

**CHENNAI – PONDICHERRY**

**Toward Comprehensible Software Fault Prediction Models Using Bayesian Network Classifiers**

**Abstract**

Software testing is a crucial activity during software development and fault prediction models assist practitioners herein by providing an upfront identification of faulty software code by drawing upon the machine learning literature. While especially the Naive Bayes classifier is often applied in this regard, citing predictive performance and comprehensibility as its major strengths, a number of alternative Bayesian algorithms that boost the possibility of constructing simpler networks with fewer nodes and arcs remain unexplored. This study contributes to the literature by considering 15 different Bayesian Network (BN) classifiers and comparing them to other popular machine learning techniques. Furthermore, the applicability of the Markov blanket principle for feature selection, which is a natural extension to BN theory, is investigated. The results, both in terms of the AUC and the recently introduced H-measure, are rigorously tested using the statistical framework of Demšar. It is concluded that simple and comprehensible networks with less nodes can be constructed using BN classifiers other than the Naive Bayes classifier. Furthermore, it is found that the aspects of comprehensibility and predictive performance need to be balanced out, and also the development context is an item which should be taken into account during model selection.