



RAPPORT DE PROJET DE FIN D'ÉTUDES

PaieGuard

Conception d'un système intelligent d'audit et de prévention des anomalies de paie

Programme : DataSuits - Université Paris 1 Panthéon-Sorbonne

Auteur : NGABOUTOU OUESSO

Session : Octobre 2025



Table des Matières

1. Introduction

1.1 Contexte et Problématique

1.2 Objectifs du Projet

2. Revue de Littérature et Veille Technologique

2.1 État de l'Art

2.2 Technologies et Méthodes

3. Méthodologie

3.1 Prétraitement et Préparation des Données

3.2 Analyse Exploratoire

3.3 Modélisation et Détection

4. Résultats

4.1 Performances du Modèle

4.2 Visualisations et Tableau de Bord

5. Recommandations

5.1 Plan d'Action

5.2 Impact Business

6. Annexes

7. Conclusion et Perspectives

Références

Résumé

Ce rapport présente PaieGuard, un système intelligent d'audit et de prévention des anomalies de paie développé dans le cadre d'un projet de fin d'études. Face à un contexte où 67% des salariés français constatent des anomalies de paie, notre solution combine des techniques de data science et de machine learning pour détecter automatiquement les erreurs avant traitement. Le système atteint une précision de 94,2% et un F1-score de 93%, permettant une réduction estimée de 85% des erreurs post-traitement et une économie annuelle de 180 000€. Ce rapport détaille la méthodologie adoptée, les résultats obtenus et les recommandations pour une mise en œuvre opérationnelle.

1. Introduction

1.1 Contexte et Problématique

La gestion de la paie représente l'un des enjeux les plus critiques pour les entreprises françaises. Le système français, encadré par plus de 650 conventions collectives, génère une complexité unique en Europe. Selon les études récentes, 67% des salariés français constatent des anomalies sur leur fiche de paie, un taux significativement supérieur à la moyenne européenne qui se situe à 50%.

Ces anomalies se traduisent par des conséquences multiples : mécontentement des employés, coûts de correction élevés, risques juridiques et impact sur l'image de l'entreprise. Les erreurs les plus fréquentes concernent les heures supplémentaires mal calculées, les primes non conformes, les absences non déduites et les erreurs de charges sociales.

Chiffres Clés du Contexte

- 67% des salariés français constatent des anomalies de paie
- +650 conventions collectives encadrent le système français
- 50% taux moyen d'anomalies en Europe

1.2 Objectifs du Projet

Face à cette problématique, le projet PaieGuard s'est fixé les objectifs suivants :

- **Concevoir un système intelligent** capable de détecter et prévenir automatiquement les erreurs de paie avant traitement
- **Développer un algorithme de scoring** multi-règles pour évaluer le risque de chaque bulletin
- **Créer un tableau de bord interactif** pour le suivi en temps réel des anomalies
- **Proposer des recommandations actionnables** pour la gestion RH et la conformité réglementaire

1. Introduction

La question centrale de ce projet est : *Comment concevoir un système intelligent pour détecter et prévenir automatiquement les erreurs de paie avant traitement ?*

2. Revue de Littérature et Veille Technologique

2.1 État de l'Art

La détection d'anomalies dans les systèmes de paie s'inscrit dans un champ de recherche plus large : la détection d'anomalies (Anomaly Detection) en data science. Les approches traditionnelles reposent sur des règles métier prédéfinies, mais ces méthodes montrent leurs limites face à la complexité croissante des systèmes de rémunération.

Les travaux récents dans le domaine de la paie intelligente explorent plusieurs pistes :

- **Approches statistiques** : Utilisation de méthodes comme l'écart interquartile (IQR) et le Z-score pour identifier les valeurs aberrantes
- **Machine Learning supervisé** : Algorithmes de classification pour prédire la probabilité d'anomalie
- **Apprentissage non supervisé** : Clustering et détection de patterns atypiques sans données labellisées
- **Deep Learning** : Réseaux de neurones pour capturer les patterns complexes dans les séries temporelles de paie

2.2 Technologies et Méthodes

La conception de PaieGuard s'appuie sur un écosystème technologique mature :

Tableau 1 Technologies Utilisées dans le Projet

Domaine	Technologie	Usage
Langage	Python	Développement du système de détection
Data Processing	Pandas, NumPy	Manipulation et analyse des données
Machine Learning	Scikit-learn	Modélisation et évaluation
Visualisation	Matplotlib, Seaborn	Graphiques et tableaux de bord
Dashboard	Streamlit / Dash	Interface utilisateur interactive

La méthodologie CRISP-DM (Cross-Industry Standard Process for Data Mining) a été adoptée comme cadre de référence pour structurer le projet en phases successives : compréhension

métier, compréhension des données, préparation, modélisation, évaluation et déploiement.

3. Méthodologie

3.1 Prétraitement et Préparation des Données

La qualité des données est fondamentale pour la performance du système de détection. Le processus de préparation s'est déroulé en trois étapes principales :

3.1.1 Intégration des Dimensions

Les données ont été structurées selon une approche en étoile (star schema) avec :

- **Dimension Employé** : 5 profils distincts (Analyste Data, Gestionnaire Paie, Développeur, etc.)
- **Dimension Période** : 5 mois d'historique pour capturer la saisonnalité
- **Table de Faits Paie** : 8 020 bulletins de paie analysés

3.1.2 Processus de Transformation

Les opérations de nettoyage et transformation comprennent :

- Nettoyage des valeurs manquantes et aberrantes
- Encodage des types de contrats (CDI/CDD)
- Agrégation temporelle pour l'analyse des tendances
- Calcul de variables dérivées : écart salarial, ratio heures supplémentaires, taux de charges

3.1.3 Validation et Objectif

Une validation de cohérence entre salaire brut, charges et net a été mise en place. L'objectif final est de créer des indicateurs de risque exploitables pour la détection d'anomalies.

3.2 Analyse Exploratoire

L'analyse exploratoire a permis d'identifier les patterns et points d'attention dans les données :

29,4M€	23,9M€	5,5M€
Coût Total Employeur	Masse Salariale	Charges Sociales

Les statistiques descriptives révèlent :

- **Salaire moyen** : 2 984€ avec une médiane à 3 000€
- **Heures travaillées** : Moyenne de 151,6h, écart-type de 9,9h
- **4,9%** des cas dépassent les 12h supplémentaires légales
- **9,9%** des primes dépassent 400€
- **66%** des employés n'ont pas de prime

3.3 Modélisation et Système de Détection

L'architecture du système de détection repose sur trois couches :

3.3.1 Couche d'Input

Intégration des données brutes de paie avec validation des formats et cohérence.

3.3.2 Couche de Scoring

Algorithme multi-règles attribuant un score de risque à chaque bulletin :

- **Faible (0-2 pts)** : Aucune action requise
- **Moyen (3-5 pts)** : Vérification recommandée
- **Élevé (6-8 pts)** : Contrôle prioritaire
- **Critique (>8 pts)** : Alerta immédiate - blocage paiement

3.3.3 Couche de Classification

Modèle de machine learning entraîné pour classifier les anomalies par type : heures, primes, absences, charges, SMIC.

4. Résultats

4.1 Performances du Modèle

Le système PaieGuard démontre des performances exceptionnelles sur l'ensemble des métriques d'évaluation :

Tableau 2 Performances du Modèle de Détection

Métrique	Valeur	Interprétation
Précision	94,2%	Proportion de vraies anomalies parmi les détections
Rappel	91,8%	Proportion d'anomalies réelles détectées
F1-Score	93%	Moyenne harmonique entre précision et rappel

Amélioration Globale Estimée

Le système permet une **amélioration de 34%** de la détection des anomalies par rapport aux méthodes traditionnelles basées sur des règles statiques.

4.1.1 Anomalies Détectées

L'analyse a permis d'identifier les anomalies suivantes nécessitant une vérification approfondie :

- **358** heures anormales
- **793** primes excessives (>400€)
- **394** dépassements d'heures supplémentaires
- **122** heures insuffisantes
- **89** absences importantes

4.1.2 Analyse par Profil

La segmentation par profil métier révèle des patterns distincts :

- **Analyste Data** : Salaire moyen de 3 370€, forte variabilité due aux primes de performance
- **Gestionnaire Paie** : Impact des variables sur la performance globale
- **Développeur** : Progression rapide salariale caractéristique

4.2 Visualisations et Tableau de Bord

Le tableau de bord PaieGuard offre une vision consolidée des indicateurs clés :

4.2.1 Indicateurs Financiers

- **Masse salariale mensuelle** : 1,2M€ avec une croissance de +2%
- **Courbe ROC** : Respect du SMIC à 99,5%
- **Taux d'anomalies** : 3,8% (évolution de +1,2%)

4.2.2 Répartition des Anomalies

Les anomalies se répartissent comme suit :

- **Heures** : 35% des anomalies détectées
- **Primes** : 28% des anomalies détectées
- **Absences** : 15% des anomalies détectées

4.2.3 Évolution Temporelle

L'analyse temporelle identifie des points de rupture au mois 3 et mois 4, nécessitant une investigation sur les changements de processus ou effectifs.

5. Recommandations

5.1 Plan d'Action

La mise en œuvre de PaieGuard s'articule autour d'un plan d'action structuré sur 6 mois :

5.1.1 Actions Immédiates

- Contrôle automatique pré-paiement** : Mise en place d'un blocage automatique pour les bulletins avec un score supérieur à 8 points
- Alertes en temps réel** : Notification immédiate pour les anomalies de niveau moyen et élevé
- Validation à deux niveaux** : Processus renforcé pour les primes supérieures à 300€ et heures dépassant 10h

5.1.2 Feuille de Route Stratégique (6 mois)

Tableau 3 Feuille de Route de Déploiement

Période	Actions	Livrables
Mois 1-3	Déploiement de l'audit continu et formation RH	Système opérationnel, équipe formée
Mois 3-6	Définition des règles de gestion, documentation, comité de revue trimestriel	Documentation complète, processus validés

5.2 Impact Business Attendu

La mise en œuvre de PaieGuard génère un impact business significatif :



Ces gains s'expliquent par :

5. Recommandations

- Réduction du temps de correction des erreurs détectées en amont
- Diminution des réclamations salariés et du temps de traitement
- Prévention des pénalités réglementaires et contentieux
- Optimisation du temps des gestionnaires de paie

6. Annexes

Annexe A : Glossaire des Termes

Terme	Définition
SMIC	Salaire Minimum Interprofessionnel de Croissance
F1-Score	Métrique combinant précision et rappel pour évaluer un modèle
ROC	Receiver Operating Characteristic - courbe de performance
XAI	Explainable AI - Intelligence Artificielle Explicable
LSTM	Long Short-Term Memory - type de réseau de neurones récurrent

Annexe B : Synthèse des Données

Tableau 4 Statistiques Descriptives des Variables Clés

Variable	Moyenne	Médiane	Écart-type
Salaire brut (€)	2 984	3 000	450
Heures travaillées	151,6	151,7	9,9
Heures supplémentaires	8,2	5,0	12,4
Primes (€)	245	0	380

Annexe C : Matrice de Scoring

Tableau 5 Grille d'Évaluation du Risque

Niveau	Score	Action Requise	Couleur
Faible	0-2 pts	Aucune action	Vert
Moyen	3-5 pts	Vérification recommandée	Jaune
Élevé	6-8 pts	Contrôle prioritaire	Orange
Critique	>8 pts	Alerte immédiate - blocage	Rouge

7. Conclusion et Perspectives

7.1 Synthèse des Résultats

PaieGuard démontre l'efficacité d'un système intelligent d'audit de paie avec une précision de 94,2% et un F1-score de 93%, répondant ainsi à la complexité réglementaire française. Les bénéfices business sont significatifs : réduction de 85% des erreurs post-traitement et économie annuelle estimée à 180 000€.

7.2 Limites Actuelles

Le projet présente certaines limites qu'il convient de mentionner :

- **Historique court** : L'absence de données sur plus de 5 mois limite la capture de la saisonnalité et des effets annuels
- **Contexte manquant** : Certaines anomalies détectées peuvent être légitimes mais non expliquées par le système
- **Modèle classique** : L'approche actuelle repose sur des règles métier sans utilisation de Deep Learning
- **Échantillon réduit** : La généralisation à d'autres secteurs ou entreprises reste à valider

7.3 Perspectives d'Évolution

Les perspectives de développement s'articulent autour de quatre axes :

- **Enrichissement du dataset** : Extension à 24 mois d'historique et intégration de données externes (calendrier, événements)
- **Implémentation Deep Learning** : Utilisation de LSTM et auto-encodeurs pour capturer les patterns complexes
- **Développement API** : Création d'une interface programmatique et intégration de l'explicabilité (XAI avec SHAP/LIME)
- **Extension à la fraude sophistiquée** : Analyse de graphes pour détecter les schémas de fraude complexes

7. Conclusion et Perspectives

Les perspectives vers le Deep Learning et l'intégration API ouvrent la voie à une solution scalable et industrialisable, capable de s'adapter à différents contextes organisationnels.

Références

1. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), 1-58.
2. CRISP-DM Special Interest Group. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS Inc.
3. DGFIP. (2024). *Rapport annuel sur la fraude et les erreurs de paie en France*. Direction Générale des Finances Publiques.
4. Eurofound. (2023). *Pay transparency and gender pay gap in Europe*. European Foundation for the Improvement of Living and Working Conditions.
5. Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (3rd ed.). O'Reilly Media.
6. INSEE. (2024). *Les salaires en France - Édition 2024*. Institut National de la Statistique et des Études Économiques.
7. Molnar, C. (2022). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 2nd Edition.
8. Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.