Chapter 4

Intraday Prediction

4.1. Introduction

The very nature of high risks in stock markets trading makes the limited forecasting an important activity. In practice this is very difficult to realize, and more over the opportunities are short lived, but the reality is that these opportunities exists. Hence the investor's main aim will be to capture these unpredictable opportunities at the right time. Obviously, excess profits are not available to all the market participants, and these profits are no way related to the growth, and underlying worth of companies. If someone makes money in the market, it comes from some one who looses it; hence an investor in an intraday trading tries to make profits, even if they are small. In the process he uses what ever little help he can get in forecasting the price. The primary intension of any investor in the stock market is to catch the market trends at an early stage and accordingly transact (Buy or Sell) at right time. Though stock market data is convertible into some form of multiple time series, it is difficult to process, analyze and mine manually. Researchers have proposed several methods to predict the future price of the stocks. In this paper, we proposed a method to predict the intraday price of a stock using the historic data. Given the time stamped transactions, the stock data is mined for pattern records using similarity profiled temporal association mining using a pattern database created with patterns of reference to a cutoff value. Using the support value for different price gain and the opening price of the stock for the day, we extract all the significant pattern records from the pattern database. Using the current trend of the stock, we project the future prices from time to time for the day. Wipro stock data from 2005 to 2009 are used for experimental evaluation of our approach. Expected price for various days are agreed to an extent of 98% with actual transaction prices. In the light of this, it is easy to see that even some very limited forecasting ability could result in large gains for the investors.

Data mining comprises of techniques for discovering patterns and hence knowledge discovery is useful for decision making. It is becoming essential for the
people in both business and research fields to use the novel methods and tools related to Data Mining for better returns. For the domains having high commercial value, researchers aimed to provide scalable algorithms which can use huge historical data to produce accurate results in less time.

A stock market facilitates the transaction of stocks among the investors willing to trade. The Stock Exchange provides certain information to the investors regarding the stocks from time to time in very short time intervals ranging 2 or 3 minutes. This time series data is captured at equal time intervals with the information in the form of records sent through network. Each record is characterized by variables like stock_id, date and time of transaction, opening price of the stock, closing price of the stock, volume transacted, minimum (low) and maximum (high) price of the stock in the corresponding time interval. The information is sent continuously by the Stock Exchanges through internet to the brokerage firms who facilitate the transaction of the stocks, from the time of opening of the markets to the close of it in the day. In the current scenario, anyone having an on-line account for brokerage, can access this data.

The objective of this paper is to use similarity profiled temporal association for identifying patterns in the stock market data. We use the available historic data to create day-wise table with equal time slots containing the opening price, gain on higher and lower side of the price within the time slot. In the process for the prediction of the price we also prepare a support table using day-wise table, representing the support for a specific gain in each time slot. Using this day-wise table we create a pattern table containing upward, downward and neutral trends of scrip. For the current day’s opening price of the stock, gain on opening is calculated for the day and similar pattern records are selected from the existing pattern table. From the selected pattern records similarity profile of upward, downward and neutral trends are identified for each and every time slot. Finally, using support table values we project expected price that the stock reaches in each time slot. With these projected prices in various slots, an investor can choose an interesting time slot and know possible appropriate price to make a transaction.
4.2. Related work

The complexity of predicting the stock prices is high due to several financial and stock trading factors influence the investor's decision in making a transaction. Initially it was opined that the behavior of an investor depends on the size of the owner company (blue chips-middle-small), but further investigation has brought forward several factors like annual earning price per stock, volume transacted, the sector to which the stock belongs, historical behavior, financial position of a company, rumors in the industry, and several other unpredictable factors [10]. In general the annual reports published by the companies give the details of financial position and health of the company. Such information influences the investors once (as soon as the financial year ends and the reports are available). At other instants, external factors like rumors whip up sentiments among the investors, make him to take decisions are not predictable. Hence the historical behavior of the market, if identified can give an idea about the future of the markets in the absence of unpredictable factors. Most of the researchers in the stock market are interested in on identifying the fluctuation of a stock over time, and then predict the future price of the stock.

In recent years applications of Data Mining in business areas have found increasing acceptance, as it handles large amounts of data for analysis and finds the undiscovered knowledge for better decision making. The importance of the Stock Market for the economy, and the complexity of it in the business environment had led to analyzing the markets for better prediction of the trends. There are several different types of methods applied to predict the Stock Market returns all of these fall in to two categories viz., Fundamental Analysis and Technical Analysis. The Efficient Market Theory proposed by Eugene Fama [8] and [28] specifies that a price of a stock reflects all known information about it at that point of time. The random walk theory corresponds to the belief that markets are efficient, and that it is not possible to beat or predict the market because stock prices reflect all available information and the occurrence of new information is seemingly random as well. It contends that a stock’s future price can be forecasted based on historical information through observing chart patterns and technical
indicators. Academicians cannot conclusively agree or contradict this theory as there are published studies that support both sides of the issue [2, 18, 23].

Several researchers in the field of economics and management sciences are still working on methods using fundamental analysis. Their main area of interest in stock markets is to identify the accounting attributes which can be used for fundamental analysis. Identifying historical accounting signals that can be used to improve the entire distribution of future returns earned by an investor, whether the investor is doing business in primary, secondary markets or is a short-seller [13, 1]. There were researchers who were with the view that the stock market does not reflect the current earning of the company fully and hence the future trends cannot be predicted. They argued that the financial markets depend on too many factors which makes it impossible to be predicted by the current accounting attributes of a company [4, 9].

DOW Theory is believed to be the foundation for technical analysis and it has showed that there are, three movements simultaneously progress in stock market. The major is the primary movement bull or bear market. Concurrently with the primary and secondary movement of the market, and constant throughout, there obviously was, the underlying fluctuations from day to day (minor trends). This theory in turn has been developed from the candlestick charting used by Japanese from 18th century [33, 26, 7, 27, 33]. Elliott Wave Principle is one of technical analysis methods in the domain of stock market prediction. Investors use to forecast trends in the financial markets by identifying extremes in investor’s psychology, high and lows in price movement, and other collective activities [24].

Temporal data mining analysis however is of a more recent origin with slightly different constraints and objectives than the traditional time series data. One main difference lies in the size of data sets and the way it is collected. Often the methods must be capable of analyzing large data sets. The second major difference lies in the kind of information that we want to extract from the data like, trends and patterns in the data which are easily interpretable. This model uses past samples of data and predicts the
future values. In order to do this, one needs to build a predictive model for the data. If the data is static, auto regressive family of models are used to predict a future value as a linear combination of earlier sample values [3, 5, 12]. Linear stationary models like ARIMA models have also been found useful in many economic and industrial applications where some suitable variants of the process can be assumed to be stationary. There are many nonlinear models for time series predictions such as neural networks [30, 31, 16].

There is lot of work done on predicting intra day price of a stock. K Senthamarai Kannan et al. in their work [29] have used five methods of analyzing stocks namely Typical Price (TP), Bollinger Bands, Relative Strength Index (RSI), CMI and Moving Averages (MA) to predict if the day’s closing price would increase or decrease. The method adopted by them could predict the price accurately at least 50% of times. In their paper Jo Ting et. al [15] have used pattern-back stock data mining approach transforming numeric data into symbolic sequences and carried out sequential and non-sequential association analysis in classifying and predicting the further price movement. This was one of the motivator for our work. The authors Paul D Yoo et al. in their paper [19] have presented developments in stock market prediction models and discussed their advantages and disadvantages. Mainly the traditional time series prediction, Neural Networks, Support Vector machines and Case based reasoning methods were looked into. They have also proposed an event extraction mechanism to enhance the prediction efficiency using neural network model. Mark O Afolabi et al. have used emerging methods like Fuzzy logic, Neural Networks and Hybridized methods such as hybrid Kohonen Self Organizing Maps (SOM), etc., for predicting stock prices for a day. In their paper they have studied the efficiency of prediction using Back propagation, Kohonen SOM, and a hybrid Kohonen SOM. The method was proved to be fast and efficient with less classification errors.

Marketos et al. [10] have used temporal data mining technique to extract significant patterns for trading strategies. In [10], an intelligent tool is designed to generate scenarios of stocks depending on these patterns and help an investor in taking
appropriate decision. Nayak et. al. [25] have adopted a strategy of applying a brute force approach to the prediction of stock prices based on the formation of a cluster around the query sequence. The prediction is then applied in a model designed to capitalize on the derived prediction [25]. Povinelli in his work [22] has used temporal pattern matching to predict the events in a financial time series. He uses an even characterization function which addresses the required prediction goals. Extending his work, teaming up with Higgs [21], have developed an approach to increase the returns and overcome transaction costs, the idea is to predict the prices for the week and allowing the investor to take a decision based on it. The work is further extended in 2003 teaming with Feng [20] to form a system for predictions that are based loosely around a desired outcome by an investor and determine whether the current input stream is likely to produce that event.

4.3. Problem Description

In the dynamic and complex domain of Stock market, carrying out the predictions on the movement of stock price in near future is an extremely complicated one and hence we need to carefully consider our choice of the multitude of factors leading to movement in the stock price [6]. The factors that need to be considered are;

i. Stock price: It is identified by many researchers that the rules governing the movement in the stock price is hidden in the historical sequences of the stock price. To perform prediction on the movement of future stock price, the prerequisite is to perform scientific classification on the historical sequence of stock price.

ii. The Investor: Stock price movement at the end of the day is the result of buying or selling on part of the investor. An investor’s anticipation for the future stock price depends on investor’s ambitious behavior. However, incorporating this factor is highly subjective and hence is not presently considered in this work.

iii. Time: We need to identify the stock price time sequence for the day, week, month and quarter and use them for prediction. The chosen time gap should be appropriate for identifying an interesting sequence in it. Studies by numerous researchers have shown that investors are selective on the time gap to invest in stocks.
iv. Basic elements: Basic elements such as economic factors, earning per share, rate of dividend, rate of interest, rate of inflation, market factors, etc. constitute the foundation for identifying the price of a stock. Over a period of time it is observed that the stock price at any time point in the market represents the result of the psychological balancing point between actual price and expected price with the comprehensive consideration of the various basic aspects taken into account. Here, we assume that such factors influence for a day after they start influencing the price of scrip. On all subsequent days, the traders look the closing price of the previous day and start trading on any day. Hence, we focus on proposing a method for predicting future prices or finding similarity patterns without taking such external factors in to account.

Government of India has a regulatory authority to balance the stock market movement whenever any of the above factors go beyond the control of the market, especially the investor’s behavior and the basic elements. However Stock markets throughout the world are regulated by some regulating authority or the other. For example American Stock Market is regulated by US Securities and exchanges commission, China has China Securities Regulatory Commission, Germany has Federal Financial Supervisory Authority, Japan has Securities and Exchange Surveillance Commission to name a few. A complete list is provided by Wikipedia. That leaves us the factors stock price and time to consider for the future prediction.

In this work we use similarity profiled temporal association mining for predicting the stock price from time to time in a day.

**Problem Statement**

Given the time stamped transactions and investor defined stock, predicting the future price of that stock at each time slot during the current day so as to enable the investor to make a decision on price and time at which he/she can trade. Future price of any stock at a time slot is obtained from “support tables” and “pattern database” formulated using similarity profiled temporal association mining of stock transaction data.
4.4. Proposed Method

The method followed here is in accordance with Similarity Profiled Temporal Association Mining (SPTAM) proposed by Jin Soung Yoo (Jin Soung Yoo et al., 2009) applying on historical data of a stock. In many applications, similarity profiled temporal association patterns can reveal interesting relationships among data items which co-occur with a particular event over time. The stock transaction database has time stamped transactions of various stocks. Each time-stamped transaction contains data for the variables like stock_Id, date and time of transaction, opening price of the stock, closing price of the stock, volume transacted, minimum and maximum price of the stock in the corresponding time interval.

Using the time stamped transaction data, we create two tables. The first table is a “day-wise table” consisting of 23 time slots, within which a pattern of price change of the stock is recorded. For simplicity this pattern is represented with highest price and lowest price that the stock reaches in a time slot. Depending on the gain/loss of price in each time slot, we create the second table “pattern table” representing the upward, downward and neutral trend in the price in each of the time slot for the day. Similarities in these patterns are identified by grouping similar gain records among all the historical data. The profile of these similar pattern records are represented as support tables for different gain ranges in different time slot. In the process of predicting the price for a day we use the opening gain for the day and identify the possible gain range for each of the time slot.

Entire process is done in following steps:

i. Data Preprocessing.
ii. Tables Creation.
iii. Price Prediction.
iv. Improved mechanism.
4.4.1 Data Preprocessing

It is observed that on majority (at least 90%) of the stock marketing working days the transactions of stocks follow certain patterns on a day. On each day, the transactions and the volume pickup slowly and will be at feverish pitch within an hour, and around mid-day it falls down and picks-up again as end of that day approaches. Particularly, during the last one hour before the end of the day it peaks up. Similar to this trend the price of a stock fluctuates and shows a pattern. It can be observed that the index of stock market in a day may vary drastically and volatile in a small unit of time. However, the price of a specific stock may not vary drastically with the pace of index.

Historical data from different companies from January, 2005 to June, 2009 is collected. However, we have chosen WIPRO data for this paper. For the purpose of tables and processing data from January, 2005 to May, 2009 is used, and the June 2009 data is set aside for the sake of comparison with the predicted data.

It is important to note that a transaction record is included in to transaction database only when there is a substantial change in the price of a stock, or the end of the time period. Thus the total number of transactions in a day is not a fixed number but varies with number of substantial changes happen in a day. However, for the proposed approach, we wish to have a fixed number of transactions on any day. Determining this fixed number and choosing the transaction records is part of Data Preprocessing step.

We need to transform these chosen records into a structured format for further processing. Stock variation in a specific time slot can be extracted from the structure of the record. For our experiment, we divide a day into 23 time slots with each slot lasting for 18 minutes and hence uniformly we got 23 transactions on any day. The structure of the record in a stock transaction database T is as below:

1. S_ID (Scrip ID) - Identifies the scrip of the company.
2. DOT - Date of transaction.
3. Open - opening price of the scrip for the day and the specific time slot.
4. High - highest price of the scrip in the time interval.
5. Low - lowest price of the scrip in the time interval.
6. Close - closing value of the scrip in the time interval.
7. Volume - volume of scrip transacted in the time interval.
8. TS - Time slot number

Table 4.1. Sample Data

<table>
<thead>
<tr>
<th>SID</th>
<th>DOT</th>
<th>OPEN</th>
<th>HIGH</th>
<th>LOW</th>
<th>CLOSE</th>
<th>VOLUME</th>
<th>TS</th>
</tr>
</thead>
<tbody>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>340.05</td>
<td>340.05</td>
<td>340.05</td>
<td>340.05</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>339.5</td>
<td>340.28</td>
<td>335.37</td>
<td>335.37</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>336.19</td>
<td>339.93</td>
<td>336.12</td>
<td>339.93</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>339.92</td>
<td>340.26</td>
<td>338.03</td>
<td>338.03</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>338.32</td>
<td>339.09</td>
<td>337.89</td>
<td>338.63</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>339.22</td>
<td>339.53</td>
<td>338.13</td>
<td>338.25</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>338.66</td>
<td>339.1</td>
<td>336.76</td>
<td>337.15</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>336.63</td>
<td>337.89</td>
<td>336.63</td>
<td>337.89</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>337.76</td>
<td>339.28</td>
<td>337.37</td>
<td>339.28</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>339.28</td>
<td>340.74</td>
<td>339.05</td>
<td>339.71</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>339.8</td>
<td>339.91</td>
<td>339.32</td>
<td>339.7</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>339.89</td>
<td>341.56</td>
<td>339.5</td>
<td>340.21</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>339.24</td>
<td>340.07</td>
<td>338.07</td>
<td>339.85</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>339.83</td>
<td>341.52</td>
<td>339.43</td>
<td>341.52</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>341.71</td>
<td>341.71</td>
<td>340.93</td>
<td>341.24</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>340.99</td>
<td>342.63</td>
<td>340.84</td>
<td>342.42</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>342.27</td>
<td>342.49</td>
<td>338.3</td>
<td>339.03</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>338.63</td>
<td>339.13</td>
<td>336.6</td>
<td>338.71</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>338.31</td>
<td>339.35</td>
<td>336.89</td>
<td>337.76</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>337.88</td>
<td>338.85</td>
<td>337.64</td>
<td>338.85</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>339.19</td>
<td>339.66</td>
<td>338.77</td>
<td>339.66</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>339.59</td>
<td>339.59</td>
<td>338.19</td>
<td>338.19</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/11/2005</td>
<td>338.38</td>
<td>338.94</td>
<td>338.38</td>
<td>338.76</td>
<td>0</td>
<td>23</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/14/2005</td>
<td>335</td>
<td>335</td>
<td>335</td>
<td>335</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/14/2005</td>
<td>327.91</td>
<td>333.36</td>
<td>327.91</td>
<td>332.45</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/14/2005</td>
<td>331.86</td>
<td>332.98</td>
<td>329.72</td>
<td>329.72</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/14/2005</td>
<td>329.41</td>
<td>330.15</td>
<td>323.42</td>
<td>323.42</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/14/2005</td>
<td>323.33</td>
<td>326.39</td>
<td>323.33</td>
<td>324.93</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>WIPRO</td>
<td>1/14/2005</td>
<td>325.66</td>
<td>327.21</td>
<td>325.66</td>
<td>327.21</td>
<td>0</td>
<td>6</td>
</tr>
</tbody>
</table>

Over a period of time the price of a scrip varies drastically (for the WIPRO data that is collected from 2005 to 2009 the price has varied between 340-540-240), and using
it directly for any prediction becomes difficult. Hence, we wish to transform the stock transaction database T further to note the difference in price over a time slot.

Having the high, low prices for each time interval, we need to calculate the high gain (there could be a negative value for some of the records with price fluctuation below the opening price of the scrip for the day) and low gain (there could be both positive and negative value, depending on the movement of the price with the opening price of the scrip for the day). These records will depict the price change in a day for all the 23 time intervals. Significantly the high gain (HG) for a time interval represents the highest possible gain an investor gets if he sells his scrip in that time interval. When the value of high gain in an interval is negative, it specifies that the scrip price is fluctuating below the opening price and the price is the least loss the investor may need to bear if he wants to sell the scrip. Similarly the low gain (LG) for a time interval represent the lowest possible gain he gets if he sells his scrip in that time interval, and may in general be negative. Subsequently when it is positive it specifies that the price fluctuations are above the opening price.

From T, we create day-wise table D with the following structure for this purpose:

<table>
<thead>
<tr>
<th>No.</th>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>DOT</td>
<td>date of transaction.</td>
</tr>
<tr>
<td>2.</td>
<td>Open</td>
<td>Opening price of the scrip for the day.</td>
</tr>
<tr>
<td>3.</td>
<td>GOO</td>
<td>Gain on opening from the previous day closing price.</td>
</tr>
<tr>
<td>4.</td>
<td>HG1</td>
<td>High gain, gain from high price to the opening of the day in the time slot 1.</td>
</tr>
<tr>
<td>5.</td>
<td>LG1</td>
<td>Low gain, gain from low price to the opening of the day in the time slot 1.</td>
</tr>
<tr>
<td>7.</td>
<td>V1</td>
<td>Volume transacted in the time slot 1.</td>
</tr>
<tr>
<td>8.</td>
<td>HG2</td>
<td>High gain, gain from high price to the opening of the day in the time slot 2.</td>
</tr>
<tr>
<td>9.</td>
<td>LG2</td>
<td>Low gain, gain from low price to the opening of the day in the</td>
</tr>
</tbody>
</table>
11. V2 - Volume transacted in the time slot 2.

91. HG23 - High gain, gain from high price to the opening of the day in the time slot 23.
92. LG23 - Low gain, gain from low price to the opening of the day in the time slot 23.
93. G23 - Gain in the time slot 23.
95. Close - Closing price of the scrip for the day.
96. Gain - Overall gain for the scrip on that day.

Table 4.2. Day-wise data table

<table>
<thead>
<tr>
<th>DOT</th>
<th>OPEN</th>
<th>GOO</th>
<th>HG1</th>
<th>LG1</th>
<th>V1</th>
<th>HG23</th>
<th>LG23</th>
<th>G23</th>
<th>CLOSE</th>
<th>GAIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>3/1/2006</td>
<td>519.9</td>
<td>0.8</td>
<td>0</td>
<td>-3.8</td>
<td>48822</td>
<td>1.25</td>
<td>-0.5</td>
<td>-0.2</td>
<td>526.3</td>
<td>6.4</td>
</tr>
<tr>
<td>3/2/2006</td>
<td>533.3</td>
<td>-7</td>
<td>5.65</td>
<td>0</td>
<td>271275</td>
<td>1</td>
<td>-0.75</td>
<td>0.25</td>
<td>528.25</td>
<td>-5</td>
</tr>
<tr>
<td>3/3/2006</td>
<td>530.4</td>
<td>-2.2</td>
<td>0</td>
<td>-8.2</td>
<td>87960</td>
<td>2.6</td>
<td>-0.05</td>
<td>1.05</td>
<td>530.35</td>
<td>-0.05</td>
</tr>
<tr>
<td>3/6/2006</td>
<td>531</td>
<td>-0.7</td>
<td>0</td>
<td>-5.8</td>
<td>34069</td>
<td>0</td>
<td>-0.8</td>
<td>-0.6</td>
<td>532.5</td>
<td>1.5</td>
</tr>
<tr>
<td>3/7/2006</td>
<td>527</td>
<td>5.5</td>
<td>1</td>
<td>-4.5</td>
<td>122102</td>
<td></td>
<td></td>
<td></td>
<td>524</td>
<td>-3</td>
</tr>
<tr>
<td>3/8/2006</td>
<td>521.5</td>
<td>2.5</td>
<td>0.5</td>
<td>-4.8</td>
<td>68594</td>
<td></td>
<td></td>
<td></td>
<td>511</td>
<td>-10.5</td>
</tr>
<tr>
<td>3/9/2006</td>
<td>507.4</td>
<td>3.6</td>
<td>3.4</td>
<td>-6.4</td>
<td>183727</td>
<td></td>
<td></td>
<td></td>
<td>507</td>
<td>-0.4</td>
</tr>
</tbody>
</table>

Thus, each record in Day-wise Table D contains consolidated transaction information over 23 time slots of a day.

4.4.2 Tables Creation

In this phase, initially we create “support value table” for high gain, low gain, high loss and low loss. If high gain is negative, we record its absolute value as high loss. Similarly, if low gain is positive, we record its absolute value as low loss.

Change in price of a scrip in any time interval normally falls in the range of 1 to 11 rupees and hence support tables are prepared with 11 attributes. In support table titled HGSUP, the attributes are RG1, RG2, .... , RG11 representing High Gain up to Rs.1,
High Gain in the range of (Rs.1, Rs.2]......High Gain in the range of (Rs.10, Rs.11] respectively. We prepare four such tables viz., HGSUP and LGSUP each with 11 attributes. We refer these attributes as range attributes as RG \( j \) represents the range \((j-1)\) to \(j\).

We define “support value” for a time slot \( t_i \) \((i = 1,......23)\) and a range attribute \(RG_j\) \((j= 1,....11)\), as the ratio of the number of records of \(D\) in the time interval \(t_i\) and \(RG_j\) to the total number of records in the Day-wise table \(D\).

The structure of a support value table is as follows:

1. \(T1\) - Indicates 1\(^{st}\) time slot of the domain values where the measure of gain/loss is taken.
2. \(T2\) - Indicates 2\(^{nd}\) time slot.
3. ...
4. ...

23. \(T23\) - Indicates 23\(^{rd}\) time slot.

Table 4.3. Support value table (HGSUP)

<table>
<thead>
<tr>
<th>RGN</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
<th>T8</th>
<th>T9</th>
<th>T10</th>
<th>T11</th>
<th>T12</th>
<th>T13</th>
<th>T14</th>
<th>T15</th>
<th>T16</th>
<th>T17</th>
<th>T18</th>
<th>T19</th>
<th>T20</th>
<th>T21</th>
<th>T22</th>
<th>T23</th>
</tr>
</thead>
<tbody>
<tr>
<td>RG1</td>
<td>0.34082</td>
<td>0.28315</td>
<td>0.27491</td>
<td>0.27491</td>
<td>0.26517</td>
<td>0.26517</td>
<td>0.25919</td>
<td>0.25919</td>
<td>0.25543</td>
<td>0.25543</td>
<td>0.25742</td>
<td>0.25742</td>
<td>0.41124</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG2</td>
<td>0.09663</td>
<td>0.07191</td>
<td>0.06442</td>
<td>0.06442</td>
<td>0.05593</td>
<td>0.05593</td>
<td>0.05169</td>
<td>0.05169</td>
<td>0.04869</td>
<td>0.04869</td>
<td>0.04045</td>
<td>0.04045</td>
<td>0.03596</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG3</td>
<td>0.07266</td>
<td>0.06292</td>
<td>0.05693</td>
<td>0.05693</td>
<td>0.05434</td>
<td>0.05434</td>
<td>0.04419</td>
<td>0.04419</td>
<td>0.04949</td>
<td>0.04949</td>
<td>0.03221</td>
<td>0.03221</td>
<td>0.03895</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG4</td>
<td>0.07191</td>
<td>0.06742</td>
<td>0.05393</td>
<td>0.05393</td>
<td>0.04669</td>
<td>0.04669</td>
<td>0.0412</td>
<td>0.0412</td>
<td>0.03745</td>
<td>0.03745</td>
<td>0.0382</td>
<td>0.0382</td>
<td>0.04794</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG5</td>
<td>0.05768</td>
<td>0.0367</td>
<td>0.0397</td>
<td>0.0397</td>
<td>0.04345</td>
<td>0.04345</td>
<td>0.03745</td>
<td>0.03745</td>
<td>0.0382</td>
<td>0.0382</td>
<td>0.03596</td>
<td>0.03596</td>
<td>0.03521</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG6</td>
<td>0.05243</td>
<td>0.03146</td>
<td>0.03446</td>
<td>0.03446</td>
<td>0.02846</td>
<td>0.02846</td>
<td>0.02397</td>
<td>0.02397</td>
<td>0.02996</td>
<td>0.02996</td>
<td>0.02397</td>
<td>0.02397</td>
<td>0.02622</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG7</td>
<td>0.04045</td>
<td>0.02697</td>
<td>0.02697</td>
<td>0.03331</td>
<td>0.03331</td>
<td>0.02547</td>
<td>0.02547</td>
<td>0.02772</td>
<td>0.02772</td>
<td>0.02247</td>
<td>0.02247</td>
<td>0.02247</td>
<td>0.02247</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG8</td>
<td>0.02846</td>
<td>0.01873</td>
<td>0.02472</td>
<td>0.02472</td>
<td>0.02097</td>
<td>0.02097</td>
<td>0.01723</td>
<td>0.01723</td>
<td>0.01873</td>
<td>0.01873</td>
<td>0.02022</td>
<td>0.02022</td>
<td>0.02172</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG9</td>
<td>0.01723</td>
<td>0.01423</td>
<td>0.00749</td>
<td>0.00749</td>
<td>0.00824</td>
<td>0.00824</td>
<td>0.01199</td>
<td>0.01199</td>
<td>0.01498</td>
<td>0.01498</td>
<td>0.01498</td>
<td>0.01498</td>
<td>0.01498</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG10</td>
<td>0.01573</td>
<td>0.01423</td>
<td>0.01124</td>
<td>0.01124</td>
<td>0.00674</td>
<td>0.00674</td>
<td>0.01348</td>
<td>0.01348</td>
<td>0.01199</td>
<td>0.01199</td>
<td>0.01348</td>
<td>0.01348</td>
<td>0.02022</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG11</td>
<td>0.01423</td>
<td>0.00674</td>
<td>0.00974</td>
<td>0.00974</td>
<td>0.00524</td>
<td>0.00524</td>
<td>0.00899</td>
<td>0.00899</td>
<td>0.00824</td>
<td>0.00824</td>
<td>0.01124</td>
<td>0.01124</td>
<td>0.01124</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Similarly all the four tables indicate the support values for a set of ranges in 23
time slots. These tables are used to identify the support for a range in a time slot
depending on the similar trend records.

Now, we create Pattern table P using Day-wise table D. This table contains
variations in the price in each time slot compared with its predecessor. With a threshold
of Rs.2, any gain greater than Rs.2 in a time slot, compared with its predecessor, is
treated as an upward trend and is represented as “U”. Similarly, a loss less than Rs.2 in a
time slot, is treated as downward trend and is shown as “D”. In all other cases, it is
treated as neutral trend and is represented as “N”.

Using these notations, we create pattern table containing the following structure:

1. DOT - Date of Transaction.
2. Open - Opening price of the scrip for the day.
3. GOO - Gain on opening from the previous day’s opening.
4. HP1 - Trend U, D or N pertaining to high price in time slot t1.
5. LP1 - First time slot low price period,
   which indicates the trend of U, N or D.
6. G1 - Gain/Loss/Neutral (G/L/N) in the first time slot.
7. HP2 - Second time slots high price period.
8. LP2 - Second time slots low price period.
9. G2 - Gain/Loss/Neutral (G/L/N) in the second time slot.
10. ... ......
11. ... ......

67. HP23 - Twenty Third time slot high price period.
68. LP23 - Twenty third time slot low price period.
69. G23 - Gain/Loss/Neutral (G/L/N) in the twenty third time slot.
70. Close - Closing price of the scrip for the day.
71. Gain - Overall gain G (or loss L) from the opening of the day.

For Pattern Table computation, we employ the following:

GOO is taken as \((OP_i - CP_j)\) where
\(OP_i\) - Opening price of the scrip on ith day
\(CP_j\) - Closing price of the scrip on (i-1)th day.

- **HP\(_j\)** is taken as
  a) U whenever \(HG\_j\) of Day-Wise table > 2;
  b) D whenever \(HG\_j\) of Day-Wise table < 2;
  c) N otherwise

- **LP\(_j\)** is taken as
  a) U whenever \(LG\_j\) of Day-Wise table > 2;
  b) D whenever \(LG\_j\) of Day-Wise table < 2;
  c) N otherwise

- **G\(_j\)** is filled with
  a) G whenever \(G\_j\) of Day-wise table > 2;
  b) L whenever \(G\_j\) of Day-wise table < 2;
  c) N otherwise.

**Table 4.5. Sample of a Pattern table**

<table>
<thead>
<tr>
<th>DOT</th>
<th>GOO</th>
<th>OPEN</th>
<th>HP1</th>
<th>LP1</th>
<th>G1</th>
<th>HP2</th>
<th>LP2</th>
<th>G2</th>
<th>HP3</th>
<th>LP3</th>
<th>G3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/11/2005</td>
<td>-1.29</td>
<td>340.05</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>D</td>
<td>L</td>
<td>N</td>
<td>D</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>1/14/2005</td>
<td>3.76</td>
<td>335.00</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>D</td>
<td>L</td>
<td>D</td>
<td>D</td>
<td>L</td>
</tr>
<tr>
<td>1/17/2005</td>
<td>-5.05</td>
<td>335.00</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>D</td>
<td>N</td>
<td>D</td>
<td>L</td>
<td></td>
</tr>
<tr>
<td>1/18/2005</td>
<td>-5.34</td>
<td>341.28</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>D</td>
<td>L</td>
<td>D</td>
<td>D</td>
<td>L</td>
</tr>
<tr>
<td>1/19/2005</td>
<td>-2.77</td>
<td>342.00</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>U</td>
<td>N</td>
<td>N</td>
<td>U</td>
<td>N</td>
<td>G</td>
</tr>
<tr>
<td>1/20/2005</td>
<td>-1.52</td>
<td>345.00</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>D</td>
<td>D</td>
<td>L</td>
<td>D</td>
<td>D</td>
<td>L</td>
</tr>
<tr>
<td>1/24/2005</td>
<td>0.71</td>
<td>335.53</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>U</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>1/25/2005</td>
<td>2.67</td>
<td>322.00</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>D</td>
<td>L</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>1/27/2005</td>
<td>0.31</td>
<td>328.00</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>U</td>
<td>U</td>
<td>G</td>
<td>U</td>
<td>U</td>
<td>G</td>
</tr>
<tr>
<td>1/28/2005</td>
<td>1.74</td>
<td>336.55</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>U</td>
<td>N</td>
<td>G</td>
<td>U</td>
<td>U</td>
<td>G</td>
</tr>
</tbody>
</table>
This table helps us in identifying the trends in a day without considering the actual increase or decrease of the price.

### 4.4.3 Price Prediction

#### Derivation of Prediction Equations

Support Table contains support values listed for each time slots $T_1, T_2, \ldots, T_{23}$ and for each range $R_{G1}, R_{G2}, \ldots, R_{G11}$, where support value under $T_i$ against $R_{Gj}$ is ratio of number of transactions having gain in the range of $(R_{S,j-1}, R_{S,j})$ to the total number of transactions in the database. Hence, it indicates support enjoyed by an arbitrary transaction (irrespective of Gain on Opening for day of transaction) belongs to time slot $T_i$.

However, on a particular chosen day, we are interested to project and hence predict the future price of the scrip at a time slot $T_i$. Here, we start computation based on Gain on Opening ($GOO = \text{closing price on the previous day} - \text{opening price on the current day}$) and collection of records having same GOO but may belong to any day of business. In the example, GOO being 8.1, we collected 18 records from Pattern Table belong to various days of business having GOO in the range of $(Rs.8, Rs.9)$. Each of these records has values such as U, D, N for each of $HP_i$ and $LP_i$ of the time slot $T_i$. Focusing on a particular $HP_i$ (in case we wish to predict the possible high price), we find number of U, D, N respectively as $u_i$, $d_i$, $n_i$.

Now, $(u_i - d_i)$ indicates strength of upward movement in the collected records in $HP_i$ and

$$(u_i - d_i)/ttl$$

indicates strength of upward movement for an arbitrary record among the collected records in $HP_i$.

Since the original support table consists of $R_{G1}, R_{G2}, \ldots, R_{G11}$ ranges, we wish to have index pointer “br” pointing to appropriate range, hence we take

$$br = (u_i - d_i)/ttl \times nrng \quad \rightarrow \text{used step 7 of the algorithm}$$

where $nrng$ is the number of ranges i.e. 11 in our example.

(This prediction equation helps us in predicting best range in the support table)
Having computed $br$ value, we find the sorted list of support values under $T_i$ and identify $br^\text{th}$ member in the sorted list. Range against $br^\text{th}$, say $RG_j$, is the requisite range indicates that a profit in the range of $(Rs.j-1, Rs.j)$ is possible for the chosen day as the entire process identified the patterns relevant to GOO, analyzed the pattern of U,D,N, obtained an index pointer $br$ to corresponding range in support table.

Now, we predicted a profit in the range of $(Rs.j-1, Rs.j)$ but for a set of ttl records having same pattern.

Hence, net gain for an arbitrary record having strength of upward movement $(u_i - d_i)/\text{ttl}$ is

$$\text{net gain} = j \times \frac{(u_i - d_i)}{\text{ttl}} \rightarrow \text{used step 9 of the algorithm}$$

Having prepared all the required tables, we proceed to predict the price of scrip for a chosen day. The basis for prediction is the opening price of the scrip on the chosen day. The prediction process goes on through the following steps:

1. We calculate the gain on opening (OG) from the opening price for the chosen day and the closing price of the scrip on the previous day.

2. Depending on the value of OG we extract records from the pattern table $P$ with $P\text{.GOO}$ value lies in $\lfloor \text{int}(OG) \ldots \text{int}(OG)+1 \rfloor$ and save in to a temporary table $\text{Temp}$.

3. Let $u_i$, $d_i$, $n_i$ be the number of records in table $\text{Temp}$ having U, D, N respectively in a time slot $t_i$. Let the ranges chosen for predictions be $\text{rng}$ and let the Previous High Price $\text{PHP}_{t_i-1}$ be initialized with the opening price of the scrip for the 1$^{\text{st}}$ time slot.

4. As the neutral trends (N) do not affect the price, we take only upward and downward trends into consideration. It is observed that number of records with upward trends (U) is more than D or N for High Price (HP) part of a time slot. Similarly number of records having D dominates for Low Price (LP) part of a time slot. However downward trend in HP to some extent neutralizes upward trends. Hence the effect of HP trend in time slot $t_i$ is represented by $(u_i-d_i)$ i.e. the difference between the counts of
upward trend records and the downward trend records. Similarly, \((d_i - u_i)\) for LP trend for \(t_i\).

5. Now we calculate the effective trend in the time slot, as \(\frac{(u_i - d_i)}{[\text{Temp}]}\) where \([\text{Temp}]\) is total number of records in Temp. Higher upward trend records in table Temp leads to higher support value for higher range gain.

6. For each time slot \(t_i\), the best range \(b_i\) is computed using
\[
b_i \leftarrow \left[ \frac{(u_i - d_i)}{[\text{Temp}]} \right] \times \text{nrng} \quad (\text{where nrng is the No. of ranges chosen for calculation}).
\]

7. From the support value table HGSUP, we extract the support values under \(t_i\) and first column containing various RG\(_j\)'s. Then, we sort the support values in ascending order. Choose \(b_i^{th}\) smallest value in the sorted column and corresponding range RG\(_j\) by looking at the first column.

8. calculate the gain as:
\[
gain \leftarrow \text{Upper limit}[\text{RG}_j] \times \frac{(u_i - d_i)}{\text{rng}}.
\]

9. The predicted High Price is
\[
\text{Predicted High Price} \leftarrow \text{PHP}_{t_i-1} + \text{gain}.
\]

10. The prediction can go on for the next time slot as:
\[
\text{PHP}_{t_i} \leftarrow \text{PHP}_{t_i-1}
\]

11. Similarly Prediction of Low Price can also be done.

---

**Figure 4.1. Algorithm for Predicting the Prices for a day.**

**Inputs:**
1. Opening price OP of interested scrip for the day of prediction.
2. Four support tables HGSUP, LGSUP.
3. Pattern Table (P).

**Outputs:**
1. Predicted price for each time slot \(t_i\), \(i = 1, 2, ..., 23\) of the chosen day.

**Variables used:**
1. \(\text{ttl}\) – Total number of records.
2. \(\text{HP}_{t_i}\) – High Price for the time slot \(t_i\).
3. \(\text{LP}_{t_i}\) – Low Price for the time slot \(t_i\).
4. \(u_i, d_i, n_i\) – no. of Downward, Neutral and Upward trend records in time slot \(t_i\)
5. \( nmg \) - Total Number of ranges.
   (In the above example, \( nmg = 11 \))
6. \( EG \) - Effective Gain
7. \( br \) - \( br \) indicates the best range.
8. \( PHP_{t_i-1} \) - Previous High Price.
9. \( PLP_{t_i-1} \) - Previous Low Price.

Initializations:
1. \( PHP_{t_i-1} \leftarrow OP \)
2. \( PLP_{t_i-1} \leftarrow OP \)

**Algorithm:**
1. Calculate \( OG \) for \( OP \).
2. For each record in \( P \), if GOO value lies in \( [\text{int}(OG), \text{int}(OG) + 1] \)
   copy the record into a temporary database table \( Temp \).
3. \( ttl \leftarrow \) Total number of records in \( Temp \).
4. For each time slots \( t_i \), \( i = 1, 2, \ldots, 23 \) do step up to step 15

   // **Calculate High Gain for a time slot:**
6. Count \( u, d \) for the gain period HP\( t_i \) from \( Temp \)
   // here \( u \) is the number of U's and \( d \) is the number of D's
7. \( br \leftarrow (u_r-d_r)/ttl * nmg. \)
8. Identify an \( RG_j \) from HGSUP table whose support value is \( (br)^{th} \) best
    under \( t_i \).
9. \( \text{Gain} \leftarrow j^* (u_r-d_r) / ttl \)
10. expected high price in time slot \( t_i \leftarrow \text{Gain} + PHP_{t_i-1}. \)

   // **Calculate Low Gain for a time slot:**
11. Count \( u, d \) for the gain period LP\( t_i \) from \( Temp \)
    // here \( u \) is the number of U's and \( d \) is the number of D's
12. \( br \leftarrow (d_r-u_r)/ttl * nmg. \)
Identify an RG\(_j\) from HLSUP table whose support value is (br)\(^{th}\) best under \(t_i\).

Gain \(\leftarrow j^* \frac{(d_i-u_i)}{\text{ttl.}}\)

Expected low price in time slot \(t_i\) \(\leftarrow\) Gain + PLP\(_{t_{i-1}}\).

**Sample computation**

Let us workout the process for a day 01-06-2009. The data is for Wipro scrip, and on the chosen day the opening price of the scrip is Rs. 387. The previous day its closing price was Rs. 378.90. Since the support for the higher gain ranges is too small we have chosen the ranges from 1 to 11 for this data. The gain on opening (GOO) is Rs 8.1, and the records from pattern table with GOO between 8.0 and 9 are copied to the temporary table Temp. Total number of records in the table Temp is 18.

In the 1\(^{st}\) time slot:

The count of patterns in high gain for "U" is 9
"D" is 0
"N" is 9.

The best range \(br_i\) is computed using

\[
br_i \leftarrow \left\lfloor \frac{(u_i-d_i)}{\text{No. of records in table Temp}} \right\rfloor \times \text{No. of ranges}
\]

\[
br_i = \left\lfloor \frac{(9-0)}{18} \right\rfloor \times 11 = 5.5
\]

Hence the range to be selected is the 6\(^{th}\) smallest sorted value under T1 of support table.

And it is 5-6. The upper limit is 6. Hence the gain is:

Gain = 6*9/18 = 3.0

Hence the projected High Price in the 1\(^{st}\) time slot is 387+3.0 = 390.00

It is observed that the actual high price in the slot is 390.20

Similarly

The count of patterns in low gain for "U" is 0
"D" is 3
"N" is 15.

The best range \( b_r \) is computed using

\[
\begin{align*}
\text{br} & = \left( \frac{d_i - u_i}{\text{No. of records in table Temp}} \right) \times \text{No. of ranges} \\
\text{bi} & = \left( \frac{3 - 0}{18} \right) \times 1.83
\end{align*}
\]

Hence the range to be selected is the 2\(^{nd}\) smallest sorted value under T1 of support table.

And it is 0 to -1. The lower limit is -1. Hence the gain is:

\[
\text{Gain} = -1 \times \frac{3}{18} = -0.16
\]

Hence the projected Low Price in the 1\(^{st}\) time slot is 387 - 0.16 = 386.84

It is observed that the actual low price in the slot is 384.00

Further values are calculated and represented in the below table.

Table 4.5. Sample Prediction Table

<table>
<thead>
<tr>
<th>DOT</th>
<th>GOO</th>
<th>OPEN</th>
<th>P11</th>
<th>P12</th>
<th>P21</th>
<th>P22</th>
<th>P31</th>
<th>P32</th>
<th>P41</th>
<th>P42</th>
<th>P51</th>
</tr>
</thead>
<tbody>
<tr>
<td>01-Jun-09</td>
<td>8.1</td>
<td>387</td>
<td>390.51</td>
<td>387</td>
<td>391.1</td>
<td>386.8</td>
<td>391.7</td>
<td>386.7</td>
<td>392.3</td>
<td>386.6</td>
<td>393.2</td>
</tr>
<tr>
<td>05-Jun-09</td>
<td>-3.3</td>
<td>393</td>
<td>393.13</td>
<td>392.9</td>
<td>393.4</td>
<td>392.8</td>
<td>393.9</td>
<td>392.8</td>
<td>394.3</td>
<td>392.8</td>
<td>394.8</td>
</tr>
<tr>
<td>08-Jun-09</td>
<td>0.9</td>
<td>395</td>
<td>398.51</td>
<td>395</td>
<td>399.1</td>
<td>394.8</td>
<td>399.7</td>
<td>394.7</td>
<td>400.3</td>
<td>394.6</td>
<td>401.2</td>
</tr>
<tr>
<td>11-Jun-09</td>
<td>-9.9</td>
<td>431.9</td>
<td>431.95</td>
<td>431.9</td>
<td>432.2</td>
<td>431.9</td>
<td>432.3</td>
<td>431.9</td>
<td>432.4</td>
<td>431.9</td>
<td>432.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P52</th>
<th>P61</th>
<th>P62</th>
<th>P71</th>
<th>P72</th>
<th>P81</th>
<th>P82</th>
<th>P91</th>
<th>P92</th>
<th>P101</th>
<th>P102</th>
<th>P111</th>
</tr>
</thead>
<tbody>
<tr>
<td>386.58</td>
<td>193.8</td>
<td>386.58</td>
<td>394.33</td>
<td>386.6</td>
<td>394.9</td>
<td>386.6</td>
<td>395.7</td>
<td>386.5</td>
<td>396.4</td>
<td>386.5</td>
<td>397</td>
</tr>
<tr>
<td>392.78</td>
<td>395.3</td>
<td>392.72</td>
<td>395.7</td>
<td>392.8</td>
<td>396.2</td>
<td>392.9</td>
<td>396.6</td>
<td>392.9</td>
<td>397.1</td>
<td>392.7</td>
<td>397.7</td>
</tr>
<tr>
<td>394.58</td>
<td>401.8</td>
<td>394.58</td>
<td>402.33</td>
<td>394.6</td>
<td>402.9</td>
<td>394.6</td>
<td>403.7</td>
<td>394.5</td>
<td>404.4</td>
<td>394.5</td>
<td>405</td>
</tr>
<tr>
<td>431.86</td>
<td>432.5</td>
<td>431.86</td>
<td>432.62</td>
<td>431.9</td>
<td>432.7</td>
<td>431.9</td>
<td>432.8</td>
<td>431.9</td>
<td>432.9</td>
<td>431.8</td>
<td>433</td>
</tr>
</tbody>
</table>

The table shows a sample prediction done for 1\(^{st}\), 5\(^{th}\), 8\(^{th}\), and 11\(^{th}\) June 2009. Let us compare the predicted values with the actual price of the WIPRO data for the selected days.
Table 4.6. Sample Comparison of Predicted and Actual Prices for few days.

<table>
<thead>
<tr>
<th>Date</th>
<th>DOT</th>
<th>GOO</th>
<th>OPEN</th>
<th>P11</th>
<th>P12</th>
<th>P31</th>
<th>P32</th>
<th>P231</th>
<th>P232</th>
<th>CLOSE</th>
<th>GAIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>6/1/2009</td>
<td>8.1</td>
<td>387.00</td>
<td>390.51</td>
<td>386.97</td>
<td>391.70</td>
<td>386.73</td>
<td>404.95</td>
<td>386.57</td>
<td>406.30</td>
<td>19.30</td>
<td></td>
</tr>
<tr>
<td>Actual Price</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6/5/2009</td>
<td>-3.30</td>
<td>393.00</td>
<td>393.13</td>
<td>392.93</td>
<td>393.86</td>
<td>392.83</td>
<td>403.02</td>
<td>392.93</td>
<td>395.90</td>
<td>2.90</td>
<td></td>
</tr>
<tr>
<td>Actual Price</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6/8/2009</td>
<td>0.90</td>
<td>395.00</td>
<td>398.51</td>
<td>394.96</td>
<td>399.70</td>
<td>394.73</td>
<td>412.95</td>
<td>394.57</td>
<td>399.65</td>
<td>4.65</td>
<td></td>
</tr>
<tr>
<td>Actual Price</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6/11/2009</td>
<td>-9.90</td>
<td>431.90</td>
<td>431.95</td>
<td>431.89</td>
<td>432.32</td>
<td>431.86</td>
<td>433.77</td>
<td>431.89</td>
<td>424.90</td>
<td>-7.00</td>
<td></td>
</tr>
<tr>
<td>Actual Price</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Observations

The observation of support table values justifies the following:

Support for a specific gain starts at higher value for the beginning few time slot and reaches a stable value, and starts increasing as we approach the end of the day.

Figure 4.2. Observation 1
Figure 4.3. Observation 2
Support value for higher ranges goes on decreasing in all time slots.

A careful study of the table values leads to the price prediction function, where

Predicted Price = f(time interval, support range, opening price of the scrip).

Opening price of WIPRO scrip on 1st June 2009 is 387=00. Using the algorithm discussed above the expected price in the 1st time interval is 390.51 and the actual price reached is 390.20. Similarly in the 2nd interval the expected price calculated is 386.97 and the actual price reached is 384.00. Like wise it is observed that the difference in predicted value to the actual price reached is around ±2% overall.
4.4.4 Improved mechanism

In the entire process of prediction we have solely depended on the opening price of the scrip for the day. As the day progresses we will be knowing the exact price of the scrip from time to time. For prediction of High and Low prices for the time slot \( t \) we can use the actual High and Low prices that the scrip has reached in the time slot \( t-1 \), thereby improving the efficiency of the prediction further. By fine tuning the method, we can reduce the error to less than 1% by considering the actual price of the scrip from the previous time interval. That is the expected price in the time interval \( t \) will be:

\[
\text{Expected Price}(t) = \text{Actual price} \ (t-1) + \text{gain/loss calculated} \ (t).
\]

This same work may be extended for “On-demand forecasting of stock price” in real-time. In this mechanism an investor is asked to give a reference price of the selected stock for which he wants to transact (Buy or Sell). From the method we have adopted earlier, for the given day we project the predicted price for all the time slots. This allows the investor to choose the timing of his transaction. The system can also help an investor to identify whether the stock reaches his expected price for the day or not.

4.5. Accuracy of Prediction

Accuracy of the method is done through predicting 23 time slot prices of Wipro stock for the month of June, 2009. Deviation of Rs.1 being taken as threshold it is found that the accuracy is about 99.5%.

4.6. Conclusions

Proposed approach is useful in predicting the market trends in their beginnings of the transactions. We proposed an approach to predict the intraday price of a stock using the historic data. Considerably large data is used for experimental verification of our approach to predict the future prices of the stocks. Given the time stamped transactions, the stock data is represented as pattern database and similarity profiled temporal
association mining is used to discover all associated pattern records that are of relevance. Using the support value for different price gain and the opening price of the stock for the day, we extracted all the significant pattern records from the pattern database. Using the current trend of the stock, we could project the future prices from time to time for the day as efficiently as possible. Having created the pattern table, we are also working on sequential patterns, and on classification of these sequential patterns to come out with classifiers and there by predict the future patterns for a week or for a specific day.

The algorithm uses existing tables of pattern table and support value tables, and creates a temporary table which will have similar value of day opening gain records. This table at the most will contain few ten's of records. The memory requirement is mainly for the existing tables, it is observed that the existing table created for WIPRO data for 5 years (2005 to 2009) required 0.45MB and the support table required 4KB. Even if we consider an application for 100 stocks simultaneously being handled the required memory will not be too large in this method. Hence a specific study on the memory requirement is not done.

In future, we wish to extend this work by considering the frequency at which these trends are taking place in a day for a specific gain/loss.