

Master's Thesis : Machine Learning Techniques for Money Laundering Detection

Auteur : L'hoest, Julien

Promoteur(s) : Geurts, Pierre

Faculté : Faculté des Sciences appliquées

Diplôme : Master : ingénieur civil en informatique, à finalité spécialisée en "management"

Année académique : 2019-2020

URI/URL : <http://hdl.handle.net/2268.2/10720>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

Machine Learning Techniques for Money Laundering Detection - Summary

Author:
L'hoest Julien

Supervisor:
Marcos Alejandro

Section:
Computer Engineering - Management

Academic year:
2019-2020

Summary

Some economic phenomena have a significant negative impact on financial institutions. Money laundering is one of them. Money laundering is the processing of illicit funds into the financial system to make them appeared from legitimate sources. It is usually detected through rule based monitoring. Unfortunately, it does not exploit the potential of data. Machine leaning approaches learn automatically from data patterns that cannot be captured by rule-based ones. Money laundering in transaction logs are defined as an imbalance problem with regard to machine learning. The challenge is to detect few fraudulent transactions over a huge population which compose a financial database.

This thesis presents the design of a simulator for transaction data generation to answer to the lack of available data needed for machine learning algorithms. Then a theoretical research has been done in order to identify machine learning algorithm designed for classification problem with skew class distribution. The machine learning models which have been selected from this research are the Hellinger Distance Decision Tree (HDDT), the Isolation Forest (iForest), the One-Class SVM (ocSVM) and an approach based on successive hypothesis testing using the Fisher, George and Pearson methods.

These approaches have been tested over synthetic data from the simulator. The model with the highest F1 score is a Random Forest composing of fully expended decision trees using the Gini split criterion. The second best performing model was the Isolation Forest.