

# Jordan University of Science and Technology

## A Guided Oversampling Technique to Improve The Prediction of Software Fault-Proneness for Imbalanced Data

**Authors:** Raed Shatnawi and Ziad Al-Sharif

**Abstract:** Fault-proneness is one of the most tackled quality factors in the field of software quality. Predicting the probability of the faulty classes is necessary information to guide developers in their endeavour to improve the software quality and to reduce the costs of testing and maintenance. The performance of the fault prediction models suffers greatly from the imbalance of fault distribution, i.e., the majority of modules are not faulty whereas the minority are only faulty. The imbalanced distribution of faults affects the efficiency of prediction models greatly. In this paper, we discuss many oversampling techniques that are used to improve the performance of prediction models. We propose to guide the oversampling process using the fault content (i.e., the number of faults in a module). This study is conducted on a large object-oriented system - Eclipse. The proposed oversampling is tested on ten classifiers. The results of this work shows that using fault content in sampling has better prediction performance than other traditional oversampling techniques. The decision trees and nearest neighbours have shown outstanding performance whereas other classifiers have shown acceptable performance.