IDENTIFY SUSPECIOUS TRANSACTIONS

M W Jagath Renuka Punyarathna

MSc

2021

Abstract

In the banking industry, there are many types of risks in money transferring activities such as terrorist financing, money laundering, fraudulence transactions related to business activities and use of non-business platforms for business related transactions to avoid taxes, documentations etc. These suspicious financial transactions are usually detected by the use of classification algorithms and continuous monitoring activities of different business units of the organization. Even-though, various AML software and screening methods available in financial institutions, still the fraudulent transactions are generating through remittance channel.

However, in the classification, the main challenge is the skewed distribution of classes which is simply demonstrated as class imbalanced issue due to limited availability of data of the minor class (in this project, suspicious transactions) compared to the major class (legitimate transactions). In this case, special data mining approaches are used along with regular classification approaches to overcome afore mentioned issue.

While considering the available dataset which is used to construct the training models, particularly a small portion of minority group (suspicious transactions) is obtainable. Further, it is more complicated to correctly classify the minority group compared to the majority with the limited data which is extracted from the core banking system. Classification of a transaction as fraudulence is danger than classification of the same as a legitimate due to consequences.

Accordingly, eight machine learning models (Random Forest, Randomly Sampling, KNN, Naïve Bayes, Logistic Regression, Decision Tree and Support Vector Machine) have been applied to detect fraudulent cases. These models are used to detect already identified fraudulent cases in the dataset in order to prove the accuracy and acceptability of a suitable model/s out of selected 8 models.

During the evaluation of the performances of each models, various accuracy levels and sensitivity ratios have been identified and after a comparison of ratios, Random Under Sampling model and Randomly Over Sampling Example (ROSE) have been identified as the suitable models for further studies.

Key Words : Machine Learning, Workers' Remittances, Fraudulent Transactions, Suspicious Transactions, Business Intelligence