Stock Market Prediction with Neural Network Method

Edy Fradinata, ST, MT
Industrial Engineering Department of
Serambi Mekkah University
Banda Aceh, Indonesia
edinata69@gmail.com

Abstract—This paper explores neural network method for predicting the stock market which much-needed accuracy. There are three predictions that can be made in the stock market, the stock price index prediction, prediction of the direction of the stock and the stock return prediction. Stock market predictions discussed in this paper are limited to the stock in open price trend but it shown also about stock index. To deal with stock market data are time series is required some modifications to the neural network used to predict the stock market, while the stock index prediction will yield specific trend will be able to determine more precisely, it has been limited one month for the data prediction. By getting the proper method will be able to reduce the uncertainty of the position of the price of a value separation both in short term and long term. In this study the resulting of the prediction the next day with Neural Network got the error 0.0002 with bias 3.90500 it shown that the trend is so accurate and this is the best method to prediction of the value with the very small error and bias.

Keywords—Neural Network, trend, Stock Market, Price value. Timeseries.

I. INTRODUCTION

In the stock market or the capital market, there is no certainty about the position of a stock index, both for short term and long term. To be able to take the right decision in trading in the stock market, the trader will need to reduce the existing uncertainties. One of the means used to do so is the stock market prediction [3]. Prediction of stock market conducted by using data mining technology. Data mining for the stock market is an interesting topic and a lot of research because it is one example of a next generation data mining applications. Next-generation data mining is a term that refers to data mining for complex data. For example, spatial data, temporal data, multimedia data, and others [7]. The data in the stock market, including temporal data, also known as time-series data.

A variety of data mining technique can be used to predict the stock market [1], [2], [9], [10]. This paper will discuss in particular the use of the technique neural network method to predict the direction of the stock market. The discussion in this paper will follow the following systematic. Part 2 will describe the time series data and stock market data. Section 3 will describe the use of neural network methods for stock market predictions. Section 4 contains a summary of all the discussion that has been done.

II. TIME SERIES AND STOCK MARKET DATA

Data Analysis and Theory takes the Fourier transform of a stretch of time series data as the basic quantity to work with and shows the power of that approach. It considers second-and higher-order parameters and estimates them equally, thereby handling non-Gaussian series and nonlinear systems directly. The included proofs, which are generally short, are based on cumulates [11].

Time series data is a sequence of data whose value is changed at regular intervals. Time series data can be presented in the form of a graph or curve showing the data to the function of a variable unit of time. Graphs are constructed can be used to analyze trends or patterns in time-series data. There are two main objectives, namely to model the trend analysis of time series data and to predict the time series data. What is meant by modeling time series data is finding mechanisms or factors that lead to the formation of a time series. Referred to predict the time series data is predicting the value of a variable time series at a time that will come. Data is also a stock market time series data. Changes in the value of the stock price index can be seen in daily intervals. [1], [3]. Data characteristics of the stock market have had a lot of attributes. Attributes possessed was the price, the opening price, closing price, volume, price change and percentage change, the maximum price and the minimum price [2].

Several research ([2], [4], [8], [10]) note that the price of a stock does not stand alone, but depends on certain factors. These factors include the financial position of a company, the business sector, the general economic trend, the history stock, and others. The analysis takes into account these factors are called fundamental analysis. The opposite of fundamental analysis is an analysis technique, which studies market
behavior based on historical data on market prices and volume.

There are three things that can be predicted from the stock market. The first is the level or stock price. The second is the direction of the stock. The third is the return obtained from a stock. Some research suggests that the development of stock trading strategies that are based on predictions of the direction of the stock is relatively more effective and profitable [4].

III. NEURAL NETWORK

Neural Network is in the category of supervised learning. In this category, we need to train a network to find the model parameters w and b the best. Furthermore, by using models found we need to perform the prediction task. The value of the parameter w and b are usually not globally optimal value but local optimum. Since the cost function in a non-linear neural network which eventually produces a solution that is not globally each trained neural network. This is the weakness of the neural network. But in practice, although resulting is a local optimal solution, neural networks provide fairly accurate prediction solutions [10]. Neural network consists of a set of input and output units are connected to each other and each unit has a weight relationships. Each input and output units in the network are part of a layer / layer in the network. A neural network may have three or more layers, one input layer, one or more hidden layers and one output layer. To see a picture of a neural network, the reader can refer to [3], [5] or [7].

The neural network learning phase involves the evaluation and readjustment of the relationship between the weight of each unit in the network so that the data tuples into the network can be given the proper class or label. Neural networks can be used both to predict the level or price of the stock index, the direction of the stock, and the return obtained from stock. Predictions will be discussed in this paper is the prediction level or price of the stock index at a specified time and the prediction of the direction of the stock. The first prediction is a prediction that will be level or price of the stock index at a specific time in the future. Data used to make these predictions is the Athens stock trade data from 1998 to 2005 [3].

There are so many types of neural networks design, known in the literature of neural networks such as perceptron, back propagation, recurrent neural networks, LVQ [10]. Neural network is used to predict the level or price of the stock index is backpropagation network. The first step is to build a neural network... Based on the training data is used, the input layer of the network consists of only three units of neurons. Output layer of the network consists of nine neurons. That is, there are nine possible values predicted results that may be obtained from the three values of the input data to the input layer. To determine the number of neurons in hidden layer network, use Kolmogorov theorem. [3]. In accordance with the backpropagation algorithm [7], after the data is processed in the network and out at the output layer, the error was calculated from the output layer to the input layer continues. Of calculation error, the weight on each connection

between neurons repaired. Usually in the calculation of the new weights on the network used Widrow-Hoff rule or regulation Least Mean Square. However, in experiments with time series data, the rule used is Delta Rule with momentum \( \alpha \). Momentum \( \alpha \) calculations of these states a constant weight. If the weight of recent changes lead to a certain direction, then the momentum \( \alpha \) will have a tendency to make subsequent weight changes lead in the same direction. To see details of the formula calculation error and changing weights in the neural network for stock market data, the reader can refer to [3]. Based on the results of experiments using learning rate \( \eta = 1 \) and momentum \( \alpha = 0.7 \) while training, obtain a fairly accurate prediction results. Graph comparing the predicted results with actual data can be found in [3].

In addition to the backpropagation network, the predicted level or price of the stock indexes can also take advantage of other neural network models. Another model that can be used is a feed forward network, multilayer perceptron network and generalized feed forward network. Experiments to compare the three neural network models using data from the Central Bank of the Republic of Turkey since July 2, 2001 until February 28, 2003.

IV. METHODOLOGY

Time series is a sequence of data points, measured typically at successive time instants spaced at uniform time intervals. Examples of time series are the daily closing value of the Dow Jones index or the annual flow volume of the Nile River at Aswan. Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values. Time series are very frequently plotted via line-charts.

Consider the one-layer network in Figure 1. The set of equations relating inputs and outputs is given by

\[ y_j = f_j \left( \sum_{i=0}^{I} w_{ji} \cdot x_i \right) \]

where \( i = 1,2,\ldots,I \), \( j = 1,2,\ldots,J \), \( S \),

where \( I \) is the number of inputs, \( J \) the number of outputs, \( x_{0i} = 1, w_{ji} \) are the weights associated with neuron \( j \) and \( S \) is the number of data points.
Fig 1: One-layer feedforward neural network

To learn the weights $w_k$, the following sum of squared errors between the real and the desired output of the networks is usually minimized.

**Time Series** is a sequence of data points, measured typically at successive time instants spaced at uniform time intervals. Examples of time series are the daily closing value of the Dow Jones index or the annual flow volume of the Nile River at Aswan. Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values. Time series are very frequently plotted via line charts. A line chart or line graph is a type of chart which displays information as a series of data points connected by straight line segments. It is a basic type of chart common in many fields. It is an extension of a scatter graph, and is created by connecting a series of points that represent individual measurements with line segments. A line chart is often used to visualize a trend in data over intervals of time – a time series – thus the line is often drawn chronologically.

Time Series data have a natural temporal ordering. This makes time series analysis distinct from other common data analysis problems, in which there is no natural ordering of the observations (e.g. explaining people's wages by reference to their respective education levels, where the individuals' data could be entered in any order). Time series analysis is also distinct from spatial data analysis where the observations typically relate to geographical locations (e.g. accounting for house prices by the location as well as the intrinsic characteristics of the houses).

A stochastic model for a time series will generally reflect the fact that observations close together in time will be more closely related than observations further apart. In addition, time series models will often make use of the natural one-way ordering of time so that values for a given period will be expressed as deriving in some way from past values, rather than from future values see time reversibility [11]. Methods for time series analyses may be divided into two classes: frequency-domain methods and time-domain methods. The former include spectral analysis and recently wavelet analysis; the latter include auto-correlation and cross-correlation analysis.

Additionally time series analysis techniques may be divided into parametric and non-parametric methods. The parametric approaches assume that the underlying stationary Stochastic process has a certain structure which can be described using a small number of parameters (for example, using an autoregressive or moving average model). In these approaches, the task is to estimate the parameters of the model that describes the stochastic process. By contrast, non-parametric approaches explicitly estimate the covariance or the spectrum of the process without assuming that the process has any particular structure. Additionally methods of time series analysis may be divided into linear and non-linear, univariate and multivariate. Time series analysis can be applied to: real-valued, continuous data, discrete numeric data, discrete symbolic data.

Fig 2. Simple graph shows data over intervals with connected points.

Normally, though, we will try to stick to "unmixed" models with either only-AR or only-MA terms, because including both kinds of terms in the same model sometimes leads to overfitting of the data and non-uniqueness of the coefficients. In addition, we have illustrated the accuracy of one-period-ahead forecasts in the test period. In Figure 4, we have forecasted only at the out-of-sample period. For forecasting performance, we have utilized the last 10 points of the data and calculate quantities such as root mean squared error (MSE) and mean absolute percentage error (MAPE). The resulting root MSE and MAPE for out-of-sample data are 0.4561 and 0.29, respectively for Dow Jones Utilities Index. A comparison of this result to traditional models can be found next section.

Fig 3. Out-of-sample forecasting for Dow Jones Utilities Index

V. APPLICATION

The Data that is used for these predictions is coming from the one of stock exchange data. Time series are data vectors sampled over time, in order, often at regular intervals. They are distinguished from randomly sampled data that form the basis of many other data analyses. Time series represent the time-evolution of a dynamic population or process. The linear ordering of time series gives them a distinctive place in data analysis, with a specialized set of techniques. Time series
analysis is concerned with: Identifying patterns, Modeling patterns and Forecasting values.

<table>
<thead>
<tr>
<th>Day</th>
<th>Prev</th>
<th>Open</th>
<th>High</th>
<th>Low</th>
<th>Close</th>
<th>change</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>9/9/2012</td>
<td>21950</td>
<td>22000</td>
<td>22100</td>
<td>21900</td>
<td>22100</td>
<td>150</td>
<td>1273500</td>
</tr>
<tr>
<td>10/9/2012</td>
<td>22100</td>
<td>22100</td>
<td>21950</td>
<td>22050</td>
<td>22100</td>
<td>50</td>
<td>990000</td>
</tr>
<tr>
<td>11/9/2012</td>
<td>22050</td>
<td>21300</td>
<td>21300</td>
<td>20650</td>
<td>20950</td>
<td>1100</td>
<td>4943500</td>
</tr>
<tr>
<td>12/9/2012</td>
<td>20950</td>
<td>21350</td>
<td>21400</td>
<td>20900</td>
<td>20950</td>
<td>0</td>
<td>677000</td>
</tr>
<tr>
<td>13/09/2012</td>
<td>20950</td>
<td>21250</td>
<td>21250</td>
<td>21000</td>
<td>21200</td>
<td>250</td>
<td>336500</td>
</tr>
<tr>
<td>14/09/2012</td>
<td>21200</td>
<td>21200</td>
<td>21200</td>
<td>20500</td>
<td>21050</td>
<td>50</td>
<td>990000</td>
</tr>
<tr>
<td>17/09/2012</td>
<td>20600</td>
<td>20650</td>
<td>20350</td>
<td>20400</td>
<td>20400</td>
<td>200</td>
<td>705500</td>
</tr>
<tr>
<td>13/09/2012</td>
<td>20400</td>
<td>20300</td>
<td>20200</td>
<td>20750</td>
<td>20750</td>
<td>350</td>
<td>459500</td>
</tr>
<tr>
<td>19/09/2012</td>
<td>20750</td>
<td>20750</td>
<td>20550</td>
<td>20600</td>
<td>20600</td>
<td>600</td>
<td>320500</td>
</tr>
<tr>
<td>20/09/2012</td>
<td>20600</td>
<td>20300</td>
<td>21200</td>
<td>20300</td>
<td>21150</td>
<td>1100</td>
<td>653500</td>
</tr>
<tr>
<td>23/09/2012</td>
<td>21150</td>
<td>21150</td>
<td>21200</td>
<td>20900</td>
<td>21150</td>
<td>0</td>
<td>201500</td>
</tr>
<tr>
<td>24/09/2012</td>
<td>21150</td>
<td>21200</td>
<td>21300</td>
<td>21050</td>
<td>21050</td>
<td>100</td>
<td>546500</td>
</tr>
<tr>
<td>25/09/2012</td>
<td>21050</td>
<td>21050</td>
<td>21100</td>
<td>20650</td>
<td>20700</td>
<td>350</td>
<td>504500</td>
</tr>
<tr>
<td>26/09/2012</td>
<td>20700</td>
<td>20750</td>
<td>20350</td>
<td>20700</td>
<td>20350</td>
<td>150</td>
<td>471000</td>
</tr>
<tr>
<td>27/09/2012</td>
<td>20850</td>
<td>20600</td>
<td>20300</td>
<td>20500</td>
<td>20700</td>
<td>150</td>
<td>290000</td>
</tr>
<tr>
<td>Risk 2</td>
<td>20616</td>
<td>20614</td>
<td>20606</td>
<td>20353</td>
<td>20501</td>
<td>97</td>
<td>375900</td>
</tr>
<tr>
<td>Risk 1</td>
<td>20533</td>
<td>20523</td>
<td>20512</td>
<td>20255</td>
<td>20401</td>
<td>94</td>
<td>319300</td>
</tr>
<tr>
<td>Risk 1</td>
<td>20450</td>
<td>20442</td>
<td>20413</td>
<td>20153</td>
<td>20302</td>
<td>90</td>
<td>263700</td>
</tr>
<tr>
<td>Forecast</td>
<td>20366</td>
<td>20355</td>
<td>20325</td>
<td>20061</td>
<td>20202</td>
<td>37</td>
<td>207601</td>
</tr>
</tbody>
</table>

When it is plotted a member of a time series collection, its time units display on the x-axis and its data units display on the y-axis. and the plot title is displayed as 'Time Series Plot: member name. If you use the same figure to plot a different member of the collection, no annotations display. The time series plot method does not attempt to update labels and titles when hold is on because the descriptors for the series can be different. To describe multiple series, add a legend. Set the Display Name property of the line series to label each member, as follows:

![Fig 4. The real data make in line and include the average data](image)

![Fig 5. Target Error average training](image)

During training, the progress is constantly updated in the training window. Of most interest are the performance, the magnitude of the gradient of performance and the number of validation checks. The New Network is used for the creation and growth of a new neural network from the values in the Grid. During optimization a new network is created every time the period of cycles or seconds completes. The number of hidden nodes in each layer that is checked grows from minimum to maximum. Each network is checked by training it for a short period. The network that produces the lowest error is selected for further training and validating.
From the network the error is 0.0000 and the bias is 3.90500 with 6 nodes detail this is an optimal condition of error and bias. Validating cycles only occur when validating examples are included in the Grid. A number of training examples can be randomly changed to validating examples using the Select control or they can be entered directly into the Grid. The cycle values are preset so that 100 learning cycles occur for every validating cycle. The number of learning cycles per the first validating cycles and the number of learning cycles per validating cycle can be changed to any positive values. Learning can be set to stop when the validating target is reached. The validating target can be set to any value up to 100%. The validating test can be either when the validating results are within a specific range or when the results are correct after rounding to the nearest whole number. Validating Rules can be used for rounding to the nearest multiple. The validating range can be set to any value from 0 to 50%.

The prediction of the open value from the stock exchange data could be shown under, it is starting predict from 28/9/2012 data and it start to increase for the next day and should be read on 20355 (it shown in table 1. Also). The point following the trend line which is decrease from the first its point in around 0.59 point, it draws as above.

VI. CONCLUSION

Neural network method is excellent when used to predict something like the above case because the neural network to read the data and perform some of the new or the data to be used as learning data and then predict based on the trend that has been calculated For the above case the error is generated by 0.000 so nearly perfect despite saying that the prediction is not necessarily always correct in prediction entirely but at least it will be very close to the expected value of the prediction. In the experiment above the point to predict the September 28th 2012 produce accurate data that is 22.355, while the actual data for the same date is 22.350 it means the the error rate of 0.0002 as well as other data that can be read in table 1 above.

REFERENCES
