Applying Cost-Sensitive Classification for Financial Fraud Detection under High Class-Imbalance

Stephen O. Moepya ∗†, Sharat S. Akhoury†, Fulufhelo V. Nelwamondo∗†,

∗Department of Electrical and Electronic Engineering Science, University of Johannesburg, Johannesburg, South Africa

†Council for Scientific and Industrial Research, Pretoria, South Africa

Email: smoepya,sakhoury,fnelwamondo@csir.co.za

Abstract

In recent years, data mining techniques have been used to identify companies who issue fraudulent financial statements. However, most of the research conducted thus far use datasets that are balanced. This does not always represent reality, especially in fraud applications. In this paper, we demonstrate the effectiveness of cost-sensitive classifiers to detect financial statement fraud using South African market data. The study also shows how different levels of cost affect overall accuracy, sensitivity, specificity, recall and precision using PCA and Factor Analysis. Weighted Support Vector Machines (SVM) were shown superior to the cost-sensitive Naive Bayes (NB) and K-Nearest Neighbors classifiers.