



## **Robust and Explainable Ensemble Learning for Fraud Detection in Financial Audit: From Model to Dashboard**

### **Contexte et partenariat :**

Ce projet s'inscrit dans la continuité directe du travail mené en 2024-2025, et bénéficie du soutien d'un partenariat actif avec l'entreprise Zacca International, spécialisée dans l'audit. Zacca fournit des données financières réelles et anonymisées ainsi qu'un appui métier précieux pour évaluer la pertinence des résultats.

Les travaux antérieurs ont comparé plusieurs modèles (Random Forest, XGBoost, Autoencoders...) sur des données déséquilibrées, révélant la nécessité d'un cadre plus intégré, explicable et opérationnel pour la détection de fraude.

**Mots clés :** Fraud Detection, Audit Automation, Hybrid Models, Explainable AI, Autoencoders, Ensemble models, ML, deepLearning

### **Objectifs du projet :**

1. Concevoir un pipeline de détection de fraude complet en Python, intégrant prétraitement, détection, visualisation et export.
2. Construire un méta-modèle robuste par stacking de plusieurs classifieurs supervisés et non supervisés.
3. Intégrer des outils d'explicabilité (SHAP, LIME) pour rendre les résultats compréhensibles par les auditeurs.
4. Développer un dashboard interactif (via Streamlit ou Dash) pour permettre une analyse visuelle des cas détectés, facilitant ainsi l'aide à la décision.

---

### **Méthodologies envisagées :**

- Préparation des données (nettoyage, normalisation, réduction de dimension).
- Rééquilibrage (SMOTE, Tomek Links) et évaluation croisée.
- Entraînement de modèles (RF, XGBoost, SVM, MLP, Isolation Forest).
- Conception du dashboard avec visualisation des métriques, importance des variables, et explications locales (XAI).

---

### **Livrables attendus :**

- Un article scientifique en anglais à soumettre dans une conférence ou revue spécialisée.

## Projet de recherche 2025-2026

- Un pipeline Python modulaire documenté, basé sur scikit-learn, imbalanced-learn, SHAP, Plotly.
- Un dashboard prêt à l'emploi pour Zacca International, avec scénarios de démonstration.
- Un rapport final incluant les recommandations pour un déploiement opérationnel en audit.

### Encadrement et partenariat :

- Supervision académique : Nancy CHENDEB (nancy.chendeb@devinci.fr)
- Partenaire industriel : Abdallah Ziade, Zacca International – données réelles + appui métier

### Plan sur 6 mois

Mois	Phase principale	Activités clés	Livrables clés
1	Préparation & cadrage	- Compréhension des données Zacca - Nettoyage, normalisation, réduction de dimension	Dataset prêt + rapport d'analyse exploratoire
2	Modélisation de base & rééquilibrage	- Entraînement des modèles individuels (RF, XGBoost, MLP, etc.) - Application de SMOTE, Tomek	Modèles individuels évalués + scripts
3	Construction du méta-modèle	- Implémentation du stacking (supervisé et non supervisé) - Validation croisée et robustesse	Méta-modèle robuste + comparatif de scores
4	Explicabilité & interprétation (XAI)	- Application de SHAP, LIME sur les prédictions - Analyse des variables influentes	Rapport XAI + visualisations interactives
5	Développement du dashboard	- Création d'une interface Streamlit/Dash - Visualisation : scores, variables, XAI	Prototype de dashboard fonctionnel
6	Tests finaux & valorisation	- Tests utilisateur avec auditeurs - Ajustements du pipeline - Rédaction du rapport final	Rapport final + dashboard + support article + dépôt GitHub