**元 智 大 學**

資 訊 工 程 學 系

碩 士 論 文

**Using Candlestick Chart Representation to Predict Taiwan and Indonesia Stock Market**

研 究 生：郭祿丁

指導教授：歐昱言 博士

中華民國 107 年 7 月

**Using Candlestick Chart Representation to Predict Taiwan and Indonesia Stock Market**

研 究 生 ：郭祿丁 Student : Rosdyana Mangir Irawan Kusuma

指 導 教 授 ：歐 昱 言 博士 Advisor : Dr. Yu-Yen Ou

元 智 大 學

資 訊 工 程 學 系

碩 士 論 文

A Thesis

Submitted to Department of Computer Science and Engineering

Yuan Ze University

in Partial Fulfillment of the Requirements

for the Degree of Master of Science

in

Computer Science and Engineering

July 2018

Chungli, Taiwan, Republic of China.

中華民國 107 年 7 月

Using Candlestick Chart Representation to Predict Taiwan and Indonesia Stock Market

**Student: Rosdyana Mangir Irawan Kusuma Advisor: Dr. Yu-Yen Ou**

Submitted to Department of Computer Science and Engineering

College of Informatics

Yuan Ze University

# ABSTRACT

This thesis explores predictability in the stock market using Deep Convolutional Network and candlestick charts. The outcome is to design a decision support framework that can be used by traders to provide suggested indications of future stock price direction. We perform this work using some neural network like convolutional neural network, residual network and visual geometry group network. From stock market historical data, we converted to candlestick chart for the model analyzing the pattern. Using Taiwan and Indonesian stock market historical time series data we can achieve an average 83 % for 5 trading days’ period, 87 % for 10 trading days’ period and 90 % for 20 trading days’ period of binary classification accuracy.

This thesis also addresses problems specific to learning with stock market historical time series data, model performance due to sliding windows period days and analyzing hidden pattern in candlestick chart.

Keywords: *Stock Market Prediction, Neural Network, Residual Network, Candlestick Chart*

# Acknowledgements

I would like to dedicate this thesis to Allah Subhanahu wa ta’ala, for your help through all the difficulties. And a special thanks to my wife, my family, my lab mates, all the Yuan Ze University teachers and friends that I have ever met during my study in Taiwan.

I would like to express my gratitude for the supervision of my advisor, Dr. Yu-Yen Ou who suggested an interesting research topic to me. And gave me clear direction to made this paper possible. You have been a tremendous teacher for me during my study especially for your support, guidance, corrections, encouragement and advices. I would also like to thank my thesis committee for the insightful comments and advice.

Finally, I would like to acknowledge the generous financial support of Yuan Ze University’s Department of Computer Science and Engineering. Which has illuminated my long-lasting dream of obtaining a Master Degree in Computer Science and Engineering.

# Table of Contents

[ABSTRACT iii](#_Toc518554473)

[Acknowledgements iv](#_Toc518554474)

[Table of Contents v](#_Toc518554475)

[Table of Figures vii](#_Toc518554476)

[Table of Tables viii](#_Toc518554477)

[Chapter 1 Introduction 1](#_Toc518554478)

[1.1. Background 1](#_Toc518554479)

[1.2. Related work 2](#_Toc518554480)

[Chapter 2 Data Collection 4](#_Toc518554481)

[2.1. Data Collection using Yahoo! Finance 4](#_Toc518554482)

[2.2. Feature Investigation 7](#_Toc518554483)

[2.2.1. Opening Price 8](#_Toc518554484)

[2.2.2. Closing Price 8](#_Toc518554485)

[2.2.3. Highest Price 8](#_Toc518554486)

[2.2.4. Lowest Price 9](#_Toc518554487)

[2.2.5. Volume 9](#_Toc518554488)

[2.3. Data Preprocessing 10](#_Toc518554489)

[Chapter 3 Methodology 11](#_Toc518554490)

[3.1. Chart Encoding 12](#_Toc518554491)

[3.2. Binary Classification 13](#_Toc518554492)

[3.3. Learning Algorithm 13](#_Toc518554493)

[3.3.1. Convolutional Neural Network 13](#_Toc518554494)

[3.3.2. Residual Network 14](#_Toc518554495)

[3.3.3 VGG Network 16](#_Toc518554496)

[3.3.4 Random Forest 16](#_Toc518554497)

[3.3.5 K-Nearest Neighbors 17](#_Toc518554498)

[3.4. Performance Evaluation 17](#_Toc518554499)

[Chapter 4 Experimental Results and Discussion 19](#_Toc518554500)

[4.1. Classification for Each Stock Market 19](#_Toc518554501)

[4.2. Classification for Taiwan 50 19](#_Toc518554502)

[4.3. Classification for Indonesia 10 23](#_Toc518554503)

[4.4. Real Environment Prediction 28](#_Toc518554504)

[4.5. Comparison 28](#_Toc518554505)

[Chapter 5 Conclusion and Future Works 31](#_Toc518554506)

[References 32](#_Toc518554507)

# Table of Figures

**Figure 1**- Yahoo! Finance website, provide an API to download historical time series data 7

**Figure 2** - General way to visualizing the candlestick chart 10

**Figure 3** – Our methodology design 11

**Figure 4** – Proposed candlestick chart without volume indicator with different period time and size. 12

**Figure 5** - Proposed candlestick chart with volume indicator with different period time and size. 12

**Figure 6** - Logic statement of our binary classification 13

**Figure 7** - A regular 3-layer Neural Network 14

**Figure 8** – ResNet Architecture shows many layers for different configuration 15

**Figure 9** - Residual learning: a building block 16

# Table of Tables

**Table 1** – List of 50 companies from Taiwan stock market, this group of companies called TW50. 4

**Table 2** – List of 10 companies of Indonesia stock market 6

**Table 3 -** The period time of our dataset, separated between the training and testing 6

**Table 4** – Number of dataset following their period of trading days 10

**Table 5** - Residual Network proved that this network have the most minimum error-rate 16

**Table 6** – Result of 5 period with volume indicator in 50 image dimension for Taiwan 50. 19

**Table 7** – Result of 10 period with volume indicator in 50 image dimension for Taiwan 50. 20

**Table 8** – Result of 20 period with volume indicator in 50 image dimension for Taiwan 50. 20

**Table 9** – Result of 5 period with volume indicator in 20 image dimension for Taiwan 50. 20

**Table 10** – Result of 10 period with volume indicator in 20 image dimension for Taiwan 50. 20

**Table 11** – Result of 20 period with volume indicator in 20 image dimension for Taiwan 50. 21

**Table 12** – Result of 5 period without volume indicator in 50 image dimension for Taiwan 50. 21

**Table 13** – Result of 10 period without volume indicator in 50 image dimension for Taiwan 50. 21

**Table 14** – Result of 20 period without volume indicator in 50 image dimension for Taiwan 50. 22

**Table 15** – Result of 5 period without volume indicator in 20 image dimension for Taiwan 50. 22

**Table 16** – Result of 10 period without volume indicator in 20 image dimension for Taiwan 50. 22

**Table 17** – Result of 20 period without volume indicator in 20 image dimension for Taiwan 50. 22

**Table 18** – Summary result of Taiwan 50 with their best classifier for each trading days and image dimension with volume indicator. 23

**Table 19** – Summary result of Taiwan 50 with their best classifier for each trading days and image dimension without volume indicator. 23

**Table 20** – Result of 5 period with volume indicator in 50 image dimension for Indonesia 10. 24

**Table 21** – Result of 10 period with volume indicator in 50 image dimension for Indonesia 10. 24

**Table 22** – Result of 20 period with volume indicator in 50 image dimension for Indonesia 10. 24

**Table 23** – Result of 5 period with volume indicator in 20 image dimension for Indonesia 10. 25

**Table 24** – Result of 10 period with volume indicator in 20 image dimension for Indonesia 10. 25

**Table 25** – Result of 20 period with volume indicator in 20 image dimension for Indonesia 10. 25

**Table 26** – Result of 5 period without volume indicator in 50 image dimension for Indonesia 10. 25

**Table 27** – Result of 10 period without volume indicator in 50 image dimension for Indonesia 10. 26

**Table 28** – Result of 20 period without volume indicator in 50 image dimension for Indonesia 10. 26

**Table 29** – Result of 5 period without volume indicator in 20 image dimension for Indonesia 10. 26

**Table 30** – Result of 10 period without volume indicator in 20 image dimension for Indonesia 10. 27

**Table 31** – Result of 20 period without volume indicator in 20 image dimension for Indonesia 10. 27

**Table 32** – Summary result of Indonesia 10 with their best classifier for each trading days and image dimension with volume indicator. 27

**Table 33** – Summary result of Indonesia 10 with their best classifier for each trading days and image dimension without volume indicator. 28

**Table 34** – Comparison result with Khaidem, Saha et al. 2016 for Samsung stock market. 28

**Table 35** – Comparison result with Khaidem, Saha et al. 2016 for Apple stock market. 29

**Table 36** – Comparison result with Khaidem, Saha et al. 2016 for GE stock market. 29

**Table 37** – Comparison result with Patel, Shah et al. 2015 29

**Table 38** – Comparison result with Zhang, Zhang et al. 2018 30

# Chapter 1 Introduction

## 1.1. Background

Stock market is something that cannot be separated from modern human life. Investment in stock market is a natural thing done by people around the world. They set aside their income to try their luck by investing in stock market to generate more profit. Traders are more likely to buy a stock whose value is expected to increase in the future. On the other hand, traders are likely to refrain from buying a stock whose value is expected to fall in the future. So, there is a need for accurately predicting the trends in stock market prices in order to maximize capital gain and minimize loss. This is also one of our reasons for doing research and try to make predictions into the field of stock market. Besides, research on stock market is still classified as a difficult thing to do considering the price changes in the stock market can happen suddenly. With the current technological advances, machine learning is a breakthrough in aspects of human life today. In this case we use machine learning to make predictions on the stock market. We choose several learning algorithms in machine learning to enhance our performance result using convolutional neural network, residual network, virtual geometry group network, k-nearest neighborhood and random forest.

Dataset format in machine learning can be various. Many kind of dataset format such as text sequence, image, audio, video, from 1D (one dimension) to 3D (three dimension) can be applicable for machine learning. Taken as an example for the use of images as inputs from machine learning, not only as inputs to classify an animal, items or the other thing, but also as an input to predict a condition, we take the example of Google DeepMind in their research in Alpha Go. Recently, they are successfully get a lot of attention in the research field. By using the image as their input, where the image represents a Go game board, which later this image dataset is used to predict the next step of the opponent in the Go game(Silver, Huang et al. 2016).

On the other occasion, from historical data of stock market converted into audio wavelength using deep convolutional wave net architecture can be applied to forecast the stock market movement(Borovykh, Bohte et al.).

Our proposed method in this thesis is using a candlestick chart from Taiwan and Indonesia stock market to predict the price movement. We are using three trading period time to analyze the correlation between those period time with the result. Our proposed candlestick chart will represent the sequence of time series with and without the daily volume stock data. Our experiments conduct two kind of image size for candlestick chart, 50 and 20 dimension to analyze the correlation of hidden pattern in various image size. Which later our dataset will be input for several learning algorithms, random forest and k-nearest neighborhood as traditional machine learning and CNN, residual network, VGG network as our modern machine learning. the goal is to analyze the correlation of some parameter such as period time, image size, feature set with the movement of stock market will be going up or going down in the next day.

## 1.2. Related work

Basically, there are many researchers who already done doing the research on the stock market predictions. (Schöneburg 1990) conducted a study using data from a randomly selected German stock market, then using the back-propagation method for their machine learning architecture(!!! INVALID CITATION !!! {}). To our knowledge, stock market data consist of open price data, close price data, high price data, low price data and also volume of the daily movement activity. In addition to using historical time series data from the stock market, some researchers in this field of stock market predictions began to penetrate the method of sentiment analysis to predict and analyze movements in the stock market.

(Bollen, Mao et al. 2011) used their sentiment analysis method by taking data from one of the famous microblogging site Twitter to predict the Dow Jones Industrial Average (DJIA) stock market movements. There are more studies on stock market predictions, they use the input data not only by using elements of historical time series data, but also processing the data into other different forms. (Borovykh, Bohte et al.) tried to use the deep convolutional wave net architecture method to perform analysis and prediction using data from S & P500 and CBOE.

We also found some related work using candlestick charts in their research. (Hu, Hu et al. 2017) in their research using candlestick chart to build a decision-making system in stock market investment. They use the convolutional encoder to learn the patterns contained in the candlestick chart. While (Tsai and Quan 2014) using candlestick chart in his research to be combined with seven different wavelet-based texture to analyze candlestick chart. Next, (do Prado, Ferneda et al. 2013) using candlestick chart to learn the pattern contained in Brazilian stock market by using sixteen candlestick pattern.

Traditional machine learning like Random Forest has been applied to predict the stock market with a good result. (Khaidem, Saha et al. 2016) using Random Forest combine with technical indicator such as Relative Strength Index (RSI) shown a good performance. Adding more feature set can be one of the way to enrich your dataset and enhance the result of classification. (Patel, Shah et al. 2015) using ten technical parameters from stock trading data for their input data and compare four prediction models, Artificial Neural Network (ANN), Support Vector Machine (SVM), random forest and naïve-Bayes. According to (Zhang, Zhang et al. 2018) input data not only from historical stock trading data, a financial news and users’ sentiments from social media can correlated to predict the movement in stock market.

To summarize, different from most of existing studies that only consider stock trading data, news events or sentiments in their models, our proposal is using a representation of candlestick chart images to analyze and predict the movement of stock market with a novel to compare modern and traditional neural network.

# Chapter 2 Data Collection

## 2.1. Data Collection using Yahoo! Finance

We choose two different stock market country, Taiwan and Indonesia. We collected 50 companies stock market data for Taiwan and 10 for Indonesia. Listed on table 1 is Taiwan 50 and table 2 is Indonesia 10.

**Table 1** – List of 50 companies from Taiwan stock market, this group of companies called TW50.

|  |  |  |
| --- | --- | --- |
| No | Name | Ticker |
| 1 | Advanced Semiconductor Engineering | 2311.TW |
| 2 | Advantech | 2395.TW |
| 3 | Asia Cement | 1102.TW |
| 4 | Asustek Computer Inc | 2357.TW |
| 5 | AU Optronics | 2409.TW |
| 6 | Catcher Technology | 2474.TW |
| 7 | Cathay Financial Holding | 2882.TW |
| 8 | Chang Hwa Commercial Bank | 2801.TW |
| 9 | Cheng Shin Rubber Industry | 2105.TW |
| 10 | China Development Financial Holdings | 2883.TW |
| 11 | China Life Insurance | 2628.HK |
| 12 | China Steel | 2002.TW |
| 13 | Chunghwa Telecom | 2412.TW |
| 14 | Compal Electronics | 2324.TW |
| 15 | CTBC Financial Holding | 2891.TW |
| 16 | Delta Electronics | 2308.TW |
| 17 | E.Sun Financial Holding | 2884.TW |
| 18 | Far Eastern New Century Corporation | 1402.TW |
| 19 | Far EasTone Telecommunications | 4904.TW |
| 20 | First Financial Holding | 2892.TW |
| 21 | Formosa Chemicals & Fibre | 1326.TW |
| 22 | Formosa Petrochemical | 1301.TW |
| 23 | Formosa Plastics Corp | 1301.TW |
| 24 | Foxconn Technology | 2354.TW |
| 25 | Fubon Financial Holdings | 2881.TW |
| 26 | Hon Hai Precision Industry | 2317.TW |
| 27 | Hotai Motor | 2207.TW |
| 28 | Hua Nan Financial Holdings | 2880.TW |
| 29 | Innolux | 3481.TW |
| 30 | Largan Precision | 3008.TW |
| 31 | Lite-On Technology | 2301.TW |
| 32 | MediaTek | 2454.TW |
| 33 | Mega Financial Holding | 2886.TW |
| 34 | Nan Ya Plastics | 1303.TW |
| 35 | Nanya Technology | 2408.TW |
| 36 | Pegatron | 4938.TW |
| 37 | Pou Chen | 9904.TW |
| 38 | President Chain Store | 2912.TW |
| 39 | Quanta Computer | 2382.TW |
| 40 | Siliconware Precision Industries | 2325.TW |
| 41 | SinoPac Financial Holdings Co. Ltd. | 2890.TW |
| 42 | Taishin Financial Holdings | 2887.TW |
| 43 | Taiwan Cement | 1101.TW |
| 44 | Taiwan Cooperative Financial Holding | 5880.TW |
| 45 | Taiwan High Speed Rail | 2633.TW |
| 46 | Taiwan Mobile | 3045.TW |
| 47 | Taiwan Semiconductor Manufacturing | 2330.TW |
| 48 | Uni-president Enterprises | 1216.TW |
| 49 | United Microelectronics | 2303.TW |
| 50 | Yuanta Financial Holding | 2885.TW |

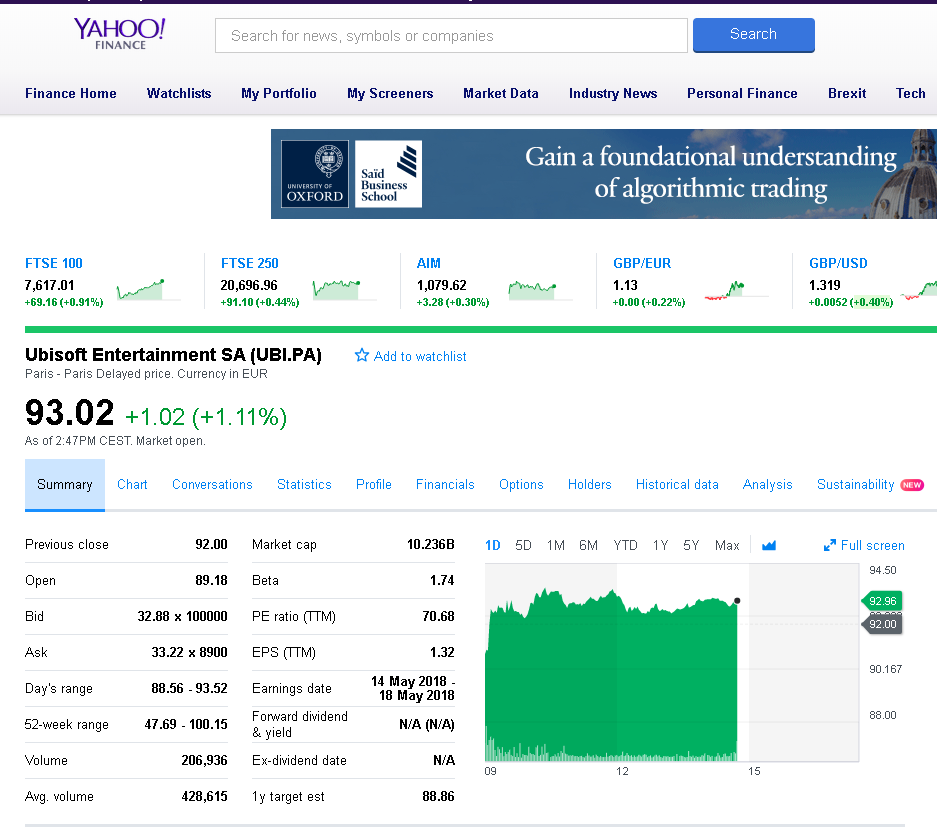
**Table 2** – List of 10 companies of Indonesia stock market

|  |  |  |
| --- | --- | --- |
| No | Name | Ticker |
| 1 | Perusahaan Perseroan (Persero) PT Telekomunikasi Indonesia Tbk | TLKM.JK |
| 2 | PT Bank Central Asia Tbk | BBCA.JK |
| 3 | PT Bank Central Asia Tbk | HMSP.JK |
| 4 | PT Bank Rakyat Indonesia (Persero) Tbk | BBRI.JK |
| 5 | PT Bank Rakyat Indonesia (Persero) Tbk | ASII.JK |
| 6 | PT Bank Mandiri (Persero) Tbk | BMRI.JK |
| 7 | PT Unilever Indonesia Tbk | UNVR.JK |
| 8 | PT Gudang Garam Tbk | GGRM.JK |
| 9 | PT Bank Negara Indonesia (Persero) Tbk | BBNI.JK |
| 10 | PT United Tractors Tbk | UNTR.JK |

In this data collection we are using application program interface (API) service from Yahoo! Finance to get historical time series data for each stock market that we are using for the dataset (Figure 1). From the time period that we have been set in the following table 3, we certainly get some trading days data, starting from Monday until Friday is the trading days period.

**Table 3 -** The period time of our dataset, separated between the training and testing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| STOCK DATA | TRAINING DATA | | TESTING DATA | |
| TW 50 | 2000/01/01 | 2016/12/31 | 2017/01/01 | 2018/06/14 |
| ID 10 | 2000/01/01 | 2016/12/31 | 2017/01/01 | 2018/06/14 |



**Figure 1**- Yahoo! Finance website, provide an API to download historical time series data

Segregation of data based on predetermined time for data training and data testing is important, where some studies make mistakes by scrambling data, this is certainly fatal because the data we use is time-series.

## 2.2. Feature Investigation

The data we get from the data collection is in the form of time-series data. This time series of data contains several elements in daily stock market activity. Some of these elements are open price, close price, high price, low price and volume. The element can be said as feature set, where we will describe one by one feature set.

### 2.2.1. Opening Price

Opening price is the first price in daily activity of the stock market that is noted when the stock market is open in the specified time period. In this case for example Taiwan stock market and Indonesian stock market will open at 09:00.

The price of the first trade for any listed stock is its daily opening price. There are several day-trading strategies based on the opening of a market. Gap Fade and Fill. "Traders attempt to profit from the price correction that usually takes place subsequent to a sizable price gap at the opening. Another popular strategy is to fade a stock at the open that is showing strong pre-market indication contrary to the rest of the market, or to similar stocks in a common sector or index.

### 2.2.2. Closing Price

The closing price is the final price at which a security is traded on a given trading day. The closing price represent the most up-to-date valuation of a security until trading commences on the next trading day. Taiwan stock market close precisely at 3:00 p.m. while Indonesian stock market will be close precisely at 4:00 p.m.

Closing prices do not reflect corporate actions, which may skew returns significantly. For example, on June 9, 2014, Apple Inc. (NASDAQ: AAPL) issued a seven-for-one stock split. Therefore, Apple's shares outstanding was increased by a multiple of seven, while its closing share price was divided by seven. On June 6, 2014, prior to Apple's stock split, it had a closing price of $ 645.57 per share. After Apple's seven-for-one stock split, the stock had a closing price of $ 93.70 per share on June 9, 2014. Since the closing price does not include adjustments for corporate actions, the calculation of Apple's returns based on closing prices would have indicated a return of -85.49%, or ($ 93.70 - $ 645.56) / $ 645.57, in just one trading day.

### 2.2.3. Highest Price

High price or today’s high is the highest price at which a stock traded during the course of the day. Today’s high is typically higher than the closing or opening price. More often than not this is higher than the closing price.

When you look at a stock quote, you can find today's high by looking at the second number listed next to "Range." One way that day traders and technical analysts use today's high, along with today's low, is to help them identify gaps or sudden jumps up or down in a stock's price with no trading in between those two prices. For example, if today's low is $25 and the previous day's high is $20, there is gap. The identification of a gap, along with other market signals such as changes in trading volume and overall bullish or bearish sentiment, helps market analysts generate buy and sell signals for particular stocks.

### 2.2.4. Lowest Price

Today’s low or low price is the lowest price at which a stock trades over the course of a trading day. Today’s low is typically lower than the opening or closing price.

When you look at a stock quote, you can find today's low by looking at the first number listed next to "Range." Today's low and today's high are important to day traders and technical analysts, who seek to earn profits from a security's short-term price movements and identify and track trends. One way that day traders use today's low along with today's high is to identify gaps, or sudden jumps up or down in a stock's price with no trading in between. Gaps are used in technical analysis to identify directional movement, average true range/price volatility, candlestick patterns and more. Traders then analyze these patterns to determine profitable entry and exit points.

### 2.2.5. Volume

Volume is the number of shares or contracts traded in a security or an entire market during a given period of time.

For every buyer, there is a seller, and each transaction contributes to the count of total volume. That is, when buyers and sellers agree to make a transaction at a certain price, it is considered one transaction. If only five transactions occur in a day, the volume for the day is five.

Volume is an important indicator in technical analysis as it is used to measure the relative worth of a market move. If the markets make a strong price movement, then the strength of that movement depends on the volume for that period. The higher the volume during the price move, the more significant the move.

## 2.3. Data Preprocessing

We are processing our time series data using library Matplotlib(Hunter 2007) in python programming to convert from the historical data that we have prepared into a candlestick chart. We divide the time period which used to create candlestick chart based on 5 trading days’ data, 10 trading days’ data and also 20 trading days’ data.

The amount of data (table 4) can be different number because we will only generate a candlestick chart that qualify based on the time period that we have set in the following table 3.

**Table 4** – Number of dataset following their period of trading days

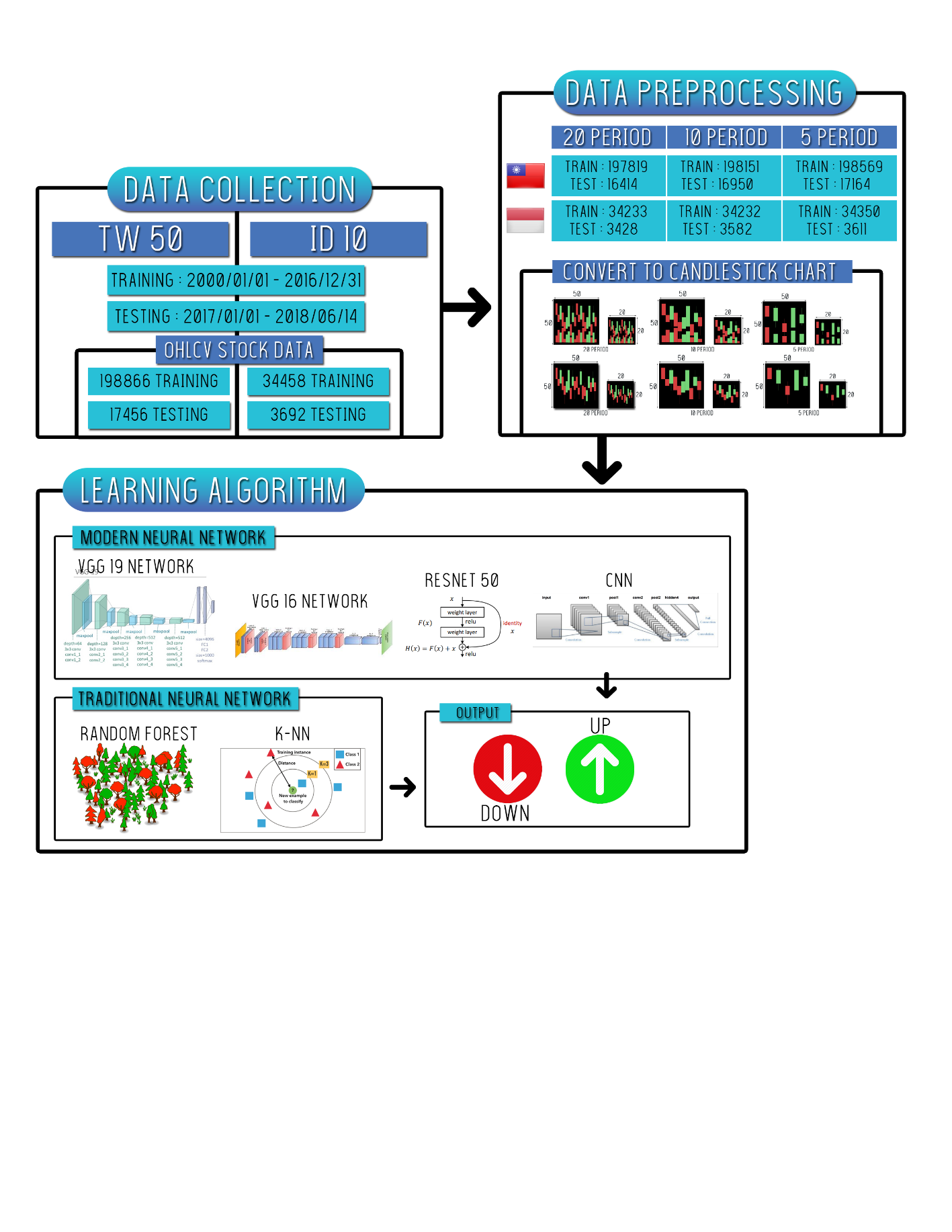
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 5 PERIOD | | 10 PERIOD | | 20 PERIOD | |
| STOCK DATA | TRAINING | TESTING | TRAINING | TESTING | TRAINING | TESTING |
| TW 50 | 198569 | 17164 | 198151 | 16950 | 197819 | 16414 |
| ID 10 | 34350 | 3611 | 34232 | 3582 | 34233 | 3482 |

Besides the period time, we also divided our candlestick chart with and without volume indicator. The general candlestick chart usually only consists of time series data such as open price, close price, low price and high price (figure.2). Adding a volume indicator into candlestick chart is one of our parameter to find out correlation between enrich candlestick chart information and prediction result.



**Figure 2** - General way to visualizing the candlestick chart

# Chapter 3 Methodology



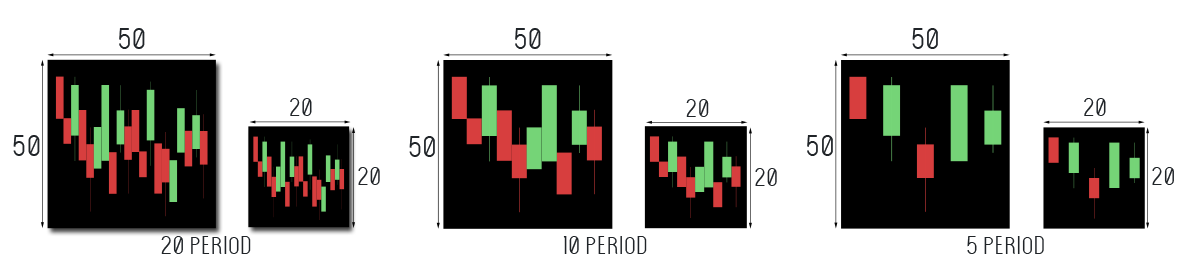
**Figure 3** – Our methodology design

Those figure below (Figure 3), shows about the design architecture of our proposed stock market prediction. The process is started with collecting stock market historical data using Yahoo! Finance API and then using sliding window to generate the period data before using computer graphic technique to generate the candlestick chart images.

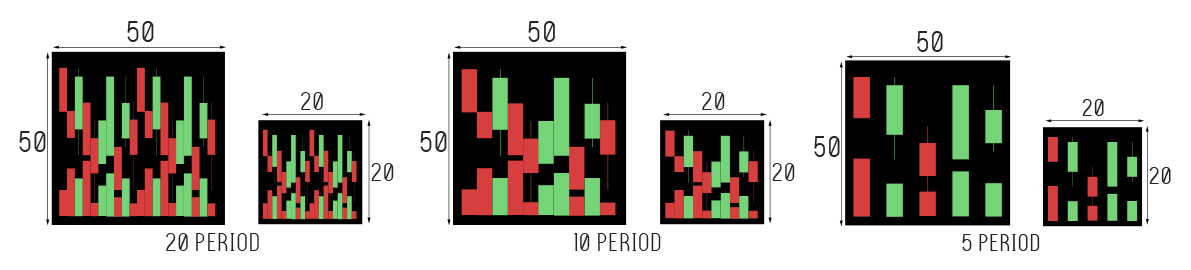
## 3.1. Chart Encoding

Candlestick chart show that emotion by visually representing the size of price moves with different colors. Traders use the candlestick chart to make trading decision based on regularly occurring patterns that help forecast the short-term direction of the price. Just like a bar chart, a candlestick chart shows the stock market’s open price, high, low, and close price during those period time. The candlestick chart has a wide part, which is called body, the body represents the price range between the open and close of that day’s trading. When the body is filled in red, it means the close was lower than the open price. If the body is filled in green, it means the close was higher than the open price.

We use computer graphics techniques with the help of a python library called Matplotlib(Hunter 2007) to convert this time series data into a candlestick image size as 50x50 and 20x20 dimension with RGB(Red Green Blue) channel.

****Figure 4 describe our candlestick chart representation with volume whether figure 5 describe our candlestick chart without volume.

**Figure 4** – Proposed candlestick chart without volume indicator with different period time and size.

**Figure 5** - Proposed candlestick chart with volume indicator with different period time and size.

We are using black color as our candlestick chart background, and then for each candlestick chart we are using green color for indicator if its closing price is higher than open price, and red color if its closing price is lower than opening price.

## 3.2. Binary Classification

Our goal here is to perform binary classification of stock market movements by analyzing and find the hidden pattern inside candlestick chart. Where we will use the value of "1" as an indicator that there is an increase or closing price will rise in the future with t = n + 1, while "0" is the opposite, if the closing price will decrease in the future. The diagram for making this label is available in figure 6.



**Figure 6** - Logic statement of our binary classification

## 3.3. Learning Algorithm

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

### 3.3.1. Convolutional Neural Network

In figure 7, Neural Networks receive an input (a single vector), and transform it through a series of hidden layers. Each hidden layer is made up of a set of neurons, where each neuron is fully connected to all neurons in the previous layer, and where neurons in a single layer function completely independently and do not share any connections. The last fully-connected layer is called the “output layer” and in classification settings it represents the class scores.



**Figure 7** - A regular 3-layer Neural Network

### 3.3.2. Residual Network

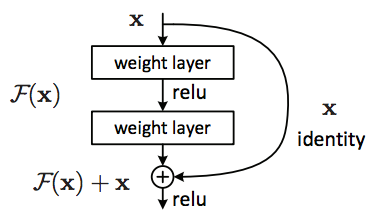
Developed by (He, Zhang et al. 2016) was the winner of ILSVRC 2015. It features special skip connections and a heavy use of batch normalization. The architecture is also missing fully connected layers at the end of the network.



**Figure 8** – ResNet Architecture shows many layers for different configuration

As per what we have seen so far, increasing the depth should increase the accuracy of the network, as long as overfitting is taken care of. But the problem with increased depth is that the signal required to change the weights, which arises from the end of the network by comparing ground-truth and prediction becomes very small at the earlier layers, because of increased depth. It essentially means that earlier layers are almost negligible learned. This is called vanishing gradient.

The second problem with training the deeper networks is, performing the optimization on huge parameter space and therefore naively adding the layers leading to higher training error. Residual networks allow training of such deep networks by constructing the network through modules called residual models as shown in the figure 8. This is called degradation problem.



**Figure 9** - Residual learning: a building block

By comparing with other CNNs architecture, residual network has been proving with the most minimum error rate according to table 3.

**Table 5** - Residual Network proved that this network have the most minimum error-rate

|  |  |  |  |
| --- | --- | --- | --- |
| Year | CNN | Developed by | Top-5-error-rate |
| 2012 | AlexNet | (Krizhevsky, Sutskever et al. 2012) | 15.3 % |
| 2013 | ZFNet | (Zeiler and Fergus 2014) | 14.8 % |
| 2014 | GoogLeNet | (Szegedy, Ioffe et al. 2017) | 6.67 % |
| 2014 | VGG Net | (Simonyan and Zisserman 2014) | 7.3 % |
| 2015 | ResNet | (He, Zhang et al. 2016) | 3.57 % |

### 3.3.3 VGG Network

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

### 3.3.4 Random Forest

Random Forest classifier is a classifier with Consist of many decision trees and adopted the technique of random decision forest prioritizes predictive performance by using multiple learning algorithms (ensemble learning). In general, Decision trees are a learning methods used in data search technique. The method used by the idea of combining the "bagging" idea or called "Bootstrap Aggregating" (reduce variance) and the random selection of features in the training sets (classification and regression tree).

We applied our random forest algorithm from a machine learning python library called skicit-learn(Pedregosa, Varoquaux et al. 2011).

### 3.3.5 K-Nearest Neighbors

K-Nearest Neighbors (KNN) is a classifier with based on the Lazy learning and Instance-based (IBk) learning algorithms (selection K based value based on model evaluation method or cross validation). Further, Lazy learning is a learning method with the purposed to store training data and enables the training data is used when there is a query request is made (waits until it is given a test) by the system. Similarity measure applied to the KNN with the aim to compare every new case with available cases (training data) that has been previously saved. Conversely different with eager learning, eager learning is a learning method with the intention of preparing training process data earlier, then wait for the query request (a test). KNN implemented lazy learning method which has the distinct advantage that it can solve the problem by comparing the problem with similar past problem (case-based reasoning). KNN adopted a supervised learning approach by utilizing the data in this case must have class/label and this learning model of the algorithm can be used for classification and regression predictive problems.

We also using skicit-learn python library for our KNN classifier. Furthermore, we used a K-D Tree algorithm in our KNN to perform prediction with default parameter from scikit-learn library.

## 3.4. Performance Evaluation

There are some statistical measure of the performance evaluation to evaluate the result of all the classifiers (binary classification test) by measuring the sensitivity (true positive rate or recall), specificity (true negative rate), accuracy and Matthew's correlation coefficient (MCC)**.** In general, TP is true positive or correctly identified, FP is false positive or incorrectly identified, TN is true negative or correctly rejected and FN is false negative or incorrectly rejected. Formulated as follows:

Sensitivity is called true positive rate or recall measures the performance of positives data are correctly identified.

Otherwise, to measure the proposition of negative rate the specificity formula is used during the prediction result and performance all classifiers.

The accuracy formula measures the quality all classifiers with based on the true value or maximum predicted values compared with measurement results.

Then, Matthews’s correlation coefficient or MCC is used to predict binary (two class) classifications and focus on the quality of predicted binary. During the prediction results MCC returns a value between -1 and +1. If the correlation value closer to +1 indicates perfect prediction, and otherwise if the correlation value closer to -1 indicates total disagreement between prediction and observation.

# Chapter 4 Experimental Results and Discussion

## 4.1. Data Analysis

During our experiments, we are using combination of stock market data for Taiwan and Indonesia. We have Taiwan 50 and Indonesia 10 for both data list. We tried to mapping our data into some statistical analysis, this step will help us to understand more about our data.

## 4.2. Classification for Each Stock Market

In this study we try to make stock market predictions by using binary classification. Where the value 1 on the label means price increasing on the next day, while the value 0 is the reverse of it.

We try to apply this binary classification to each stock market data we have prepared. There are 50 sources for Taiwan and 10 sources for Indonesia of stock market data that we use. We also divide the retrieval period based on the duration of the 5 days, 10 days and also 20 days of trading days to create a sequence of sliding windows that will be converted to candlestick chart. Another hands, we also generate the candlestick chart with and without volume indicator and with 2 different image size, 50 and 20 dimension. Here is a table of results from binary classification predictions that we do with several machine learning algorithms.

### 4.2.1 Classification for Taiwan 50

Our first experiment is to compare several classifiers using Taiwan 50 dataset with different trading days and candlestick chart image dimension. Table 6, 7, and 8 show our result of Taiwan 50 in different trading days’ period by 50 dimension of candlestick chart with combination of volume price indicator. Table 7 shown that CNN aim better than the other classifier with 0.915 accuracy in 10 trading days’ period by 50 dimension of candlestick chart with combination of volume price indicator.

**Table 6** – Result of 5 period with volume indicator in 50 image dimension for Taiwan 50.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 7090 | 1763 | 6927 | 1384 | 0.797 | 0.837 | 0.817 | 0.634 |
| Resnet50 | 6792 | 1792 | 6898 | 1682 | 0.794 | 0.802 | 0.798 | 0.595 |
| VGG16 | 6841 | 1625 | 7065 | 1633 | 0.813 | 0.807 | 0.81 | 0.62 |
| VGG19 | 6718 | 1742 | 6948 | 1756 | 0.8 | 0.793 | 0.796 | 0.592 |
| CNN | 7100 | 1458 | 7232 | 1374 | 0.832 | 0.838 | 0.835 | 0.67 |

**Table 7** – Result of 10 period with volume indicator in 50 image dimension for Taiwan 50.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 6678 | 1428 | 7801 | 1043 | 0.845 | 0.865 | 0.854 | 0.708 |
| Resnet50 | 6299 | 837 | 8524 | 790 | 0.911 | 0.889 | 0.901 | 0.799 |
| VGG16 | 6298 | 887 | 8474 | 791 | 0.905 | 0.888 | 0.898 | 0.792 |
| VGG19 | 6345 | 874 | 8487 | 744 | 0.907 | 0.895 | 0.902 | 0.8 |
| CNN | 6469 | 785 | 8576 | 620 | 0.916 | 0.913 | 0.915 | 0.827 |

**Table 8** – Result of 20 period with volume indicator in 50 image dimension for Taiwan 50.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 6354 | 1013 | 8348 | 735 | 0.892 | 0.896 | 0.894 | 0.785 |
| Resnet50 | 6792 | 1792 | 6898 | 1682 | 0.794 | 0.802 | 0.798 | 0.595 |
| VGG16 | 6841 | 1625 | 7065 | 1633 | 0.813 | 0.807 | 0.81 | 0.62 |
| VGG19 | 6718 | 1742 | 6948 | 1756 | 0.8 | 0.793 | 0.796 | 0.592 |
| CNN | 7100 | 1458 | 7232 | 1374 | 0.832 | 0.838 | 0.835 | 0.67 |

Table 9, 10, and 11 shown our result for Taiwan 50 with 20 image dimension. Table 11 conclude that 20 image dimension and volume indicator is aim better result than the other with 0.906 accuracy.

**Table 9** – Result of 5 period with volume indicator in 20 image dimension for Taiwan 50.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 7160 | 1692 | 6998 | 1314 | 0.805 | 0.845 | 0.825 | 0.651 |
| KNN | 6095 | 2251 | 6439 | 2379 | 0.741 | 0.719 | 0.73 | 0.46 |
| Resnet50 | 6792 | 1603 | 7087 | 1682 | 0.816 | 0.802 | 0.809 | 0.617 |
| VGG16 | 6755 | 1561 | 7129 | 1719 | 0.82 | 0.797 | 0.809 | 0.618 |
| VGG19 | 6750 | 1611 | 7079 | 1724 | 0.815 | 0.797 | 0.806 | 0.611 |
| CNN | 7005 | 1396 | 7294 | 1469 | 0.839 | 0.827 | 0.833 | 0.666 |

**Table 10** – Result of 10 period with volume indicator in 20 image dimension for Taiwan 50.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 6810 | 1200 | 8004 | 900 | 0.87 | 0.883 | 0.876 | 0.751 |
| KNN | 5902 | 2123 | 7106 | 1819 | 0.77 | 0.764 | 0.767 | 0.533 |
| Resnet50 | 6638 | 1192 | 8012 | 1072 | 0.87 | 0.861 | 0.806 | 0.731 |
| VGG16 | 6657 | 1275 | 7929 | 1053 | 0.861 | 0.863 | 0.862 | 0.723 |
| VGG19 | 6440 | 1025 | 8179 | 1270 | 0.889 | 0.835 | 0.864 | 0.726 |
| CNN | 6781 | 1242 | 7987 | 940 | 0.865 | 0.878 | 0.871 | 0.742 |

**Table 11** – Result of 20 period with volume indicator in 20 image dimension for Taiwan 50.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 6390 | 870 | 8471 | 683 | 0.907 | 0.903 | 0.905 | 0.808 |
| KNN | 5199 | 1849 | 7512 | 1890 | 0.802 | 0.733 | 0.733 | 0.536 |
| Resnet50 | 6232 | 757 | 8584 | 841 | 0.919 | 0.881 | 0.903 | 0.801 |
| VGG16 | 6315 | 806 | 8535 | 758 | 0.914 | 0.893 | 0.905 | 0.806 |
| VGG19 | 6280 | 813 | 8528 | 793 | 0.913 | 0.888 | 0.902 | 0.801 |
| CNN | 6397 | 860 | 8501 | 692 | 0.908 | 0.902 | 0.906 | 0.808 |

As we said in our experiments method about find correlation between with or without volume indicator to enhance our result. Table 12, 13, 14 are our result for Taiwan 50 in 5, 10, and 20 trading days’ period by 50 image dimension. Table 14 shown that with 20 periods can aim better than the other with 0.922 accuracy.

**Table 12** – Result of 5 period without volume indicator in 50 image dimension for Taiwan 50.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 7217 | 1650 | 7040 | 1257 | 0.81 | 0.852 | 0.831 | 0.662 |
| Resnet50 | 6643 | 1612 | 7078 | 1831 | 0.814 | 0.784 | 0.799 | 0.599 |
| VGG16 | 6772 | 1585 | 7105 | 1702 | 0.818 | 0.799 | 0.808 | 0.617 |
| VGG19 | 6737 | 1577 | 7113 | 1737 | 0.819 | 0.795 | 0.807 | 0.614 |
| CNN | 7213 | 1423 | 7267 | 1261 | 0.836 | 0.851 | 0.844 | 0.687 |

**Table 13** – Result of 10 period without volume indicator in 50 image dimension for Taiwan 50.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 6828 | 1198 | 8006 | 882 | 0.87 | 0.886 | 0.886 | 0.753 |
| Resnet50 | 6535 | 1132 | 8072 | 1175 | 0.877 | 0.848 | 0.864 | 0.725 |
| VGG16 | 6660 | 1197 | 8007 | 1050 | 0.87 | 0.864 | 0.867 | 0.733 |
| VGG19 | 6311 | 1002 | 8202 | 1399 | 0.891 | 0.819 | 0.858 | 0.713 |
| CNN | 6794 | 994 | 8210 | 916 | 0.892 | 0.881 | 0.887 | 0.773 |

**Table 14** – Result of 20 period without volume indicator in 50 image dimension for Taiwan 50.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 6403 | 897 | 8444 | 670 | 0.904 | 0.905 | 0.905 | 0.806 |
| Resnet50 | 6348 | 908 | 8433 | 725 | 0.903 | 0.807 | 0.901 | 0.798 |
| VGG16 | 6361 | 799 | 8542 | 712 | 0.914 | 0.899 | 0.908 | 0.813 |
| VGG19 | 6316 | 798 | 8543 | 757 | 0.915 | 0.893 | 0.905 | 0.807 |
| CNN | 6415 | 629 | 8712 | 658 | 0.933 | 0.907 | 0.922 | 0.84 |

Table 15, 16, and 17 are result of Taiwan 50 in 5, 10, and 20 trading days’ period without volume indicator by 20 image dimension. Both of VGG19 and CNN in table 17 shown good performance with 0.91 accuracy.

**Table 15** – Result of 5 period without volume indicator in 20 image dimension for Taiwan 50.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 7227 | 1649 | 7041 | 1247 | 0.81 | 0.853 | 0.831 | 0.663 |
| KNN | 6793 | 1502 | 7188 | 1681 | 0.827 | 0.802 | 0.815 | 0.629 |
| Resnet50 | 7116 | 1685 | 7005 | 1358 | 0.806 | 0.84 | 0.823 | 0.646 |
| VGG16 | 6937 | 1719 | 6971 | 1537 | 0.802 | 0.819 | 0.81 | 0.621 |
| VGG19 | 6733 | 1727 | 6963 | 1741 | 0.801 | 0.795 | 0.798 | 0.596 |
| CNN | 7032 | 1321 | 7369 | 1442 | 0.848 | 0.83 | 0.839 | 0.678 |

**Table 16** – Result of 10 period without volume indicator in 20 image dimension for Taiwan 50.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 6825 | 1227 | 7977 | 885 | 0.867 | 0.885 | 0.875 | 0.75 |
| KNN | 6449 | 1182 | 8022 | 1261 | 0.872 | 0.836 | 0.856 | 0.709 |
| Resnet50 | 6557 | 1179 | 8025 | 1153 | 0.872 | 0.85 | 0.862 | 0.722 |
| VGG16 | 6711 | 1277 | 7927 | 999 | 0.861 | 0.87 | 0.865 | 0.73 |
| VGG19 | 6484 | 1142 | 8062 | 1226 | 0.876 | 0.841 | 0.86 | 0.718 |
| CNN | 6801 | 1105 | 8099 | 909 | 0.88 | 0.882 | 0.881 | 0.761 |

**Table 17** – Result of 20 period without volume indicator in 20 image dimension for Taiwan 50.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 6377 | 871 | 8470 | 696 | 0.907 | 0.902 | 0.905 | 0.806 |
| KNN | 6060 | 905 | 8436 | 1013 | 0.903 | 0.857 | 0.883 | 0.761 |
| Resnet50 | 6181 | 771 | 8570 | 892 | 0.917 | 0.874 | 0.899 | 0.793 |
| VGG16 | 6212 | 754 | 8587 | 861 | 0.919 | 0.878 | 0.902 | 0.799 |
| VGG19 | 6366 | 772 | 8569 | 707 | 0.917 | 0.9 | 0.91 | 0.817 |
| CNN | 6402 | 805 | 8536 | 671 | 0.817 | 0.914 | 0.91 | 0.817 |

From all experiments about Taiwan 50, we conclude a summary result with and without volume indicator for different trading days’ period and image dimension result. Table 18 shown that CNN in 20 trading days’ period with 50 dimension and volume indicator is better than the other with 0.915 accuracy. And without volume indicator for Taiwan 50, CNN in 20 trading days’ period with 50 dimension also perform better than the other with 0.922 accuracy. Both of those experiment, we learnt that without volume indicator and using longer trading day’s period can achieve a good result in Taiwan 50.

**Table 18** – Summary result of Taiwan 50 with their best classifier for each trading days and image dimension with volume indicator.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Classifier | Period | Dimension | Sensitivity | Specificity | Accuracy | MCC |
| CNN | 5 | 50 | 0.832 | 0.838 | 0.835 | 0.67 |
| CNN | 10 | 50 | 0.886 | 0.873 | 0.88 | 0.758 |
| CNN | 20 | 50 | 0.916 | 0.913 | 0.915 | 0.827 |
| CNN | 5 | 20 | 0.839 | 0.827 | 0.833 | 0.666 |
| Random Forest | 10 | 20 | 0.87 | 0.883 | 0.876 | 0.751 |
| CNN | 20 | 20 | 0.908 | 0.902 | 0.906 | 0.808 |

**Table 19** – Summary result of Taiwan 50 with their best classifier for each trading days and image dimension without volume indicator.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Classifier | Period | Dimension | Sensitivity | Specificity | Accuracy | MCC |
| CNN | 5 | 50 | 0.836 | 0.851 | 0.844 | 0.687 |
| CNN | 10 | 50 | 0.892 | 0.881 | 0.887 | 0.773 |
| CNN | 20 | 50 | 0.933 | 0.907 | 0.922 | 0.84 |
| CNN | 5 | 20 | 0.848 | 0.83 | 0.839 | 0.678 |
| CNN | 10 | 20 | 0.88 | 0.882 | 0.881 | 0.761 |
| CNN | 20 | 20 | 0.817 | 0.914 | 0.91 | 0.817 |

### 4.2.2 Classification for Indonesia 10

Our next experiment is performing our proposal method in Indonesia stock market. Indonesia is a promising country with good growth for their gross domestic product(Wongbangpo and Sharma 2002). Table 20, 21, 22 are our result for Indonesia 10 with volume indicator by 50 image dimension. CNN with 20 trading days’ period aim better result than the other with 0.861 accuracy.

**Table 20** – Result of 5 period with volume indicator in 50 image dimension for Indonesia 10.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 1554 | 445 | 1334 | 278 | 0.75 | 0.848 | 0.8 | 0.602 |
| KNN | 1290 | 370 | 1306 | 301 | 0.779 | 0.811 | 0.779 | 0.59 |
| Resnet50 | 1564 | 344 | 1435 | 268 | 0.807 | 0.854 | 0.831 | 0.661 |
| VGG16 | 1447 | 307 | 1472 | 385 | 0.827 | 0.79 | 0.808 | 0.617 |
| VGG19 | 1477 | 366 | 1413 | 355 | 0.794 | 0.806 | 0.8 | 0.601 |
| CNN | 1575 | 389 | 2390 | 257 | 0.781 | 0.86 | 0.821 | 0.643 |

**Table 21** – Result of 10 period with volume indicator in 50 image dimension for Indonesia 10.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 1450 | 394 | 1511 | 227 | 0.793 | 0.865 | 0.827 | 0.657 |
| KNN | 1158 | 362 | 1543 | 519 | 0.81 | 0.691 | 0.754 | 0.505 |
| Resnet50 | 1483 | 217 | 1688 | 194 | 0.886 | 0.884 | 0.885 | 0.77 |
| VGG16 | 1404 | 268 | 1637 | 273 | 0.859 | 0.837 | 0.849 | 0.697 |
| VGG19 | 1424 | 271 | 1634 | 253 | 0.858 | 0.849 | 0.854 | 0.706 |
| CNN | 1423 | 245 | 1660 | 254 | 0.871 | 0.849 | 0.861 | 0.72 |

**Table 22** – Result of 20 period with volume indicator in 50 image dimension for Indonesia 10.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 1370 | 289 | 1652 | 171 | 0.851 | 0.889 | 0.868 | 0.736 |
| KNN | 1061 | 213 | 1728 | 480 | 0.89 | 0.689 | 0.801 | 0.597 |
| Resnet50 | 1364 | 212 | 1729 | 177 | 0.891 | 0.885 | 0.888 | 0.774 |
| VGG16 | 1346 | 194 | 1747 | 195 | 0.9 | 0.873 | 0.888 | 0.774 |
| VGG19 | 1335 | 194 | 17147 | 206 | 0.9 | 0.866 | 0.885 | 0.767 |
| CNN | 1388 | 195 | 1746 | 153 | 0.9 | 0.901 | 0.9 | 0.798 |

Table 23, 24, 25 are our result for Indonesia 10 with volume indicator in 20 image dimension. from those table result, table CNN with 20 trading days’ period achieve better result with 0.871 accuracy.

**Table 23** – Result of 5 period with volume indicator in 20 image dimension for Indonesia 10.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 1541 | 471 | 1308 | 291 | 0.735 | 0.841 | 0.789 | 0.58 |
| KNN | 1297 | 368 | 1308 | 294 | 0.78 | 0.815 | 0.797 | 0.596 |
| Resnet50 | 1508 | 377 | 1402 | 324 | 0.788 | 0.823 | 0.806 | 0.612 |
| VGG16 | 1403 | 361 | 1418 | 429 | 0.797 | 0.766 | 0.781 | 0.563 |
| VGG19 | 1416 | 382 | 1397 | 416 | 0.785 | 0.773 | 0.779 | 0.558 |
| CNN | 1563 | 455 | 1324 | 269 | 0.744 | 0.853 | 0.8 | 0.602 |

**Table 24** – Result of 10 period with volume indicator in 20 image dimension for Indonesia 10.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 1445 | 404 | 1501 | 232 | 0.788 | 0.862 | 0.822 | 0.649 |
| KNN | 1139 | 363 | 1542 | 538 | 0.809 | 0.679 | 0.748 | 0.494 |
| Resnet50 | 1330 | 335 | 1570 | 347 | 0.824 | 0.793 | 0.81 | 0.618 |
| VGG16 | 1357 | 298 | 1607 | 320 | 0.844 | 0.809 | 0.827 | 0.653 |
| VGG19 | 1336 | 291 | 1614 | 341 | 0.847 | 0.797 | 0.824 | 0.645 |
| CNN | 1432 | 318 | 1587 | 245 | 0.833 | 0.854 | 0.843 | 0.686 |

**Table 25** – Result of 20 period with volume indicator in 20 image dimension for Indonesia 10.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 1341 | 287 | 1654 | 200 | 0.852 | 0.87 | 0.86 | 0.719 |
| KNN | 1014 | 232 | 1709 | 527 | 0.88 | 0.658 | 0.782 | 0.558 |
| Resnet50 | 1328 | 318 | 1623 | 213 | 0.836 | 0.862 | 0.848 | 0.694 |
| VGG16 | 1360 | 285 | 1656 | 181 | 0.853 | 0.883 | 0.866 | 0.732 |
| VGG19 | 1315 | 291 | 1650 | 226 | 0.85 | 0.853 | 0.852 | 0.701 |
| CNN | 1303 | 211 | 1730 | 238 | 0.891 | 0.846 | 0.871 | 0.738 |

The result without volume indicator for Indonesia 10 with50 image dimension shown in table 26, 27, 28. The better result achieved by CNN in 20 trading days’ period with 0.921 accuracy.

**Table 26** – Result of 5 period without volume indicator in 50 image dimension for Indonesia 10.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 1358 | 426 | 1250 | 233 | 0.746 | 0.854 | 0.798 | 0.602 |
| KNN | 1290 | 370 | 1306 | 301 | 0.779 | 0.811 | 0.795 | 0.59 |
| Resnet50 | 1398 | 351 | 1325 | 193 | 0.791 | 0.879 | 0.833 | 0.671 |
| VGG16 | 1278 | 320 | 1356 | 313 | 0.809 | 0.803 | 0.806 | 0.612 |
| VGG19 | 1274 | 338 | 1338 | 317 | 0.798 | 0.801 | 0.8 | 0.599 |
| CNN | 1362 | 342 | 1334 | 229 | 0.796 | 0.856 | 0.825 | 0.625 |

**Table 27** – Result of 10 period without volume indicator in 50 image dimension for Indonesia 10.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 1248 | 345 | 1456 | 173 | 0.808 | 0.878 | 0.839 | 0.682 |
| KNN | 1153 | 270 | 1531 | 268 | 0.85 | 0.811 | 0.833 | 0.661 |
| Resnet50 | 1241 | 344 | 1457 | 180 | 0.809 | 0.873 | 0.837 | 0.678 |
| VGG16 | 1183 | 205 | 1596 | 238 | 0.886 | 0.833 | 0.863 | 0.721 |
| VGG19 | 1191 | 241 | 1560 | 230 | 0.866 | 0.838 | 0.854 | 0.704 |
| CNN | 1231 | 225 | 1576 | 190 | 0.875 | 0.866 | 0.871 | 0.74 |

**Table 28** – Result of 20 period without volume indicator in 50 image dimension for Indonesia 10.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 1245 | 256 | 1839 | 140 | 0.878 | 0.899 | 0.886 | 0.768 |
| KNN | 1165 | 220 | 1875 | 220 | 0.895 | 0.841 | 0.874 | 0.736 |
| Resnet50 | 1206 | 144 | 1951 | 179 | 0.931 | 0.871 | 0.907 | 0.806 |
| VGG16 | 1267 | 184 | 1911 | 118 | 0.912 | 0.915 | 0.913 | 0.821 |
| VGG19 | 1255 | 169 | 1926 | 130 | 0.919 | 0.906 | 0.914 | 0.822 |
| CNN | 1276 | 165 | 1930 | 109 | 0.921 | 0.921 | 0.921 | 0.837 |

Table 29, 30, 31 are our result for 20 image dimension without volume indicator in Indonesia 10. Table 31 shown that VGG16 with 0.907 accuracy is better with the other result.

**Table 29** – Result of 5 period without volume indicator in 20 image dimension for Indonesia 10.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 1365 | 402 | 1274 | 226 | 0.76 | 0.858 | 0.808 | 0.62 |
| KNN | 1297 | 368 | 1308 | 294 | 0.78 | 0.815 | 0.797 | 0.596 |
| Resnet50 | 1234 | 338 | 1338 | 348 | 0.798 | 0.781 | 0.79 | 0.58 |
| VGG16 | 1328 | 410 | 1266 | 263 | 0.755 | 0.835 | 0.794 | 0.591 |
| VGG19 | 1280 | 335 | 1341 | 311 | 0.8 | 0.805 | 0.802 | 0.604 |
| CNN | 1311 | 279 | 1397 | 380 | 0.834 | 0.824 | 0.829 | 0.658 |

**Table 30** – Result of 10 period without volume indicator in 20 image dimension for Indonesia 10.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 1228 | 334 | 1467 | 193 | 0.815 | 0.864 | 0.836 | 0.674 |
| KNN | 1171 | 265 | 1536 | 250 | 0.853 | 0.824 | 0.84 | 0.676 |
| Resnet50 | 1201 | 275 | 1526 | 220 | 0.847 | 0.845 | 0.846 | 0.69 |
| VGG16 | 1196 | 278 | 1523 | 225 | 0.846 | 0.842 | 0.844 | 0.685 |
| VGG19 | 1242 | 326 | 1475 | 179 | 0.819 | 0.874 | 0.843 | 0.688 |
| CNN | 1217 | 263 | 1538 | 204 | 0.854 | 0.856 | 0.855 | 0.708 |

**Table 31** – Result of 20 period without volume indicator in 20 image dimension for Indonesia 10.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | TN | FN | TP | FP | Sensitivity | Specificity | Accuracy | MCC |
| Random Forest | 1230 | 259 | 1836 | 155 | 0.876 | 0.888 | 0.881 | 0.756 |
| KNN | 1297 | 368 | 1308 | 294 | 0.68 | 0.815 | 0.797 | 0.596 |
| Resnet50 | 1223 | 203 | 1892 | 162 | 0.903 | 0.883 | 0.895 | 0.782 |
| VGG16 | 1242 | 179 | 1916 | 143 | 0.915 | 0.897 | 0.907 | 0.808 |
| VGG19 | 1268 | 212 | 1883 | 117 | 0.899 | 0.916 | 0.905 | 0.806 |
| CNN | 1234 | 178 | 1917 | 151 | 0.915 | 0.891 | 0.905 | 0.803 |

For all experiment results in Indonesia 10, we conclude a summary result with or without volume indicator in table 32 and 33. CNN 20 trading days’ period in 50 dimension with volume indicator shown the best result with 0.9 accuracy in table 32. While VGG16 in 20 trading days’ period with 20 image dimension without volume performed better result with 0.907 accuracy.

**Table 32** – Summary result of Indonesia 10 with their best classifier for each trading days and image dimension with volume indicator.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Classifier | Period | Dimension | Sensitivity | Specificity | Accuracy | MCC |
| Resnet50 | 5 | 50 | 0.807 | 0.854 | 0.831 | 0.661 |
| Resnet50 | 10 | 50 | 0.886 | 0.884 | 0.885 | 0.77 |
| CNN | 20 | 50 | 0.9 | 0.901 | 0.9 | 0.798 |
| Resnet50 | 5 | 20 | 0.788 | 0.823 | 0.806 | 0.612 |
| CNN | 10 | 20 | 0.833 | 0.854 | 0.843 | 0.686 |
| CNN | 20 | 20 | 0.891 | 0.846 | 0.871 | 0.738 |

**Table 33** – Summary result of Indonesia 10 with their best classifier for each trading days and image dimension without volume indicator.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Classifier | Period | Dimension | Sensitivity | Specificity | Accuracy | MCC |
| Resnet50 | 5 | 50 | 0.791 | 0.879 | 0.833 | 0.671 |
| CNN | 10 | 50 | 0.875 | 0.866 | 0.871 | 0.74 |
| CNN | 20 | 50 | 0.875 | 0.866 | 0.871 | 0.74 |
| CNN | 5 | 20 | 0.834 | 0.824 | 0.829 | 0.658 |
| CNN | 10 | 20 | 0.854 | 0.856 | 0.855 | 0.708 |
| VGG16 | 20 | 20 | 0.915 | 0.897 | 0.907 | 0.808 |

## 4.3. Real Environment Prediction

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

## 4.4. Comparison

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

**Table 34** – Comparison result with Khaidem, Saha et al. 2016 for Samsung stock market.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| (Khaidem, Saha et al. 2016) - Samsung | | | | | |
| Name | Trading Period | ACC | Precision | Recall | Specificity |
| Khaidem, Saha et al. 2016 | 1 month | 0.868 | 0.881 | 0.87 | 0.865 |
| Our | 1 month | 0.875 | 0.88 | 0.87 | 0.891 |
| Khaidem, Saha et al. 2016 | 2 month | 0.906 | 0.91 | 0.925 | 0.88 |
| Our | 2 month | 0.942 | 0.94 | 0.94 | 0.862 |
| Khaidem, Saha et al. 2016 | 3 month | 0.939 | 0.924 | 0.95 | 0.926 |
| Our | 3 month | 0.945 | 0.94 | 0.95 | 0.882 |

**Table 35** – Comparison result with Khaidem, Saha et al. 2016 for Apple stock market.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| (Khaidem, Saha et al. 2016) - Apple | | | | | |
| Name | Trading Period | ACC | Precision | Recall | Specificity |
| Khaidem, Saha et al. 2016 | 1 month | 0.882 | 0.892 | 0.907 | 0.848 |
| Our | 1 month | 0.896 | 0.9 | 0.9 | 0.863 |
| Khaidem, Saha et al. 2016 | 2 month | 0.93 | 0.941 | 0.938 | 0.919 |
| Our | 2 month | 0.936 | 0.94 | 0.94 | 0.877 |
| Khaidem, Saha et al. 2016 | 3 month | 0.945 | 0.945 | 0.961 | 0.923 |
| Our | 3 month | 0.956 | 0.96 | 0.96 | 0.885 |

**Table 36** – Comparison result with Khaidem, Saha et al. 2016 for GE stock market.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| (Khaidem, Saha et al. 2016) - GE | | | | | |
| Name | Trading Period | ACC | Precision | Recall | Specificity |
| Khaidem, Saha et al. 2016 | 1 month | 0.847 | 0.855 | 0.876 | 0.809 |
| Our | 1 month | 0.902 | 0.9 | 0.9 | 0.86 |
| Khaidem, Saha et al. 2016 | 2 month | 0.908 | 0.913 | 0.93 | 0.876 |
| Our | 2 month | 0.978 | 0.98 | 0.98 | 0.993 |
| Khaidem, Saha et al. 2016 | 3 month | 0.925 | 0.931 | 0.945 | 0.895 |
| Our | 3 month | 0.974 | 0.98 | 0.98 | 0.983 |

**Table 37** – Comparison result with Patel, Shah et al. 2015

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| (Patel, Shah et al. 2015) –S&P BSE SENSEX | | | (Patel, Shah et al. 2015) – NIFTY 50 | | |
|  | ACC | F-measure |  | ACC | F-measure |
| Patel, Shah et al. 2015 | 0.8984 | 9.9026 | Patel, Shah et al. 2015 | 0.8952 | 0.8935 |
| Our | 0.972 | 0.97 | Our | 0.934 | 0.93 |
| (Patel, Shah et al. 2015) – Reliance Industry | | | (Patel, Shah et al. 2015) - Infosys | | |
| Patel, Shah et al. 2015 | 0.9222 | 0.9234 | Patel, Shah et al. 2015 | 0.9001 | 0.9017 |
| Our | 0.939 | 0.94 | Our | 0.939 | 9.94 |

**Table 38** – Comparison result with Zhang, Zhang et al. 2018

|  |  |  |
| --- | --- | --- |
| (Zhang, Zhang et al. 2018) – Hong Kong 13 | | |
|  | Accuracy | MCC |
| Zhang, Zhang et al. 2018 | 0.617 | 0.331 |
| Our | 0.926 | 0.846 |

# Chapter 5 Conclusion and Future Works

After doing some experiments to do the stock market prediction, among other experiments it is using several different data periods. The data period is five days trading data, ten days trading data, and also twenty days trading data. In addition, we also use some kind of residual network that is distinguished by the number of layers in our architecture. Among other is the residual network with the number 50 layers, one 101 layers, and also 152 layers. With some experiments, the conclusion with the best result is X. It shows that experiment with X can produce X.

Given that the results may not be satisfactory enough, for the future we want to try to improve those results and also add features to this research. Using candlestick chart image on input data is new, the possibility of doing more research on image processing handling will further improve prediction results. To make it easier for readers other than the source code available in github, the possibility of building a web server would be great.

# References

Bollen, J., et al. (2011). "Twitter mood predicts the stock market." Journal of computational science **2**(1): 1-8.

Borovykh, A., et al. "Dilated Convolutional Neural Networks for Time Series Forecasting."

do Prado, H. A., et al. (2013). "On the effectiveness of candlestick chart analysis for the Brazilian stock market." Procedia Computer Science **22**: 1136-1145.

He, K., et al. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition.

Hu, G., et al. (2017). "Deep Stock Representation Learning: From Candlestick Charts to Investment Decisions." arXiv preprint arXiv:1709.03803.

Hunter, J. D. (2007). "Matplotlib: A 2D graphics environment." Computing in science & engineering **9**(3): 90-95.

Khaidem, L., et al. (2016). "Predicting the direction of stock market prices using random forest." arXiv preprint arXiv:1605.00003.

Krizhevsky, A., et al. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems.

Patel, J., et al. (2015). "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques." Expert Systems with Applications **42**(1): 259-268.

Pedregosa, F., et al. (2011). "Scikit-learn: Machine learning in Python." Journal of machine learning research **12**(Oct): 2825-2830.

Schöneburg, E. (1990). "Stock price prediction using neural networks: A project report." Neurocomputing **2**(1): 17-27.

Silver, D., et al. (2016). "Mastering the game of Go with deep neural networks and tree search." nature **529**(7587): 484-489.

Simonyan, K. and A. Zisserman (2014). "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556.

Szegedy, C., et al. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. AAAI.

Tsai, C.-F. and Z.-Y. Quan (2014). "Stock prediction by searching for similarities in candlestick charts." ACM Transactions on Management Information Systems (TMIS) **5**(2): 9.

Wongbangpo, P. and S. C. Sharma (2002). "Stock market and macroeconomic fundamental dynamic interactions: ASEAN-5 countries." Journal of Asian Economics **13**(1): 27-51.

Zeiler, M. D. and R. Fergus (2014). Visualizing and understanding convolutional networks. European conference on computer vision, Springer.

Zhang, X., et al. (2018). "Improving stock market prediction via heterogeneous information fusion." Knowledge-Based Systems **143**: 236-247.