

Context-Aware and Explainable Anomaly Detection in Financial Transactions: Building a Meta-Model for Trustworthy Audit Automation

Contexte et partenariat :

À la suite du projet mené en 2024–2025, qui a exploré l’analyse sémantique et la détection non supervisée d’anomalies financières (via Isolation Forest, One-Class SVM et incohérences de taux de change), ce sujet vise à aller plus loin sur trois dimensions :

1. L’enrichissement contextuel des données (temps, nature de la dépense),
2. L’explicabilité des décisions algorithmiques,
3. L’agrégation intelligente des modèles via un méta-modèle.

Le projet est réalisé en partenariat avec Zacca International, entreprise d’audit, qui apporte les données réelles et le cadre métier.

Mots clés : Audit Automation, Anomaly Detection, Semantic Features, Unsupervised Learning, contextually enriched dataset

Objectifs du projet :

1. Construire un méta-modèle de détection d’anomalies basé sur les résultats de plusieurs détecteurs (Isolation Forest, OCSVM, LOF, autoencoders).
2. Intégrer des variables contextuelles (temporelles, structurelles, sémantiques) dans les modèles.
3. Mettre en œuvre une pipeline Python modulaire allant de la préparation des données à la priorisation des cas.
4. Développer un dashboard interactif en Python (Streamlit ou Dash) intégrant :
 - la visualisation des anomalies,
 - les scores des différents modèles,
 - les explications locales via SHAP/LIME,
 - un module d’historique pour auditabilité.

Méthodologies envisagées :

- Extraction de caractéristiques : nature du flux, variation temporelle, profils utilisateur.
- Modèles de base : Isolation Forest, One-Class SVM, Local Outlier Factor, autoencoder.
- Méta-modèle (niveau 2) : logistic regression ou arbre de décision entraîné sur les scores produits par les modèles de base.
- Explicabilité : SHAP pour la vision globale, LIME pour les cas individuels.
- Construction du dashboard avec filtres dynamiques, export Excel/PDF, API REST éventuelle.

Livrables attendus :

- Un article scientifique sur les approches hybrides explicables en détection d'anomalies financières.
- Un pipeline modulaire open-source réutilisable.
- Un dashboard professionnel pour Zacca International, prêt à être intégré en phase de test.

Encadrement et partenariat :

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

Plan 6 mois

Mois	Phase principale	Activités clés	Livrables clés
1	Cadrage & enrichissement contextuel	- Analyse métier avec Zacca - Intégration de variables contextuelles (temps, nature, structure...)	Dataset enrichi + document de spécification
2	Modélisation des détecteurs unitaires	- Implémentation d'algos : Isolation Forest, OCSVM, LOF, Autoencoders - Comparaison initiale	Résultats comparatifs des détecteurs + scripts
3	Construction du méta-modèle	- Fusion des scores via stacking ou vote pondéré - Validation croisée - Optimisation des seuils	Méta-modèle validé + benchmark
4	Explicabilité & auditabilité (XAI)	- Application de SHAP et LIME - Visualisation des explications locales - Préparation des logs	Rapport XAI + module d'historique pour audit
5	Développement du dashboard	- Conception d'une interface interactive (Streamlit/Dash) - Intégration : visualisation, scores, XAI	Dashboard fonctionnel + documentation utilisateur
6	Tests finaux & valorisation	- Tests utilisateurs avec Zacca - Ajustements pipeline + interface - Dépôt GitHub + rapport final	Rapport final + dépôt GitHub + démonstration