

N/A

Auteur : Saulas, Adrien

Promoteur(s) : Debruyne, Christophe

Faculté : Faculté des Sciences appliquées

Diplôme : Master : ingénieur civil en science des données, à finalité spécialisée

Année académique : 2023-2024

URI/URL : <http://hdl.handle.net/2268.2/21035>

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.



UNIVERSITÉ DE LIÈGE

FACULTÉ DES SCIENCES APPLIQUÉES

Balancing Durability, Performance, and Interpretability in Unbalanced Data as Fraud Detection

ATFE9009-1 - Master thesis

Author:

SAULAS Adrien

Supervisor:

Prof. DEBRUYNE

Christophe

Intech Advisor:

NOGATCHEWSKY

Matthieu

August 21, 2024

Abstract

The problem of fraud detection is one of the most discussed topics in the field of machine learning. This study addresses four key areas essential for a fraud detection platform: prediction accuracy in imbalanced datasets, interpretability of predictions, deployment and sustainability of the platform, monetary costs associated with model errors. To tackle these issues, we first conducted extensive research in the field, then proposed and evaluated our solutions. We introduce methods such as using a WCGAN (Wasserstein Conditional Generative Adversarial Network) for sampling or cost-sensitive learning with new models like Light Gradient Boosting, employing interpretable models like Explainable Boosting, deploying and automating training processes with Kubernetes and Kubeflow, and utilizing approaches like thresholding or tuning metrics that account for monetary costs. Each of these solutions shows promising results and improves upon existing research in the field.