CHAPTER 6
CONCLUSION AND FUTURE WORK

6.1 Conclusion

This work investigated prediction of defects in software modules using data mining techniques and software metrics. The research goal namely

- To study data mining techniques for software defect prediction

- To propose a novel method to preprocess the software metrics based on the defect class.

- To propose a novel Neural Network classifiers using a fuzzy hidden layer to further improve the classification, could be achieved

KC1 dataset from the promise repository was used in the investigation Probability based classification algorithm - Bayesian Logistic Regression, Decision Tree based classification algorithms - Classification and Regression Tree (CART) and ensemble of classification tree - Random tree were investigated for the accuracy of its classification, measured in the range of 81-86%. The area under the ROC curve varies in the range of 0.5 to 0.6 and hence highly ineffective.
To overcome the shortcomings, a novel pre processing technique based on the cumulative distributed function and probability density function on the class label was proposed. The same classifiers were used to find the classification accuracy. The output obtained was in the range of 94-97%.

A novel neural network classifier based on the Multi layer perceptron model was proposed with a novel fuzzy bell hidden layer to squash the data. The classification accuracy of the proposed system for defect prediction increased to 98.2%. The proposed method is compared with other work in the literature with good results.

Defect prediction module plays a very crucial role in the industry. This research address the improvement in the classification accuracy as even as small change in percentage improves the quality of service. A typical software project module has 1000’s of modules and an improvement in accuracy by 1% gives significant improvement in managing the project.

6.2 Future work

Investigations using the proposed method need to be carried out in software modules obtained from real time projects developed by a large number of users. Though such real time datasets are currently not available in the public domain, the proposed technique in thesis may be extended for investigation and study in the following datasets.
• KC2, CM1, JM1, PC1 dataset from NASA data metrics program

• The proposed Neural Model can be studied and extended to the pre processing work done by


  ▪ Song et al., (2011)

  ▪ Menzies (2007)

Haider et al., (2008) proposed a novel Estimation of Defects based on Defect Decay Model ($ED^3M$) which estimate of the total number of defects in an ongoing testing process. Since attribute selection technique is different from regularly available data in literature, it could be interesting to study the proposed work with this work.

Song et al., (2011) observed that selecting attributes based on historical data is useful for software defect prediction and obtained classification accuracy of upto 90.4% on different datsets. Investigating her attribute selection method with our classification method might throw new light on the capability of our algorithm.

Menzies (2007) showed improvement in classification accuracy by ranking the importance of attributes using a novel technique. The proposed method can reduce the computation time for defect prediction.