Preface

Software defects are the major issue of any software industry. Software defects not only reduce the software quality, increase costing but also suspend the development schedule. New technologies, new software releases are upcoming speedy. The current scenario of the software industry which focuses on the above issue has motivated me to work on the research, titled “Software Defect Prediction Using Data Mining Techniques”.

Many data mining techniques and dataset repository are available to predict the software defect. To accomplish the research goal, the data is collected from Open Software repositories. For my research Promise repository is taken for the experiment as it is a publicly available, authentic repository. The product metrics, that is Chidamber and Kemerer metrics suite have been used and the software modules named as Ant, Ivy, Tomcat, Berek, Camel, Lucene, POI , Synapse and Velocity were taken to discover the most optimal / significant software metrics. I have used regression technique for the analysis as my response variable (dependent) is continuous data. In a larger dataset all the variables are not important; it is always desirable to reduce the variable and to include the important variable in a dataset. For the research wrapper and filter method like Boruta, regsubset, FSelector (Random Forest, Information Gain, Linear Correlation and Rank Correlation) were used to get the best software metrics. The Machine learning technique is an important branch of computers for software bug prediction. Machine learning can be classified as supervised learning and unsupervised learning. Supervised learning can be categorized as Regression and Classification. Regression technique were applied on machine learning models like Linear Regression, Decision Tree, Random Forest, Neural Network, Support Vector Machine and Decision Stump to predict the best model for software bug prediction. Various measures were also discussed to prevent defects in future software version releases.