Stock Price Forecasting Dashboard using Data Mining

"ForecastFuture"

By

Fakhri Akhdan Akhiar

17821

Dissertation submitted in partial fulfilment of

the requirement for the

Bachelor of Technology (Hons)

(Business Information System)

MAY 2015

Universiti Teknologi PETRONAS

Bandar Seri Iskandar

31750 Tronoh

Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Stock Price Forecasting Dashboard using Data Mining
"ForecastFuture"

By

Fakhri Akhdan Akhiar
17821

A project dissertation submitted to the
Business Information System Programme
Universiti Teknologi PETRONAS
In partial fulfilment of the requirements for the
BACHELOR OF TECHNOLOGY (Hons)
(BUSINESS INFORMATION SYSTEM)

Approved by,

_____

(Ahmad Izzuddin B Zainal Abidin)

UNIVERSITI TEKNOLOGI PETRONAS

TRONOH, PERAK

May 2015

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for work submitted in this project, that the original work is my own expect as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

_____

FAKHRI AKHDAN AKHIAR

ABSTRACT

Stock price forecasting Dashboard is an application that enable the user to predict he price of particular stock quote and help them to make decision making with fundamental source of information. The dashboard is mainly targeted to the early investor or individual with low level of understand to the stock investment and the prediction. User is only need to write the stock and the country code at the text box entitled which the application later will find the historical data at the live database. From the historical data, a plot of historical and future price will be plotted added with news from twitter newsfeed to compliment the result. This report intends to technical analysis with time series analysis equipped with the statistical model and data mining techniques to predict the stock price trends.

Findings indicate that implementation data mining onto stock price historical data to apply on the statistical model is able to provide fast and better result of prediction. Later, Fundamental added into this application will suggest the user to make right decision after the prediction has been plotted in time series format. As stock market known for its risk and uncertainty, attempt to forecast the price is required further studies and readiness to accept the future realities.

ACKNOWLEDGEMENT

I take this opportunity to express my profound gratitude and deep regards to my Final Year Project Supervisor Mr. Ahmad Izzuddin B Zainal Abidin for his exemplary guidance, monitoring and constant encouragement throughout the course of this thesis. The blessing, help and guidance given by him time to time shall carry me a long way in the journey of life on which I am about to embark.

I also take this opportunity to express a deep sense of gratitude to Co-supervisor Mr. Amzairi B Amar and Final Year Project Coordinator Mr. Ahmad Sobri Hashim for their cordial support, valuable information and guidance, which helped me in completing this task through various stages.

I am obliged to also express gratitude to this project expert Ms. Nor Atiqah Zainal Abidin for her knowledge sharing, advices, and support so that I am confidently choose this topic and write my report.

Lastly, I thank almighty, my parents, brother, and friends for their constant encouragement without which this assignment would not be possible.

Table of Contents

List of Equation

List of Equation

List of Figure

CHAPTER 1
INTRODUCTION

1.1. Background of Study

Over the last decades, security market investment has become one of the most attractive investments compared to its peers. Stock market, one of the investment vehicles is one of the investments that promise a high return in line with the investor's risk profile. Due to the enormous interest towards the stock market, stockbroker and investment bank match each other to provide the best practice and best service to their customers and investors. To make sure the company is provides necessary supports to the individual or corporate investor, the bank or financial institution set up a research department or separate entity to prepare such kind of report to predict value of the stock in the future.

The research department later will prepare such kind of report that lists numerous stock options that could become the guide for investors to select companies or stocks and add to their portfolio. The report complete with forecasting of particular of stock entity may only be acquired by the customers if they are willing to buy or subscribe the report that occasionally expensive. Furthermore, the report may come with complex and a very technical econometrics result rather than information that could easily engrossed by the consumer. This resulted to evidently current situation where only investors who understand to read the report or the corporate that bought this typical of report.

Trading means the process of transferring of money worth to the stock or security from the seller to the buyer (Investopedia, 2015). The money transferred for trading purposes is equally worth to the price of the single stock that has agreed by the parties. The interest of ownership of a particular security is refers from the stock or share that the investor has bought or received. As same as other

investment, stock trading followed with risk. The risk that followed investors during trading are mainly market risk and liquidity risk. This risk will determined investor to enter the market, and vice versa.

Hedge fund managers use risk management best practices onto their diversified portfolio to increase the performance of the portfolio (Basak, Shapiro, & Teplá, 2003; Fung et al., 1997). Basak et al., (2003) further explained that investor risk profile would drive investment decision and post decision evaluation. The level of confidence of investors will also mutually work on price fluctuation. The information grasped by the investors are very important to make them knowledgeable enough to the company or index profile and financial standings that will be bought in the future. When the flow of information is stranded or hardly to understand, one must happen is that investor will not invest to the stock offered in stock market.

Numerous tools stock price prediction is keeping being developed by the experts and researchers, and highly valuable. Mittermayer (2004) explained that Forecasting based on structural data is popular among other forecasting application. Furthermore, Zhai, Hsu, & Halgamuge (2007) report that publications related about stock market prediction attempt are numerous, most of them focus on time series prediction with various model approaches.

This paper proposes an interactive application that utilizes an open source tools and information from fast and abundant data from internet, which could be suggested for the solution of the current problems and maximize every opportunity for the early investor. In this application development, the paper also explore about data mining studies to be chosen and implemented to predict the share price. The application is added with the quote newsfeed related to the market and public listed companies, providing an improvement of the investor confidence level.

1.2. Problem Statement

Today, there are numerous tools to predict stock price used by investment managers, corporate or individual investors. Some of the tools that utilize sophisticated methodology and statistical model have implemented to the system or application. However, those kinds of applications are usually costly to be acquired by individual investor or beginner player in stock market. Thus, an application that utilizes an open source tools and data from fast and rich of content from internet can be suggested for the solution of the problem. The application focused on user friendliness features in context to deliver the best forecasting method and result for the users.

Inexperienced stock traders or called as early investor is usually entered the market with minimum of principle. Commonly, early investor are using salary remainder, grants, and family capital trust to start their first capital to enter the market. This resulted a high tendency of soft approach and be very careful put their money into respected stock according into investor personal risk profile. Thus, many early investor directly trust their capital to hedge funds and asset management company to manage their money and expect return in the future.

However, not all early investors opted out from directly trading in the market. Some of them are doing the trading activities by their own to evade several fees and commission to asset management companies that is renowned to be high-priced. To expose this early investor with sufficient knowledge about stock trading mechanism and information, a content management tool is importantly required. Yet, most tools relevant to stock market information or stock price forecasting come with expensive subscription fees, which excluded from the trading software fees. This situation could hardly discouraging potential traders to start trading and investing in the market.

As many factors influence the fluctuate moving stock price of a company or sector of a company, it is also affected by the incoming news articles and updates (Godbole & Srinivasaiah, 2007; Herath, 2007). This news varies from latest earning from the company, announcement related to management, dividends announcement by the company, analysis by editorial news and professional traders, and domestic and international news that related to the company. This

news has made it an important factor to be use for each trader to analyze stock of a company before make any dicision. Unfortunately, some of the current forecasting tools is not providing a relevant and most recent information about the company to use by the tools subscribers, which hinders the investors to make fundamental analysis.

ForecastFactor or Stock Price Forecasting Dashboard with Data Mining is attempt the entire problem explained before. The dashboard is aim to provide a user-friendly application that can be used by all types of investor with different objectives. An advance and proven statistical model is utilized to plot the historical and prediction of the price to ensure an accurate and reliable result to te users.

The rest of this chapter will be organized as follows: the next section will put forward the problem statement of this project. This will be followed with the objectives section and lastly scope of study.

1.3. Objectives

The main objectives of this research are:

- To study data mining approach on stock price forecasting
- To create tool that is able to forecast stock price movement fluctuation in the future

In essence, this project is to help investor to make decision to buy or sell the stock rather than need them to read long writing report that is complicated.

This project deliverables would be:

- Data mining approach that is accurate, integrity, and reliable.
- Analyze time series model approach for prediction
- A system that would able to generate forecast plot and newsfeed for decision-making purposes.

1.4. Scope of Study

This project aims is to predict public listed companies stock price listed in open market all over the world. Because the prediction tool is using data mining method, historical data of the price is very important. The project will use existing historical data that distributed freely in internet and provided by trustable provider to make sure the desired result is fully trusted and reliability.

This project could be adapted into large capital market around the world. However, certain of limitation is necessary to make sure that the scope will not be large and become time consuming, and decided to only focus on 10 stocks of the highest market capitalization of companies publicly listed in Bursa Malaysia. Furthermore, the implementation of several methods needs to be tested into different environment where the excellent desired output will be chose as the object of this project.

The historical data is downloaded from search engine and historical data provider 'Yahoo Financial'. Several developer and time series experts have used yahoo finance historical data in order to calculate the price prediction with different approaches. Other than that, the data would also be tested and analyze to ensure the accuracy and reliability of data.

The historical data that will be used in this project would be naturally come from 'closed price' without look into volume of transaction and other variable in the historical data. This is because the historical closed price is already adequate to find the pattern or plot and to predict the next pattern of future stock price.

The scope of the study would further assist in creating the accuracy and reliability that the system proposes. The historical data downloaded from not-native provider such as 'Yahoo Financial' will always be look into the real data provided by the regulators in particular country. The make sure accuracy of data, this project will also try to find the best model approach that can be used to provide an output that become solution for large communities.

CHAPTER 2
LITERATURE REVIEW

2.1. Stock Market Analysis

There are numerous methods currently use by the investment managers and brokers to evaluate stocks. The evaluation covers many elements i.e. the performance of the stock to be market index as a whole, company news and action, return analysis made by the experts. All of the evaluation later lead towards predicting the future value of the stock in short and long period of time fundamental and technical analysis are the most widely used for stock valuation analysis method (Smart, Gitman, & Joehnk, 2014). Those analyses are looking into different attributes and class of measurement of the company.

Fundamental analysis or economic analysis focuses on the financial performance and economic situation. This method is mutually exclusive monitored the financial performance of the particular company towards the stock price in the security market. One most assured that the indication of the stock value whether it is overvalued or undervalued is reflected through the fundamental value of the company. In conclusion, fundamental analysis is important to investors to help get insight of the economic value of and the potential reflection to the share price (Smart et al., 2014).

In conclusion, fundamental analysis is the company evaluation of the financial performance and their current competition state in which influence the stock price in the market. Hence, it will automatically change the investor behavior into the stock price. The situation of undervalued is arbitrated when fundamental or the economic value of the company is lower to value of stock price, which eventually strive the line of graph to down to adjust the value consecutively. As

in the case of overvalued the share price value is much higher than the economic value, and led the investor to bullish the stock and lower down the price until the equilibrium achieved.

On the other hand, technical analysis focuses onto the stock behavior in the market rather than the economic value occur in corporation. (Yu & Cheng, 2007) explained that technical analysis method deals with determination of the stock price based on the past pattern of the stock (using time-series analysis), performed by technical analysis. Technical analysis requires basic investment knowledge and understanding of historical data and certain type of formula. Studies that covered on the technical analysis covered from dividend return, yield, until earnings per share. Most of the time, the analysis used by the beginner level of investor are the historical price, volume, buy or sell behavior in the market. By looking into those analysis, the pattern will be established which help investor to determine the investor respond.

It is a common knowledge that each stock price trade in the market is volatile. The entire stock are relates to each other as an event occurred. The trend of stock price is may not be like a climb of the price up to the highest level and go downturn to the lowest level and back again to medium level or up the ladder again. The technical approach is use to acknowledge this behavior and try to disseminate all of the information. Investor are usually extracted all of the historical data within time range and try to predict the future value and the movement. This movement of stock price can prevent the investor from big downturn and loss. This approach is one of the most the most reasonable and effective way to predict the price.

Several financial experts believe that technical approach is the best way to measure and predict the share price rather than to use the fundamental analysis. This because fundamental analysis requires the analyzer to farther lot of information, data processing, and knowledge in orders to prepare the report. Thus, fundamental analysis might not be suitable for this project, as it would take larger scope of the project and more investigation of the company's intrinsic value. However, the approach of the fundamental might not be left off. On the final phase, fundamental analysis will be using to help investor confidence on

their decision. Such a newsfeeds absorb from the internet will be useful for this approach.

2.2. Stock Market Prediction

As investing in stock has become one of the common tools of investors to gain massive return, stock price prediction has become a common ways to attain maximum return. Technical analysis has been widely known of its superiority to other analysis to predict stock prices. Numerous attempts have influence the breadth of the study of stock price prediction, this attempt only aims to create an accurate and proven model that provide a valid result.

Forecasting with technical analysis visualized in graphical pattern or table, which explained a list down the probabilities of future trend of stock market. Time series analysis is the main branch of forecasting techniques. Time series is a trend chart of historical data that visualize the trend in series of time. There is a huge variety of ways in forecasting economy indices like stock prices, most of which are time series analysis (TSA) methods based on structured data (e.g. stock prices table) (Pai & Lin, 2005).

2.2.1. Time Series: Definition

Time series analysis is the main branch of forecasting techniques. Time series is a trend chart of historical data that visualize the trend in series of time. (Devi, B, Sundar, & Alli, P, 2013) explained that time series is the best technique to forecast the trend of the stock price in indices, die to its volatility movement.

Time series features forecasting the share price in seasonal and long-term period. It provides a trend image of flow of the movement of share price, the growth, and the variations. The trend visualized to the predictors will eventually all-inclusive view and decision of stock in the market.

The general trend towards stock market among the society is that it is highly risky for investment or not suitable for trade (Devi, B et al., 2013). The seasonal

variance and steady flow of any index will help both for existing and naïve investor to understand and make a decision to invest in the stock market. To solve these types of problems, the time series analysis will be the best tool for forecast and also to predict the trend (Brockwell & Davis, 1991).

Time series technique is insufficient to be use when predictors want to define between demands of particular stock and other factors such as economic situation and changes in the price of a company to the others, but it is better when it is used to visualize volatility in time desired (Moon, Mentzer, Smith, & Garver, 1998). Nevertheless, time series analysis is still considered as the most relevant tools to forecast stock price (Brockwell & Davis, 1991).

Time series visualize single or many data series at the homogenous distance of time sequence. Time series are widely use to visualize financial history or its prediction weather, temperature, etc. Typically, time series plotted in line chart, time series given the user the output of trend based on the data given. Time series chart is as shown in Figure 1.

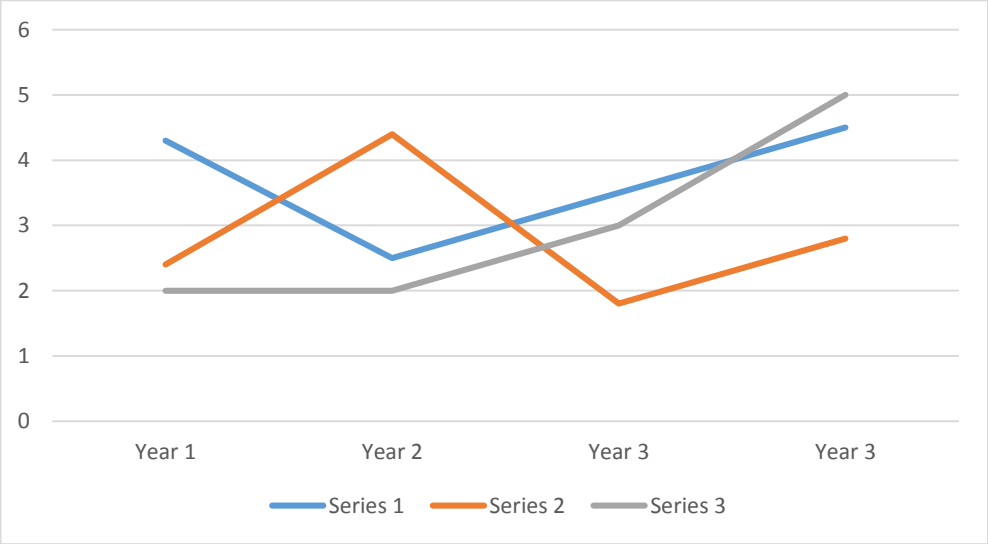

Figure 2-1 Time Series Chart

According to Sekar, Kannan, Sathik, & Arumugam (2010) prediction of financial time series is the same difficult as stock market prediction. It is important to consider both of the prediction at the same time to prevent work repetition and work burden.

Financial time series forecasting is using the historical data for later forecast the future based on the time interval that the predictor wants. After the prediction visualized, the predictor will eventually could see the trend and relate the historical and future each other. Later they will use the trend to make financial decision-making onto buy or sell final decision.

2.3. Data Mining for Financial Forecasting

Devi, B et al., (2013) explained that data Mining is an approach where analytically designed to discover data (market related or business oriented) and in search of reliable patterns and/or logical relationships between variables. Appropriate validation is used to find the patterns with the detected patterns among the data set. Data mining is also usually call as data discovery or knowledge discovery. Data mining enhance the revenue and reduce the cost incurred for the exploration of data

At the study of financial forecasting, the ability to predict and make the right decision is most desired. It prevents the investor or company from making wring decision and take away loss. The ability to forecast the future requires the skill to manage data and formations, which influence the decision making from the predictor perspective.

Financial institutions and brokers have now started to create massive data stored electronically to start making sustain groundwork of financial forecasting that widely known to be complex and dynamic. The tools that featured in these activities are to combine data mining tools. The data stored is historical data of the particular subject that to be forecasted. Furthermore, this data storing will eventually become bigger as more information and data is being stored.

Although data mining and data storing is very important, they come with unprecedented disadvantages. Researchers showed that many financial institution has not yet fully take full advantage of the data despite data storage is keeping increased (Sekar et al., 2010).

In financial modelling, data mining is the most useful in terms of finding useful patterns and correlations (Kovalerchuk & Vityaev, n.d.). It is common to observe financial modelling use data mining method and other computerized method. In conclusion, data mining is one of the method and most important to implement model of forecasting.

To measure the success of financial data mining, is by measuring the standard deviation between forecast result and actual values on testing data (Kovalerchuk & Vityaev, n.d.). Standard deviation however used to complete the mission to forecast the data for more brief of forecasting model.

Almost every financial forecasting study has been modeled by the researchers and being computerized. This project will only mention some of them that have been widely used and recently being continued to be explored. Most of these works focus on time series prediction with various model such Artificial Intelligent models i.e. Artificial Neural Network (ANN), Genetic Algorithm, Fuzzy System and statistical techniques such as Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedicity (GARCH).

Type of Method	Author	Research Overview
ARTIFICIAL INTELLIGENCE		
ANN	(Refenes, Burgess, & Bentz, 1997)	One of the most suited financial time-series forecasting • Non-linear models • Effective learning algorithm • Handle noisy data • Inputs of different kind of data
Genetic Algorithm	Armano, Marchesi, & Murru (2005)	A computational method come from natural evolution.

		• Population based model by using selection and recombination operators to generate new sample point in a search space. • GA deals with population of 'chromosomes' as potential solution in binary form. • Gas have been used in broad different task of prediction.
Fuzzy System	Wong, Wang, & Goh (1991)	Fuzzy System has been developed since 1920s, proposed by Lukaciewicz, • The degree of appropriateness is looking from the membership functions and outputs referred as membership degrees. • Investment analysis overcomes as Information's delivered as probabilities.
STATISTICAL		
ARIMA	(Sutheebanjard & Premchaiswadi, 2010)	Introduced by G. Box and G. Jenkins in the early of 1970s. • Patterns like stationary, non-stationary, and seasonal or periodic is able to be taken to be

		modelled with approach of time series. • ARIMA approached is complex and has limited in accuracy over other refined method.
GARCH	McDonald (2013)	Introduce by Bollerslev in 1986, GARCH is variant of Autoregresive Condirtional Heteroskedasticity that allows infinite lags with small parameters. • Model focus on maximum likelihood.

2.3.1. Autoregressive Integrated Moving Average (ARIMA)

Mondal et al. (2014) argued that stock prices could be forecasted due to they are not random generated values relatively they are a time series model and can be plotted in trend perspective. Forecasting usually performed by the analyst to make financial plans. Since having an excellent knowledge about share price in the future serves the confidence of the professional traders and investors.

The forecaster is forecasting to opt and conclude either enter or exit the market. This to make sure a high return without compromising cents of loss during the process. Since the aim to provide a necessary information and data for investor to make decision, a model of forecast to show the trend is important. ARIMA is an algorithm approach to transform the series, which is better compare to forecasting directly, and gives more accurate results (Al Wadi, Ismail, Altaher, & Karim, 2010; Devi, B et al., 2013).

ARIMA, or also known as Box-Jenkins methodology is powerful method with complex structure (Han & Kamber, 2006). The complexity of the method made

this approach result to different quality of result based on the user experience to the model. ARIMA is model that based from its predecessor ARMA model. The difference between them is that ARIMA attempted to convert non-stationary to stationary data before start to create the time series analysis (Mondal et al., 2014).

ARIMA model is classifies as ARIMA (p,d,q) where p signifies autoregressive of the data, d signifies to the integer of the data, lastly q as the moving average of the data set.

Equation 2-1 ARIMA Model (Devi, B et al., 2013)

$$Y_t = C + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \ldots + \theta_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_p \varepsilon_{t-q}$$

C – Constant term

AR – Non seasonal Auto Regressive coefficients ($\emptyset_1, \ldots \emptyset_p$)

MA- Non seasonal Moving Average coefficient ($\emptyset_1, \ldots \emptyset_p$)

ARLags - Lags corresponding to nonzero, non-seasonal AR coefficients

MALags – Lags corresponding to nonzero, nonseasonal MA coefficients

D - Degree of no seasonal differencing, D [if D has value 0 meaning no seasonal integration]

Variance – Scalar variance of the innovation process (σ_ε^2)

Distribution - Distribution of the innovation process

2.4. Existing Financial System

Bases on the findings, there are few existing product of dashboard offering to large community easily access in internet or standalone application that require user to install in personal computer. There are (2) existing dashboard chosen for this comparison study, that all of them come from developer origin from United States of America (USA). The dashboard is FactSet and Bloomberg Forecast by Bloomberg Terminal.

2.4.1. FactSet Research Systems

FactSet is software that provide financial information and analytic of key investment around the world. The system is not only taking direct data from the security market but also providing fundamental and corporate actions distributed at the news media.

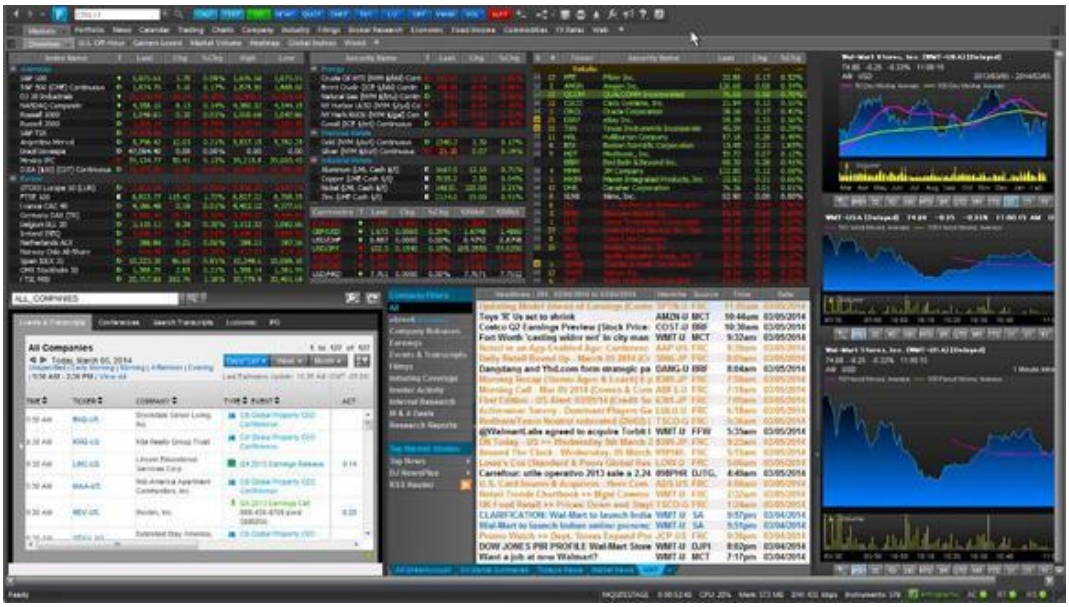

Figure 2-2 FactSet Software

2.4.2. Bloomberg Terminal

Bloomberg Terminal is computer system provided by news media giant Bloomberg that enable users to analyze real time financial market data and trade to it at the same time. Bloomberg Terminal added with the feature of forecasting with its own method. Usually the terminal can show the trend of the currency market and user can pick some prediction based on the conversion.

Bloomberg is current costing the use of the terminal in quarterly and yearly basis amounted up to US$ 24,000 a year for single use of the terminal.

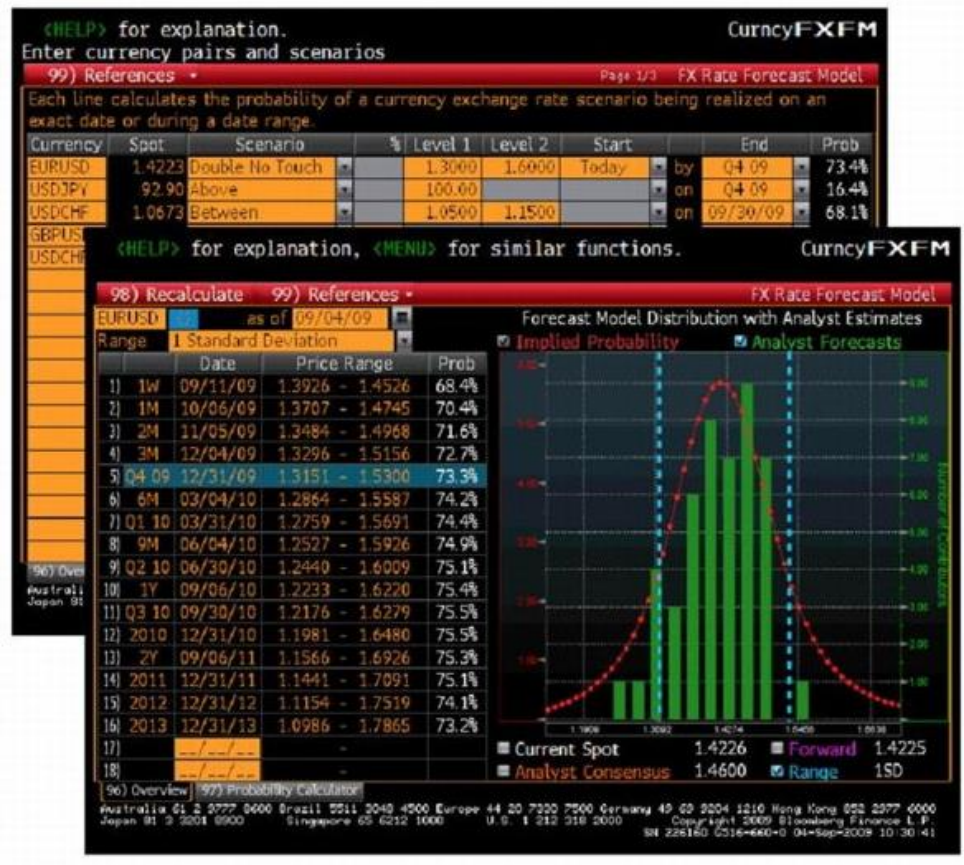

Figure 2-3 Bloomberg Terminal

CHAPTER 3
RESEARCH METHODOLOGY

In conducting this project, various processes and activities required before delivering the project and the complete application proposed. Figure 3-1 illustrate phases involved and deliverables.

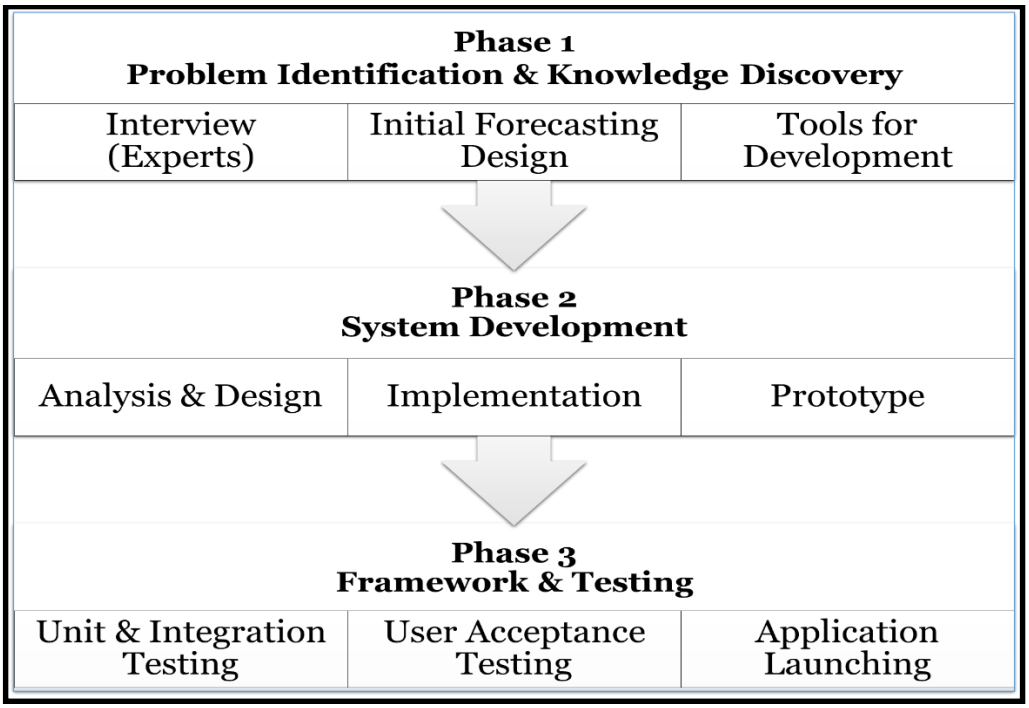

Figure 3-1 Application Development Phases

3.1. Phase 1: Problem Identification & Knowledge Discovery

Early before these phases started, a problem identification and objective are gathered and discussed. Various data and current state of the forecasting methodology is compiled and selected. Those data are grouped based on the

group of stock price investment as general, investment, stock price prediction, and data mining technique and model.

3.1.1. Interview

Interview to the expert is conducted to find related findings and analyze the need of the application being proposed. The expert selected for this study is **Ms. Nor Atiqah Zainal Abidin**, who has high experience on stock price forecasting and data mining technique. She also use and perform all of the above terms during her job at one of the Malaysia Government Linked Company.

3.1.2. Initial Forecasting Design

Forecasting design phase is the process to determine tools, resources, and data that will be useful for the project development and deliverables. Research and interview to the expert conducted to receive inputs regarding methodology and design that will be proper to implement in the dashboard.

3.1.3. Tools for Development

i. R Programming (R Studio)

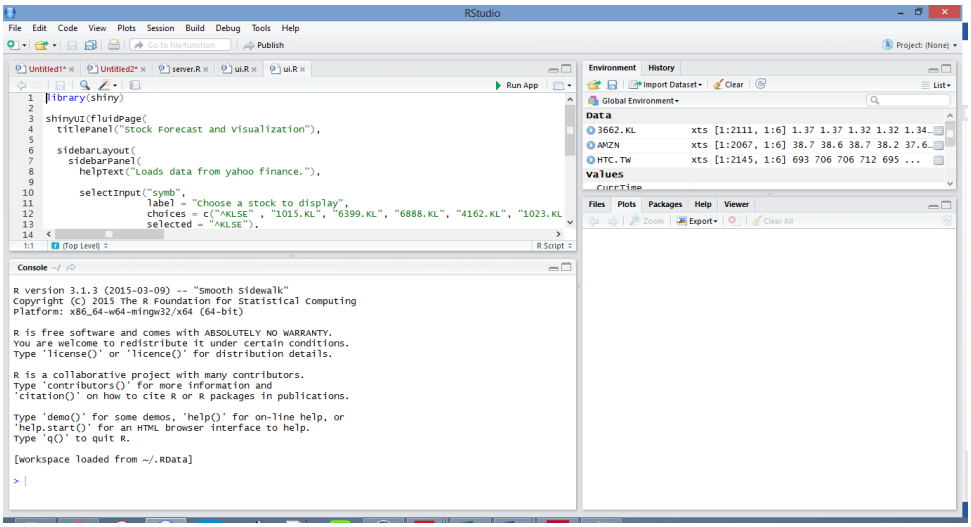

Figure 3-2 R Studio Interface

R is a programming language and software environment for statistical computing graphics. The R language is widely used among statisticians and data miners for developing statistical software. R programming is a very common tool to be used for forecasting, as many financial institutions currently required their employees to fully understand this software.

R programming is previously only widely used for statistical studies, but now it has been widely used for financial application to have some plotting and forecast. Other tools that can be used than R Programming is to utilized the same function tools such as Mat Lab, Python, etc.

R is a very important tool to develop this project. R is an open source software that full with capabilities and packages to ease the user in doing several work with it. This project utilized several packages such as "**quantmod**", "**ggplot2**", "**TTR**", and "**shiny**". Package "**quantmod**" used to extract data from Yahoo Finance directly. "**ggplot2**" and "**TTR**" utilized to plot the graph and time series of the historical data and the forecasted data. Lastly, "**shiny**" is being used to create a User Interface (UI) from R Studio.

ii. HTML Web Developer

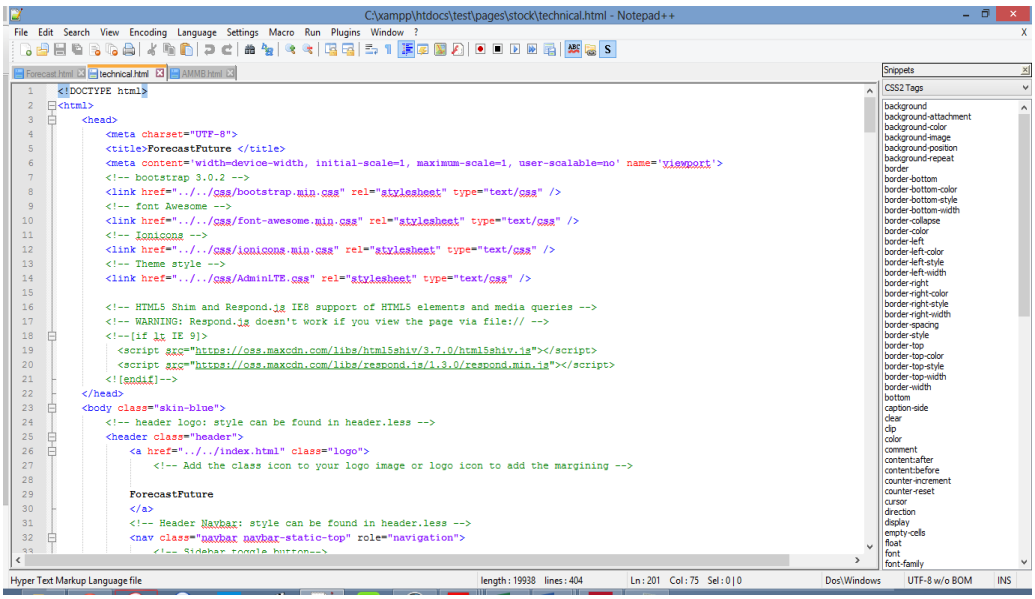

Figure 3-3 HTML Web Developer - Notepad++

HTML Web Developer is tools that being used to develop the dashboard and functions. HTML chosen due to its ability to launch online and interconnected to the cyber world of internet. Notepad++ used to code the HTML and gather all of the resources of back end to the dashboard. XAMPP utilized to reveal the HTML code and API widgets connected directly to internet.

To make the widget of the Googne News and Yahoo News newsfeed and twitter Tweets, several APIs from Twitter Dev, Yahoo Finance, Yahoo Pipes, etc. applied and setting into necessary requirements.

3.1.4. Equipment and Materials

This project requires information and data that will be using in the planning and development phase. The most important of materials is stock price historical data, a computerized stock quote from the market that has been group into raw data. Currently, there are many ways to locate and extract the historical data from internet. Such websites FinancialContent, Yahoo Finance, Google Finance, and AOL are providing a real time historical data for paid subscribing and free. This project utilize a feature from R Programming with one of its toolbox to take real time data from Yahoo Finance.

Furthermore, as explained before Twitter newsfeed used for fundamental analysis for user decision making after user successfully analyze the price and predict the quote. The algorithm of Twitter will be carried by only allowing user to input the search criteria based on the company's name or quote stock such as Google GOOG or Apple AAPL. CNBC (2013) reported that some of the big hedge funds and investors are now using twitter and other social media for their trade strategies or only to maintain watching on corporate action.

3.2. Phase 2: Prototyping Methodology

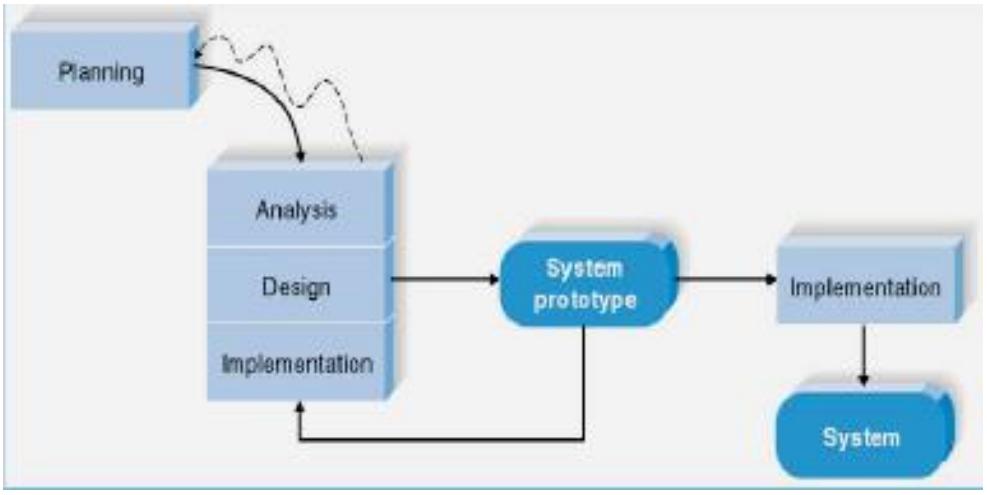

Figure 3-4 Prototyping Methodology (Box, Jenkins, & Reinsel, 2013)

Prototyping-based Methodology is a development lifecycle designed to give much faster development and higher-quality results than those achieved with traditional lifecycle. Prototyping has been widely known to make developers experience of software development is easier than other methodology. Provides a system for the users to interact with the product although it may not finished yet. This

Prototyping picked as the method of system lifecycle to conduct this project due to its abilities to adapt with developer time that has been set previously. Although the major steps such as analysis and quick design, testing, and implementation blocked into time. Prototyping helps developer to take care any problem in testing phase because of substantial prototyping cycles. These cycles can help the developer to make sure the analysis is correct and enable to re-build the analysis without compromising the final phase and product.

3.2.1. Requirement Gathering and Planning

The project topic and title to the final year project committee with the endorsement from supervisor proposed has been completed in the second week of FYP 1. After the topic, approved students are allowed to continue making the project and discuss with respective supervisor to conduct a requirement and

scope analysis. Phases occurred during the completion of Final Year Project are displayed in Gantt chart and key milestone so that students can easily monitor the process and key deliverables. The Gantt chart and key milestone is shown Table 3-1 – 3-2 below.

Table 3-1 Gantt chart FYP 1

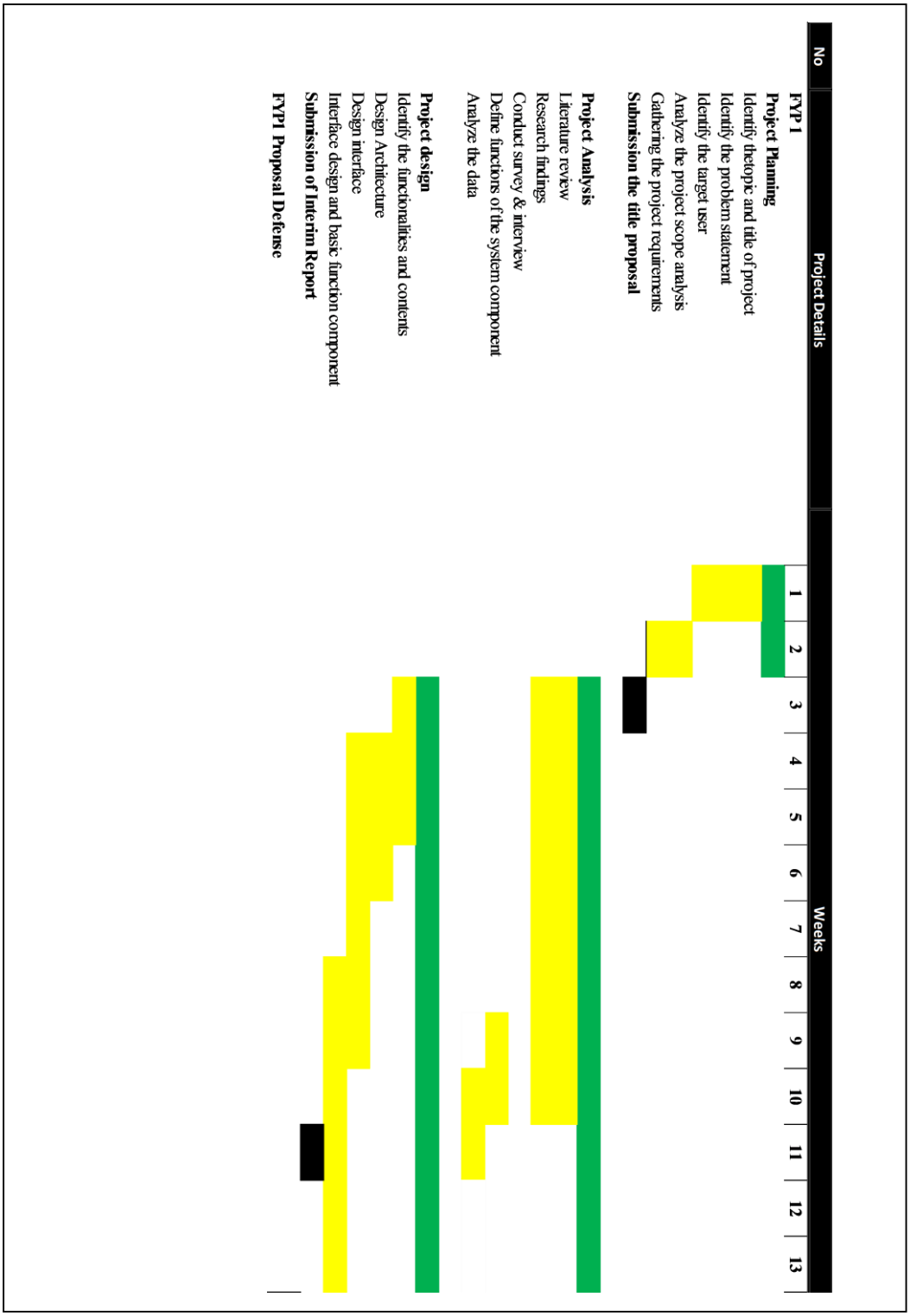

Table 3-2 Gantt chart FYP 2

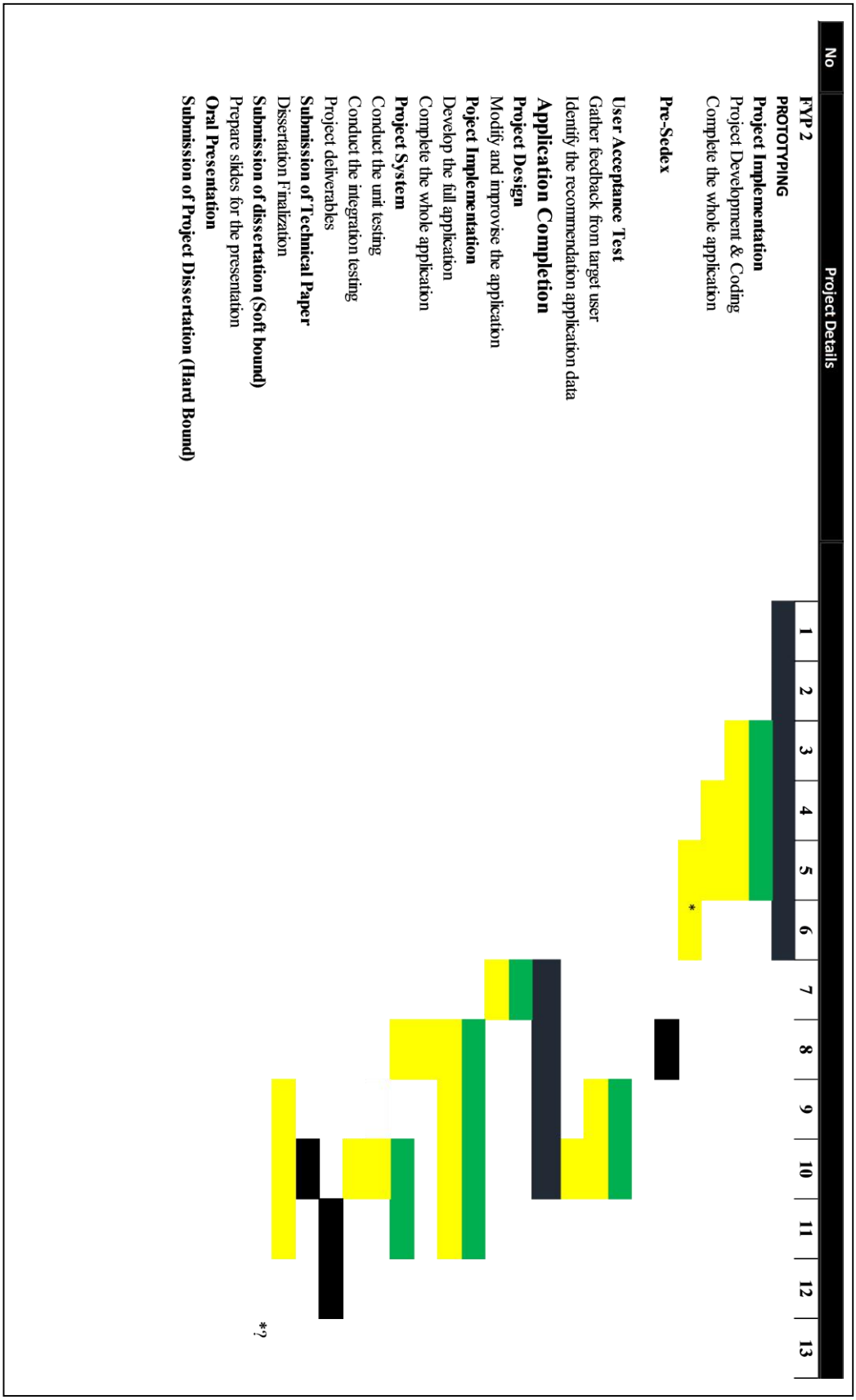

Table 3-3 Project Milestone

Milestone		Week
FYP 1		
Project Title & Topic Selection	3%	2
Project Title & Topic Acceptance	5%	3
Interim Report	40%	12
Proposal Defence	50%	14
FYP 2		
Prototyping	35%	3-9
Pre-Sedex	10%	8
Technical Paper	10%	9
Application Launching	5%	TBC
Presentation	20%	TBC
Hardbound submission	20%	TBC

3.2.2. Prototyping Phase

The prototyping phase starts with gathering all equipment, materials, and resources required. The historical data extracted from Yahoo Finance, the statistical software selected, R Programming, and models to be implemented for forecasting the stock quote. The historical price are either extracted automatically by sing one of the tool box from R, or by downloading it from the Yahoo Finance site. Furthermore, the forecasting method is designed by passively state the level of confidence of risk made by the developer.

The prototype made by combining the method use in the R Programming and the HTML Developer, which is expected to, completed at the same time. The building of the statistical algorithm will be the widely held in the phase of FYP one, while the combining method will be briefly done in FYP 2. Several attempt has been done towards find the algorithm and proper model that will be chosen. Analyzing the stock price by plotting it into time series as shown in Figure 3-2, is one of the steps to complete the project.

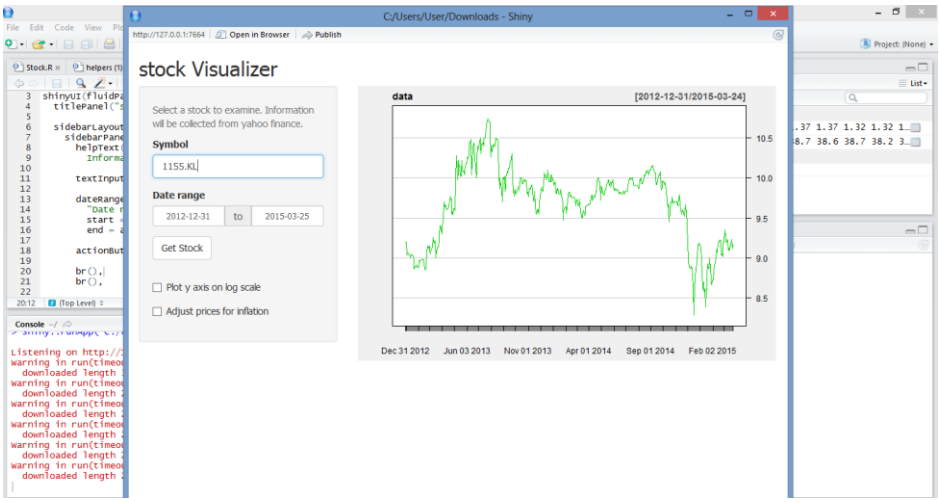

Figure 3-5 Stock Visualization

3.2.3. Testing Phase

Testing will be performed after the entire project modules have been completed and ready to test. Testing is expected to be done by the series of attempt by the experts and several students who interested to study this project. The testing will eventually result to some kind of record and feedback from the users those benefits to the form. The evaluation is also given and recorded for getting insight and feedback from the users as well as any shortcomings in requirements or project deliverables.

3.2.4. Deployment Phase

The final phase of the prototyping is to deploy the project after completing series of testing and improvement in terms of design and functional capabilities. The dashboard will later be ready to be installed and use by the users for their own investing strategy.

3.3. Phase 3: Proposed Framework

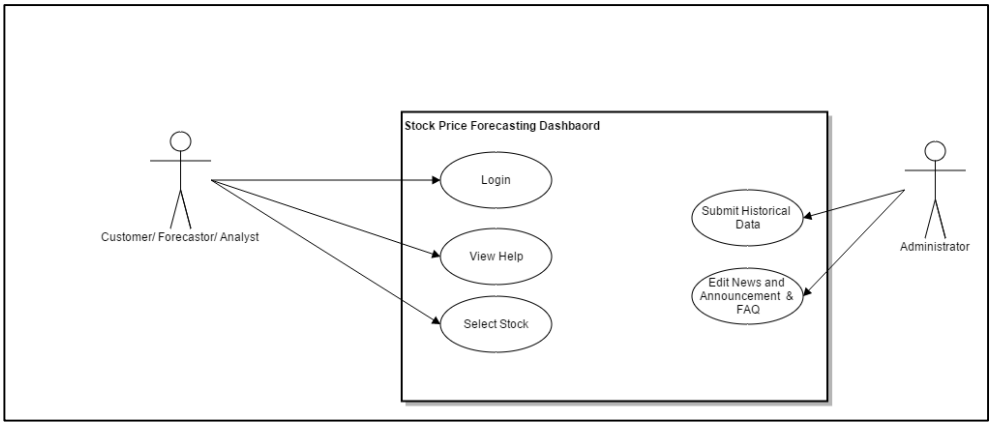

Figure 3-6 Use Case Diagram

Use Case above explained the user and admin ability towards the application. User is able to login to the application to make sure the security of the application guaranteed. View Help is the pop up window to explain the application brief and guidelines of ‘how-to’ when they open to the dashboard. The help window is also filled with the explanation of the copyright, the dashboard creator details and the contact details. Lastly, the main function of the dashboard is that user is able to select stock or write the stock quote of the market to the tet box that will be searched and forecast by the machine learning.

The administrator is the owner to the application. The admin has authority to submit the historical data to the database or submit the new list of quote and edit the help windows.

3.3.1. Flow Chart

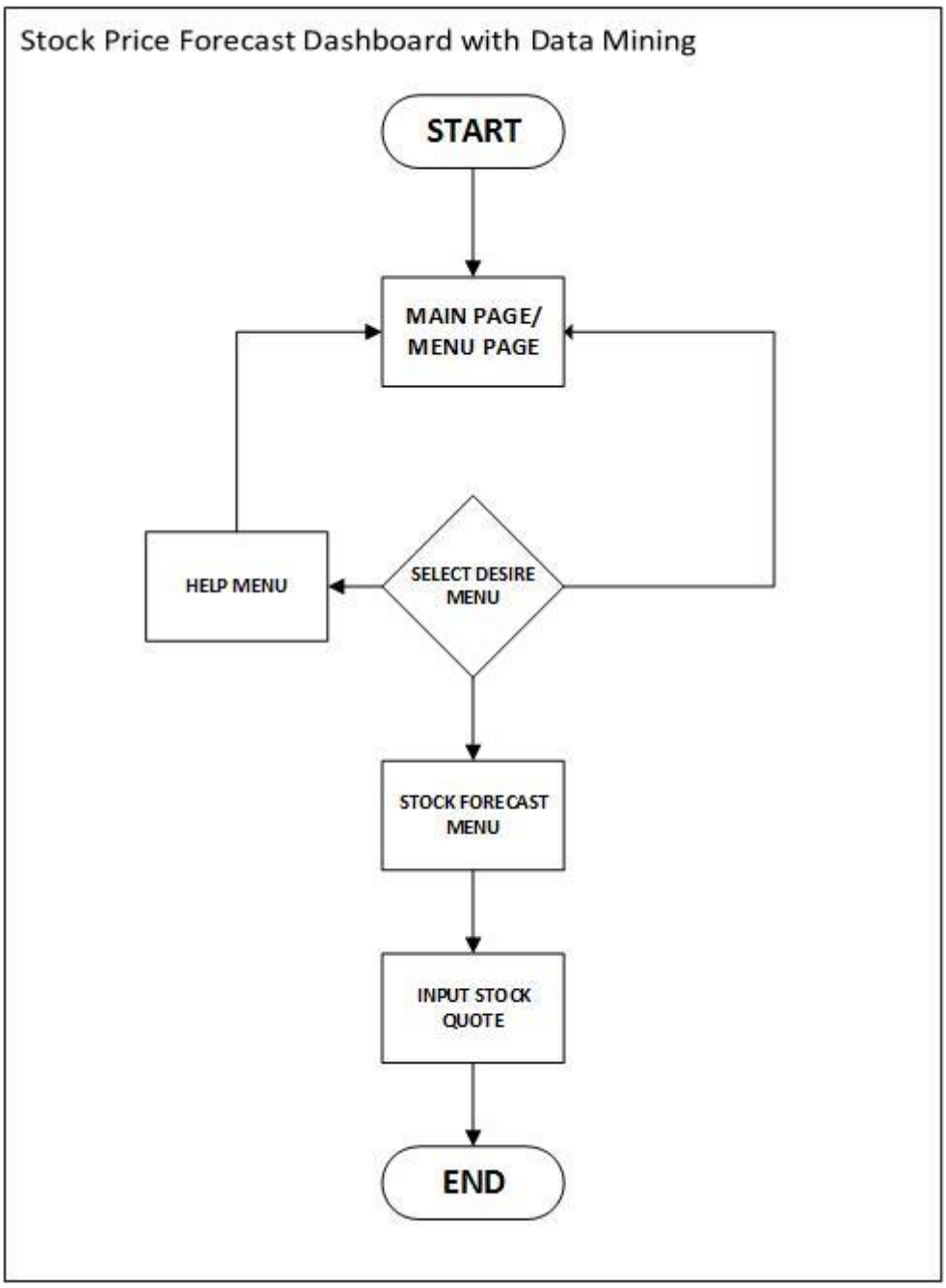

Figure 3-7 Flow Chart

3.3.2. System Architecture

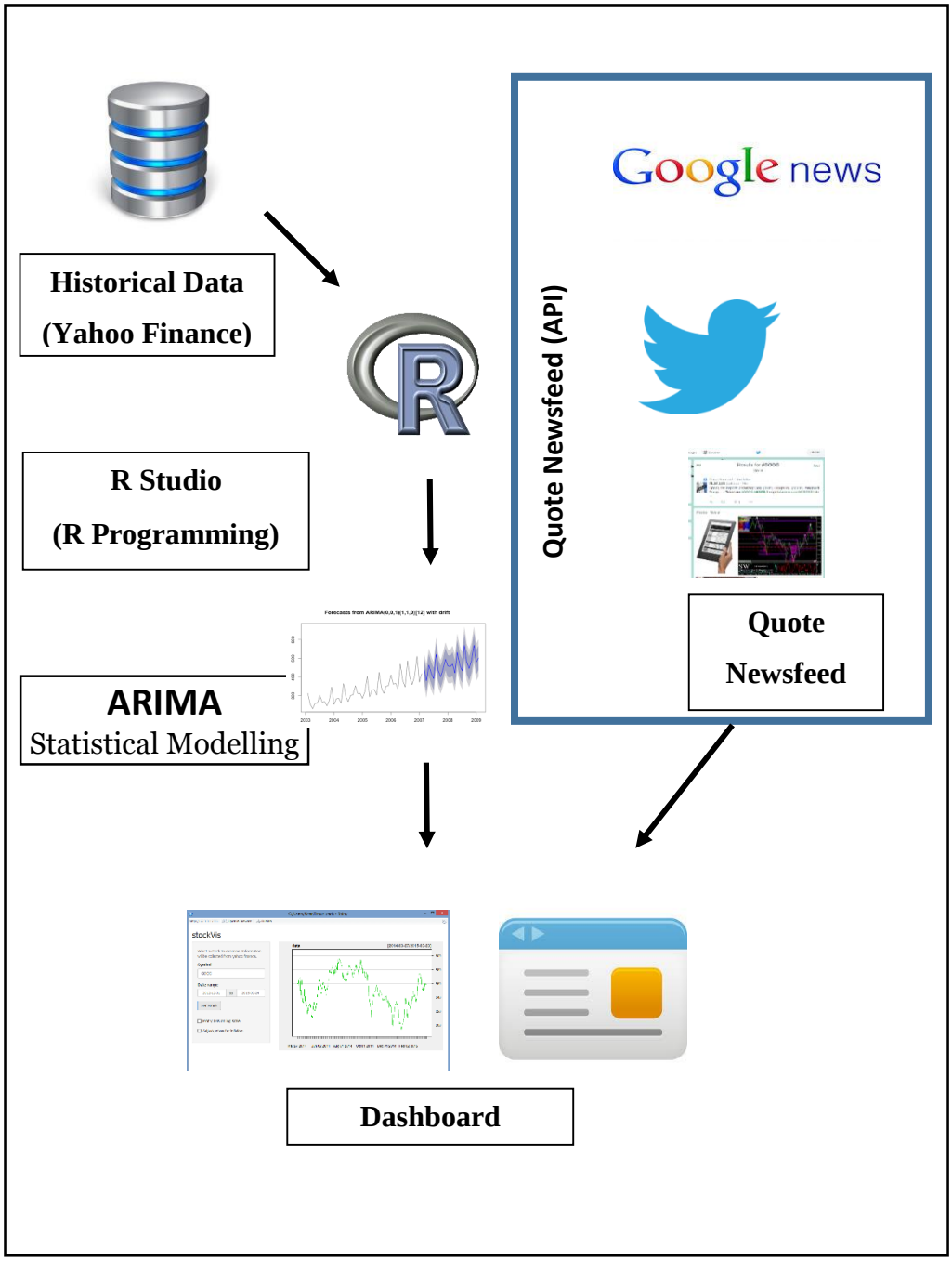

Figure 3-8 System Architecture

CHAPTER 4
RESULT AND DISCUSSION

4.1. Introduction

This chapter explained the process and result of research performed to uphold the project objectives. The method of research conducted are expert interview and experimental. An interview was conducted by interviewing one of the expert that has this study background and perform day-to-day basis on statistical occupation and depiction to R Programming. Furthermore, experimental method conducted by attempt series of experiment of stock price forecasting with model chosen to investigate and decide the best implementation for the project deliverables.

4.2. Qualitative Method: Interview

The main objective to conduct interview to the experts is to gain deep understanding on the stock price forecast and existing forecasting tools, which use by the expert or by the company. Furthermore, by interviewing the expert, several inputs and suggestion towards the project and the dashboard. Nineteen questions were delivered to the expert that consist of stock price forecasting fundamental, the forecasting model, and the dashboard.

4.2.1. Stock Price Fundamental

Several question asked to the experts to receive necessary point of information to accomplish the project. The experts conveyed her opinion that stock forecasting is an attempt to learn and analyze the movement of the price trend with either fundamental or technical or quantitatively approach. Most of the investors are using technical approach, as it is only required short time and lesser resources compare to the other.

She said that fundamental approach usually look into the balance sheet of the company or the financial statement released by the company. The company's movement of Return on Investment (ROI), management direction and economic situation can also be used to predict the company's stock performance in the future. Moreover, quantitative approach is trying to avoid fundamental factors and solely treat current and historical price as the subject of method by using statistical or other technique.

Next, the expert agrees that all forecaster need to understand the technical of the forecasting method. The expert believes that the forecaster needs to understand the model and the technique so that when they start to do some forecasting they can easily interact with the model if they want to change the parameters to find the proper and fit to their profiles without affect the whole method applied.

On the topic about who the most party benefits to stock forecasting, she explained that the person who most likely do forecasting for stock price range from equity analysts, equity investment, sales representative or broker and personal or group investor.

Onto the current issue that the broker or security companies are currently only provide technical paper of bundle of report of technical analysis which not included with from the commission to the broker or trading fees the experts believes that it should not be supposed to. She believes that the companies pay researchers and analysts' salary to analyze and forecast the stock has already come together and not need to add separate cost to consumers. Therefore, it is not justifiable to add certain of money for the investor if they want to subscribe the forecasting report.

The expert later added that there are numbers of method currently use to forecast share price, such as ARIMA, Smart Beta, econometrics, and machine learning. She said she prefers to apply ARIMA as the method to forecast stock price, if the forecaster is still new in these activities. She said ARIMA package is mostly available, easy to get, and do not require the user to do some math.

When author was trying to proposed the application that want to developed, she personally agrees and encourage the developer to also make it in subscribe basis to keep the value exist. She responded that if the dashboard is distributed to

everyone especially beginner investor, it may increase number of investor in the future, as the forecasting is become easier.

Based on the interview stated above, it can be concluded that the dashboard proposed is in line with the demand in the market. As more people are currently starting to trade in stock market, more knowledge and instant update of information is truly important. The dashboard method to utilize ARIMA as its model is also in a row to the analyst and statistician use right now, as it is ways and suitable for beginner predictors.

4.3. Experimental

Several attempts done to find the most suitable model being use for this project. Overall, most of the attempt are using ARIMA model and use necessary package that is free distributed in the application. From the perspective of forecasting ability, author has been successfully implemented ARIMA model to forecast the price.

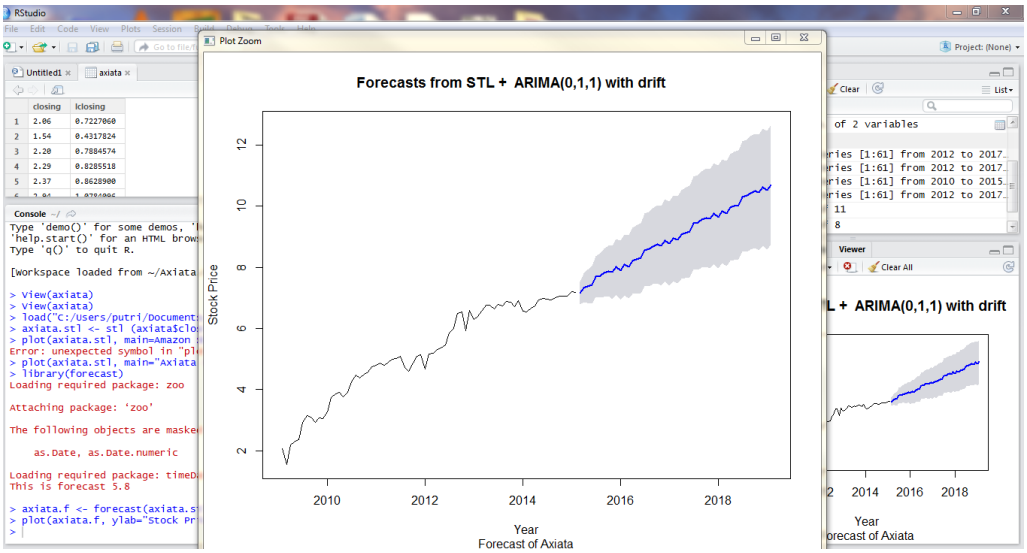

Figure 4-1 Stock Price Forecasting with ARIMA

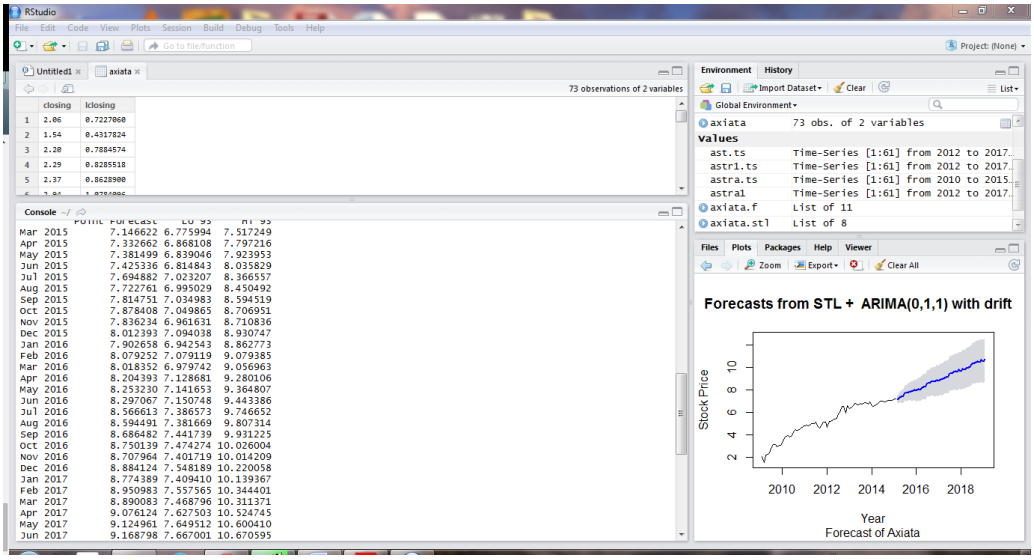

Figure 4-2 Stock Price Forecasting with ARIMA

4.4. Prototype: Stock Price Forecasting Design Dashboard

Figure - shown proposed design of the dashboard that will be developed during developing phase. The proposed design has been compare to certain of example of existing dashboard in the market. The dashboard will be using full HTML web page by combining result from R Programming and Twitter and Google News API as the objects. Twitter API will be fascinate from the technique of bootstraps that offers free-cost to developers especially in HTML platform.

These user interfaces is presented on the dashboard that being live on localhost. The application developed to forecast stock price of an index or a company in Bursa Malaysia. The main idea of this dashboard is to become an alternative for the early or ready-to-enter investor to analyze information and forecast the stock price before entering the market. There are mainly 4 activities which are "**Dashboard**" or homepage, "**Forecast**", "**Technical Analysis**", and "**Companies Profile**".

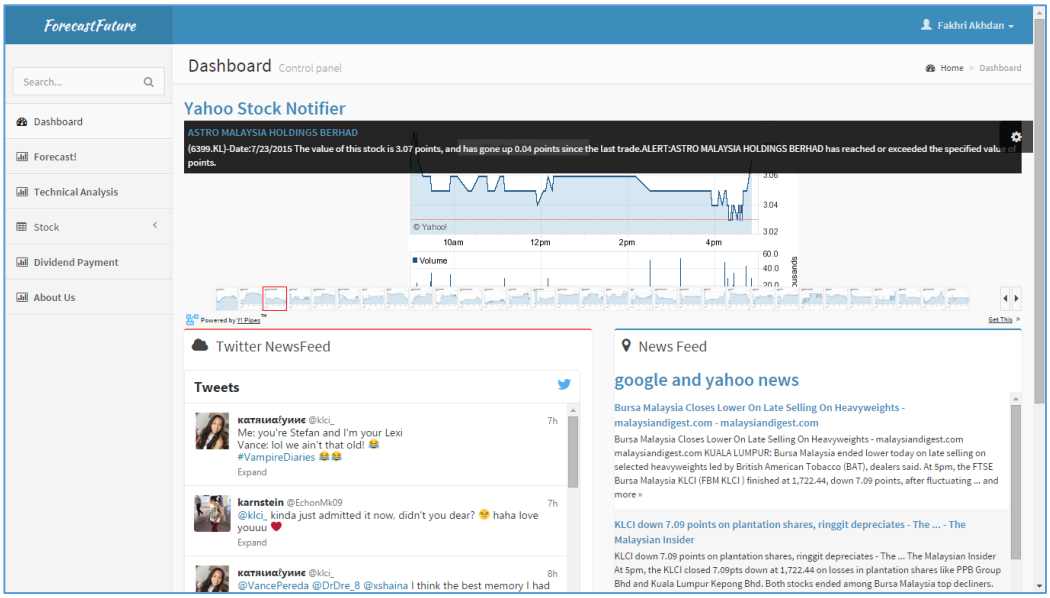

Figure 4-3 Homepage Page

Home Page or Dashboard, as in Figure 3, consists of options "Forecast!" "**Technical Analysis**", "**Stock Profile**", "**About Us**". For "**Forecast!**" button, the system will proceed to forecasting tool page. "Technical Analysis" will open to the page of technical analysis. "Stock Profile" button will show and display the stock profile by options of the tree-view. Lastly, "About Us" will proceed to the dashboard disclaimer and profile.

On the dashboard or the homepage, it is displaying a live Stock Yahoo Notifier from the Yahoo Finance of top 30 largest companies on the Bursa Malaysia by market capitalization called as Kuala Lumpur Composite Index (KLCI) (Bursa Malaysia, n.d.). Added with the twitter and google newsfeed portraying live tweet and news quote of the KLCI and market subject.

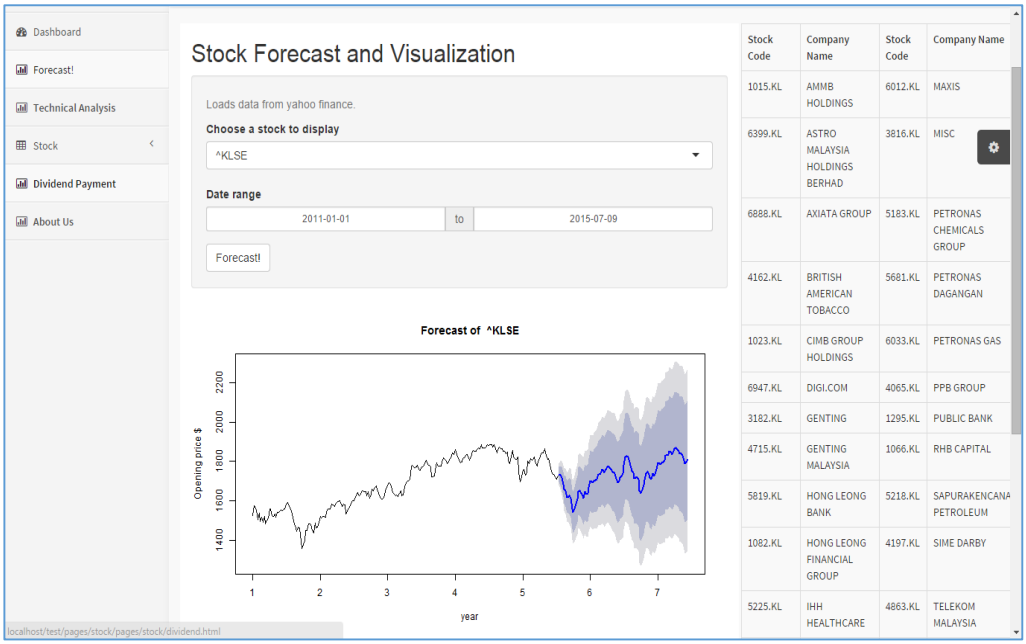

Figure 4-4 Stock Price Forecast & Visualization

At the "**Forecast!**" page, user is able to pick any of the KLCI's 30 biggest market capitalization in Bursa Malaysia by choosing the stock code in the drop down menu. The user could refer to the table on the right to find the desired company or index code to the names of the stock company. After the code has been chosen, a date of parameters can be an optional choose if the user want to analyze the historical price less or more than 5 years.

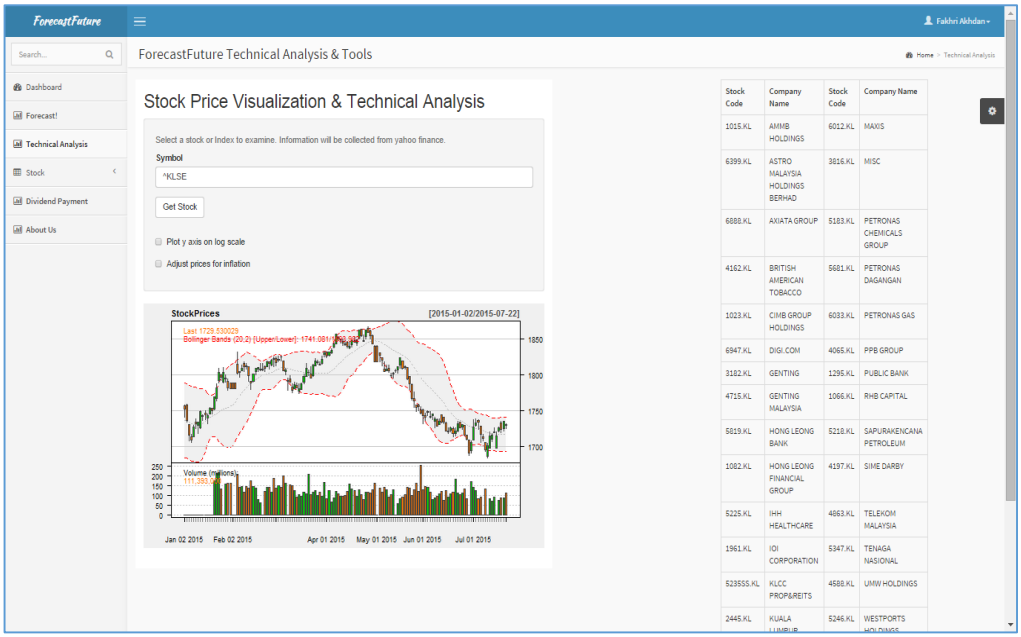

Figure 4-5 Stock Price Visualization & Technical Analysis

"**Technical Analysis**" page is able to displaying a technical analysis-using graph and chart technique that usually use by professional trader to analyze pattern and predict the price. In this page, user is not only limited to list 30 companies in KLCI on the next to the right table of information, but also enable user to analyze every stock quote that registered in Yahoo Finance. The date parameters is also changeable to analyze lesser and longer time of period.

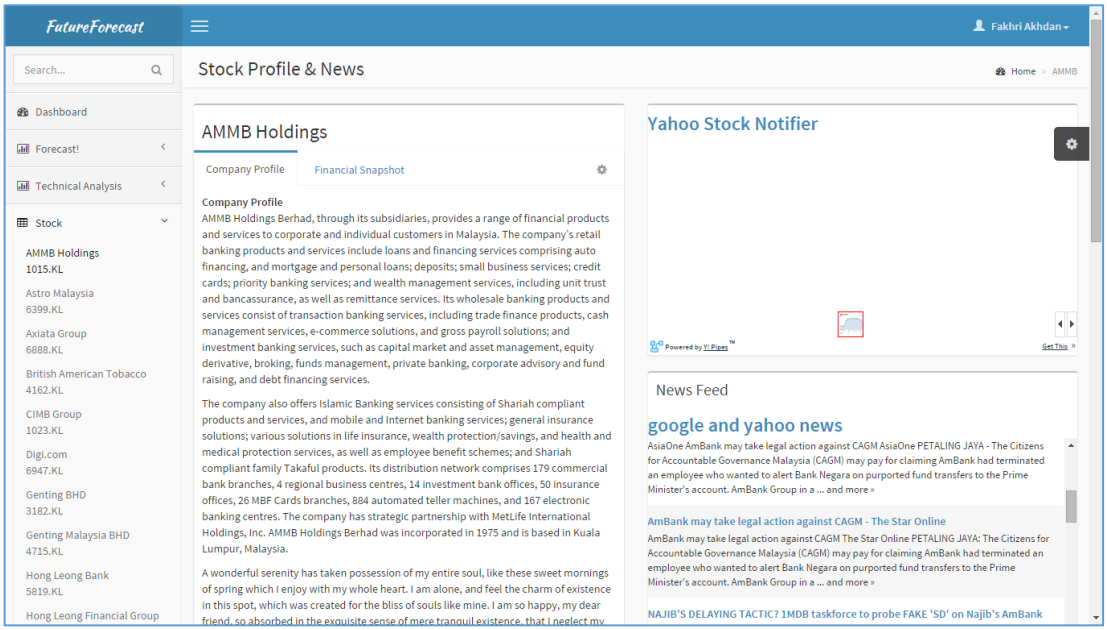

Figure 4-6 Stock Profile & News

At "**Stock Profile**" page, user is served with the stock of companies profile and financial standing of the particular company. Yahoo Stock Notifier is also added to provide an updated stock price while reading the stock profile. Under the notifier, the newsfeed from Google news and Yahoo is helping the user to analyze the summary of business action related to the company.

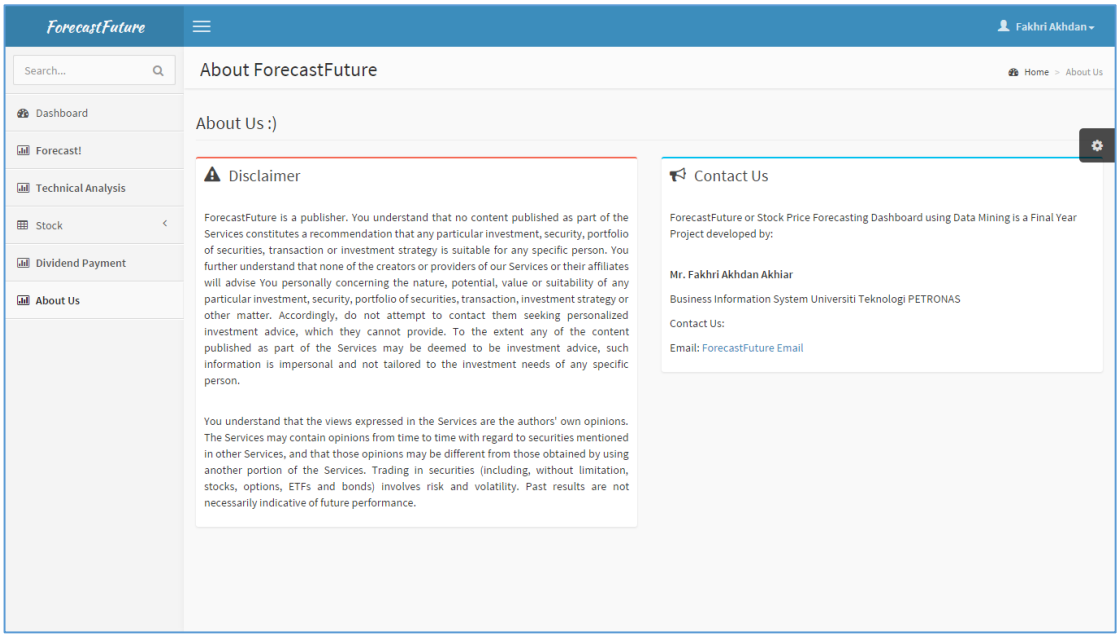

Figure 4-7 About Us

"**About Us**" page consists of the developer information by explaining the dashboard developer and the purpose of the application. The page is also explaining about the disclaimer of the dashboard of the use and all content features in the website.

4.5. Discussion on Survey Result

Set of questionnaire were distributed to 40 sample as the target user of the application. These samples are generally grouped all society level without restrict into age group, job group, and races group. The early investor or ready-to-enter investor is being generalized as a persona who has never opt in as an investor before the questionnaire distributed to them. Results of the survey are discussed in the next point under this section. The results were analyzed from the question and being categorized into category below.

4.5.1. General Knowledge on Stock Investment

Question 1: Have you ever had invested in stock market?

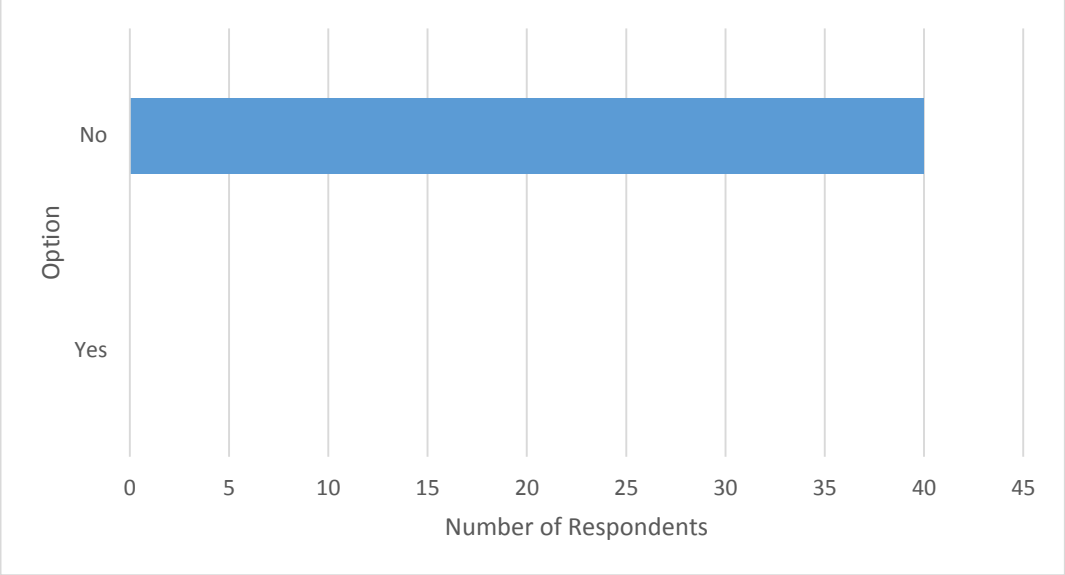

Figure 4-8 General Question on Stock Investment Experience

The result were to determine the respondent experience onto investing in stock market. This question will result to the further analysis that will be used at the next questions. Based on the graph of respondents, the number of respondents who never had invested in stock market is 40 (100%) compared to the zero (0%) of the correspondents who have ever trade in stock market. From this result, it could be said that the 40 samples are never invest in stock market despite recent trends in society.

Question 2: from these five of terms (stock trading, dividends, stock exchange, options, and KLCI), how many that you are most understand and able to explain to anyone?

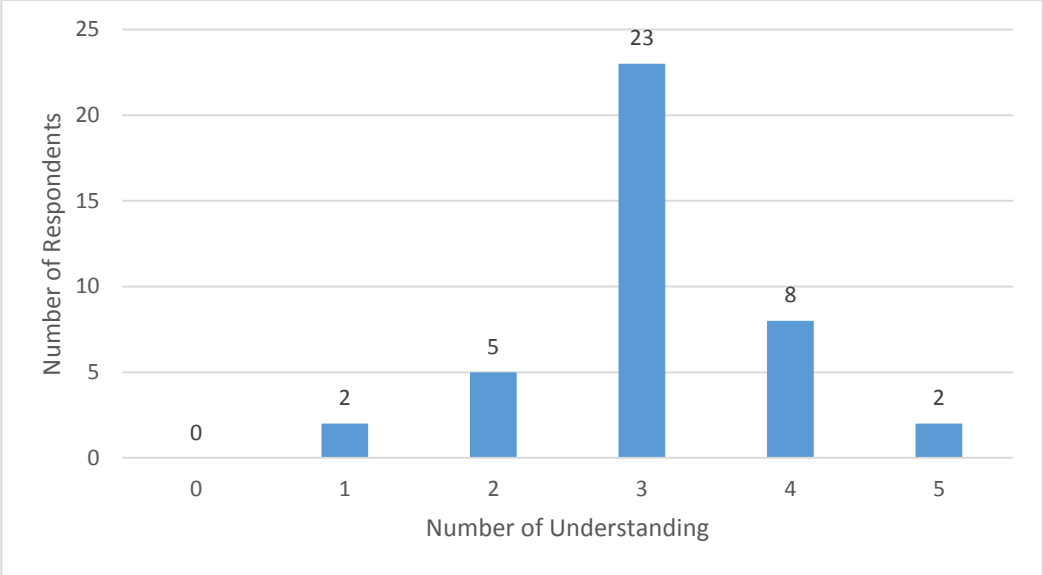

Figure 4-9 Understanding on Stock Investment Terms

The result shown in graphs that the over half of the respondents understand and able to explain about these five stock investment terms to anyone. This result shown that majority of the respondents are able to define these terms whenever they crossed and able to articulate to anybody with their own language.

4.5.2. ForecastFuture: Interest

Question 3: what is your first impression of this dashboard?

(Please circle the following option with 1 being not interested and 5 very interested)

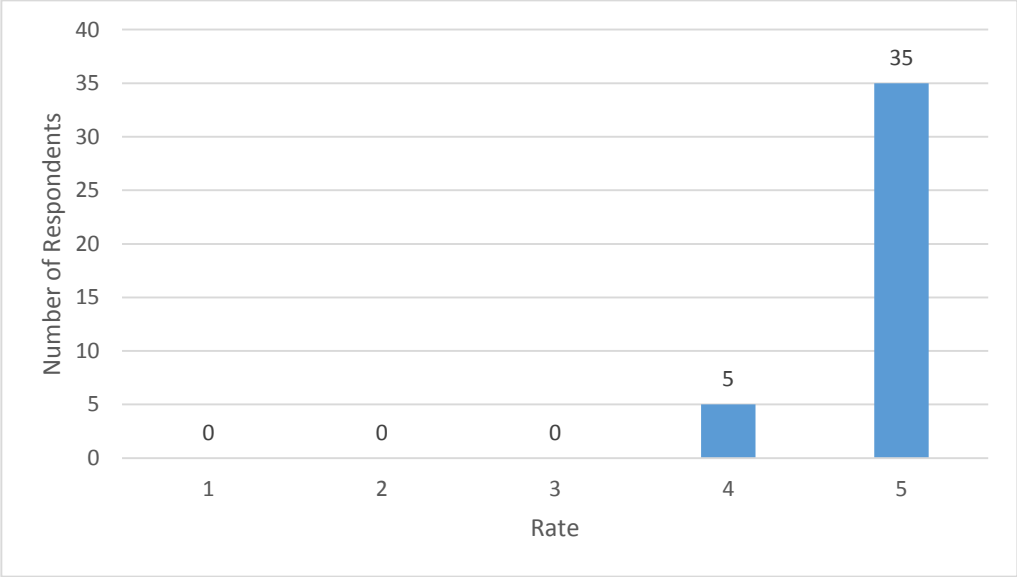

Figure 4-10 Question on ForecastFuture First Impression

The result shown in graphs that most of the respondents are interested into the application. By looking 35 (87.5%) rated very interested, ForecastFuture contents and functionalities are expected to serve the early investor for their need.

4.5.3. ForecastFuture: Usability

Question 4: Please rate the ease of navigation of the dashboard.

(Please circle the following option with 1 being very complicated and 5 very user-friendly)

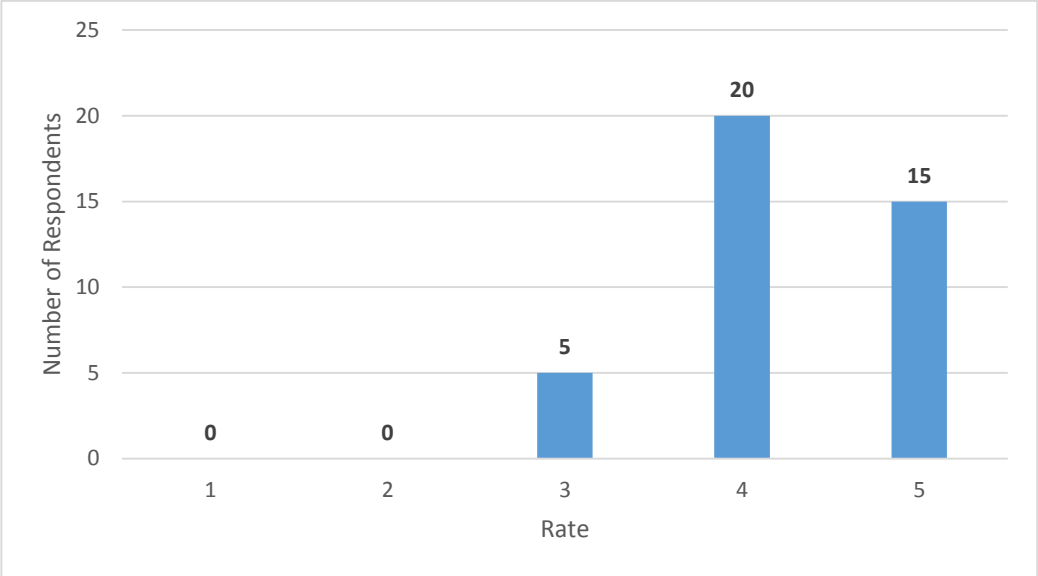

Figure 4-11 Question on application ease of navigation

Graph above suggests that majority of the responders consider ForecastFuture user friendliness by showing 35 responds (87.5%) of user-friendly given rate.

Question 5: Please rate the user interface of the dashboard

(Please circle the following option with 1 being inconvenient and 5 very intuitive)

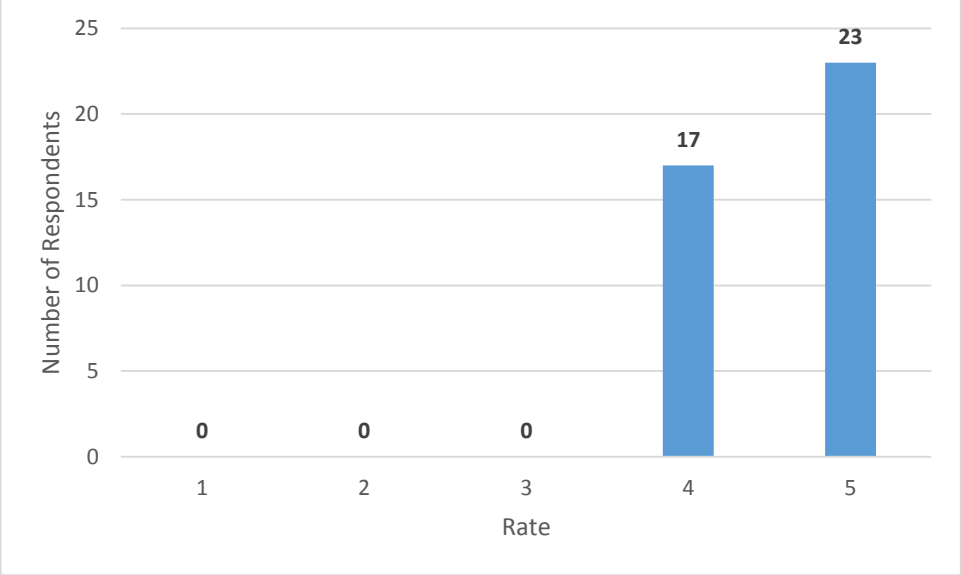

Figure 4-12 Question on application user interface

Question about user interface is being questioned to receive inputs and necessary change onto the design. Based on the graph above, the responders that the interface of the application is intuitive and interesting.

4.5.1. ForecastFuture: Content

Question 6: Please rate the accuracy level of the forecast result to the actual data.

(Please circle the following option with 1 being not acceptable and 5 very acceptable)

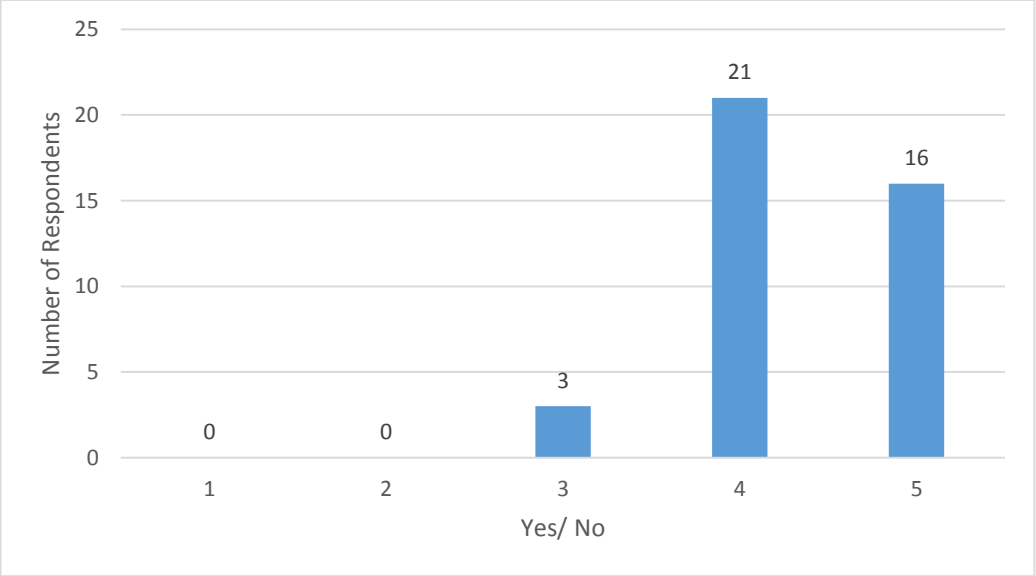

Figure 4-13 Question on forecasting accuracy level

Level of accuracy is become one of the important deliverables of the dashboard. The question is being asked to measure the accuracy and the reliability of the forecasting price with the actual data or price with certain level of error margin. And based on the graph show above, majority of the responders believe that the result of forecasting is acceptable and under the error margin tolerancy.

Question 7: What did you like the most about the application?

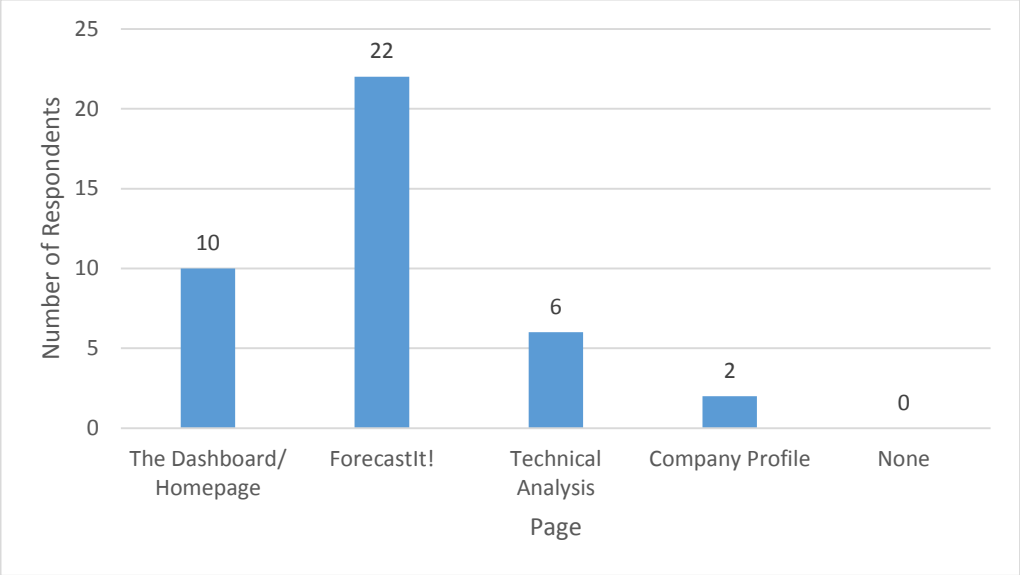

Figure 4-14 Question on application interest

The content and function offered by the application is important to maximize the value of the dashboard and satisfy the need of every user. Each of the user might end up only to use certain of content for their personal purpose. The result will become guidelines for future works to maximize the most highly interest and demand by the users. And after the survey conducted, most of the responders is highly interesting to the “Forecast!” page or the tools to forecast the stock price, followed with the home page or called as “The Dashboard” which displaying data about KLCI Index.

Question 8: Would you recommend this application to a friend/ colleagues if the function/ content were relevant to them?

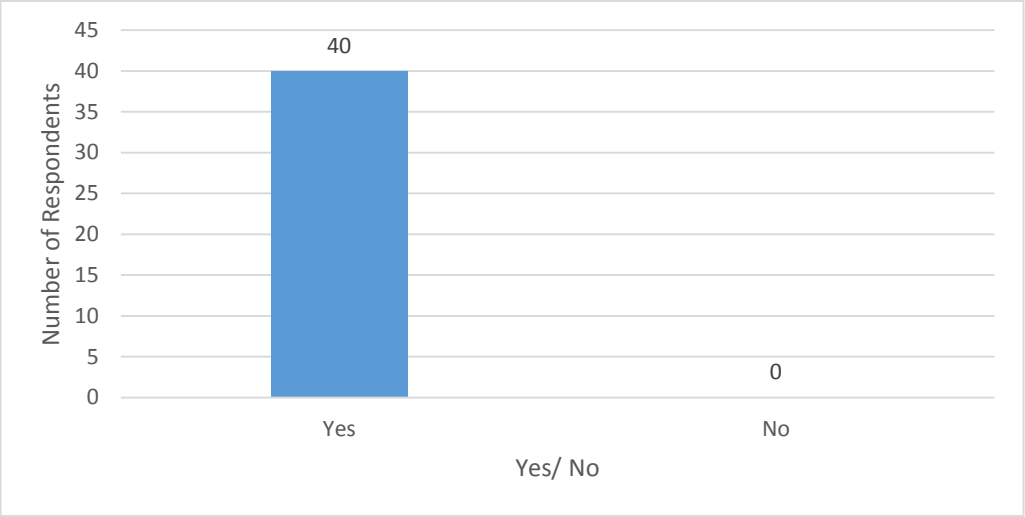

Figure 4-15 Question on application recommendation

Question 9: Will you be more confident to start to invest in stock market after using this application?

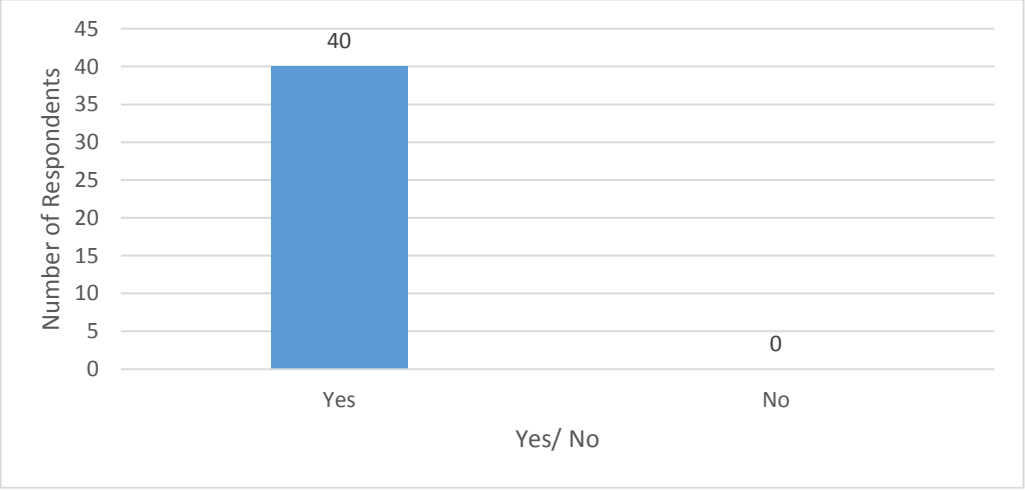

Figure 4-16 Question on application acceptance

Two graph above shown that the application is obtaining the acceptability from the the responders. Forty or 100% of the responders conveyed that they will recommend ForecastFuture if they meet friends who will be relevant to the application. This means that the sample has shown the interest over the application and would endorse the application due to the functionalities that being offered to users.

Question on investment confidence after the use of the application asked to gather information whether the application might achieve its objective and purpose. Based on the responders responds, the application might increase the confidence of the early investor to start the trade and invest in the stock market. This would mark the achieved objectives and purpose of the ForecastFuture.

Based on the survey in the activity, it can be concluded that sample of users can use and accept the application very well. Users can use and utilize all functionalities in well-off. Furthermore, the result also showed that all the users agreed that ForecastFuture is very helpful in order help them analyze, profile, and forecast the stock price. This seen on positive responds from the sample of users during the User Acceptance Test. These accomplishments is hoped to highlight the achieved and realized functions and objectives of this project.

CHAPTER 5
CONCLUSION & RECOMMENDATION

The research on this paper includes with the study of the forecasting of stock price and application development to provide a user-friendly application for early investors to analyze stock price and prediction for financial decision purposes. This dashboard aims to serve the increasing need of stock evaluation and stock market investment penetration in societies. The key findings of the study and the dashboard development that integrated stock price forecasting tool with news feed related will ultimate benefit the early investor to be prepared before decide to enter the market and help them to make a careful decision before buy or sale shares of company.

ForecastFuture is an alternative technology offers to the investor especially in Malaysia, in order to find relevant of information about the market activity and companies. Also as quickly predict the share price of a company using the pattern presents from the historical data with the data mining methodology that is proven to be effective.

This research and development can be extended some ways. For example a dashboard that also ready to be used in smart phone or tablet to ensure user-friendly and ease of use of the application and to answer the mobile application demand in the market nowadays. Furthermore, further research could also work to improve the statistical model suggested in this paper to find result that is more accurate to the user of the application

REFERENCES

Al Wadi, S., Ismail, M. T., Altaher, A. M., & Karim, S. A. A. (2010). Forecasting volatility data based on Wavelet transforms and ARIMA model. In *CSSR 2010 - 2010 International Conference on Science and Social Research* (pp. 86–90). http://doi.org/10.1109/CSSR.2010.5773909

Armano, G., Marchesi, M., & Murru, a. (2005). A hybrid genetic-neural architecture for stock indexes forecasting. *Information Sciences*, *170*, 3–33. http://doi.org/10.1016/j.ins.2003.03.023

Basak, S., Shapiro, A., & Teplá, L. (2003). Department of Finance Risk Management with Benchmarking, (December).

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2013). Time Series Analysis: Forecasting and Control. *Journal of Time Series Analysis*, *3*, 746.

Brockwell, P. J., & Davis, R. A. (1991). *Time Series: Theory and Methods*. *Technometrics* (Vol. 31). Springer. http://doi.org/10.1007/978-1-4419-0320-4

Bursa Malaysia. (n.d.). KLCI Overview. Retrieved July 24, 2015, from http://www.bursamalaysia.com/market/products-services/indices/ftse-bursa-malaysia-indices/overview/

CNBC. (2013). Traders turn to Twitter for stock tips. Retrieved March 24, 2015, from http://www.cnbc.com/id/101106590

Devi, B, U., Sundar, D., & Alli, P, D. (2013). An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50. *International Journal of Data Mining & Knowledge Management Process*, *3*(1), 65–78. http://doi.org/10.5121/ijdkp.2013.3106

Fung, W., Fung, W., Hsieh, D., Hsieh, D., ‘, & ‘. (1997). Empirical characteristics of dynamic trading strategies: the case of hedge funds. *’ Review of Financial Studies, Vol.*, *10*(2), 275–302.

Godbole, N., & Srinivasaiah, M. (2007). Large-scale sentiment analysis for news and blogs. *Conference on Weblogs and Social Media (ICWSM 2007)*, 219–222. http://doi.org/10.1177/01461079070370040501

Han, J., & Kamber, M. (2006). Data Mining: Concepts and Techniques. *Annals of Physics*, *54*, 770. http://doi.org/10.5860/CHOICE.49-3305

Herath, R. G. and S. (2007). The volatility of the stock market and news. *International Research Journal of Finance and Economics*.

Investopedia. (2015). Stock Market Definition | Investopedia. Retrieved July 24, 2015, from http://www.investopedia.com/terms/s/stockmarket.asp

Kovalerchuk, B., & Vityaev, E. (n.d.). Chapter 1: Data Mining For Financial Applications, 1–14.

McDonald, R. L. (2013). *Derivatives markets* (Third Edit). New Jersey: Pearson Education Limited. Retrieved from http://eprints.lse.ac.uk/43074/

Mittermayer, M.-A. (2004). Forecasting Intraday stock price trends with text mining techniques. *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*.

Mondal, P., Shit, L., & Goswami, S. (2014). Study Of Effectiveness of Time Series Modeling (ARIMA) in Forecasting Stock Prices. *International Journal of Computer Science, Engineering and Applications (IJCSEA)*, *4*(2), 2–43.

Moon, M. a., Mentzer, J. T., Smith, C. D., & Garver, M. S. (1998). Seven keys to better forecasting. *Business Horizons*, *41*(October), 44–52. http://doi.org/10.1016/S0007-6813(98)90077-5

Pai, P.-F., & Lin, C.-S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*. http://doi.org/10.1016/j.omega.2004.07.024

Refenes, A. N., Burgess, A. N., & Bentz, Y. (1997). Neural networks in financial engineering: a study in methodology. *IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council*, *8*, 1222–1267. http://doi.org/10.1109/72.641449

Sekar, P. S., Kannan, K. S., Sathik, M. M., & Arumugam, P. (2010). Financial stock market forecast using data mining techniques. *Proceedings of the International MultiConference of Engineers and Computer Scientists*, *I*, 5. Retrieved from http://www.iaeng.org/publication/IMECS2010/IMECS2010_pp555-559.pdf

Smart, S., Gitman, L., & Joehnk, M. (2014). *Fundamentals of Investing* (12th ed.). Harlow: Pearson Education Limited.

Sutheebanjard, P., & Premchaiswadi, W. (2010). Forecasting the Thailand Stock Market. *Exchange Organizational Behavior Teaching Journal*, *6*(2), 85–114.

Wong, F. S., Wang, P. Z., & Goh, T. H. (1991). Fuzzy neural systems for decision making. *[Proceedings] 1991 IEEE International Joint Conference on Neural Networks*. http://doi.org/10.1109/IJCNN.1991.170361

Yu, K., & Cheng, Y. (2007). Machine Learning Techniques for Stock Prediction. *Talanta*, *71*, 676–82. http://doi.org/10.1016/j.talanta.2006.05.013

Zhai, Y., Hsu, A., & Halgamuge, S. K. (2007). Daily Stock Price Trends Prediction, 1087–1096.

Appendices

Acceptance Testing Survey

Acceptance Testing Survey for "ForceastFuture"

Final Year Project

The objective of this questionnaire is to determine the fulfilment of functional requirements of the mobile application, "**ForceastFuture: Stock Price Forecasting Dashboard Using Data Mining**". With these user evaluations, future enhancement for the application will be structured.

a. Name :
b. Age :
c. Sex :
d. Occupation :

Survey Questions

Section 1: General Knowledge on Stock Investment

Q1: Have you ever had invested in stock market?

Yes No

Q2: From these five of terms (stock trading, dividends, stock exchange, options, and KLCI), how many that you are most understand and able to explain to anyone?

0 1 2 3 4 5

Section 2: Interest

Q3: what is your first impression of this dashboard?

(Please circle the following option with 1 being not interested and 5 very interested)

1 2 3 4 5

Q4: Please rate the ease of navigation of the dashboard.

(Please circle the following option with 1 being very complicated and 5 very user-friendly)

1 2 3 4 5

Q5: Please rate the user interface of the dashboard

(Please circle the following option with 1 being inconvenient and 5 very intuitive)

Section 3: Content

Q6: Please rate the accuracy level of the forecast result to the actual data.

(Please circle the following option with 1 being not acceptable and 5 very acceptable)

1 2 3 4 5

Q7: What did you like the most about the application?

The Dashboard/ Homepage	Forecast!	Technical Analysis	Company Profile	None

Q8: Would you recommend this application to a friend/ colleagues if the function/ content were relevant to them?

Yes No

Question 9: Will you be more confident to start to invest in stock market after using this application?

Yes No

Please kindly provide any comments or feedback relevant to your experience using ForecastFuture.