CHAPTER 7
CONCLUSION AND FUTURE SCOPE

The rise of social media has fuelled interest in sentiment classification. Promptly and correctly classifying sentiment from the text has become an important task for individuals and companies. In the development of prediction models to classify the reviews, more reliable approaches are expected to reduce the misclassifications. In this study, the results of various hybrid methods are empirically evaluated on datasets of different size for use in sentiment mining. Among the methods used, hybrid ensemble method (HEM1) is highly robust in nature for balanced data models I, II and III, which is studied through various quality parameters. The analysis also shows that the compound combination of unigram, bigram and trigram performs better for almost all the prediction methods. To handle imbalance data distribution in real time applications, it is observed from the results that using SVMs for class prediction can be influenced by the data imbalance, although SVMs can adjust itself well to some degree of data imbalance. To cope with the problem, rebalancing the data is chosen as a promising direction, but both under sampling and over sampling have limitations. Through extensive experiments with benchmark and real application datasets, the proposed modified bagging method is shown to be effective and superior to several other methods with different data sampling methods. The results also proved that the PCA is a suitable dimension reduction method for hybrid methods for both balanced and imbalanced datasets.

In future, the effect of various other feature reduction techniques like latent dirchlet allocation can be investigated. Further experiments should be conducted in the future to evaluate the
impact of various domain and region specific parameters. Extending sentiment mining to other domains may lead to interesting new results. In future, the use of more combination of n-grams and feature weighting that gives a better accuracy level than this can be considered. The work done in this research is only related to classification sentiment into two of the classes (binary classification) that is a positive class and negative class. In the future development, a multiclass of sentiment classification such as positive, negative, neutral and so on might be taken into consideration. In this work, the focus is on finding features that appear explicitly as nouns or noun phrases in the reviews. The finding of implicit features is left to future work. As ensemble learning methods need a lot of computing time, parallel computing techniques should be explored to tackle this problem. A major limitation of ensemble learning methods is the lack of interpretability of the results and the knowledge learned by ensembles is difficult for humans to understand. Therefore improving the interpretability of ensembles is another important research direction.

Future opinion-mining systems need broader and deeper common and commonsense knowledge bases. This will lead to a better understanding of natural language opinions and will more efficiently bridge the gap between multimodal information and machine processable data. Blending scientific theories of emotion with the practical engineering goals of analyzing sentiments in natural language text will lead to more bio-inspired approaches to the design of intelligent opinion-mining systems capable of handling semantic knowledge, making analogies, learning new affective knowledge, and detecting, perceiving, and “feeling” emotions.