CHAPTER III

PROBLEM DEFINITION AND SYSTEM MODEL

3.1 PROBLEM STATEMENT

Precise definition of our problem is, “converting the data into information for decision support”. Broadly, data lying in the news articles about a company can be processed to get the information for predicting the stock prices of that company and the news articles reflect the information that is not available in the current price of the stock.

The investors who invest in stock markets are usually unaware of the stock market behaviour. Many are facing the problem of stock trading as they do not know which stocks to buy and which to sell in order to gain more profits. And moreover they understood that the behaviour of the stock market depends a lot, on the relevant news items. But they need to go through a lot of news articles manually and analyse them in a very short span of time. But analysis of such amount of financial news and articles in order to extract useful knowledge exceeds human capabilities. Text mining techniques can help them automatically extract the useful knowledge out of textual resources. Many researchers showed that news articles have a greater impact on predictions of the stock market. At the same time, analysis of past price developments and the traditional time series analysis, where predictions are made solely based on the technical and fundamental data, cannot be ignored.

The thesis aims at developing a system which is able to provide the investors, an informed decision, regarding the prediction of stocks by exploiting the advantage of both traditional technical analysis and sentiment analysis. Traditional technical analysis is based on the historic stock prices and Sentiment analysis is done with techniques which automatically classify the news articles. For technical analysis moving average indicator and for sentiment analysis lexicon method and NBC were used. This system is used to model the reaction of the stock market and predict their direction. By doing so, the investors are able to foresee the future behaviour of their stocks when relevant news are released and acted upon immediately. As input we use real-time news articles and intra-day stock prices of some companies in Bombay Stock Exchange.
We also built a system that uses neural network which takes the historical data and sentiment analysis of news articles and predicts the future stock price. We further investigate the role of the negative events on the direction of stock market. The attempt will be made to take last 7 years of price points of all stocks on BSE (app. 16,80,000), 7 years of news articles from primary sources (app. 5,00,00,000 from times of India, money control, governmental websites etc.), 7 years of company announcement data from BSE India website (app. 1,87,000 to include stock split announcements, insider trading disclosure, dividend declarations, and other announcements).

3.2 OBJECTIVES

To solve the above defined problem, there are three main objectives of our system:

1. To provide evidence that the stocks have a correlation with the sentiments in the news items by building a real time system to predict the future stock price of a company. Stock price of company is based mainly on three aspects: the company’s profile, the historical prices and news articles of that company published in the various media. Our work aims at considering the last two aspects. A database has been created by using the historical data. The news articles of previous the day and the historical prices will be given as inputs to the system which learns the patterns and produces the predicted price of that stock for the next day.

2. To establish a system which gives overall performance of a financial institute based on technical and sentiment analysis. Once an investor decides to invest in a stock, he/she wants to know the health of a company. Many researchers used either technical analysis or sentiment analysis to check the performance of the company’s profile. Our system combines goodness of both the worlds to calibrate the company’s performance.

3. To build a system that would analyze and predict the variations in stock prices over a timeline, based on the sequence of events. The work done so far in the stock market prediction took only the single event into account. The stock price movement of the next day was predicted based on the event that occurred on the previous day. Our system is built around the stock market movements at the time the multiple events occur.
3.3 SYSTEM MODEL

The overall system architecture is shown in the Figure 3.1. The main components of the system architecture are data extraction, data pre-processing, Sentiment classification and the Extraction of historical values.

The stock prediction models make use of the websites in which the news articles are published. The other input is the set of historical stock data which is also available in the websites. The historical stock prices are extracted and kept in the database. The system crawls on to the specified websites and the related news articles are extracted. These news articles are stored in a file. The file of news articles is given as input to sentiment extraction module which gives the overall sentiment of news articles. These, along with the historical stock prices, are used to predict the rise or fall of the stock prices.

The main modules of the system are:

- Neural Network Predictor
- Combined Technical and Sentiment Analysis
- Price prediction Model

In the first module Neural Network Predictor, the artificial neural network has been used which takes the inputs of historical stock prices and the sentiment score of the news articles and produces the predicted prices.

The following Figure 3.2 shows work flow in the module neural network predictor.

![Figure 3.2: Work flow of neural network predictor module](image)

The module, combined technical and sentiment analysis exploits the advantages of both technical analysis and sentiment analysis and gives the overall health of a company. This gives the investor, an informed decision over, whether or not to invest in the company under consideration. The following figure Fig. shows the work flow in the module combined technical and sentimental analysis.
The price prediction model takes different events into consideration while predicting the movements of the stocks. The following figure shows the work flow of price prediction model module.