Time Series Forecasting Of Nifty Stock Market Using Weka

Raj Kumar¹, Anil Balara²

¹M.Tech, Global institute of Engineering and Technology ,Gurgaon
²Associate Professor, Global institute of Engineering and Technology,Gurgaon

Abstract:

Forecasting accuracy is the most important factor in selecting any forecasting methods. Research efforts in improving the accuracy of forecasting models are increasing since the last decade. The appropriate stock selections those are suitable for investment is a very difficult task. The key factor for each investor is to earn maximum profits on their investments. Numerous techniques used to predict stocks in which fundamental and technical analysis are one among them. In this paper latest time series forecasting feature of Weka is used. Results has been checked using for different methods. These methods applied on 3 months NIFTY data. 10 days result has been predicted. The results from SMO regression shows that it offers the ability to predict the stock prices more accurately than the other existing techniques. The results will be used to analyze the stock prices and their prediction in depth in future research efforts.

1.Introduction:-

Generally, stock market across the globe replicates the fluctuation of the market’s economy, and attracts the attention of millions of investors. The stock market is characterized by high risk and high yield; hence investors are concerned about the analysis of the stock market and are trying to forecast the trend of the stock market. To accurately predict stock market, various prediction algorithms and models have been proposed in the literature. Forecasting is a process that produces a set of output with a set of variables [9]. The variables are normally past data of the market prices. Basically, forecasting assumes that future values are based, at least in part on these past data. Past relationships can be derived through the study and observation. Neural networks have been used to perform such task. The ideas of forecasting using neural network is to find an approximation of mapping between the input and output data through training. Later to predict the future stock value trained neural network can be used [8]. The rest of the paper is organized as follows. Section 2 reviews the literature in predicting the stock market price through various methods. Section 3 focuses on the objectives of the research. Section 4 discusses about the National Stock Exchange (NSE), stock market, stock classification, forecasting, fundamental analysis and technical analysis. Section 5 explains the methodology of using Weka tool to forecast the stock prices using time series forecasting package. Section 6 draws the conclusion of the paper.

2.Literature review :-

Prediction of stock price variation is a difficult task and the price movement behaves more like a random walk and varies with time. Since the last decade, stockbrokers and future traders have relied upon various types of intelligent systems to make trading decisions. Enke, D., Thawornwong [1] has been explained the techniques of data mining and neural network needed for prediction of stock market prices and values. It has been widely accepted by many studies that non-linearity exists in the financial markets and that neural networks can be effectively used to uncover this relationship. Wang, J.L., Chan [2] uses the daily stock prices of Microsoft, Intel, and IBM to assess stock market purchasing opportunities with simple technical indicators. Sapna Jain[3] has been assimilated the knowledge about Weka tool and demonstrated with taking example of clustering method and introduce WEKA workbench, reviews the history of the project and k-means clustering execution in WEKA 3.7. Hazem M. El-Bakry[4] a new technique for fast forecasting of stock market prices is presented. Such algorithm uses new high speed time delay neural networks (HSTDNNs). Swasti Singhal [5] has explained the basic concept is used to analyze the data from different angle,
categorize it and finally to summarize it. They introduce the key principle of data pre-processing, classification, clustering and introduction of WEKA tool. Weka is a data mining tool. They have been described the steps of how to use WEKA tool for these technologies. It provides the facility to classify the data through various algorithms.

markets. Jageshwer Shriwas[6] explains the hybrid approach in stock price prediction. this research determines the feasibility and practicality of using data mining as a forecasting tool. K.K.Sureshkumar[7] use Weka 3.6.3 tool to obtain more accurate stock prediction price and to compare them with weka classifier functions such as Gaussian processes, isotonic regression, least mean square, linear regression, multilayer perceptron, pace regression, simple linear regression and SMO regression. Most of these uses data mining prediction with weka older versions. Very little is done on forecasting package of weka on the Indian stock markets.

3. Objective of the study:
The main objective of this paper is to use Weka 3.7.8 tool to obtain more accurate stock prediction price using time series forecasting package of weka. This research will examine and analyze the use of weka as a forecasting tool. Specifically some mining and forecasting ability to predict future trends of Stock Market Indices will be tested. Our objective is to analyze prediction of various methods on 3 months record of NIFTY. Analyze and compare results by plot graphs. We will forecast 10 days result using all four methods.

4. National Stock Exchange of India (NSE):
NSE was incorporated in November 1992, and received recognition as a stock exchange under the Securities Contracts (Regulation) Act, 1956 in April 1993. Since its inception in 1992, NSE of India has been at the vanguard of change in the Indian securities market. This period has seen remarkable changes in markets, from how capital is raised and traded, to how transactions are cleared and settled. The market has grown in scope and scale in a way that could not have been imagined at that time. Indian equity markets are today among the most deep and vibrant markets in the world. NSE offers a wide range of products for multiple markets, including equity shares, Exchange Traded Funds (ETF), Mutual Funds, Debt instruments, Index futures and options, Stock futures and options, Currency futures and Interest rate futures. The role of trading members at NSE is to the extent of providing only trading services to the investors; the Exchange involves trading members in the process of consultation and participation in vital inputs towards decision making [10].

4.1 Stock Market
A stock market index is a method of measuring a stock market as a whole. The most important type of market index is the broad-market index, consisting of the large, liquid stocks of the country. In most countries, a single major index dominates benchmarking, index funds, index derivatives and research applications. In addition, more specialized indices often find interesting applications. In India, we have seen situations where a dedicated industry fund uses an industry index as a benchmark. In India, where clear categories of ownership groups exist, it becomes interesting to examine the performance of classes of companies sorted by ownership group.

4.2 Stock Classification
Stocks are often classified based on the type of company it is, the company’s value, or in some cases the level of return that is expected from the company. Below is a list of classifications which are generally known to us Growth Stocks, Value Stocks, Large Cap Stocks, Mid Cap Stocks, and Small Cap Stocks. Stocks are usually classified according to their characteristics. Some are classified according to their growth potential in the long run and the others as per their current valuations. Similarly, stocks can also be classified according to their market capitalization. S&P CNX NIFTY has NIFTY (50), Junior NIFTY (50), CNX IT (20), Bank NIFTY (12), NIFTY Midcap50, CNX Realty (10) and CNX Infra (25). The sectoral distribution of NSE are Financial services or banks, Energy, Information Technology, Metals, Automobile, FMCG, Construction, Media & Entertainment, Pharma, Industrial Manufacturing, Cement, Fertilizers & Pesticides, Textiles, Power and Telecom [10].

4.3 Limitation in earlier version of Weka
Upto weka 3.7.5, forcasting feature was not supported. Only clustering, classify and associate was possible. Now version 3.7.8 and higher supports
forecasting feature which can predict values on the basis on history[11].

4.4 Time Series Forecasting

Time series analysis is the process of using statistical techniques to model and explain a time-dependent series of data points. Time series forecasting is the process of using a model to generate predictions (forecasts) for future events based on known past events. Time series data has a natural temporal ordering - this differs from typical data mining/machine learning applications where each data point is an independent example of the concept to be learned, and the ordering of data points within a data set does not matter. Examples of time series applications include: capacity planning, inventory replenishment, sales forecasting and future staffing levels. This technique is more flexible and less time consuming.

5. Research Methodology:

We will apply following four methods on 3 months data of Nifty stock exchange. These data mining methods are:

1. Gaussian processes
2. Linear regression
3. Multilayer Perceptron
4. SMOreg

5.1 Data Preprocessing

In data preprocessing, we have consider 90 days of CNX NIFTY daily data. From this, minimum price, maximum price, mean, standard deviation, distinct and unique percentage has been found. The details are given in the table 1. From the table 1, we found the Minimum, Maximum, Mean and Standard deviation for the four attributes. In view of the data, we find out in-between values for the stock prices. In table 2 we have classifications in the table and in the graphs each in four categories namely open price, high price and low price, close price. For example, consider the Table 2 open column containing the attribute data as 13 which means 13 days of data indicates stock price range lies between 6216.75 and 6519.75. Similarly for data 29.7.11. The last data 11 which denotes 11 days of Nifty stock price range lies between 7125.75 and 7428.75. Totally data of 60 instances of open price, high price and low price and close price are represented in the table 2 and 3 and lies within value of minimum and maximum.
<table>
<thead>
<tr>
<th>Data</th>
<th>LOW (6212.25-7302.6)</th>
<th>Data</th>
<th>CLOSE (6221.45-7367.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lies Between The Price</td>
<td></td>
<td>Lies between The Price</td>
</tr>
<tr>
<td>9</td>
<td>6212.25-6484.83</td>
<td>9</td>
<td>6221.45-6507.86</td>
</tr>
<tr>
<td>31</td>
<td>6484.83-6757.42</td>
<td>30</td>
<td>6507.86-6794.27</td>
</tr>
<tr>
<td>6</td>
<td>6757.42-7030.01</td>
<td>7</td>
<td>6794.27-7080.68</td>
</tr>
<tr>
<td>14</td>
<td>7030.01-7302.6</td>
<td>14</td>
<td>7080.68-7367.1</td>
</tr>
</tbody>
</table>

Table 3. Data Classification of Low Price, Close price

Table 1. Statistical values of NIFTY

Table 2. Data Classification of Open price, High Price

<table>
<thead>
<tr>
<th>Data</th>
<th>LOW (6216.75-7424.75)</th>
<th>Data</th>
<th>HIGH (6277.75-7563.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lies Between The Price</td>
<td></td>
<td>Lies between The Price</td>
</tr>
<tr>
<td>13</td>
<td>6216.75-6519.75</td>
<td>8</td>
<td>6277.75-6534.9</td>
</tr>
<tr>
<td>29</td>
<td>6519.75-6822.75</td>
<td>29</td>
<td>6534.9-6792.05</td>
</tr>
<tr>
<td>7</td>
<td>6822.75-7125.75</td>
<td>9</td>
<td>6792.05-7049.2</td>
</tr>
<tr>
<td>11</td>
<td>7125.75-7428.75</td>
<td>6</td>
<td>7049.2-7306.35</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
<td>7306.35-7563.5</td>
</tr>
</tbody>
</table>
5.2 Results and Discussion

The all four methods Gaussian process, Linear regression, Multilayer perceptron, SMO regression has been applied on 3 months of nifty stock market. All methods are applied on 60 instances and the next 10 days of data has been predicted. The forecast value and graph of 10 days data for smo method is shown in figure 3 and figure 4.

The sample data has been tested on Windows 7 operating system. We have used Weka 3.7.8 tool for preprocessing, evaluation and forecasting of the stock. From this analysis, it is found that the forecasting values of next 10 days results has shown in the figures. We acquire four functions for this analysis to forecast values and evaluate them. By comparing the results of the direction accuracy values and error percentage the SMO regression is the best suited method for forecasting the stock prices. Following are the results.

SMO regression

Figure 3:- 10 days forecasted result of SMO regression

<table>
<thead>
<tr>
<th>Method</th>
<th>Forecast Value</th>
<th>Graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multilayer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMO</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© JRPS International Journal for Research Publication & Seminar Vol 05 Issue 02 March - July 2014
**Conclusion:**

In this research, we examined and applied different forecasting techniques by using the Weka tool. We compared various prediction functions, and found that SMO regression function offers the ability to predict the stock price of NSE more accurately than the other functions such as Gaussian processes, linear regression, multilayer perceptron. This analysis can be used to reduce the error percentage in predicting the future stock prices. It increases the chances for the investors to predict the prices more accurately by reduced error percentage and hence increased profit in share markets. More Accurate results can be found if we will take data of more duration. It will help to check deviations in Values and predict more accurate results.

Future work can be done by apply more techniques and verify results with actual values. Error rate and graph can be designed to study it’s validity. After this step, by observing error rate on different method, best technique can be selected. All this steps can be repeated with different set of data.

**References:**


