ABSTRACT

Software metrics have been used to define the complexity of the software program and to estimate programming time. Extensive research has also been carried out to quantify defects in a module using software metrics. Data mining techniques have been used for defect prediction in software modules. Cyclomatic complexity, Halstead metrics and size of the software are the most commonly used metrics in the defect prediction models.

The goal of this research is to help IT team identify defects based on existing software metrics using data mining techniques and thereby improve software quality which ultimately results in reducing the software development cost in the development and maintenance phase. This research focuses on identifying defective modules and therefore the scope of software that needs to be examined for defects can be prioritized. Since the proposed methodology helps in identifying modules that require immediate attention, the reliability of the software can be improved faster by handling higher priority defects first.

This research focuses on classifying defects based on software metrics and can be broadly classified into three areas:

- Investigating existing data mining techniques for software defect prediction using software metrics.
• Propose a novel preprocessing technique based on cumulative distributed function and normalized distribution function.

• Implement a novel neural network classification algorithm, Fuzzy Bell Multi Layer Perceptron (FB- MLNN) Neural Network architecture for software defect prediction process model based on the software metrics.

Investigations show that the proposed preprocessing technique improves the classification accuracy over 10% with various classifiers producing classification accuracies of 94.55 to 96.79%. The proposed FB-MLNN achieves a classification accuracy of 98.2% which is a significant improvement over other methods. The sensitivity and specificity of the proposed method is converges which is highly essential for identifying defective modules. The proposed method is compared with other latest work in the literature using KC1 dataset and it is found that the proposed MLP neural network model in the range of 1.2% to 15.7%.