

Nom et Prénom : RTAYLI NAOUFAL

Date de soutenance : 06/07/2021

Directeur de Thèse : ENNEYA NOURDDINE

Sujet de Thèse :

Enhancing Credit Card Fraud Detection Using Machine Learning Methods: Design and Development of a Hybrid Fraud Detection Model

Résumé :

Le développement d'un modèle efficace pour la détection des transactions frauduleuses n'est pas une tâche facile. Pour surmonter les défis liés à la détection de ce type de fraude, notamment dans des données volumineuses et non équilibrées, il faut que les praticiens identifient les attributs à utiliser, la stratégie (supervisée ou non-supervisée), et l'algorithme à implémenter. Avec des transactions étiquetées de ce qui est considéré comme frauduleux et de ce qui est considéré comme normal, des modèles hybrides ont été conçues et développées durant cette thèse pour détecter avec une haute précision les transactions frauduleuses et pour minimiser les fausses alertes. Le résultat de ce travail est l'élaboration de trois modèles hybrides. Le premier est basé sur la sélection des attributs les plus prédictifs à l'aide du classifieur forêts aléatoire (RFC) et d'algorithmes d'apprentissage automatique tels que les machines à vecteurs de support (SVM), Isolation Forest (IF). Le deuxième est nommé $RFC(HPO, RFE)$, est un modèle avancé (un prototype d'un système de détection de la fraude par carte de crédit) qui vise à générer des alertes précises ; en diminuant le taux de faux positifs et à assurer la détection des transactions frauduleuses ; en réduisant le taux de faux négatifs, au prix de fausses alertes. Le troisième présente une version étendue du $RFC(HPO, RFE)$, nommé ST-BPNN. Ce modèle a exploité les performances de l'apprentissage profond (Deep learning) qui a été combiné avec une technique hybride de rééquilibrage des ensembles de données très déséquilibrées. Les trois modèles sont appliqués à de grands ensembles de données du monde réel pour évaluer la précision de leurs capacités en matière de détection de la fraude.

Abstract :

Developing an effective model for detecting fraudulent transactions is a hard task. To overcome the challenges of detecting this type of fraud, especially in large and imbalanced data, practitioners need to identify the attributes to use, the strategy (supervised or unsupervised), and the algorithm to implement. With transactions labeled as fraudulent and normal, hybrid models were designed and developed during this thesis to detect fraudulent transactions with high accuracy and to minimize false alerts. The result of this work is the development of three hybrid models. The first one is based on the selection of the most predictive attributes using the Random Forest Classifier (RFC) and machine learning algorithms such as Support Vector Machines (SVM), Isolation Forest (IF). The second is named $RFC(HPO, RFE)$, is an advanced model (a prototype of a credit card fraud detection system) that aims to generate accurate alerts; by decreasing the false positive rate as well as ensuring the detection of fraudulent transactions by reducing the false-negative rate, at the cost of false alerts. The third represents an extended version of the $RFC(HPO, RFE)$, is named ST-BPNN. This model exploited the performance of deep learning which was combined with a hybrid technique for rebalancing highly imbalanced datasets. All three models are applied to large real-world data sets to assess the accuracy of their fraud detection capabilities.

