

**FRAUD DETECTION MECHANISM FOR MOBILE
MONEY TRANSACTIONS USING MACHINE LEARNING
TECHNIQUES.**

Thimali Yasoda Midigaspege

A dissertation submitted in partial fulfilment of the requirement
for Master of Science degree in Business Analytics

Department of Computing

Informatics Institute of Technology, Sri Lanka

In collaboration with

Robert Gordon University, UK

2022

Abstract

Mobile money transactions have been rapidly growing all with universal presence of mobile phones. Using mobile phones people can perform unbanked transactions which is quick and easy. With the overwhelming growth in mobile transactions, fraudulent activities have also grown in a significant rate. Even though popularity of mobile transactions is beyond the predictions still mobile devices, apps and service providers does not provide 100% secured systems and security. Therefore, it is crucial to construct highly effective fraud detection mechanism for mobile money transactions. However, there are very few literature studies related to mobile money fraud detection where other financial fraud detections are well studied and guided. This can be due to the novelty of the technology, rapid growth and limited available data sets in the field. To address this gap, this research is conducted with objective of constructing highly effective fraud detection mechanism using machine learning techniques.

A publicly available synthetic PaySim data set has been used for the research which is generated in 2016 using a mobile service provider in Africa, and transactions have been collected over one month. Dataset includes 6,320,620 transactions where only 8213 fraud transactions are included. This makes it a highly imbalanced data set which will be a huge barrier when using classification algorithms. Classification algorithms tend to be biased for majority class when imbalance data set is present. Resampling techniques and resemble methods have been used in the research to mitigate this limitation. Synthetic minority Over-Sampling technique (SMOTE) and hybrid resampling technique SMOTE-Tomek Links removal methods are used. Ninety Six models have been designed in combination with three data sets (imbalance, SMOTE, SMOTE-Tomek Links removal), three different feature scale methods and eight different classification algorithms, namely: logistic regression, random forest, support vector machine, decision trees, naïve bayes, gradient boosting and ada boost. The performance of the models has been evaluated using confusion matrix, precision, recall, f1 score, ROC_AUC, and execution time. The experimental resulted that random forest classifier is a highly effective model in combination with SMOTE-Tomek Links removal and minmax scaler.