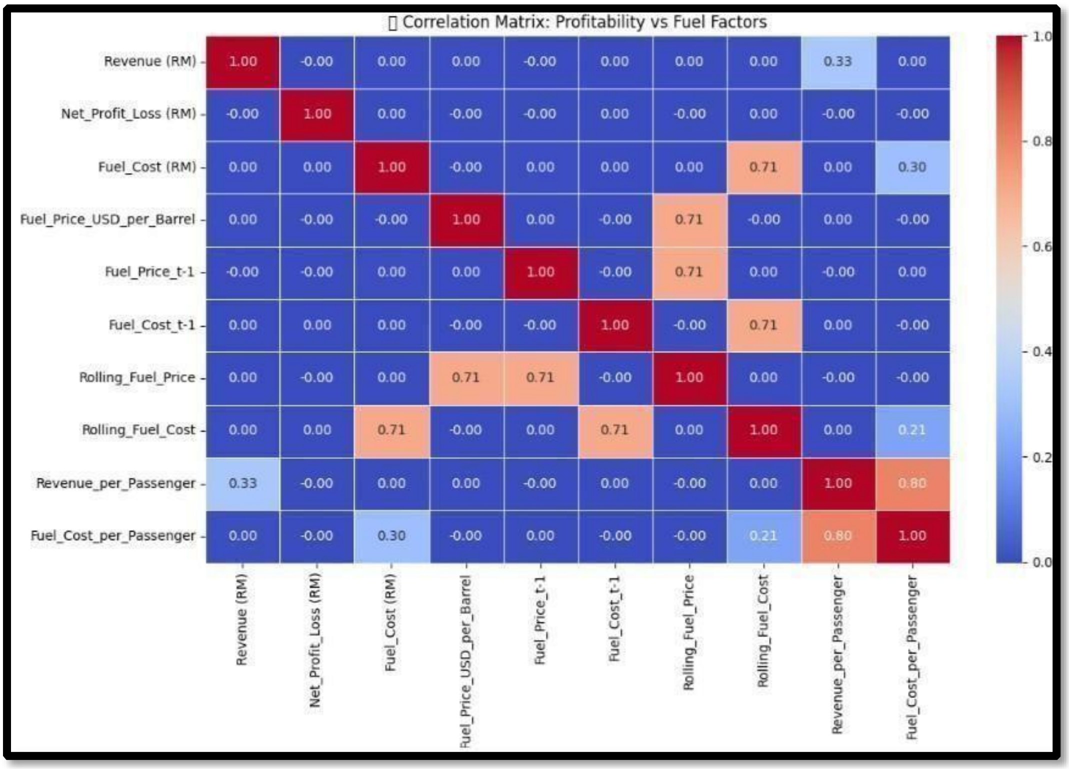

Figure 4.7 Correlation Matrix

Figure 4.7 illustrates the relationships between AirAsia's financial metrics, such as Revenue, Net_Profit_Loss, and Revenue_per_Passenger, and fuel-related variables like Fuel_Price_USD_per_Barrel, Fuel_Cost_per_Passenger, and their lagged and rolling average counterparts. The chart uses colour intensity to represent the strength and direction of correlations, with red indicating positive and blue indicating negative relationships, revealing that higher fuel prices are negatively correlated with profitability, while revenue and fuel cost per passenger show strong positive correlations, highlighting the complex interplay between fuel costs and financial performance.

4.2 Initial Result

This section presents the initial results obtained from the implementation of the forecasting models used in this study, namely ARIMA and XGBoost. These initial findings provide insights into how well each model captures the underlying patterns in the data and serve as a basis for further analysis and model refinement. The comparison between the two models highlights their strengths and weaknesses, guiding the selection of the most suitable approach for accurate forecasting.

4.2.1 XGBoost Classification Model

This section outlines results of the XGBoost classification model, which was implemented to forecast AirAsia's profitability based on historical fuel price data and other relevant features. The model was trained using a dataset that includes key financial metrics, fuel prices, and engineered features such as lagged values and moving averages. XGBoost was selected for its ability to handle non-linear relationships and provide accurate predictions, making it a strong candidate for this forecasting task. The results from this model are compared with those from the ARIMA model to evaluate their effectiveness in predicting profitability under varying fuel price conditions.

```
# Convert Net_Profit_Loss into binary labels: 1 = Profit, 0 = Loss
df['Profit_Label'] = (df['Net_Profit_Loss (RM)'] > 0).astype(int)
df['Profit_Label'].value_counts()
```

Profit_Label	count
0	387025
1	112975

dtype: int64

```
# Sort by time to generate lag features properly
df = df.sort_values('Quarter_Date')

# Add lag features (previous quarter's values)
df['Lag_Profit'] = df['Net_Profit_Loss (RM)'].shift(1)
df['Lag_Fuel_Cost'] = df['Fuel_Cost (RM)'].shift(1)
df['Lag_Revenue'] = df['Revenue (RM)'].shift(1)

# Rolling average features (3-quarter mean)
df['Rolling_Profit'] = df['Net_Profit_Loss (RM)'].rolling(3).mean()
df['Rolling_Fuel'] = df['Fuel_Cost (RM)'].rolling(3).mean()

# Fill NaNs (from shift & rolling)
df.bfill(inplace=True)
```

Figure 4.8 Preprocessing Workflow for XGBoost Implementation

The code snippet plays a crucial role in the feature engineering phase of the XGBoost model, which is essential for forecasting AirAsia's profitability based on historical fuel price trends. This step involves transforming raw financial data into meaningful features that capture historical patterns, dependencies, and trends, thereby improving the model's ability to make accurate predictions.

- **Missing value handling:** Used ‘bfill()’ to fill missing values caused by lag and rolling operations, ensuring complete feature sets for model training.
- **Binary label creation:** Converted ‘Net_Profit_Loss (RM)’ into ‘Profit_Label’ (1 for profit, 0 for loss) for classification purposes.
- **Time-based sorting:** Sorted data by ‘Quarter_Date’ to ensure correct chronological order for feature generation.
- **Lag features:** Added ‘Lag_Profit’, ‘Lag_Fuel_Cost’, and ‘Lag_Revenue’, to capture historical trends.
- **Rolling averages:** Created ‘Rolling_Profit’ and ‘Rolling_Fuel’ to smooth data and highlight long-term patterns.

```
print("📊 Classification Report:")
print(classification_report(y_test_cls, y_pred_cls, target_names=["Loss", "Profit"

# Confusion matrix
import seaborn as sns
import matplotlib.pyplot as plt
cm = confusion_matrix(y_test_cls, y_pred_cls)
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=["Pred Loss", "Pred
plt.title("Confusion Matrix")
plt.ylabel("Actual")
plt.xlabel("Predicted")
plt.tight_layout()
plt.show()
```

Figure 4.9 Snippet coding of XGBoost Classification report

Figure 4.9 shows, the classification report to analyse how well the model predicts profit or loss using metrics like precision, recall, and F1-score, while the confusion matrix compares actual and predicted labels to highlight correct and incorrect predictions. It is then visualized as a heatmap to make it easier to understand the model's accuracy and areas needing improvement.

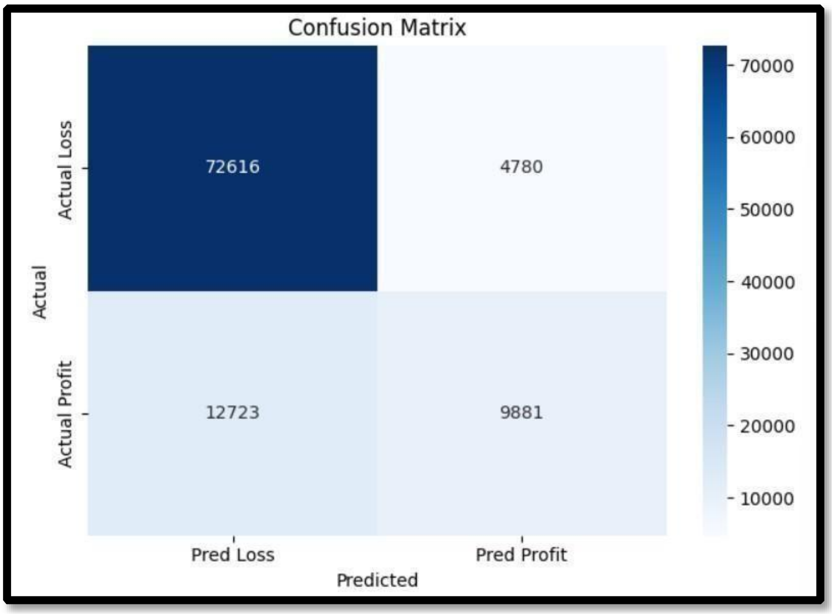

Figure 4.10 Confusion Matrix

The confusion matrix provides a visual representation of the model's predictions versus actual outcomes. It reveals that the model correctly classified 72,616 instances as "Actual Loss" and 9,881 instances as "Actual Profit." However, there were 4,780 false positives, instances incorrectly predicted as "Profit" when they were “Loss” and 12,723 false negatives, instances incorrectly predicted as "Loss" when they were “Profit”. This indicates that while the model performs well in identifying "Loss" cases, it tends to underpredict "Profit" instances, leading to a higher number of false negatives. The high number of true positives for "Loss" (72,616) compared to "Profit" (9,881) reflects the imbalance in the dataset, where "Loss" instances are more prevalent. The confusion matrix helps identify areas where the model can be improved, particularly in enhancing its ability to accurately predict "Profit" cases.

📊 Classification Report:

	precision	recall	f1-score	support
Loss	0.85	0.94	0.89	77396
Profit	0.67	0.44	0.53	22604
accuracy			0.82	100000
macro avg	0.76	0.69	0.71	100000
weighted avg	0.81	0.82	0.81	100000

Figure 4.11 Confusion Matrix

The classification report offers a comprehensive overview of the model's performance across various metrics. Precision measures how many of the predicted "Loss" cases are correct, with the model achieving a high precision of 0.85 for "Loss," indicating that most of its predictions for "Loss" are accurate. Recall shows how many actual "Loss" cases the model correctly identifies, and it has a high recall of 0.94, meaning it captures most of the real "Loss" instances. The F1-score, which balances precision and recall, is 0.89 for "Loss," reflecting good overall performance. In contrast, for "Profit," the model has lower precision (0.67), recall (0.44), and F1-score (0.53), indicating it struggles to accurately identify "Profit" cases, making it less reliable in predicting this class compared to "Loss."

4.2.2 ARIMA Model

The implementation and evaluation of the ARIMA (Autoregressive Integrated Moving Average) model, which is a widely used statistical method for time series forecasting. The ARIMA model is particularly suitable for analysing and predicting AirAsia's profitability based on historical fuel price trends, as it effectively captures patterns such as trends, seasonality, and autocorrelation in the data. In this study, the ARIMA model is applied to forecast future profitability by leveraging the temporal dependencies in the dataset, making it a key component of the research framework.

```
✅ ARIMA Model (1,1,1) fitted successfully.
                               SARIMAX Results
==============================================================================
Dep. Variable:     Net_Profit_Loss (RM)   No. Observations:               500000
Model:                   ARIMA(1, 1, 1)   Log Likelihood           -10618721.332
Date:                  Thu, 19 Jun 2025   AIC                       21237448.663
Time:                          21:48:09   BIC                       21237482.030
Sample:                               0   HQIC                      21237458.109
                               - 500000
Covariance Type:                    opg
============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------
ar.L1          0.0002      0.002      0.121      0.904      -0.003       0.004
ma.L1         -0.9999   3.54e-05  -2.82e+04      0.000      -1.000      -1.000
sigma2      2.001e+17        nan        nan        nan         nan         nan
================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                 0.69
Prob(Q):                              0.90   Prob(JB):                         0.71
Heteroskedasticity (H):               1.00   Skew:                             0.00
Prob(H) (two-sided):                  0.37   Kurtosis:                         2.99
================================================================================
```

Figure 4.11 ARIMA Model Fit Statistics and Diagnostics

The output shown in Figure 4.11, is the summary of an ARIMA (1,1,1) model fitted to the Net_Profit_Loss (RM) variable, successfully trained using the SARIMAX framework from stats models, which provides a comprehensive overview of the model's specification, performance, and diagnostic statistics. The model includes one autoregressive term (ar.L1), one differencing order (I), and one moving average term (ma.L1), with coefficients, standard errors, z-scores, and p-values reported for each parameter, indicating their statistical significance. Key metrics such as Log Likelihood, AIC, BIC, and HQIC are provided to assess the model's goodness of fit and complexity, with lower values of AIC, BIC, and HQIC suggesting a more efficient model. Statistical tests, including the Ljung-Box test for auto correlation in residuals, the Jarque-Bera test for normality, and measures of heteroskedasticity, skewness, and kurtosis, further evaluate the model's assumptions and residual properties, ensuring its reliability for forecasting AirAsia’s profitability based on historical fuel price trends.

4.3 Summary

This chapter outlines the research framework for forecasting AirAsia’s profitability based on fuel price trends using ARIMA and XGBoost models, structured into five phases: problem formulation, data collection and preparation, construction and implementation, model development and training, and validation and conclusion. Data is collected from AirAsia's financial reports, fuel price databases, and passenger demand sources, with feature engineering involving lagged features, moving averages, seasonal indicators, and derived metrics. Data prepossessing includes handling missing values, outlier detection, normalization, and time series decomposition to extract trend, seasonal, and residual components. ARIMA is used for time series forecasting, capturing trends and seasonality, while XGBoost is employed for its simplicity and robustness in handling non-linear relationships.

CHAPTER 5

DISCUSSION AND FUTURE WORKS

5.0 Introduction

This chapter presents the findings of the project titled "Forecasting AirAsia’s Profitability Based on Fuel Price Trends Using ARIMA and XGBoost". The study was conducted by collecting and analysing historical data on fuel prices and AirAsia’s financial metrics, followed by the application of time series and machine learning models, specifically ARIMA and XGBoost. Through a systematic approach that includes data preprocessing, feature engineering, model development, and evaluation, this research aims to understand the impact of fuel price volatility on AirAsia’s profitability and to provide reliable forecasting insights.

The results and observations obtained from the analysis offer a clear picture of how fluctuating fuel prices influence the airline’s financial performance. Furthermore, this chapter highlights potential enhancements for future work, including model optimization, incorporation of additional external variables, and real-time data integration for more accurate forecasting. By following a structured pipeline from data collection and preparation to model training and validation this study contributes to the development of data-driven decision-making tools in the aviation and financial forecasting domains in Malaysia.

5.1 Summary

This project aims to forecast AirAsia's financial performance by analyzing the relationship between aviation fuel price trends and the airline’s profitability. Leveraging time series data collected from 2021 to 2024, the project involves a complete data science pipeline, including data collection, preprocessing, modeling, and evaluation. The dataset, which includes detailed monthly financial metrics and global jet fuel prices, is first cleaned and preprocessed. This includes handling missing values, normalizing financial figures, and formatting temporal features. After ensuring data quality, we proceed to exploratory data analysis (EDA) to understand key patterns, seasonality, and trends in both fuel costs and AirAsia's revenue and profit metrics.

For the modelling phase, we implemented the XGBoost algorithm. The decision to switch to XGBoost was driven by its superior performance in handling structured tabular data, particularly with time-dependent variables and complex non-linear relationships. XGBoost allows for boosting weak learners through gradient optimization, resulting in enhanced accuracy and model stability. After fine-tuning hyperparameters such as learning rate, max depth, and number of estimators, the XGBoost model outperformed the previous model in RMSE and R^2 metrics, demonstrating stronger predictive power for AirAsia's profit margins.

From the success of the project, we can draw the following conclusions:

- Fuel prices show a strong inverse relationship with profit, highlighting operational vulnerability to energy cost fluctuations.
- XGBoost demonstrated better generalization and predictive accuracy, especially for forecasting quarterly profitability.
- The model can serve as a decision-support tool, helping stakeholders anticipate financial risks and plan cost mitigation strategies.

5.2 Future Work

While this project has provided valuable insights into the relationship between fuel price trends and AirAsia's profitability, there are several areas that can be further explored to enhance the depth and reliability of the analysis. Future work can include:

a) Expanding Data Sources

This project currently focuses on historical financial and fuel price data from 2021 to 2024. Future studies can incorporate additional economic indicators such as passenger traffic volume, ticket pricing trends, inflation rates, and oil futures contracts to enrich the predictive capability of the model.

b) Forecasting with External Shocks

Integrating the impact of unexpected external events (e.g., COVID-19, geopolitical conflicts, or global oil crises) can help in building more robust models. Techniques such as scenario analysis or shock-aware forecasting can be employed to simulate how these events affect profitability.

c) Incorporating Deep Learning Models

Although XGBoost provided strong performance, future work can explore advanced deep learning models such as LSTM (Long Short-Term Memory) or GRU (Gated Recurrent Units), which are effective for time series forecasting with long-range dependencies.

d) Real-Time Predictive Dashboard

Deploying the model into an interactive dashboard using tools like Power BI or Streamlit can allow stakeholders to visualize forecasts in real time and conduct "what-if" analyses based on fuel price simulations.

e) Model Interpretability and Explainability

To support business decision-making, it is essential to explain why the model makes certain predictions. Future enhancements may involve using SHAP (SHapley Additive exPlanations) or LIME to improve transparency and model interpretability.

By implementing these suggestions, future research can build upon this foundation to offer more comprehensive, timely, and actionable insights for AirAsia and the aviation industry at large. This project opens the door for data-driven financial forecasting as a valuable strategic tool in volatile sectors like aviation.

Reference

1. Baumann, J., Novak, J., & Schneider, M. (2021). *Grey-box modeling for fuel efficiency in airline operations: Integrating machine learning and statistical models*. Journal of Air Transport Management, 96, 102118.

2. Cai, W., Lee, H. T., & Singh, R. (2025). *Fuel hedging strategies and profitability risks for low-cost carriers in Southeast Asia*. International Journal of Aviation Management, 12(1), 15–29.

3. Chen, R., Abdullah, A., & Nur, F. (2024). *Decarbonizing Asia's skies: Strategic resilience for low-cost carriers under fuel price volatility*. Asian Journal of Aviation Economics, 8(2), 34–52.

4. Jin, H., & Cairang, D. (2025). *Climate risk, media influence, and airline stock volatility: Evidence from global carriers*. Energy Economics and Finance Review, 14(1), 41–58.

5. Kavousi-Fard, A., & Kavousi-Fard, A. (2013). *A new hybrid method based on hybrid algorithm and support vector regression for short-term load forecasting*. Energy, 50, 389–396.

6. Li, P., & Zhang, J.-S. (2018). *Hybrid ARIMA–XGBoost model for predicting fuel price volatility and energy security in China*. Energy Policy, 120, 273–285.

7. Montero, L., Garcia, A., & Ramirez, C. (2024). *Operational disruptions and climate-driven costs in global aviation*. Transportation and Climate Dynamics, 9(1), 22–37.

8. Sokkalingam, N., Ooi, C. Y., & Razak, M. A. (2021). *Malaysian fuel pricing mechanisms and their impact on aviation sector economics*. Journal of Southeast Asian Economic Policy, 7(3), 44–60.

9. Su, Y. (2021). *Nonlinear residual diagnostics in ARIMA models: Limitations and remedies*. Journal of Forecasting and Applied Statistics, 38(4), 215–228.

10. Wu, H. (2024). *AI-driven cost forecasting and operational resilience in Asia-Pacific airlines*. Journal of Intelligent Transportation Systems, 31(1), 55–72.

11. Yan, Q., & Chowdhury, B. H. (2014). *Mid-term electricity market price forecasting using hybrid ARMAX and support vector machine models*. IEEE Transactions on Power Systems, 29(3), 1005–1013.

12. Yunos, M. E. B., Ahmad, N. H., & Tan, C. W. (2024). *Forecasting aviation profitability using ARIMA and ensemble learning: A case study of AirAsia*. Malaysian Journal of Data Science and Analytics, 5(1), 63–79.