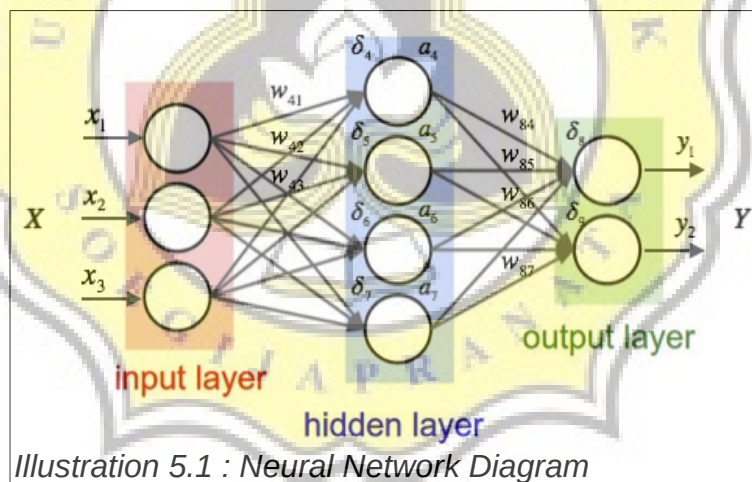


## CHAPTER 5

### IMPLEMENTATION AND TESTING

#### 5.1 Implementation

This chapter explain the algorithm and data structure used. The application is created using Java programming language. Linked list as data structure is used to save the stock data. Linked list chosen because data stock is always growth. Backpropagation algorithm is used by the variation of different hidden layer cell numbers. The variation of hidden layer cell numbers are 2, 4, 6 and 8. All of the hidden layer differences will produce different prediction price. Some experiments will be done by using this application. The result of this experiments will give the most optimal and accurate prediction.



This application will process the stock history data taken from finance.yahoo.com. Stock data to be retrieved is TLKM, HMSP, BBCA, BBNI and ASII. The data will be taken from January 2016 to November 2017. Data from January to Desember 2016 will be used as training data and Januari - November 2017 will be used as data testing. Stock data is retrieved and saved into

text file. The data that has been stored will be normalized before it is used in calculation process. The data should be normalized because the volume data is tens of millions but the open price, high, low and adjust data are only in thousands. So normalization is used to equalize the value of different data. After the data is normalized then the stage “feed forward” will be going on. The feed forward stage will produce a temporary prediction. Temporary predictions are compared with the original price and the difference will be calculated. Those differences is used for backpropagation process so that the weight will be adjusted. Then with adjusted weights, the process is repeated until the target epoch is reached.

### **Fetch Stock Data**

The first stage of this application is taking stock data from [finance.yahoo.com](http://finance.yahoo.com). The technique for retrieving data from the Internet is called webscraping. In java programming language there is a library called JSOUP that can be used for webscraping.

Before retrieving the data, the application will check the stock data that has been stored in the text file and see the last data it has. Then do the webscraping to take the newer data that has not been saved in the text file. The newer data is append into the text file.

Below is the code of webscraping :

```

1 document doc = Jsoup.connect(html).get();
2 Elements tableElements = doc.select("table");
3 Elements tableHeaderEles = tableElements.select("thead tr th");
4 System.out.println("headers");
5 for (int i = 0; i < tableHeaderEles.size(); i++)
6 {
7     System.out.println(tableHeaderEles.get(i).text());
8 }
9 System.out.println();
10
11 Elements tableRowElements = tableElements.select(":not(thead) tr");
12
13 Outerloop :
14 for (int i = 2; i < tableRowElements.size(); i++)
15 {
16     Element row = tableRowElements.get(i);
17     System.out.println("row");
18     Elements rowItems = row.select("td");
19     for (int j = 0; j < rowItems.size(); j++)
20     {
21         System.out.println(rowItems.get(j).text());
22         if(rowItems.get(j).text().equals(last))
23         {
24             break Outerloop;
25         }
26         gabungan = gabungan + (rowItems.get(j).text()) + ";";
27     }
28 }

```

*Illustration 5.2 : Code of Webscraping*

Line codes 1 - 12 take advantage of jsoup to open finance.yahoo.com and read the table containing stock data. The table is read per column and row one by one. Lines 20 - 29 detect the date to be retrieved, when the specified date is found then the string data is retrieved. The captured data is then stored into the text file in accordance with the name of the stock.

## **Normalization**

Data in linked list will entered into the process of data normalization. Normalization is used so that the different data becomes more equalize and has a range of 0 - 1. Normalization requires the minimum and maximum value of a data that must be searched before entering the formula process. The result of this process is will be saved into the linked list.

## **Feed Forward**

The normalized linked list moves into the feed forward process. In the beginning there are some weights that must be initiated such as the weight input to hidden layer, weight bias to the hidden layer, weight of the hidden layer to the output layer and weight bias to the output layer. The weight data are randomly filled with a range of -0.5 to 0.5. This weight is only as a starting number only, when it entered the stage of Backpropagation then the weight will change according to the calculation.

The feed forward stage performs the operation calculation on the hidden layer, the activation function on the hidden layer, the operation on the output layer and the activation function of the output layer. Prediction results will appear in the activation function of the output layer.

The process at this Feed Forward stage is to calculate the input with initiated weight to produce a temporary price prediction. Called a temporary prediction because in the next stage this prediction will be fixed again with better prediction results. The five inputs in the form of open price, high, low, adjust price and volume will produce an output that is close price.

### Calculate Error Rate

After the prediction results appear, calculate the error rate using RMSE formula. From the RMSE results that can be known whether the data is optimal or not. The formula to calculate RMSE :

$$\text{RMSE} = \sqrt{\sum \frac{(y_{pred} - y_{ref})^2}{N}}$$

*Illustration 5.3 : RMSE Formula*

$Y_{pred}$  : Prediction price

$Y_{ref}$  : Real price

$N$  : Data amount

Each time the neural network process runs then the predicted results will be reduced by the original price and squared. The results will continue to be added until all neural network processes are completed then the results are divided by the amount of data used.

RMSE formula calculation is to know to the accuracy of predictions. The smaller the RMSE value the more accurate prediction results with the real price. The RMSE value is considered small if the value is below 0.01.

### **Backpropagation Process**

The Backpropagation stage is to improve the weight value to become more optimal with the network. Temporary prediction price from Feed Forward stage is compared to the original price. The price difference processes to change all the weight value. The weight value will be updated so that the weight becomes more accurate. The process of adjusting the weights makes the algorithm able to learn from mistakes. When the predicted price difference with the original price is still too large, the algorithm will replace the weights until an accurate value is reached.

After all the weights are updated, the process of feed forward and backpropagation are repeated. The repeating phase is called an epoch. Therefore the algorithm requires many epoch to repeat the process. Each epoch runs, then the weight will be more adjusted so that at the end of target epochs can provide an accurate prediction results. In this study used epoch target as much as 10,000 times.

### **Calculate a Prediction**

Calculate the prediction done after the feed forward and backpropagation process runs as much as 10,000 epoch. To calculate the prediction required inputs of open price, high, low, adjust price and volume of date to be predicted. The input data is calculated using the Feed Forward stage and using the last updated weights. The Result is close price prediction.

## User Interface Design



First the user chooses the stock of the company to be predicted. Select the date to be predicted. The date to be predicted should be in the past. After that select the initial data used for training. In this research used training data for 1 year. Then choose how many cells on the hidden layer will be used to calculate. When "All Layer" are selected then will displayed four results from 2, 4, 6 and 8 cell. As for adding a new company list then fill in the name, http address and text file path then press "Tambah".

## 5.2 Testing

This sub chapter will explain the results of research that has been done. The study used stock data from PT Telekomunikasi Indonesia (TLKM), PT HM Sampoerna (HMSP), PT Bank Central Asia (BBCA), PT Bank Rakyat Indonesia (BBRI) and PT Astra International (ASII). Training data used for 1 year from January 2016 to December 2016 (amounting to about 252 data). Data testing used during the month of January 2017 - November 2017 (about 241 data). Using the epoch of 10,000 and learning rate 1.

### 1. PT Telekomunikasi Indonesia (TLKM)

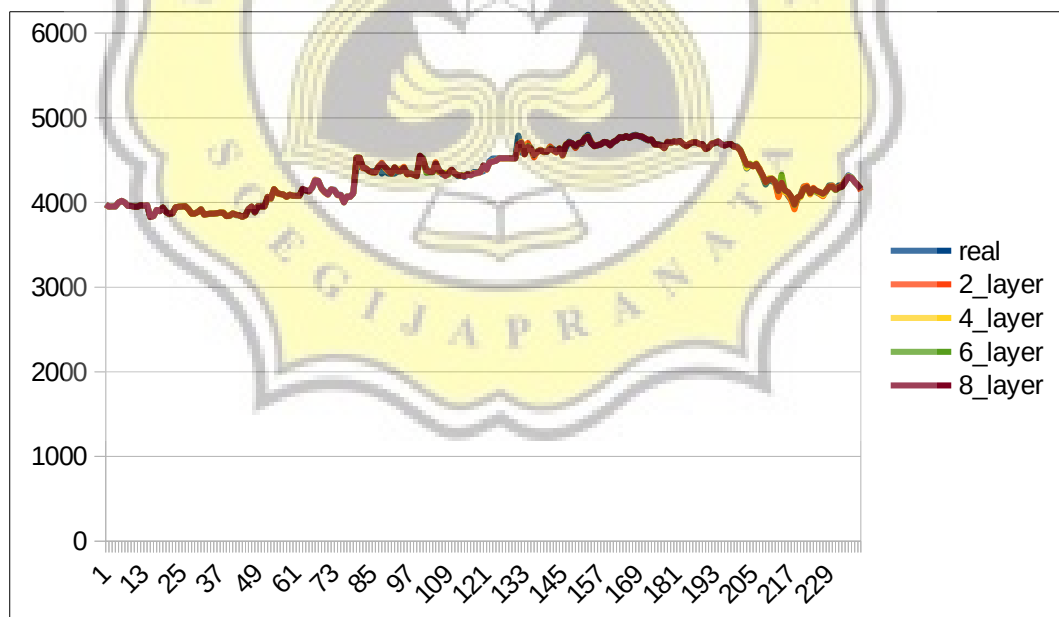


Illustration 5.5 : Chart of PT Telekomunikasi Indonesia



2. PT HM Sampoerna (HMSP)

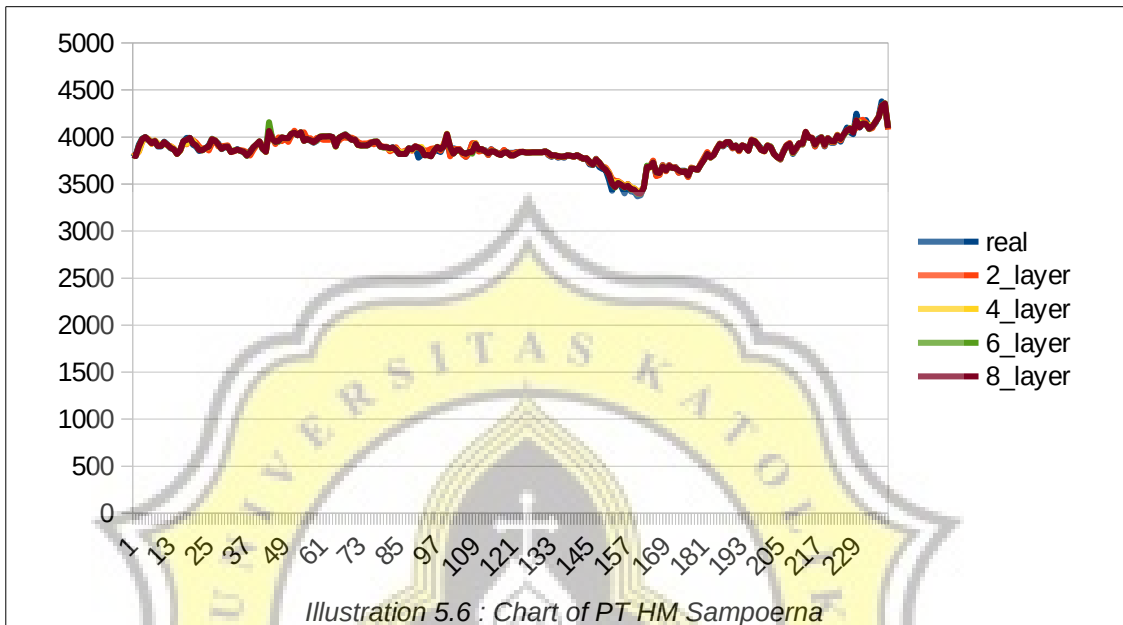


Illustration 5.6 : Chart of PT HM Sampoerna

3. PT Bank Central Asia (BBCA)

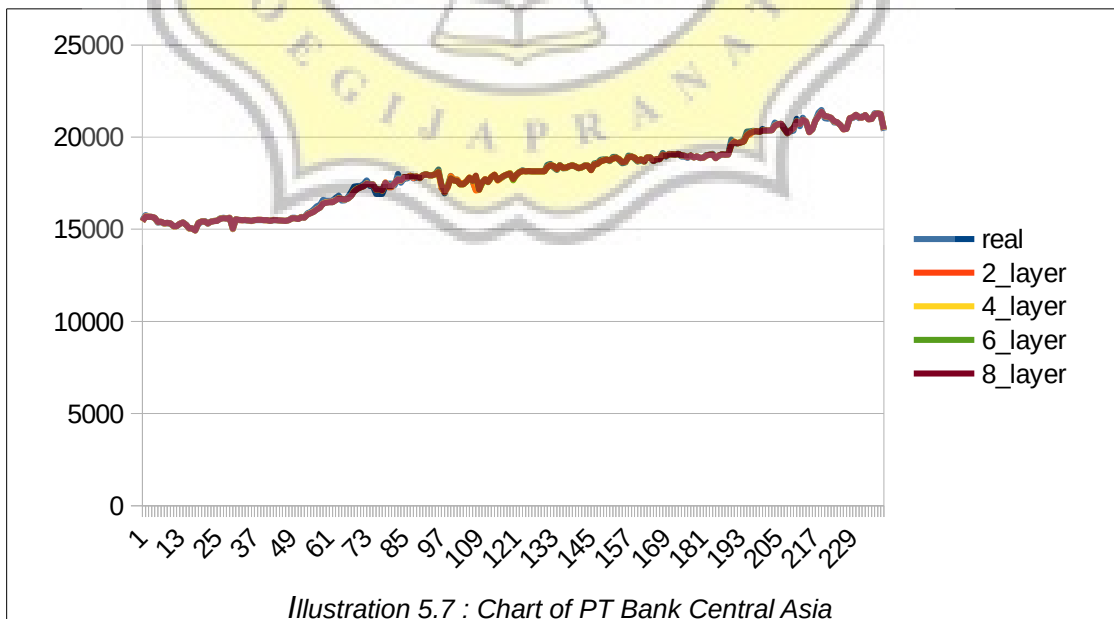


Illustration 5.7 : Chart of PT Bank Central Asia

4. PT Bank Negara Indonesia (BBRI)

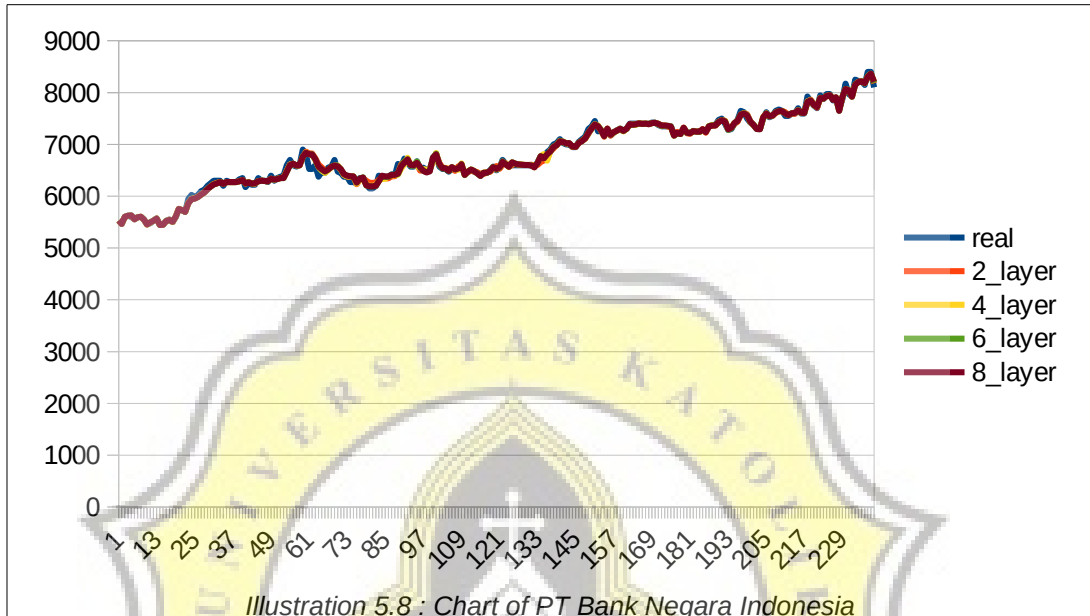


Illustration 5.8 : Chart of PT Bank Negara Indonesia

5. PT Astra International (ASII)

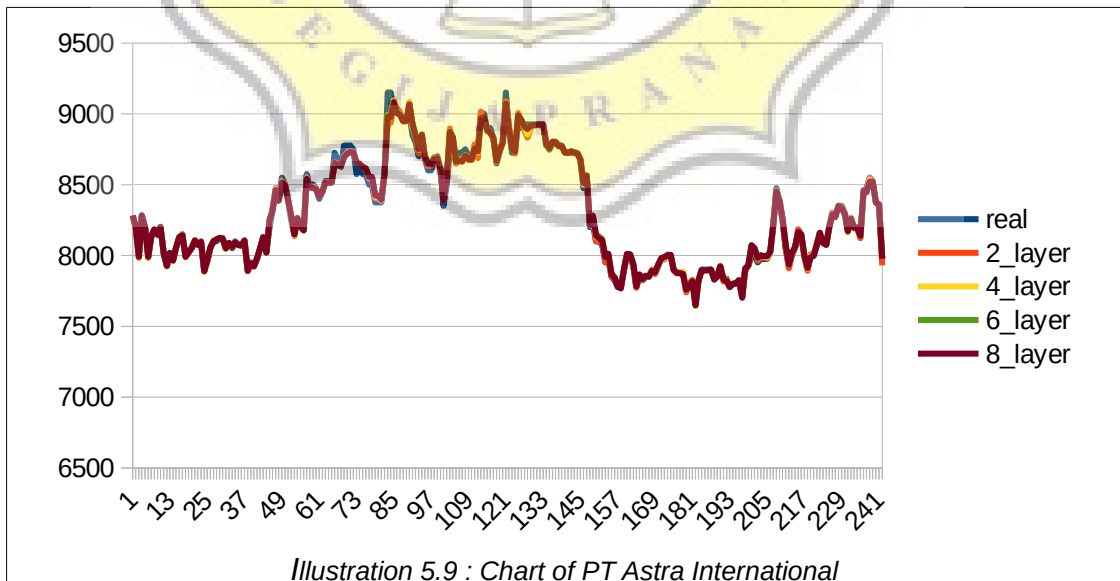


Illustration 5.9 : Chart of PT Astra International

The following is the result in percentage :

Stock	Number of Hidden Layer Cell	Minimum	Maximum	Average
PT Telekomunikasi Indonesia	2	96.60%	100%	99.75%
	4	97.21%	100%	99.75%
	6	97.21%	100%	99.76%
	8	95.84%	100%	99.74%
PT HM Sampoerna	2	96.38%	100%	99.57%
	4	96.99%	100%	99.65%
	6	97.17%	100%	99.66%
	8	97.17%	100%	99.67%
PT Bank Central Asia	2	98.31%	100%	99.72%
	4	95.31%	100%	99.71%
	6	94.26%	100%	99.71%
	8	96.72%	100%	99.71%
PT Bank Negara Indonesia	2	95.45%	100%	99.46%
	4	95.57%	100%	99.53%
	6	95.62%	100%	99.54%
	8	95.65%	100%	99.54%
PT Astra International	2	97.65%	100%	99.80%
	4	98.23%	100%	99.80%
	6	98.23%	100%	99.79%
	8	98.22%	100%	99.78%

Table 5.1: Comparison of Prediction Accuracy

According to the data training used in this application, the average result shows good results with errors of less than 1%. Because the difference error is less than 1%, therefore can not be seen of any particular layer that provides better results.