



Faculty of Computer Science and Information Technology

Forecasting Stock Price by Using Artificial Neural Networks

NUR'AZRA ALIA NISA BINTI ZULPAKAR

Bachelor of Computer Science with Honours

(Software Engineering)

2023

UNIVERSITI MALAYSIA SARAWAK

THESIS STATUS ENDORSEMENT FORM

TITLE Forecasting Stock Price by Using Artificial Neural Networks

ACADEMIC SESSION: 2022/2023

NUR'AZRA ALIA NISA BINTI ZULPAKAR

(CAPITAL LETTERS)

hereby agree that this Thesis* shall be kept at the Centre for Academic Information Services, Universiti Malaysia Sarawak, subject to the following terms and conditions:

1. The Thesis is solely owned by Universiti Malaysia Sarawak
2. The Centre for Academic Information Services is given full rights to produce copies for educational purposes only
3. The Centre for Academic Information Services is given full rights to do digitization in order to develop local content database
4. The Centre for Academic Information Services is given full rights to produce copies of this Thesis as part of its exchange item program between Higher Learning Institutions [or for the purpose of interlibrary loan between HLI]
5. ** Please tick (✓)

- | | | |
|-------------------------------------|---------------------|--|
| <input type="checkbox"/> | CONFIDENTIAL | (Contains classified information bounded by the OFFICIAL SECRETS ACT 1972) |
| <input type="checkbox"/> | RESTRICTED | (Contains restricted information as dictated by the body or organization where the research was conducted) |
| <input checked="" type="checkbox"/> | UNRESTRICTED | |

Azra

(AUTHOR'S SIGNATURE)

Permanent Address
NO 1, LORONG 22M,
TAMAN DESA ILMU,
94300, KOTA SAMARAHAN.

Date: 30/6/2023

Validated by.

Shapiee
DR SHAPIEE ABD RAHMAN
Senior Lecturer (DSS2)
Computational Science Programme
Faculty of Computer Science & Information Technology
Universiti Malaysia Sarawak

(SUPERVISOR'S SIGNATURE)

Date: 30/6/23

Note * Thesis refers to PhD, Master, and Bachelor Degree

** For Confidential or Restricted materials, please attach relevant documents from relevant organizations / authorities

Declaration

I, Nur'azra Alia Nisa binti Zulpakar, solemnly affirm that the thesis titled "Forecasting Stock Price by Using Artificial Neural Networks" is the product of my independent research, except for the inclusion of relevant quotations and proper citations. I further affirm that the thesis has not been previously or currently submitted for any other degree at Universiti Malaysia Sarawak (UNIMAS).

Signature,

A handwritten signature in black ink that reads "Azra". The letters are cursive and slightly slanted to the right.

(NUR'AZRA ALIA NISA BINTI ZULPAKAR)

Faculty of Computer Science and Information Technology

Date: 25th January 2023

University Malaysia Sarawak

Acknowledgment

This project, which marks the culmination of my final year, has been an incredibly honorable and humbling experience. I am deeply grateful for the unwavering support and guidance that have been extended to me throughout this challenging yet rewarding journey. Firstly, I would like to express my heartfelt appreciation to my project supervisor, Dr. Shapi-ee bin Abd Rahman, and my project co-supervisor, Madam Nur Fazliana binti Rahim. I would like to thank Sir Ahmad Alif bin Kamal too for guiding me throughout this project. Without their invaluable support, guidance, and mentorship, this project would not have come to fruition. I am truly grateful for their patience, encouragement, and constructive feedback, which played a vital role in the successful completion of this project. I would also like to extend a special thank you to my friends, whose unwavering support and encouragement have been a constant source of inspiration and motivation for me throughout this undertaking. Their unwavering faith in my abilities has been truly uplifting. Lastly, but certainly not least, I want to express my deepest gratitude to my family, who have been with me every step of the way. Their love and unwavering support have served as a constant reminder of how fortunate I am, and I cannot adequately express my appreciation for everything they have done for me. In conclusion, I would like to extend my sincere thanks to everyone who has been a part of this journey. Without the support, guidance, and mentorship of all the individuals mentioned above, this project would not have been possible. I am truly grateful to each and every one of you. Thank you all.

Abstract

Machine learning is widely used in predicting the stock prices. A stock market trade is an activity that requires investors to obtain accurate and timely information in order to make informed decisions. Due to the large number of stocks that are traded on a stock exchange, a variety of factors are considered in the decision-making process. In addition, it is also difficult to predict the behaviour of stock prices due to the uncertainty associated with them. There have been a number of studies conducted on the topic of forecasting stock values using machine learning. Hence, in this study, an Artificial Neural Network model is proposed as a machine learning algorithm for forecasting stock prices. This study utilizes the daily stock prices of Apple Inc. and Microsoft Corp gathered from the NASDAQ stock exchange. The processed data are then evaluated using the Root Mean Square Error (RMSE) and Absolute Error to analyse the performance of the model proposed.

Abstrak

Machine learning digunakan secara meluas dalam meramalkan harga saham. Pasaran saham ialah aktiviti yang memerlukan pelabur mendapatkan maklumat yang betul dan tepat pada masanya untuk membuat keputusan yang sewajarnya. Oleh kerana bilangan saham yang banyak didagangkan di bursa saham, pelbagai faktor perlu dititikberatkan dalam sebelum membuat keputusan berkaitan pelaburan. Selain itu, gerak-geri harga saham juga turut sukar untuk diramal kerana sifatnya yang tidak jelas. Terdapat beberapa kajian yang telah dijalankan berkaitan dengan topik peramalan nilai stok menggunakan *machine learning*. Oleh yang demikian, model *Artificial Neural Network* telah dicadangkan sebagai algoritma *machine learning* untuk meramalkan harga saham dalam kajian ini. Kajian ini menggunakan harga saham harian Apple Inc. dan Microsoft Corp yang didapatkan dari bursa saham NASDAQ. Data yang diproses kemudiannya dinilai menggunakan *Root Mean Square Error (RMSE)* dan *Absolute Error* untuk menganalisis prestasi model yang diusulkan.

Table of Contents

Declaration.....	i
Acknowledgment.....	ii
Abstract.....	iii
Chapter 1: Introduction	1
1.1 Introduction	1
1.2 Problem Statement	2
1.3 Scope	3
1.4 Aims and Objectives	3
1.5 Methodology	4
1.6 Significance of Project	6
1.7 Project Schedule.....	7
1.8 Expected Outcome	9
1.9 Summary	9
Chapter 2: Literature Review.....	10
2.1 Introduction	10
2.2 Stock Trading	10
2.3 Machine Learning in Stock Price Prediction.....	14
2.4 Review of similar research.....	17
2.4.1 Stock Price Forecasting Using ANN Method.....	18
2.4.2 Predicting Stock Prices Using Long Short-Term Memory (LSTM).....	26
2.4.3 Stocks Market Prediction Using Support Vector Machine (SVM).....	30

2.5 Comparison of Related Research Work	36
2.6 Summary	39
Chapter 3: Methodology	40
3.1 Introduction	40
3.2 Proposed Framework of Research Methodology	41
3.2.1 <i>Data Collection and Pre-processing</i>	43
3.2.3 <i>Performance Evaluation of the Model and Comparison between other Methods</i> ..	54
3.2.3.1 <i>Performance Evaluation</i>	54
3.2.3.2 <i>Comparison between other methods</i>	55
3.3 Summary	56
Chapter 4: Implementation	57
4.1 Introduction	57
4.2 Data Collection.....	57
4.3 Data Pre-processing.....	59
4.3.1 Data Cleaning	59
4.3.2 Import Data into RapidMiner	64
4.3.3 Set Role of Attributes	67
4.3.4 Split Data into Training and Testing Data.....	70
4.3.5 Feature Selection	72
4.4 Cross Validation.....	74
4.5 Processes in RapidMiner.....	82

4.6 Conclusion.....	83
Chapter 5: Result.....	84
5.1 Introduction	84
5.2 Result and Discussion	85
5.2.1 AAPL Inc.....	86
5.2.2 MSFT Corp.....	93
5.2.3 Comparison of RMSE and Absolute Error Values for AAPL and MSFT Datasets	98
5.2.4 Comparison with Other Method	100
5.3 Conclusion.....	103
Chapter 6: Conclusion and Future Works.....	104
6.1 Introduction	104
6.2 Conclusion.....	105
6.3 Future Works.....	106
References.....	109

List of Tables

Table 2.1: Supervised and Unsupervised Machine Learning and Its Algorithm.....	177
Table 2.2: Results of Stock Return Prediction.....	29
Table 2.3: Description of Variables in Dataset.....	31
Table 2.4: The Results of The SVM Model's Classification on Training Samples.....	32
Table 2.5: Classification of Test Samples by SVM Classifier.....	33
Table 2.6: Accuracy Rates.....	33
Table 2.7: Prediction Error.....	35
Table 5.1: Table of close prices for dates in training and testing data for AAPL.....	89
Table 5.2: Table of close prices for dates in training and testing data for MSFT.....	95

List of Figures

Figure 1.1: Framework of the study.....	4
Figure 1.2: Gantt Chart of FYP1.....	7
Figure 1.3: Gantt Chart of FYP2.....	8
Figure 2.1: E*TRADE Sign In page (E*TRADE, 2023).....	12
Figure 2.2: Fidelity page (Fidelity, 2023).....	12
Figure 2.3: Robinhood page (Robinhood, 2023).....	13
Figure 2.4: Illustration of a feedforward MLP network.....	19
Figure 2.5: Network based on a comparison of the output the target.....	20
Figure 2.6: Neural Network Training.....	21
Figure 2.7: Training Data Set.....	22
Figure 2.8: Performance curve for training.....	23
Figure 2.9: Error Histogram.....	24
Figure 2.10: Regression plot for training.....	25
Figure 2.11: Target and Estimated data of annually NASDAQ data.....	25
Figure 2.12: Accuracy for High/Low/Open/close with 500 epochs.....	28
Figure 2.13: Final Evaluation of Actual and Predicted Values.....	34
Figure 3.1: Framework of the study.....	42
Figure 3.2: Flowchart of Data Collection and Pre-processing phase.....	43
Figure 3.3: Flowchart of Model Development phase.....	46
Figure 3.4: Basic ANN Model.....	51
Figure 4.1: Historical data of MSFT on Yahoo Finance.....	57
Figure 4.2: Historical data of AAPL on Yahoo Finance.....	57
Figure 4.3: Date column for AAPL data.....	59
Figure 4.4: Date column for MSFT data.....	59

Figure 4.5: Date column for AAPL data after replacing the values.....	59
Figure 4.6: Date column for MSFT data after replacing the values.....	59
Figure 4.7: Adjusted close price and volume columns for AAPL data.....	60
Figure 4.8: Adjusted close price and volume columns for MSFT data.....	60
Figure 4.9: AAPL dataset after eliminating irrelevant columns.....	61
Figure 4.10: MSFT dataset after eliminating irrelevant columns.....	61
Figure 4.11: Added previous close price column in AAPL data.....	62
Figure 4.12: Added previous close price column in MSFT data.....	62
Figure 4.13: "Read Excel" operator.....	63
Figure 4.14: "Read Operator" is dragged onto the canvas.....	64
Figure 4.15: Choosing AAPL.xls to import be imported.....	64
Figure 4.16: Preview of AAPL dataset.....	65
Figure 4.17: Format columns for all attributes.....	66
Figure 4.18: "Set Role" operator.....	67
Figure 4.19: Assign role to attributes.....	67
Figure 4.20: Connection between "Read Excel" operator and "Set Role" operator.....	69
Figure 4.21: "Split Data" operator.....	70
Figure 4.22: Relation of "Read Excel" operator, "Set Role" operator and "Split Data" operator.....	70
Figure 4.23: Assign partitions value for splitting data.....	71

Figure 4.24: "Select Attributes" operator.....	72
Figure 4.25: Connection between "Split Data" operator and "Select Attributes" operator.....	72
Figure 4.26: Parameters configuration for "Select Attributes" operator.....	73
Figure 4.27: Connection of previous operators with "Cross Validation" operator.....	74
Figure 4.28: Configuration of "Cross Validation" operator.....	75
Figure 4.29: "Neural Net" operator in Training set.....	76
Figure 4.30: Configuration for "Neural Net" operator.....	76
Figure 4.31: "Apply Model" and "Performance" operators in testing set.....	79
Figure 4.32: Process in RapidMiner.....	81
Figure 5.1: RMSE value for AAPL Inc dataset.....	86
Figure 5.2: Absolute Error value for AAPL Inc dataset.....	87
Figure 5.3: Table of close price column, prediction of close price column and date column for AAPL Inc dataset.....	88
Figure 5.4: Graph of Predicted Stock Price and Actual Stock Price of AAPL Inc.....	91
Figure 5.5: RMSE value for MSFT Corp. dataset.....	93
Figure 5.6: Absolute Error value for MSFT Corp. dataset.....	94
Figure 5.7: Table of close price column, prediction of close price column and date column for MSFT Corp. dataset.....	95
Figure 5.8: Graph of Predicted Stock Price and Actual Stock Price of MSFT Corp.....	98

List of Equation

Equation 2.1: Equation of optimal classifier.....	34
Equation 3.1: Equation of Root Mean Square Error.....	54
Equation 3.2: Equation of Absolute Error.....	54

Chapter 1: Introduction

1.1 Introduction

A stock price represents the cost of acquiring one share of a company and is subject to daily fluctuations driven by market forces. These changes in share prices are influenced by the balance between supply and demand. When there is a higher demand for a stock compared to its supply, the price tends to rise. Conversely, if there is more supply than demand, as more investors want to sell the stock than buy it, the price tends to decrease. Consequently, the stock market is recognised for its dynamic and non-linear nature. Predicting stock prices has historically posed a significant challenge for business analysts and researchers (Roondiwala et al., 2015). Investors find the prediction of stock prices intriguing, as it aids them in making profitable investment decisions by providing insights into future market conditions. Given that the stock market involves time series data, researchers have extensively investigated and proposed various models for its analysis (Ghosh et al., 2019). Although many papers have been published on nonlinear statistical modelling of stock prices, most studies require the specification of a nonlinear model prior to estimation (Gurjar et al., 2018). Considering the nature of the stock market, Artificial Neural Networks (ANNs) have emerged as a more suitable method for capturing the complex relationship between a stock's performance and its determining factors compared to other statistical methods (Gurjar et al., 2018). Previous research has utilised different sets of input variables to forecast stock returns, ranging from individual time series data to incorporating heterogeneous market data and macroeconomic variables. The identification of nonlinear patterns in stock market index returns has directed researchers' attention towards the nonlinear prediction of stock prices. A study by Rasel et al. demonstrates that the Artificial Neural Network model outperforms other models such as Support Vector Machine (SVM) and Logistic Regression (LR) in terms of higher accuracy when using multiple attributes, including date as a special attribute, followed by regular attributes like open price, high price, low price, and close price (2015). Therefore, this project

aims to employ the machine learning technique of Artificial Neural Networks to forecast stock prices. Section 1.2 addresses the problem statement, Section 1.3 discusses the project's scope, Section 1.4 outlines the aims and objectives, Section 1.5 describes the methodology, Section 1.6 emphasizes the significance of the project, Section 1.7 presents the project schedule, and Section 1.8 highlights the expected project outcome.

1.2 Problem Statement

The phrase "buy low, sell high" is relatable to investors, but it does not provide sufficient context for making actual investment decisions. Before investing in stocks, an investor must understand how the stock market operates. Investing in good stock at the wrong time can be devastating, whereas purchasing a mediocre stock at the right moment can be profitable. Today's financial investors are struggling with trading issues since they are unable to determine which stocks to buy or sell to earn excellent profit due to many factors affecting the stock price such as inflation, exchange rate and political affairs across the globe. Without a proper strategic approach to the market, investors tend to lose money in the long run. A positive rate of return can be achieved if an efficient algorithm can be designed to forecast the stock price since the data in stock price analysis is huge and non-linear. In addition, the efficient algorithm must be able to handle the data variation to analyse the hidden patterns and relationships within the data. Compared with previous approaches, machine learning technique has shown to increase efficiency by 60% to 80% in this field (Vijh et al., 2020). ANNs are capable of identifying hidden features through self-learning process which can be used to discover patterns and relationships in a huge data set. ANNs have acquired widespread recognition due to their abilities to solve complex problems, identify non-linear patterns between data that are notoriously hard to model using traditional techniques, ability to generalise and learn, adaptability, universal function approximation and parallel data

processing (Wanjawa & Muchemi, 2014). ANNs consists of three layers which are input layer, hidden layers, and output layer (Moghar & Hamiche, 2020). Therefore, artificial neural networks-machine learning method is used widely to predict stock prices as investing is becoming more crucial in the financial market to outpace inflation as well as having the potential for healthy long-term returns.

1.3 Scope

The primary objective of this project is to model and assess the effectiveness of artificial neural network (ANN) models for predicting stock prices on companies listed in the United States stock exchange. The dataset used for analysis will comprise daily stock market information, including the opening price, closing price, low price, high price, and previous close price. To evaluate the performance of the ANN models, a five-year dataset will be gathered and analysed. The data will be collected from Yahoo Finance, specifically focusing on stocks from Apple Inc. (AAPL) and Microsoft Corporation (MSFT).

1.4 Aims and Objectives

The aim of this project is to forecast stock prices by using artificial neural networks on stocks listed on the US stock exchange. The objectives of this project are as follows:

- To collect data on a number of public listed companies on the US stock market.
- To explore a popular machine learning method called Artificial Neural Network for forecasting stock prices.
- To validate the forecasting accuracy using artificial neural network model based on various performance metrics.

1.5 Methodology

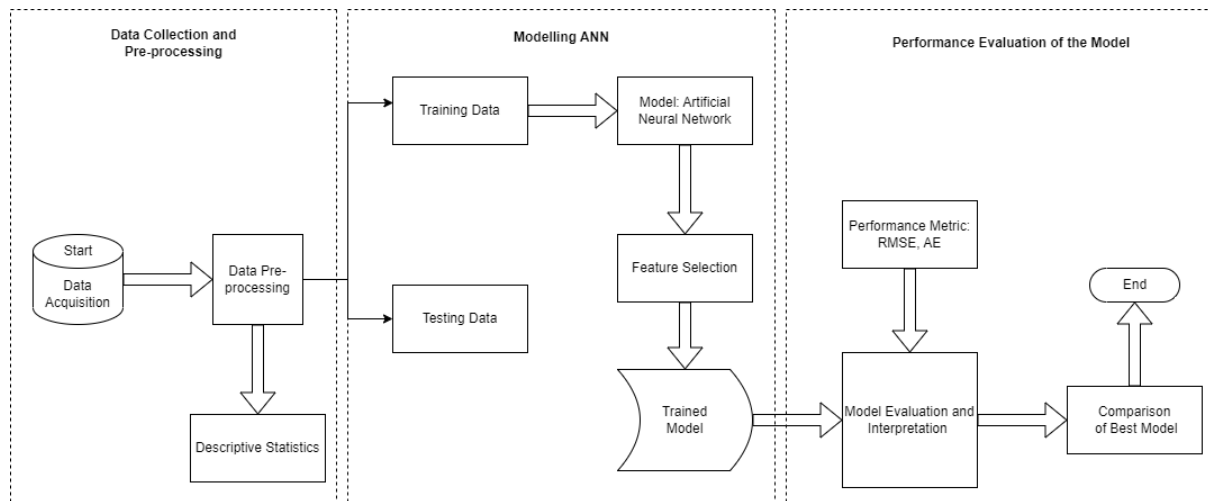


Figure 1.1 Framework of the study.

Phase 1: Data Collection and Pre-processing

To conduct this project, an initial dataset containing historical stock prices will be acquired. During this stage, data from the selected companies will be collected, spanning a period of five years. Several essential indicators can be employed to identify fluctuations in stock prices, such as the initial stock price, the final stock price, the lowest price observed during the day, the highest price observed during the day, and the total trading volume of shares for that day (Nivethitha et al., 2019).

The subsequent stage is a crucial aspect of this project which is data pre-processing. It is essential to ensure accurate results by addressing issues that may arise from improper data collection methods, such as out-of-range values and missing data. Therefore, it is important to examine the representation and quality of the data before conducting any analysis. Data pre-processing encompasses various steps, including data cleaning, feature selection, and data splitting. Data cleaning involves identifying and resolving problems such as missing values, inconsistencies, and errors in the dataset, improving its overall quality and reliability. Data splitting involves dividing the dataset into two subsets, a training set and a testing set. The training set is used to train the model, while the testing set is used to evaluate its performance

on unseen data. This separation ensures an unbiased assessment of the model's generalisation ability. Additionally, feature selection techniques are employed to identify the most relevant and informative features from the dataset. By reducing the dimensionality and focusing on key features, the model's performance and interpretability are improved, while minimising the impact of noise and redundancy.

Phase 2: Modelling ANN

During modelling ANN phase, the project focuses on creating an accurate and reliable predictive model. The chosen model for this project is an ANN, a versatile and powerful algorithm known for capturing complex patterns in data. Through iterative training, the ANN learns from the training data, adjusting its internal parameters to minimise prediction errors and optimise performance. After training the ANN model with the chosen features and the training dataset, its performance is assessed using the testing set. Throughout the modelling phase, iterative experimentation and fine-tuning may occur to optimise the ANN model's performance. This may involve adjusting hyperparameters, exploring different architectures, or employing regularization techniques to prevent overfitting.

Phase 3: Model Performance Evaluation.

The performance evaluation phase of this project involves the assessment of the developed predictive model's effectiveness and accuracy. Model evaluation focuses on measuring the predictive performance using specific evaluation metrics. In this project, the chosen metrics are Root Mean Squared Error (RMSE) and absolute error. RMSE is widely used in stock prediction tasks and represents the average magnitude of prediction errors. A lower RMSE indicates better prediction accuracy. Additionally, absolute error is used to assess the magnitude of prediction errors without considering their direction. During the interpretation phase, the evaluation metrics are analysed to gain insights into the model's performance. The

aim is to understand the strengths and weaknesses, identify patterns or biases in the prediction errors, and assess the model's predictive power. Furthermore, the model's performance can be compared with other models or benchmarks to evaluate its relative effectiveness. This comparison helps determine whether the model outperforms existing approaches or provides competitive results.

1.6 Significance of Project

The benefits that could be gained from this study of stock market forecasting is that it significantly increases the chances of making profit. Before knowing how to predict, investors frequently commit the mistake of not doing their analysis appropriately, which results in incorrect predictions. As a result, they invest their money in stocks based on gut instincts or random guesses in the hope that the prices move in their favours. However, most of the time, they end up incurring losses. After mastering the proper prediction strategies and applying them effectively, one can minimize losses. Making well-informed decisions by using reliable stock market prediction strategies can substantially increase their profits. For instance, investors can visualise the data produced by ANNs model which can help them gain better understanding of the underlying trends and patterns in the data. It is particularly useful for ordinary investors who are less familiar with the technical details of artificial neural networks, as it makes understanding the model's predictions easier.

Besides, stock price forecasts are closely related to economic forecasts as stock prices are often used as indicators of the health of an economy. Furthermore, economic forecasting can be used to anticipate future policy decisions, such as changes in interest rates, government spending, and taxes, which may affect stock prices. By analysing stock prices, economists and analysts can gain insight into the overall economic situation, such as the level of economic growth, inflation, and employment.

1.7 Project Schedule

Figure 1.2 shows the Gantt Chart for final year project 1.

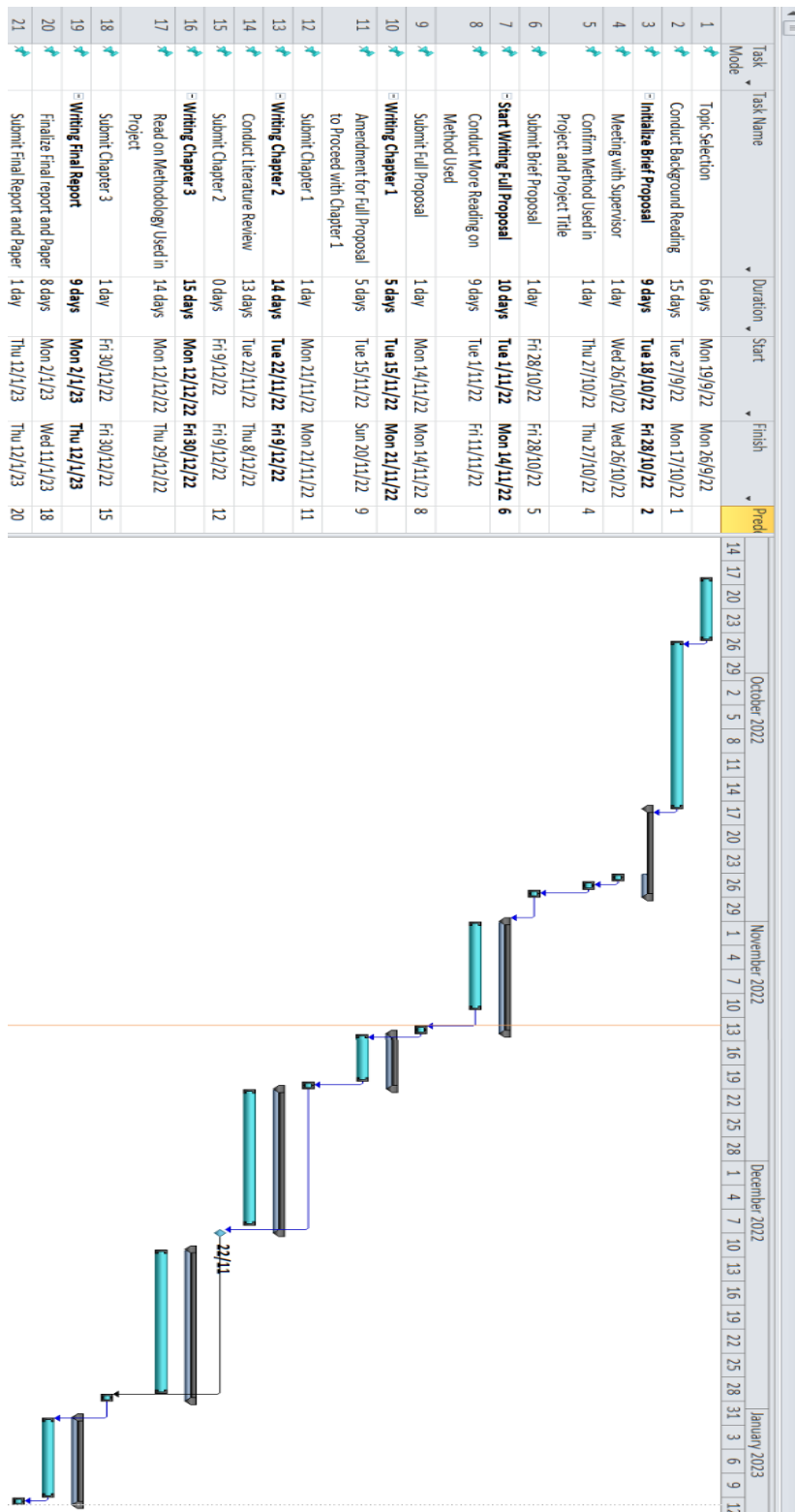


Figure 1.2: Gantt Chart of FYP1.

Figure 1.3 shows the Gantt chart of final year project 2.

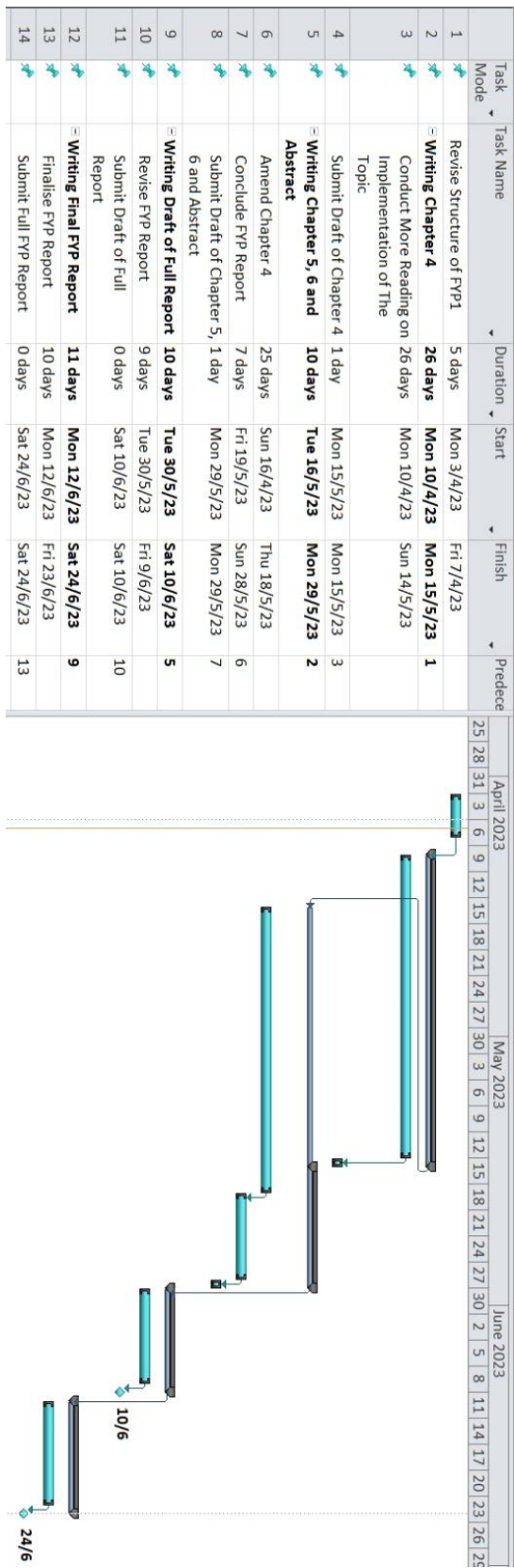


Figure 1.3: Gantt Chart of FYP2.

1.8 Expected Outcome

By the end of this project, the developed ANN model is able to help investors to make smart investment decisions based on the stock price forecast. Reliable stock price forecasts can aid investors in achieving consistency in terms of return on investment (ROI) over the long term.

1.9 Summary

As analysts and investors attempt to outperform the market, the use of neural networks to predict stock market prices will continue to be an important topic of study. The main objective is to improve the investment's yield. Studies has previously demonstrated that evaluating the return on investment in stock market using any of the fundamental methods is a difficult, costly, and time-consuming as it involves gathering and analysing financial data from a variety of sources, including financial statements, annual reports, and these data should be reviewed as well as understood by the analysts to perform a comprehensive fundamental analysis. Hence, researchers are encouraged to develop efficient approaches to forecast future stock prices. Studies have shown that if the ANN model is trained with a larger set of input data, it will produce more accurate forecast prices. Therefore, with the proposed ANNs model, we can achieve our goal by generating more accurate predictions as we train the model over time, providing us with more accurate prediction prices.

Chapter 2: Literature Review

2.1 Introduction

Literature review is the most important chapter that act as supporting materials to this research work. Besides acting as a supporting material, this chapter also include an overview and summary of previous research that are related to this research work, which is forecasting stock prices. Through this chapter, the details of stock price and how it works will be discussed. In addition, the different types of machine learning will be elaborated as well. Some literature reviews of previous research are studied and compared in the following section of this chapter.

2.2 Stock Trading

A stock is a financial token that signifies a fractional ownership interest in a company. Each stock has its own value which depends on the company's worth. When an individual purchases shares of a firm's stock, he is purchasing a small portion of that company. Investors who purchase the stock in companies believe that the stock value will rise in the future subsequently can be sold for a profit. An individual who owns stock in a company is called a shareholder as he receives a share of the company's profit.

Companies offer their shares to the public as a means of generating funds. In today's stock market, the value of a company is often determined based on its future potential earnings. Consequently, even relatively small companies have the potential to amass significant profits if investors believe in their future success. To facilitate the sale of shares, companies require a platform known as the stock exchange. By conducting an initial public offering (IPO) on an exchange, companies transition from being privately owned to becoming publicly traded entities. An IPO allows company founders to either sell their ownership stakes or raise additional capital. Once a company's stocks are listed on an exchange, they can be traded by the general public. Price fluctuations are typically influenced by public sentiment, but more substantial trends are usually linked to a company's earnings and operational performance.