Chapter VIII

CONCLUSION
8.1 Findings and suggestions

The research work has been carried out to enhance the accuracy of prediction by using web Sentiment analysis techniques. In Web Mining, data are preprocessed before any mining techniques are applied in the discovery of knowledge. Web mining techniques are affected by the presence of irrelevant, redundant and insignificant features in the high dimensional data. These features affect the accuracy of prediction, clustering, classification, regression and others. Therefore, there arises a need to enhance accuracy of the prediction and other learning approaches. Sentiment Analysis or Opinion Mining is the computational treatment of opinions, sentiments and subjectivity of text. Sentiment analysis, a branch of digital analytics aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. Opinion Mining is a process of extraction of knowledge from the opinion of others about some particular topic or a problem. The ability to leverage such voice of the web to gain consumer, brand reputation and market insights can be truly differentiating and valuable in today’s organizations. This work investigated whether sentiment analysis of public mood derived from larger scale Twitter feeds can be used to identify important events and predict movements of stock prices. The researcher analyzed the overall sentiment score of the companies to predict future movements of its stock prices.
The research work analyzed the correlation between Twitter sentiment and stock prices using different techniques. In web sentiment analysis, a number of approaches are used for prediction of the stock market price movements. But using twitter data and scoring sentiment to predict stock price approaches is new dimension for prediction of the stock price movement. The proposed systems are implemented in R tool for obtaining data from Tweeter and using Excel and SPSS for findings and improving the accuracy of prediction. The overall findings of this research work are provided here. All these aspects become important from the point of view that text categorization or polarity of the text is an emerging and a very relevant field in Information Technology. Following are the overall findings of this research work: Twitter APIs can be accessed only via authenticated requests. Twitter uses Open Authentication and each request must be signed with valid Twitter user credentials. This Streaming API is the best versatile tool to access data from twitter. Twitter API is a vital pre-processing step for the Information Retrieval. Only if a better API is adopted along with OAuth authentication it is possible to get a high quality data for classification. The new algorithm is proposed to obtain the significant features used for analyzing the overall sentiment for each object by computing the weighted average for all the sentiments in the textual data. The company sentiment score and their stock price have strong relationship between them. The overall prediction is significant positive correlation between overall sentiment of a firm and its respective stock prices and when research predicts the closing prices of a firm by using a combination of opening prices and overall sentiment, the information criterion is low, thus it leads to higher accuracy of prediction. This study concluded that sentiment prediction is a very difficult task. It
is highly dependent on the audience whose reactions to news are being predicted. The high amount of variability in the audience to social media news, whether a particular user decides to react to news or not, the personal preferences of these audiences, all influence their reactions to the twitter news. In this research it is shown that the company sentiment score and their stock price have strong relationship between them. The overall prediction is significant positive relation between overall sentiments of a firm and its respective stock prices. The proposed algorithms have been simulated and tested with various data sets. From the results, it is evident that accuracy of prediction used to confirm the web sentiments are good indicators of future prediction of stock price movements. This is evident from experimental results on the influence of news categories on web sentiments and accuracy of prediction.

8.2 Significance of the Proposed Research Work

The significance of the research is focused on three factors namely theory, methodology and applications. With regard to theory, there are a number of theories used in the field of Web mining system. In machine learning, unsupervised learning is used to find the various component mixtures by finding the relevant features from the prediction. With regard to methodology, four different classifiers or approaches such as correlation, regression, ALM, ANN are used to predict the stock price movements and to find the irrelevant, redundant and insignificant features of prediction. The significance of these algorithms is that the convergence of the algorithm is fast using statistical approach. There is a combination of unsupervised
learning with statistical weighing that helps to identify the relevant features in the prediction. With regard to applications, prediction algorithms are applied to five different domains such as Telecommunication Services, Retail Industry, Automobiles Industry, Information Technology Industry, and FMCG Industry. The relevant predictions are identified from the different domains and used different application datasets. By identifying the relevant techniques, the accuracy of prediction is improved.

8.3 Limitations of this Research Work

In this section, the limitations of the research are listed: There are many words whose polarity changes from domain to domain. Some word may be positive in one domain and the same may be negative in another domain. This is to differentiate between opinionated and non-opinionated text. This is used to enhance the performance of the system by including a subjectivity detection module to filter out objective facts. Although there exists a number of techniques for finding relevant predictions, the following techniques namely ALM, ANN are considered in the research for prediction. The experimentation of the proposed algorithms is done only by utilizing the numerical datasets.

8.4 Directions for Future Works

The approach described in this thesis is therefore only reliably usable within the constraints of the corpus we have collected. In future work, we propose an improved system for sentiment prediction based on the lessons learned during the work on this thesis. Due to time constraint, the research has been restricted to a
sample, but in future, people could use Twitter sentiment analysis in real time to predict the price movements of any stock continuously, which will also improve the accuracy of prediction. Even researchers can identify new ways of classifying the textual data into various moods such as happiness, alertness, certainty, and calmness. In this research, Twitter is the only social media taken into account, but various platforms such as, Stock Twits, Yahoo Finance, Facebook, blogs, discussion forums can also be analyzed. The research work presented in this thesis has identified new directions for future research. This experimental work is an improvement on the accuracy prediction using three different algorithms. In this work, the five different numerical datasets are used from Learning Repository. Further, the research work may be extended and analyzed with categorical datasets. This can be extended by evaluation criterion measures for finding the relevant features and for improving the accuracy of prediction.