

Projet n°069

Titre : Hybrid and Explainable Models for Financial Fraud Detection

Description : Contexte et partenariat :

Dans la continuité du travail de recherche mené en 2024-2025, ce projet vise à approfondir les approches hybrides de détection de fraude dans des contextes d'audit réel. Il s'inscrit dans une collaboration active avec l'entreprise Zacca International, spécialisée dans l'audit, qui met à disposition un jeu de données authentique issu de ses missions d'audit financier.

Les travaux préliminaires ont démontré l'efficacité de la combinaison entre des méthodes statistiques (loi de Benford, test χ^2) et des modèles d'apprentissage automatique (XGBoost, Isolation Forest). L'enjeu pour l'année à venir est d'industrialiser, fiabiliser et expliciter ces approches pour en faire un véritable outil d'aide à la décision pour les auditeurs.

Objectifs du projet :

1. Concevoir et tester un méta-modèle de détection de fraude fondé sur le stacking de plusieurs modèles complémentaires (XGBoost, RF, OCSVM, LOF, etc.).
2. Intégrer des techniques d'explicabilité (XAI, SHAP, LIME, ou TreeExplainer) pour faciliter la prise de décision et rendre les résultats interprétables par les auditeurs.
3. Proposer une évaluation réaliste des performances en tenant compte du bruit, du déséquilibre des classes et des contextes métier spécifiques.
4. Développer un prototype logiciel interactif (un pipeline répliquable en Python avec interface exploratoire comme streamlit ou dash) basé sur les données de Zacca International, intégrant visualisation et priorisation des cas suspects.

Informations sur le projet

Compétences développées : Entraînement de modèles supervisés (XGBoost, Random Forest, MLP) et non supervisés (LOF, Isolation Forest, OCSVM). • Implémentation de stacking, blending ou voting classifieurs. • Simulation d'étiquettes manquantes + apprentissage semi-supervisé (Label Propagation, pseudo-labelling). • Utilisation de SHAP, LIME ou TreeExplainer pour l'interprétation des décisions. • Déploiement d'un dashboard (Streamlit ou Dash) pour démonstration.

Informations complémentaires : Encadrement et partenariat : • Supervision académique : Nancy CHENDEB (nancy.chendeb@devinci.fr) • Partenaire industriel : Abdallah ZIADE, Zacca International - données réelles + appui métier

Majeure(s) concernée(s) : DIA, IF, FIN, CCC

Année(s) concernée(s) : A4, A5

Acceptera une équipe A4 si aucune équipe A5 ne se positionne : Oui

Mot(s)-clé(s) concerné(s) : Audit, Data Science, Finance, IA, Machine learning, Python

Informations sur le partenaire

Entreprise / Association / École : ESILV

Nom : CHENDEB

Prénom : Nancy

