

République Algérienne Démocratique et populaire
Ministère de l'enseignement supérieur et de la recherche scientifique

École Nationale Polytechnique
Département Génie Industriel



Mémoire de Projet de Fin d'Études
Pour l'obtention du diplôme d'Ingénieur d'État en Génie Industriel
Option : Data Science & Intelligence Artificielle

Détection intelligente de la fraude dans les
sinistres d'assurance santé

Réaliser par :
ABDELLAZIZ Nour Safouane

Supervisé par :
Imene BARECHE ENP
& Mohamed Ikkal ZELEMAT AXA

Présenté et soutenu publiquement le : 30 Juin 2025

Devant le jury composé de :

<i>Président :</i>	<i>ZOUAGHI Iskander</i>	<i>MCA</i>	<i>ENP</i>
<i>Examinatrice :</i>	<i>BELDJOURDI Samia</i>	<i>MCA</i>	<i>ENP</i>
<i>Promotrice :</i>	<i>BARECHE Imene</i>	<i>MAB</i>	<i>ENP</i>

ENP 2025

République Algérienne Démocratique et populaire
Ministère de l'enseignement supérieur et de la recherche scientifique

École Nationale Polytechnique
Département Génie Industriel



Mémoire de Projet de Fin d'Études
Pour l'obtention du diplôme d'Ingénieur d'État en Génie Industriel
Option : Data Science & Intelligence Artificielle

Détection intelligente de la fraude dans les
sinistres d'assurance santé

Réaliser par :
ABDELLAZIZ Nour Safouane

Supervisé par :
Imene BARECHE ENP
& Mohamed Ikkal ZELEMAT AXA

Présenté et soutenu publiquement le : 30 Juin 2025

Devant le jury composé de :

<i>Président :</i>	<i>ZOUAGHI Iskander</i>	<i>MCA</i>	<i>ENP</i>
<i>Examinatrice :</i>	<i>BELDJOURDI Samia</i>	<i>MCA</i>	<i>ENP</i>
<i>Promotrice :</i>	<i>BARECHE Imene</i>	<i>MAB</i>	<i>ENP</i>

ENP 2025

Remerciement

Je tiens tout d'abord à exprimer ma profonde gratitude à mon tuteur de stage, Monsieur Mohamed Ikbal ZELMAT, pour son accompagnement bienveillant, sa confiance et ses conseils avisés tout au long de mon expérience chez AXA Assurance. Sa disponibilité, son exigence et sa rigueur professionnelle ont grandement contribué à la réussite de ce travail.

Je souhaite également remercier chaleureusement Madame Leïla SALMI, responsable de l'équipe Data Mining & Stratégie, pour son accueil, son soutien et l'intérêt qu'elle a porté à mon projet. Sa vision et son engagement ont été déterminants dans l'orientation et l'aboutissement de mon stage.

Mes remerciements s'adressent aussi à toute l'équipe Data Mining & Stratégie d'AXA Algérie, pour leur esprit de collaboration, leur disponibilité et leur soutien constant. Leur expertise et leur bienveillance ont rendu cette expérience aussi enrichissante qu'agréable.

Je remercie également l'équipe pédagogique de l'École Nationale Polytechnique, et tout particulièrement Mme Imene BARECHE, ma tutrice pédagogique, pour ses conseils précieux et sa disponibilité.

Enfin, je remercie du fond du cœur ma famille, mes amis et toutes les personnes qui m'ont soutenu durant cette aventure, en pensant tout particulièrement à mes parents pour leur soutien indéfectible, leur patience et leurs encouragements constants. Ce mémoire leur doit beaucoup.

الملخص

تُعدّ عمليات الاحتيال في التأمين الصحي من أبرز التحديات التي تواجه شركات التأمين، نظراً لتأثيرها المالي الكبير وتزايد تعقيد أنماطها. وفي ظل محدودية الأساليب التقليدية التي تعتمد على قواعد ثابتة، يقدم هذا البحث منهجية مبتكرة للكشف الآلي عن الاحتيال، تجمع بين تقنيات علوم البيانات، التعلم الآلي، والخبرة الميدانية. تم تنفيذ هذا العمل خلال فترة تدريب في شركة AXA للتأمين في الجزائر، بالاعتماد على بيانات حقيقية للمفاتيح تعويضات التأمين الصحي.

ترتكز المنهجية المقترحة على استراتيجية تصنيف شبه خاضعة للإشراف (semi-supervised labeling) اعتماداً على قواعد مبنية، مع بناء متغيرات سلوكية للمؤمن والمستفيد، وتقليل الأبعاد باستخدام تحليل المكونات الرئيسية (PCA) وتطبيق نماذج تعلم جماعي (Random Forest)، الانحدار اللوجستي، Voting Classifier. كما تم استخدام تقنية SMOTE لمعالجة اختلال التوازن بين الفئات، وتم تعزيز تفسير نتائج النماذج من خلال أدوات توضيحية مثل SHAP وLIME.

أظهرت النتائج التجريبية أداءً قوياً، مع قيمة AUC بلغت 0.937 ونسبة استرجاع (Recall) بلغت 79٪ لفئة الاحتيال. وتسلط المناقشة الضوء على بعض القيود، مثل جودة التصنيفات وعدم توفر بيانات علاقات (Prestataires) مع اقتراح آفاق تطوير مستقبلية، تشمل استخدام الشبكات العصبية البيانية (GNN)، التعلم النشط، وتطبيق استراتيجيات الإنتاج الذكي (MLOps). ويشكّل هذا العمل أساساً منهجياً صلباً يمكن تطويره وتكييفه للكشف الذكي عن الاحتيال في قطاع التأمين الصحي.

الكلمات المفتاحية: التأمين الصحي، الاحتيال، التعلم الآلي، التصنيف شبه الخاضع للإشراف، Random Forest، SHAP، SMOTE، GNN أكسا.

Abstract

Health insurance fraud poses a growing threat to insurance companies due to its financial impact and increasingly sophisticated patterns. Traditional rule-based approaches are often limited in their adaptability and efficiency. This thesis presents an innovative and automated methodology for fraud detection, combining data science techniques, machine learning, and domain expertise. The work was carried out during an internship at AXA Assurance Algeria, using real-world claims data from the health insurance sector.

The proposed approach is based on a semi-supervised labelling strategy using expert-defined rules, the construction of behavioral features per claim and per beneficiary, dimensionality reduction through Principal Component Analysis (PCA), and ensemble learning methods (Random Forest, logistic regression, and a VotingClassifier). Class imbalance was addressed using SMOTE, and model interpretability was enhanced with tools such as SHAP and LIME.

Experimental results show strong performance, with an AUC of 0.937 and a recall of 79% on the fraud class. A critical discussion highlights the limitations related to label noise and the absence of relational data, while suggesting improvements through the integration of Graph Neural Networks (GNN), active learning, and industrialization via MLOps pipelines. This work provides a robust and adaptable methodological foundation for developing intelligent fraud detection systems in health insurance.

Keywords : health insurance, fraud detection, machine learning, semi-supervised labelling, Random Forest, SHAP, SMOTE, GNN, AXA.

Résumé

La fraude à l'assurance santé constitue un enjeu majeur pour les compagnies d'assurance, en raison de son impact financier croissant et de la complexification des schémas frauduleux. Dans un contexte où les approches traditionnelles reposant sur des règles fixes montrent leurs limites, ce mémoire propose une méthodologie innovante de détection automatisée de la fraude, en combinant techniques de data science, apprentissage automatique et expertise métier. Le travail a été mené dans le cadre d'un stage au sein d'AXA Assurance Algérie, à partir de données réelles de gestion des sinistres santé.

L'approche développée repose sur une stratégie de labellisation semi-supervisée fondée sur des règles métier, la construction de variables comportementales par sinistre et par bénéficiaire, la réduction de dimensionnalité via l'analyse en composantes principales (PCA), et l'utilisation de modèles d'ensemble (Random Forest, régression logistique, VotingClassifier). Des techniques d'équilibrage comme SMOTE ont permis d'atténuer le déséquilibre entre classes, tandis que des outils d'explicabilité tels que SHAP et LIME ont renforcé l'interprétation des résultats.

Les résultats expérimentaux montrent une performance robuste, avec une AUC de 0,937 et un rappel de 79 % sur la classe fraude. Une discussion critique met en évidence les limites liées à la qualité des labels et à l'absence de données relationnelles, tout en proposant des pistes d'amélioration, notamment par l'intégration de modèles graphiques (GNN), l'apprentissage actif et l'industrialisation MLOps. Ce mémoire offre ainsi une base méthodologique solide et adaptable pour le développement de systèmes intelligents de détection de fraude en assurance santé.

Mots-clés : assurance santé, fraude, apprentissage automatique, labellisation semi-supervisée, Random Forest, SHAP, SMOTE, GNN, AXA.

Table des matières

Liste des tableaux	12
Liste des figures	12
Introduction générale	11
1 L'assurance santé face à la fraude : état des lieux	14
1.1 Introduction	14
1.2 Définition de l'assurance santé	14
1.2.1 Modalités de l'assurance maladie	14
1.2.2 Assurance santé dans le contexte algérien	15
1.3 Contenu du contrat d'assurance maladie et sa gestion	17
1.4 Les Systèmes de Gestion des Assurances Maladie : Fonctionnement, Acteurs et Rôles	19
1.5 Les Acteurs des Systèmes de Gestion des Assurances Maladie	20
1.6 L'organisation du secteur de l'assurance santé en Algérie	22
1.6.1 Organismes publics : CNAS et CASNOS	22
1.6.2 Émergence des assurances privées	22
1.6.3 Contexte Institutionnel : AXA Assurance Algérie	23
1.6.4 Garanties proposées par les assurances privées	23
1.7 Typologie des fraudes en assurance santé	23
1.7.1 Fraude à la souscription	24
1.7.2 La fraude par multi-assurance	25
1.7.3 La fraude au sinistre	25
1.7.4 Autres schémas frauduleux fréquents en assurance santé	26
1.8 Le positionnement de la fraude dans l'assurance	27
1.8.1 Fraude à la souscription d'un contrat d'assurance	28
1.8.2 Fraude lors de la gestion des sinistres	29
1.9 Impact économique de la fraude à l'assurance santé	30
1.9.1 Répercussions systémiques et stratégiques	30
1.9.2 Mécanismes de lutte et enjeux techniques	31
1.9.3 Considérations spécifiques à l'assurance santé en Algérie	31
1.10 Objectifs de l'étude	32
1.11 Modélisation économique de la solution : Business Model Canvas	32

1.12 Conclusion	34
2 Intelligence artificielle et détection de fraude : état de l'art	35
2.1 Introduction	35
2.2 Intelligence Artificielle	35
2.3 Catégories d'Apprentissage	36
2.4 Régression logistique	38
2.5 Prétraitement des données et gestion du déséquilibre	46
2.5.1 Cycle de vie d'un modèle d'apprentissage automatique en détection de fraude	47
2.5.2 Problématique du déséquilibre des classes	47
2.5.3 Techniques de gestion du déséquilibre	49
2.5.4 Impact sur la détection de fraude et workflow opérationnel	51
2.6 Réseaux de Neurones Artificiels (ANN)	54
2.7 Réseaux de Neurones Graphiques (GNN)	56
2.8 Applications des Graph Neural Networks (GNN)	57
2.9 Synthèse des recherches scientifiques sur la détection de la fraude en assurance santé	60
3 Méthodologie proposée pour la détection de la fraude chez AXA	76
3.1 Introduction	76
3.2 Schéma global du pipeline de détection de fraude	77
3.3 Contexte Métier de l'étude et Jeu de données	78
3.4 Structuration du jeu de données AXA et contraintes opérationnelles	78
3.5 Définition des règles métier de fraude chez AXA	82
3.6 Stratégie de labellisation semi-supervisée	83
3.6.1 Principe et intérêt	83
3.6.2 Formalisation des règles	83
3.6.3 Analyse sur la qualité des labels	83
3.7 Construction des variables comportementales	86
3.7.1 Variables comportementales par bénéficiaire	86
3.7.2 Variables comportementales par sinistre élémentaire	87
3.8 Construction des scores latents (réduction dimensionnelle)	87
3.9 Modélisation supervisée par apprentissage en ensemble	90
3.9.1 Mise en œuvre technique : interface web de détection de fraude	97
4 Évaluation et discussion des résultats	101
4.1 Introduction	101
4.2 Analyse et interprétation des résultats obtenus	101
4.2.1 Qualité de la détection et impact opérationnel	105
4.2.2 Discussion de la labellisation et le bruit	106
4.2.3 Explicabilité, adoption et confiance	106
4.3 Limites et identifiés	106
4.4 Propositions et axes d'amélioration	107

4.4.1	Industrialisation métier	109
4.5	Perspectives de recherche	110
4.6	Conclusion	113

Liste des tableaux

1.1	Estimations du coût de la fraude à l'assurance par pays et par branche	30
1.2	Business Model Canvas du projet de détection de fraude chez AXA Assurance Algérie avec chiffres réalistes	33
2.1	Définitions de l'intelligence artificielle selon différentes approches	36
2.2	Synthèse des techniques de sampling	51
2.3	Résumé des principales métriques de classification pour la détection de fraude	54
2.4	Principales performances des modèles de la littérature récente . .	67
3.1	Exemples de règles de détection de fraude en assurance santé . .	82
3.2	Variables comportementales par bénéficiaire	86
3.3	Variables comportementales par sinistre élémentaire	87
3.4	Statistiques descriptives des variables comportementales des bénéficiaires avant et après PCA	89
3.5	Statistiques descriptives des variables comportementales des sinistres élémentaires avant et après PCA	90
3.6	Distribution des classes avant et après SMOTE	92
3.7	Métriques de classification par classe	94
4.1	Scores moyens en validation croisée (5-fold) et indicateurs de robustesse. . .	102

Table des figures

1.1	Classification des fraudes à l'assurance : à la souscription, au sinistre et de multi-assurance	24
1.2	Typologie des fraudes au sinistre : par exagération, planifiée, provoquée et opportuniste	26
1.3	Processus usuel de souscription d'un contrat d'assurance	28
1.4	Processus usuel de souscription d'un contrat d'assurance	29
2.1	Catégories d'apprentissage en machine learning : supervisé, non supervisé, semi-supervisé et par renforcement	37
2.2	Methode pour traitement des données déséquilibré	48
2.3	Méthodes au niveau des données pour traiter les ensembles de données déséquilibrés	49
2.4	Architecture de réseau de neurone	55
3.1	Pipeline globale de détection de fraude.	77
3.2	Codage binaire des règles métier de détection de fraude.	83
3.3	Calcule le nombre d'indicateurs de fraude par ligne dans un DataFrame	84
3.4	Agrégation des indicateurs statistiques par bénéficiaire pour l'analyse de fraude	87
3.5	Création et agrégation des sinistres élémentaires à partir des données de soins	88
3.6	Entraînement et évaluation d'un Voting Classifier pondéré (logistique + Random Forest)	93
3.7	Interface Web <i>Detect Fraud with AXA AI</i> pour l'analyse automatisée des sinistres	97
3.8	Interface d'authentification sécurisée de Detect Fraud with AXA AI	98
3.9	Interface d'authentification sécurisée de Detect Fraud with AXA AI	99
4.1	L'importance des variables lime	103
4.2	Diagramme de fiabilité avant calibration : le modèle sous-estime la probabilité de fraude dans les cas intermédiaires	104
4.3	Les scores de la courbe d'apprentissage ROC AUC	105

Introduction générale

Contexte

La fraude à l’assurance santé représente aujourd’hui une menace importante pour les compagnies d’assurance, les organismes publics de santé et les dispositifs de protection sociale. Elle se manifeste sous des formes variées telles que les fausses déclarations, la surfacturation, les actes médicaux fictifs ou les collusions entre assurés et prestataires. D’après la *Coalition Against Insurance Fraud*, ces pratiques entraîneraient chaque année plusieurs milliards d’euros de pertes à l’échelle mondiale[1]. Cette situation engendre une pression accrue sur les primes d’assurance, fragilise les mécanismes de solidarité et compromet la viabilité des systèmes de santé.

Dans les pays en développement comme l’Algérie, le phénomène est amplifié par des facteurs structurels : fragmentation des systèmes d’information, absence de base de données centralisée, hétérogénéité des informations médicales et administratives, ainsi qu’une faiblesse des dispositifs de contrôle automatisés. Ces limites rendent les approches traditionnelles de détection – essentiellement fondées sur des règles expertes fixes – peu efficaces face à des schémas de fraude de plus en plus complexes, dynamiques et collaboratifs.

Face à ce constat, les avancées récentes en *intelligence artificielle* (IA), *machine learning*, et en *analyse de graphes* offrent des perspectives nouvelles et prometteuses pour automatiser la détection de comportements frauduleux dans des bases de données massives, même en l’absence de supervision explicite.

Objectifs et contributions

Ce mémoire s’inscrit dans cette dynamique d’innovation technologique et analytique. Réalisé dans le cadre d’un stage au sein de **la compagnie AXA Assurance Algérie**, il a pour objectif de concevoir un système intelligent de détection de fraude dans les sinistres d’assurance santé, à partir de données réelles issues de la gestion opérationnelle.

Les contributions principales de ce travail sont les suivantes :

- La mise en place d’une **stratégie de labellisation semi-supervisée**, fondée sur des règles métier établies en collaboration avec les experts d’AXA ;

- La **construction de variables comportementales** par sinistre et par bénéficiaire, permettant de capturer les habitudes, irrégularités et anomalies des assurés ;
- La **réduction de la dimensionnalité** via l’analyse en composantes principales (PCA) pour optimiser la modélisation ;
- L’utilisation de **modèles d’ensemble** (Random Forest, régression logistique, VotingClassifier) pour la classification des sinistres ;
- L’application de **techniques d’équilibrage des classes** (SMOTE) et d’**outils d’interprétabilité** (SHAP, LIME) pour garantir la transparence des décisions algorithmiques ;
- Une **discussion critique** sur les performances, les limites et les perspectives d’industrialisation (GNN, apprentissage actif, MLOps).

Plan de la thèse

Le mémoire est structuré en quatre chapitres :

- **Chapitre 1 – L’assurance santé face à la fraude : état des lieux**
Présente le fonctionnement de l’assurance santé, les types de fraude, les acteurs impliqués et les enjeux spécifiques au contexte algérien.
- **Chapitre 2 – Intelligence artificielle et détection de fraude : état de l’art**
Analyse les méthodes existantes d’IA utilisées pour la détection de fraude, en mettant l’accent sur les modèles supervisés, les approches graphiques (GNN) et les solutions adoptées dans le secteur santé.
- **Chapitre 3 – Méthodologie proposée chez AXA Assurance Algérie**
Décrit l’architecture du pipeline de détection : préparation des données, labellisation semi-supervisée, création de variables comportementales, modélisation et interprétabilité.
- **Chapitre 4 – Évaluation et discussion des résultats**
Présente les résultats expérimentaux, évalue la performance du modèle (AUC de 0,937, rappel de 79 % sur la classe fraude), et propose des pistes concrètes d’amélioration.

Conclusion

Ce travail contribue à démontrer la faisabilité d’une détection intelligente, automatisée et interprétable de la fraude en assurance santé. Il combine rigueur scientifique et applicabilité opérationnelle dans un contexte réel. Les résultats obtenus ouvrent la voie à l’industrialisation progressive de ces outils dans les processus métiers d’un assureur, tout en respectant les exigences de transparence, d’éthique et de conformité réglementaire. Enfin, les perspectives identifiées – notamment l’intégration de réseaux de neurones graphiques, l’apprentissage actif, et les pipelines MLOps – laissent entrevoir un potentiel considérable pour

l'évolution future des systèmes anti-fraude dans le secteur assurantiel, en Algérie comme ailleurs.

Chapitre 1

L'assurance santé face à la fraude : état des lieux

1.1 Introduction

L'assurance santé occupe aujourd'hui une place centrale dans la protection sociale, en offrant aux individus une couverture contre les risques financiers associés aux dépenses médicales imprévues. Elle repose sur des principes de mutualisation et de solidarité, tout en s'adaptant aux évolutions des besoins et des systèmes de soins. Toutefois, la complexité croissante des produits d'assurance, conjuguée à la diversité des acteurs et des modalités de gestion, expose le secteur à des risques accrus de fraude. Ce chapitre propose une présentation générale de l'assurance santé, en détaillant ses mécanismes, ses spécificités dans le contexte algérien, ainsi que les principaux enjeux liés à la lutte contre la fraude, enjeu majeur pour la pérennité et l'équilibre du secteur.

1.2 Définition de l'assurance santé

L'assurance santé est un mécanisme de transfert de risque qui vise à protéger les individus contre les conséquences financières imprévues liées aux dépenses de santé résultant de maladies ou d'accidents. Elle repose sur la mutualisation des risques et des ressources au sein d'un groupe d'assurés, permettant ainsi à chacun d'être protégé contre des dépenses médicales potentiellement déstabilisantes sur le plan financier [2].

1.2.1 Modalités de l'assurance maladie

L'assurance maladie peut être structurée selon deux modalités principales, chacune ayant ses spécificités en termes de fonctionnement et d'impact sur les assurés :

Les prestations en nature : Dans ce modèle, l'assureur rembourse les frais médicaux engagés par l'assuré, tels que les consultations, les médicaments, les hospitalisations, ou les soins chirurgicaux. Ce type de contrat suit le principe indemnitaire, où les remboursements sont effectués en fonction des dépenses réelles, sans excéder les coûts effectivement supportés par l'assuré. Ce système est largement adopté par les systèmes de santé publics, comme en France, où la Sécurité sociale rembourse une partie des frais médicaux. L'assuré peut bénéficier d'un complément de remboursement par le biais de mutuelles [3].

Les prestations en espèces : Dans ce modèle, l'assuré reçoit une indemnité forfaitaire en cas d'incapacité temporaire ou permanente de travail due à une maladie ou un accident. Ces indemnités visent à compenser la perte de revenus pendant la période d'invalidité. Ce modèle suit le principe forfaitaire, où les montants versés sont prédéfinis dans le contrat, indépendamment des dépenses médicales réelles. Les assurances privées sont souvent davantage tournées vers ce modèle, offrant ainsi un revenu de remplacement pour les travailleurs affectés.

La dualité des modèles : Ainsi, l'assurance maladie est souvent qualifiée de produit hybride, combinant à la fois des éléments indemnitaire et forfaitaire. Cette dualité permet de répondre à la diversité des besoins des assurés, offrant une couverture pour les frais médicaux tout en protégeant contre la perte de revenus. Cependant, cette flexibilité expose également les systèmes d'assurance à des risques de fraude.

Les défis de gestion et les risques de fraude :

La gestion des produits d'assurance maladie présente plusieurs défis, notamment la complexité des prestations et les risques de déficits techniques. Les assureurs doivent équilibrer les prestations offertes avec les primes perçues, tout en mettant en place des mécanismes de gestion des risques efficaces pour assurer la viabilité de ces produits. C'est dans ce cadre que la détection de fraude devient cruciale. Les systèmes de fraude, comme les fausses déclarations de sinistres, l'utilisation abusive de multi-DNI (numéros d'identification), ou encore les falsifications d'actes médicaux, peuvent entraîner des pertes financières importantes pour les assureurs. [4].

1.2.2 Assurance santé dans le contexte algérien

L'assurance maladie en Algérie occupe une place importante dans le paysage assurantiel national. En raison de la croissance des besoins en santé et de l'évolution des attentes des citoyens en matière de protection sociale, ce segment attire une attention croissante de la part des compagnies d'assurance. Bien que confrontée à divers défis économiques et structurels, l'assurance maladie représente un axe stratégique dans le développement du marché des assurances de personnes. Cette section examine l'intérêt spécifique que revêt l'assurance maladie en Algérie, notamment son rôle stratégique dans les portefeuilles des assureurs.

Assurance santé : un produit d'appel stratégique

En Algérie, l'assurance maladie est souvent perçue par les assureurs comme un produit d'appel. Bien qu'elle ne génère pas de profits significatifs et puisse même être déficitaire, elle joue un rôle crucial dans la stratégie commerciale des compagnies d'assurance. En offrant une couverture santé attrayante, les assureurs attirent de nouveaux clients, qui souscrivent ensuite à d'autres produits d'assurance plus rentables, tels que l'assurance automobile ou habitation. Cette approche est particulièrement pertinente dans un contexte où la branche des assurances de personnes, incluant l'assurance maladie, a connu une baisse de 7,2 % de son chiffre d'affaires au premier trimestre 2024, atteignant près de 5,2 milliards de dinars algériens . Malgré cette diminution, l'assurance maladie demeure un levier important pour fidéliser la clientèle et élargir le portefeuille des assureurs.[5]

Analyse des résultats de la branche en Algérie

Selon les données du Conseil National des Assurances (CNA), le chiffre d'affaires global du marché algérien des assurances a atteint 169,6 milliards de dinars en 2023, marquant une hausse de 3,5 % par rapport à 2022 . Cependant, la part spécifique de l'assurance maladie dans ce chiffre d'affaires n'est pas détaillée, ce qui reflète un manque de transparence et de données précises dans ce secteur[5] [6] En ce qui concerne les sinistres, le premier trimestre 2024 a enregistré une augmentation de 14,3 % en montant et de 12,3 % en nombre, totalisant 428 887 dossiers pour un montant de 20,8 milliards de dinars . Cette tendance souligne la pression croissante sur les assureurs pour gérer efficacement les risques et les coûts associés à l'assurance maladie.[5]

Sécurité et accès aux soins pour l'assuré

Pour les assurés, l'assurance maladie représente une sécurité essentielle, offrant un accès aux soins médicaux et une protection financière contre les dépenses de santé imprévues. En Algérie, la Caisse Nationale des Assurances Sociales (CNAS) rembourse 80 % des frais médicaux et pharmaceutiques pour la majorité des assurés, et jusqu'à 100 % pour les personnes atteintes de maladies chroniques ou les titulaires de pensions de retraite ou d'invalidité dont le revenu est égal ou inférieur au Salaire National Minimum Garanti (SNMG) .[7]

Cette couverture contribue à réduire l'absentéisme au travail et à améliorer la productivité, en particulier dans les entreprises où les employés bénéficient d'une assurance maladie complémentaire. Conscient de l'importance de cette protection, l'État algérien a mis en place des mécanismes pour renforcer le système de sécurité sociale et assurer une couverture santé adéquate pour l'ensemble de la population .

1.3 Contenu du contrat d'assurance maladie et sa gestion

L'assurance maladie prend en charge divers frais médicaux liés aux accidents, maladies et maternité, tout en excluant certains risques de manière spécifique ou générale [8, 9].

1. Portée des garanties

Garantie de base

La garantie principale assure le remboursement des frais engagés pour les traitements médicaux, notamment :

- frais d'hospitalisation,
- honoraires médicaux,
- analyses radiologiques,
- frais pharmaceutiques,
- certaines cures thermales.

Sont également couverts :

- les frais liés à la grossesse,
- l'accouchement (chirurgical ou non),
- les suites pathologiques [10].

Concernant les soins dentaires, seuls les actes référencés dans la nomenclature professionnelle sont pris en charge, à l'exception des actes chirurgicaux lourds (coefficient supérieur à six).

Les garanties peuvent être limitées :

- **Géographiquement** : couverture restreinte au territoire national ou étendue à certaines régions selon les coûts médicaux locaux,
- **Temporellement** : application soumise à un délai de carence (ex. : 3 mois pour la maladie, 9 mois pour la maternité, 12 mois pour la lunetterie) [11],
- **Selon la nature des prestations** : couverture limitée aux hospitalisations ou étendue aux soins dentaires.

Extensions de garantie

Certaines prestations initialement exclues peuvent être couvertes par souscription complémentaire :

- évacuation sanitaire (transport vers un établissement adapté),
- soins d'urgence à l'étranger (durant un séjour de courte durée),
- rapatriement du corps en cas de décès à l'étranger,
- frais funéraires (remboursement forfaitaire lié aux obsèques).

Ces extensions sont souvent proposées sous forme de contrats d'assistance médicale internationale [12].

Prestations en espèces

Outre les remboursements, certaines garanties prévoient le versement d'une indemnité journalière en cas d'incapacité temporaire de travail, afin de compenser la perte de revenu durant l'arrêt maladie. Cette indemnité dépend généralement d'un forfait journalier multiplié par le nombre de jours d'arrêt [13].

2. Exclusions de risques

Exclusions classiques

Ces exclusions concernent notamment les dommages liés :

- à la participation à des actes de violence ou mouvements sociaux,
- à des actes intentionnels (ex. : tentative de suicide),
- à la guerre, aux explosions nucléaires ou aux radiations accidentelles [14].

Exclusions spécifiques à l'assurance maladie

Sont notamment exclus :

- traitements esthétiques (hors suites d'accident),
- produits pharmaceutiques non homologués,
- maladies congénitales,
- traitements de stérilité et contraceptifs,
- affections antérieures à la souscription (sauf mention expresse),
- pathologies liées à l'alcoolisme ou à l'usage de drogues non prescrites,
- bilans psychologiques et traitements psychanalytiques [15].

3. Modalités de tarification de l'assurance maladie

Critères de tarification

La cotisation est influencée par plusieurs facteurs :

- **Âge de l'assuré** : augmentation progressive avec le risque,
- **Niveau de couverture** : formule de base ou étendue (optique, dentaire, international),
- **Statut professionnel** : régime collectif ou individuel,
- **Nombre de bénéficiaires** : assuré seul, couple, famille,
- **Antécédents médicaux** : parfois considérés lors de l'acceptation.

Formes de tarification

- **Forfaitaire** : prime fixe, souvent en collectif,
- **Par tranches d'âge** : grille tarifaire selon l'âge,
- **Personnalisée** : selon comportement médical ou profil de risque.

Mécanismes de modulation

- **Bonus-malus** : modulation selon la consommation médicale,
- **Révision périodique** : ajustement annuel en fonction des coûts [11].

1.4 Les Systèmes de Gestion des Assurances Maladie : Fonctionnement, Acteurs et Rôles

Les systèmes de gestion en assurance santé se regroupent principalement en deux catégories : la gestion classique (ou hors tiers payant) et le système du tiers payant. Chaque modèle implique des acteurs spécifiques et présente un mode de fonctionnement distinct. Comprendre ces systèmes est essentiel pour évaluer leur efficacité, leurs contraintes et leur impact sur les assurés, les prestataires de soins, les assureurs et les sociétés de gestion [16, 17].

La gestion classique (hors tiers payant)

Dans ce modèle, l'assuré paie intégralement ses soins médicaux ou ses produits pharmaceutiques avant de se faire rembourser par son assureur. Le remboursement est conditionné par la présentation de justificatifs précis [18] :

- Formulaire de déclaration de maladie rempli et signé,
- Reçus des consultations et prestations médicales (analyses, imagerie, soins),
- Ordonnances et preuves d'achat (tickets de caisse, vignettes des médicaments),
- En cas d'accouchement, un extrait d'acte de naissance du nouveau-né,
- Certificat médical pour les hospitalisations excédant 24 heures.

Ces documents sont soumis à l'avis d'un médecin-conseil de la compagnie, qui détermine les frais admissibles au remboursement. Le remboursement est ensuite effectué selon le taux de couverture prévu dans le contrat [19].

Le système du tiers payant

Ce système repose sur des conventions entre les assureurs ou leurs gestionnaires délégués et des centres de soins. L'assuré ne règle que le ticket modérateur (la part non couverte par l'assurance), le reste étant pris en charge directement par l'assureur ou un tiers payeur [16].

Dans certains cas, le tiers payeur est une structure indépendante spécialisée dans la gestion des risques santé. Elle contractualise avec les prestataires et gère les remboursements selon les modalités définies par l'assureur [17, 20].

1.5 Les Acteurs des Systèmes de Gestion des Assurances Maladie

Les systèmes de gestion des assurances maladie impliquent divers acteurs intervenant à différentes étapes du contrat. On distingue principalement ceux qui établissent le lien contractuel initial et ceux qui interviennent lors de l'exécution du contrat.

1. Acteurs à l'origine du contrat

La souscription d'un contrat d'assurance santé implique plusieurs parties prenantes dès l'établissement du contrat. Ces acteurs interviennent à différents niveaux de la chaîne de souscription et d'intermédiation, chacun avec des rôles juridiques, commerciaux ou techniques bien définis. Cette section décrit les fonctions respectives du souscripteur, de l'assuré, des intermédiaires (courtiers, apporteurs d'affaires) et de l'assureur dans la formation du contrat.

Le souscripteur

Le souscripteur est la personne physique ou morale qui conclut le contrat d'assurance avec l'assureur. Il s'engage à verser les primes convenues et à respecter les clauses contractuelles. Il peut s'agir, par exemple, d'un chef d'entreprise souhaitant couvrir ses employés ou d'un individu assurant sa famille.

L'assuré

L'assuré est la personne sur laquelle repose le risque couvert par le contrat. Il peut être distinct du souscripteur ou être la même personne.

Les courtiers et apporteurs d'affaires

Les courtiers agissent en tant qu'intermédiaires entre le souscripteur et l'assureur. Mandatés par le client, ils recherchent les meilleures garanties disponibles sur le marché. Contrairement aux agents généraux, les courtiers ne sont pas liés à un assureur spécifique et peuvent transférer leur portefeuille à un autre assureur selon les intérêts du client.

Les apporteurs d'affaires, quant à eux, sont mandatés par l'assureur pour prospecter de nouveaux clients. Le Code des Assurances de la CIMA reconnaît leur rôle à travers l'article 501 alinéa 3 [21].

Certains courtiers, tels que Gras Savoye, bénéficient de mandats de gestion leur permettant de gérer les contrats, régler les sinistres et transmettre les pièces justificatives à l'assureur pour remboursement [22].

L'assureur

L'assureur est une entité juridique, généralement une société anonyme ou une mutuelle, qui s'engage à indemniser l'assuré en cas de réalisation du risque couvert, en échange du paiement des primes.

2. Acteurs pendant l'exécution du contrat

Une fois le contrat d'assurance conclu, son exécution opérationnelle mobilise d'autres acteurs essentiels qui interviennent dans la mise en œuvre des garanties, la gestion des prestations et le suivi des sinistres. Cette section présente les rôles du bénéficiaire, des prestataires de soins conventionnés, ainsi que des sociétés de gestion spécialisées dans le traitement des dossiers médicaux et le contrôle des prestations.

Le bénéficiaire

Le bénéficiaire est la personne désignée pour recevoir les prestations prévues par le contrat d'assurance. Il peut être nommément identifié ou déterminé par des critères définis dans le contrat. Il est possible qu'une même personne soit à la fois souscripteur, assuré et bénéficiaire.

Les prestataires de soins

Les centres de soins et officines fournissent des services médicaux aux assurés. Dans le cadre du tiers payant, ces prestataires peuvent être liés à l'assureur par des conventions, facilitant ainsi la prise en charge directe des frais médicaux [23].

Les sociétés de gestion maladie

Les sociétés de gestion, telles que Managed Care International (MCI), jouent un rôle clé dans l'administration des contrats d'assurance maladie, notamment en tiers payant. Elles sont responsables de :

- La gestion des bénéficiaires (ajouts, retraits).
- Le traitement des dossiers médicaux et le remboursement des frais.
- Le contrôle des prestations en collaboration avec les médecins-conseils.
- La fourniture de rapports statistiques réguliers.
- L'information de l'assureur sur les sinistres importants.

Ces sociétés perçoivent des honoraires de gestion, généralement compris entre 10% et 12,5% du volume traité [24].

1.6 L'organisation du secteur de l'assurance santé en Algérie

En Algérie, l'assurance santé constitue un pilier essentiel du système de protection sociale. Elle est majoritairement gérée par deux organismes publics : la Caisse Nationale des Assurances Sociales des Travailleurs Salariés (CNAS) et la Caisse Nationale de Sécurité Sociale des Non-Salariés (CASNOS). Ce système vise à garantir un accès équitable aux soins médicaux pour l'ensemble de la population, tout en limitant l'impact financier des maladies, des hospitalisations et des accidents.

Contrairement à certains pays où l'assurance santé repose essentiellement sur un modèle privé, le système algérien est basé sur un régime public obligatoire financé par les cotisations sociales des employeurs et des travailleurs. Cependant, ce régime ne couvre qu'une partie des dépenses de santé, notamment les soins de base. C'est dans ce contexte que les compagnies d'assurance privées, comme AXA Assurance Algérie, viennent compléter ce dispositif public en proposant des contrats d'assurance santé complémentaire. Aujourd'hui, une part croissante des assurés opte pour cette complémentarité : une proportion significative du marché est ainsi partagée entre les régimes publics (CNAS/CASNOS) et les assureurs privés, dont AXA représente un acteur clé dans le paysage local.

1.6.1 Organismes publics : CNAS et CASNOS

Le système de sécurité sociale algérien repose principalement sur deux institutions publiques : La Caisse Nationale des Assurances Sociales des Travailleurs Salariés (CNAS), chargée de la couverture des travailleurs salariés pour les risques liés à la maladie, la maternité, l'invalidité, le décès, ainsi que les accidents du travail et les maladies professionnelles.

La Caisse Nationale de Sécurité Sociale des Non-Salariés (CASNOS), responsable de la protection sociale des travailleurs non salariés, tels que les commerçants, artisans, professions libérales, agriculteurs, etc. Elle couvre les risques de maladie, maternité, invalidité, décès et vieillesse .

Ces deux caisses assurent une couverture de base, financée par les cotisations sociales des employeurs et des travailleurs, visant à garantir un accès équitable aux soins médicaux pour l'ensemble de la population .

1.6.2 Émergence des assurances privées

Au cours des dernières années, les assurances privées ont connu une expansion notable dans le système de santé algérien. En complément du régime public, elles proposent désormais des contrats d'assurance santé complémentaires, offrant une meilleure prise en charge des dépenses non couvertes par la CNAS ou la CASNOS. Ces contrats prévoient une couverture extensive, incluant notamment les services d'hospitalisation, les soins dentaires et optiques, ainsi que certaines pratiques de médecine alternative.

1.6.3 Contexte Institutionnel : AXA Assurance Algérie

AXA Assurance Algérie s'inscrit dans la dynamique de modernisation et de diversification du secteur de l'assurance privé en Algérie. Filiale du groupe AXA, acteur de référence à l'échelle mondiale dans les domaines de l'assurance et de la gestion d'actifs, AXA Assurance Algérie a su adapter les standards internationaux aux spécificités du marché national. Elle se distingue par une stratégie axée sur l'innovation, l'amélioration continue de la qualité de service, et le développement de solutions d'assurance santé accessibles et adaptées aux besoins locaux.

Parmi ses offres phares figure «AXA Sahty», une formule d'assurance santé dédiée aux particuliers, qu'ils soient affiliés ou non à la sécurité sociale. Cette offre propose plusieurs niveaux de couverture, incluant le remboursement des frais médicaux courants, la prise en charge des hospitalisations, ainsi que des prestations en soins dentaires et optiques. En complément, AXA Assurance Algérie offre une gamme étendue de services d'assistance santé, visant à renforcer la protection de ses assurés au quotidien.

L'ancrage d'AXA Assurance Algérie dans une approche centrée sur la qualité et l'innovation lui a permis de devenir un acteur clé du marché de l'assurance santé dans le pays. Toutefois, comme l'ensemble du secteur, l'entreprise est confrontée à des défis majeurs, parmi lesquels figure la lutte contre la fraude aux assurances, particulièrement dans le traitement des dossiers de remboursement.

Notre immersion au sein d'AXA Assurance Algérie nous a permis d'appréhender avec précision ces défis. En effet, la fraude représente non seulement un risque financier pour l'entreprise, mais également un facteur d'érosion de la confiance entre assureur et assurés. Face à l'accroissement du volume des demandes et à la complexification des techniques de fraude, il devient essentiel d'optimiser les mécanismes de détection et de prévention

1.6.4 Garanties proposées par les assurances privées

Les assurances privées proposent notamment les garanties suivantes :

- Soins médicaux de base,
- Hospitalisation,
- Soins dentaires et optiques,
- Maternité,
- Médecines alternatives,
- Assistance santé.

1.7 Typologie des fraudes en assurance santé

La fraude à l'assurance est un phénomène complexe et polymorphe qui concerne l'ensemble des branches de l'assurance, et plus particulièrement le secteur de la santé. Si certaines fraudes peuvent paraître anodines, leur répétition, leur organisation ou leur sophistication en font une menace sérieuse pour les

assureurs, les assurés honnêtes, ainsi que pour la viabilité économique des systèmes d'assurance. Dans cette section, nous présenterons les différentes formes de fraude en assurance santé, en nous concentrant à la fois sur les fraudes commises à la souscription, les fraudes au sinistre, les pratiques de multi-assurance, ainsi que les comportements à risque ou irréguliers fréquemment observés par les assureurs.

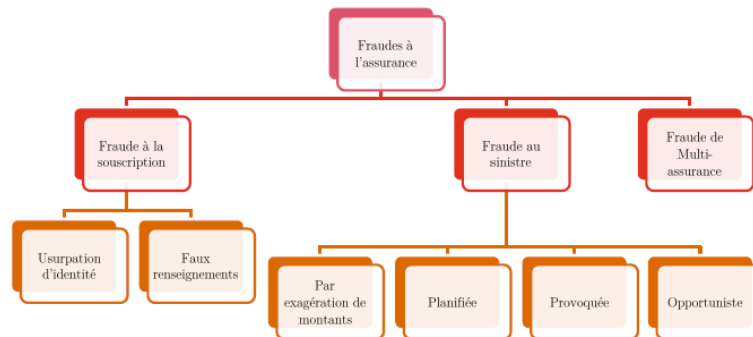


FIGURE 1.1 – Classification des fraudes à l'assurance : à la souscription, au sinistre et de multi-assurance

1.7.1 Fraude à la souscription

La fraude à la souscription constitue souvent la première étape d'un processus frauduleux. Elle a pour objectif principal de permettre à un assuré de réduire le coût de sa prime d'assurance ou d'accéder à des garanties auxquelles il ne pourrait pas prétendre en déclarant honnêtement sa situation. Ce type de fraude repose sur la fourniture intentionnelle d'informations erronées ou incomplètes au moment de la signature du contrat.

Les données falsifiées peuvent concerner :

- l'état de santé de l'assuré,
- l'identité de la personne (usurpation d'identité),
- le lieu ou les conditions de résidence,
- la nature de l'activité professionnelle,
- ou encore le nombre de personnes réellement couvertes.

La fraude à la souscription est l'une des plus difficiles à détecter car elle repose sur la déclaration unilatérale de l'assuré, et son impact n'est souvent visible qu'au moment du traitement d'un sinistre. Dans bien des cas, aucune vérification rigoureuse n'est effectuée tant qu'aucune demande de remboursement ou d'indemnisation n'est soumise.

Exemples courants :

- Déclarer une personne en bonne santé alors qu'elle souffre d'une pathologie chronique afin de souscrire à une complémentaire santé.
- Mentir sur son âge réel pour bénéficier d'une prime plus avantageuse.

- Fournir un faux bulletin de salaire ou un faux contrat de travail pour obtenir une couverture professionnelle.

- Usurper l'identité d'un tiers pour bénéficier de ses droits à l'assurance.

Des cas plus élaborés peuvent être observés, notamment à travers des réseaux organisés. Le phénomène des ghost brokers ou « courtiers fantômes » en est un exemple illustratif : ces individus ou groupes créent de faux sites d'assurance, attirent des clients via des publicités en ligne, puis émettent de fausses polices d'assurance ou falsifient des contrats réels pour offrir des tarifs attractifs. Ils peuvent également souscrire des contrats réels à bas prix grâce à des informations falsifiées, tout en vendant ces contrats à des assurés non informés. Ce type de fraude met en péril la sécurité juridique et financière des assurés concernés, qui peuvent se retrouver sans couverture effective en cas de sinistre.

1.7.2 La fraude par multi-assurance

Moins fréquente, mais tout aussi dommageable, la fraude par multi-assurance consiste à souscrire à plusieurs contrats d'assurance auprès de différentes compagnies pour un même risque, dans le but d'obtenir plusieurs indemnisations pour un sinistre unique. Cette pratique est notamment répandue en assurance santé, où certains assurés cumulent plusieurs complémentaires santé en espérant se faire rembourser au-delà de leurs dépenses réelles, ce qui est formellement interdit par la législation.

Cette fraude est difficile à détecter car elle nécessite une coordination ou un échange d'informations entre plusieurs assureurs, ce qui n'est pas toujours automatisé. De plus, les fraudeurs peuvent utiliser des variations d'identité ou des documents falsifiés pour éviter le croisement des données. Les systèmes de partage d'information inter-assureurs et les plateformes centralisées de gestion des contrats sont des outils clés pour la prévention de ce type de fraude.

1.7.3 La fraude au sinistre

La fraude au sinistre représente un large éventail de comportements malhonnêtes survenant après la souscription du contrat, lors de la déclaration d'un sinistre. On distingue généralement quatre grands types de fraude au sinistre, selon la classification classique utilisée par de nombreux assureurs :

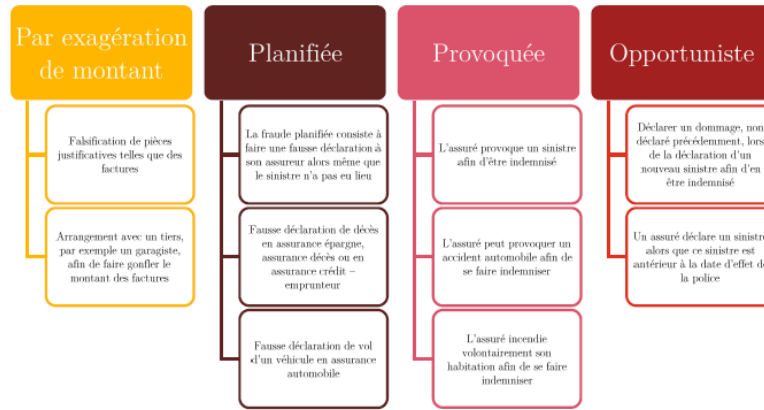


FIGURE 1.2 – Typologie des fraudes au sinistre : par exagération, planifiée, provoquée et opportuniste

Fraude exagérée : L'assuré grossit les dommages subis dans l'objectif d'augmenter le montant de l'indemnisation.

Exemple : Un assuré gonfle le montant de ses frais médicaux en collaboration avec un professionnel de santé.

Fraude planifiée : Le sinistre est entièrement prémédité par l'assuré, qui organise les faits pour simuler une situation éligible à une indemnisation.

Exemple : L'assuré déclare le vol d'un bien ou la perte d'un document alors qu'il l'a lui-même détruit ou dissimulé.

Fraude provoquée : Le sinistre est directement causé par l'assuré lui-même.

Exemple : Un assuré s'auto-administre une blessure légère pour se faire prescrire des arrêts de travail ou des médicaments.

Fraude opportuniste : Le sinistre est réel, mais il est déclaré de manière mensongère ou décalée dans le temps pour bénéficier d'une couverture.

Exemple : L'assuré a subi un vol ou un accident avant la date de début de son contrat, mais attend que le contrat soit en vigueur pour le déclarer.

1.7.4 Autres schémas frauduleux fréquents en assurance santé

Au-delà de ces typologies classiques, les compagnies d'assurance ont identifié au fil du temps un certain nombre de schémas récurrents dans les fraudes à l'assurance santé. Ces comportements, souvent fondés sur des heuristiques empiriques issues de l'expérience des experts, permettent de guider les contrôles et les analyses. Toutefois, les assureurs restent discrets quant à la diffusion de ces schémas pour éviter que les fraudeurs ne s'adaptent.

Parmi les comportements suspects les plus fréquents, on retrouve :

- **Des facturations excessives** pour des actes médicaux ou des médicaments, souvent en collaboration avec des professionnels de santé ;
- **Un nombre inhabituellement élevé de factures** pour un même assuré sur une période courte (quelques jours) ;
- **Des traitements obtenus sans paiement de cotisation**, ou avec un paiement partiel (souvent uniquement le premier versement) ;
- **L'achat de médicaments sans consultation médicale préalable** ;
- **La production de factures datées en dehors de la période de validité du contrat** ;
- **Le changement non justifié de coordonnées bancaires** d'un prestataire (clinique, pharmacie, centre médical) ;
- **Une multiplication de factures manuelles** dont les montants restent volontairement sous les seuils de déclenchement du contrôle automatique.
- **Des remboursements demandés pour des montants supérieurs à ceux mentionnés sur les factures réelles.**

La complexité de certaines fraudes réside dans leur caractère collaboratif. Elles peuvent impliquer plusieurs acteurs : l'assuré, un ou plusieurs prestataires de santé, des courtiers, voire des employés d'assurance malveillants. Dans ces cas, la détection requiert des investigations approfondies, des outils de data mining, et de plus en plus, des techniques d'intelligence artificielle.

1.8 Le positionnement de la fraude dans l'assurance

La fraude à l'assurance, notamment dans le domaine de la santé, constitue un problème croissant pour les compagnies opérant en Algérie, telles qu'AXA Assurance Algérie. Cette fraude peut survenir lors de la souscription au contrat ou au moment de la gestion du sinistre. Pour bien comprendre où se situent les vulnérabilités du système, il est nécessaire de revenir sur les fondements juridiques du contrat d'assurance.

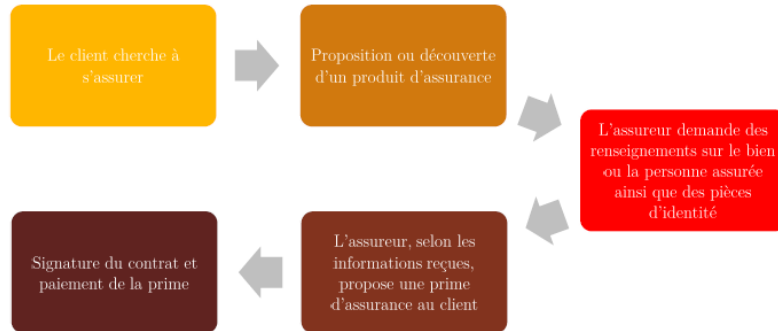


FIGURE 1.3 – Processus usuel de souscription d'un contrat d'assurance

1.8.1 Fraude à la souscription d'un contrat d'assurance

Le Code civil algérien, dans son article 54, définit le contrat comme :

« Un contrat est une convention par laquelle une ou plusieurs personnes s'obligent envers une ou plusieurs autres à donner, faire ou ne pas faire quelque chose. »[25]

Quant à l'assurance, elle est régie par l'ordonnance n° 95-07 du 25 janvier 1995, qui dispose dans son article 2 :

« L'assurance est une opération par laquelle une partie, l'assureur, s'engage envers une autre partie, le souscripteur, moyennant une prime ou cotisation, à fournir une prestation en cas de réalisation d'un risque prévu au contrat. »[26]

Le contrat d'assurance santé est donc, par nature, un contrat aléatoire, car la réalisation du risque (la maladie ou l'hospitalisation) est incertaine et échappe à la volonté des parties. Cela est conforme à l'article 107 du Code civil algérien qui prévoit que :

« Il ne peut y avoir d'obligation lorsque la condition en dépend de la seule volonté du débiteur. »[27]

Lors de la souscription, l'assuré est tenu de déclarer avec exactitude son état de santé, ses antécédents, ou tout autre élément susceptible d'influencer le montant de la prime ou l'acceptation du risque. Cette obligation d'honnêteté est posée par l'article 21 de l'ordonnance 95-07, qui stipule que :

« L'assuré est tenu de déclarer exactement, lors de la conclusion du contrat, toutes les circonstances qu'il connaît et qui sont de nature à faire apprécier par l'assureur les risques qu'il prend à sa charge. »

Tout manquement à cette obligation peut entraîner la nullité du contrat ou la réduction de l'indemnisation, comme le précise l'article 23 :

« En cas de réticence ou de fausse déclaration intentionnelle de la part de l'assuré, le contrat est nul. »

Ainsi, la fraude à la souscription se manifeste lorsque l'assuré dissimule volontairement certaines informations pour bénéficier de garanties sans contrepartie

équitable.

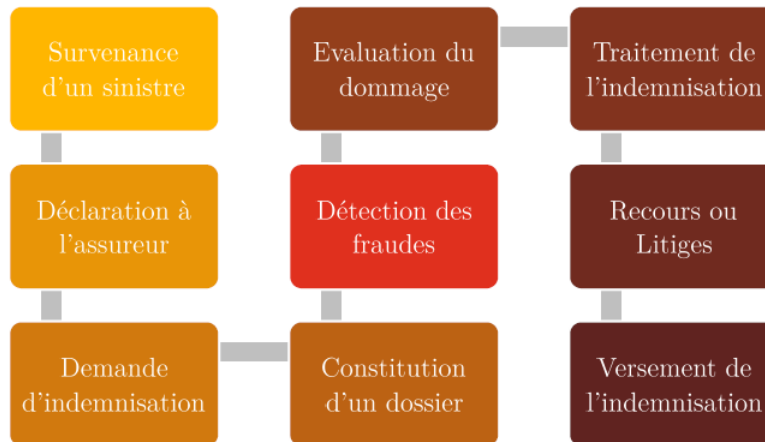


FIGURE 1.4 – Processus usuel de souscription d'un contrat d'assurance

1.8.2 Fraude lors de la gestion des sinistres

Lorsqu'un sinistre survient (hospitalisation, acte médical, etc.), l'assuré est tenu d'en informer l'assureur dans les délais impartis. L'article 26 de l'ordonnance 95-07 précise :

« L'assuré doit, sous peine de déchéance, déclarer tout sinistre à l'assureur dans un délai fixé par le contrat. »

Dans le domaine de la santé, la fraude peut consister en la présentation de factures falsifiées, la simulation de soins, ou encore la collusion avec un professionnel de santé pour gonfler le montant remboursé. L'assureur, à travers ses gestionnaires de sinistres, analyse alors les éléments du dossier pour détecter toute anomalie.

L'article 22 de la même ordonnance donne la possibilité à l'assureur de vérifier les déclarations de l'assuré et de refuser l'indemnisation en cas de fraude avérée.

Enfin, la détection des fraudes s'appuie sur l'exploitation de données contenues dans les dossiers de sinistres (dates, montants, descriptions, pièces justificatives) et sur des mécanismes de contrôle renforcés (scoring, IA, audit interne), afin de protéger l'équilibre financier de la compagnie.

1.9 Impact économique de la fraude à l'assurance santé

La fraude à l'assurance représente un enjeu économique majeur pour les compagnies d'assurance, tant au niveau mondial que local. Elle entraîne des pertes financières directes et indirectes, affectant non seulement la rentabilité des assureurs, mais également la stabilité des primes d'assurance imposées aux assurés.

Deux conséquences majeures sont généralement mises en évidence :

L'évaluation des pertes liées à la fraude ;

La répercussion de ces pertes sur les primes d'assurance.

Selon les estimations de la Fédération des sociétés d'assurance européennes Insurance Europe, la fraude représenterait en moyenne 10 % du coût total des sinistres déclarés en Europe. Ce chiffre est cependant en hausse dans les régions où la régulation est plus souple ou les dispositifs de contrôle moins développés, comme en Afrique ou en Amérique latine [28]

Voici quelques estimations indicatives du poids de la fraude selon les pays et les branches concernées :

Pays	Branches concernées	Montant estimé
Allemagne	Toutes les branches	10% du coût des sinistres
Australie	Toutes les branches	10% du coût des sinistres
Canada	Toutes les branches	10% à 15% du coût des sinistres
Espagne	Automobile	22% du coût des sinistres
Grande-Bretagne	Branches des particuliers	7% du coût des sinistres
Scandinavie	Toutes les branches	5% à 10% du coût des sinistres
États-Unis	Automobile	11% à 15% du coût des sinistres
États-Unis	Toutes les branches	10% du coût des sinistres

TABLE 1.1 – Estimations du coût de la fraude à l'assurance par pays et par branche

En Grande-Bretagne, le coût annuel de la fraude a été estimé à 2,5 milliards de dollars, selon l'Association des Assureurs Britanniques. En Allemagne, environ la moitié des déclarations concernant les vols de téléphones portables seraient fausses [28].

Il convient de noter que ces chiffres ne tiennent pas compte des fraudes commises au moment de la souscription, telles que les fausses déclarations de risque, qui aboutissent à une sous-évaluation des primes.

1.9.1 Répercussions systémiques et stratégiques

Dans un contexte économique difficile, les compagnies d'assurance comme AXA Algérie sont confrontées à une hausse potentielle des actes frauduleux. Cette augmentation impacte inévitablement :

- **Le coût des sinistres moyens ;**
- **La tarification des produits d'assurance ;**
- **La viabilité de certaines branches**, notamment l'assurance santé dans les pays en développement ;
- **L'image de marque et la confiance des assurés.**

Cette situation pousse les assureurs à renforcer leurs dispositifs de prévention et de détection, à travers des investissements accrus dans les technologies de l'information, la mise en place de systèmes prédictifs basés sur l'analyse de données et des collaborations avec les autorités judiciaires et réglementaires.

1.9.2 Mécanismes de lutte et enjeux techniques

Face à l'ampleur de la fraude, les assureurs structurent leur stratégie de lutte autour de quatre axes fondamentaux :

- **La collecte et l'exploitation des données internes ;**
- **Le développement de systèmes d'alerte et de scoring**, tels que *Risk Shield* de Gen Re ;
- **Le renforcement des contrôles internes** et la coopération inter-entreprises ;
- **La formation et la sensibilisation du personnel.**

En parallèle, les réglementations prudentielles, notamment *Solvabilité II*, imposent aux compagnies d'assurance l'identification et la gestion des risques opérationnels, dont la fraude fait partie intégrante [29].

Malgré ces avancées, plusieurs freins limitent encore l'efficacité de la lutte contre la fraude :

- **Le manque de données exploitables ;**
- **L'insuffisance des ressources humaines spécialisées ;**
- **La difficulté à apporter la preuve de la fraude**, notamment la preuve matérielle et l'intention frauduleuse.

Une enquête européenne menée par Insurance Nexus en 2016 a révélé que l'importance stratégique de la lutte contre la fraude est en forte croissance : de 11 % en 2015 à 18 % en 2016 parmi les priorités des dirigeants d'assurance [30].

1.9.3 Considérations spécifiques à l'assurance santé en Algérie

En Algérie, les assureurs tels qu'AXA Assurance Algérie sont confrontés à une double problématique :

Une réglementation encore en consolidation ;

Un écosystème de santé où les pratiques frauduleuses (exagération des soins, factures falsifiées, collusion entre professionnels et assurés) sont parfois difficilement traçables.

Cela renforce le besoin de solutions technologiques avancées (intelligence artificielle, machine learning) pour détecter les schémas inhabituels, et de partenariats interinstitutionnels pour structurer un système de reporting et de sanction plus rigoureux.

1.10 Objectifs de l'étude

Dans un contexte où la fraude à l'assurance santé représente un risque financier croissant et un défi opérationnel majeur, il devient essentiel de développer des approches innovantes capables de détecter ces comportements de manière automatisée, fiable et explicable. Ce mémoire s'inscrit dans cette dynamique, en cherchant à répondre à la problématique suivante :

Comment concevoir un système intelligent de détection de fraude en assurance santé, capable d'identifier les anomalies dans les sinistres à partir des données disponibles chez AXA Assurance Algérie, tout en répondant aux contraintes métier, techniques et réglementaires ?

Pour ce faire, notre travail poursuit les objectifs suivants :

- Comprendre les mécanismes opérationnels de l'assurance santé en Algérie, notamment le processus de traitement des sinistres, les relations entre assurés, bénéficiaires et dossiers, ainsi que les points de vulnérabilité aux abus ;
- Identifier et formaliser les règles métier utilisées pour qualifier des comportements frauduleux types, en collaboration avec les équipes d'experts métier ;
- Construire un jeu de données exploitable à partir des bases de données internes, en enrichissant les sinistres de variables comportementales pertinentes ;
- Mettre en œuvre un pipeline de détection de fraude fondé sur des techniques de machine learning, en combinant des approches supervisées, semi-supervisées et des outils d'explicabilité ;
- Évaluer la performance du modèle et sa robustesse sur des données réelles, et proposer des pistes d'amélioration en vue d'un déploiement opérationnel.

Cette orientation scientifique et opérationnelle structurera les chapitres suivants, qui exposeront d'abord les méthodes de détection de fraude existantes, puis la mise en œuvre concrète du pipeline sur les données d'AXA Algérie.

1.11 Modélisation économique de la solution : Business Model Canvas

Dans le but de synthétiser la logique de création de valeur autour de notre système de détection de fraude en assurance santé, nous proposons ci-après un Business Model Canvas. Il permet d'identifier les ressources, activités, partenaires et flux économiques impliqués dans le projet mené au sein d'AXA Assurance Algérie.

Partenaires clés	Activités clés	Proposition de valeur
— AXA Algérie – équipes IT et sinistres : 50 000 sinistres/an — Sociétés de gestion santé (MCI) : 30 000 dossiers/an — Médecins-conseils : 5 000 validations/an — Data scientists / ingénieurs IA : 2 ETP (équivalents temps plein) — Régulateurs et CNAS : 4 échanges/an	— Traitement et enrichissement de 300 000 sinistres sur 5 ans — Détection d'anomalies avec précision de 92 % — Développement d'un tableau de bord métier (PowerBI) — Collaboration avec experts métier : 4 workshops/an	— Réduction des pertes dues à la fraude : 3 M€/an — Fiabilisation des décisions : +15 % fiabilité — Renforcement de la confiance assurés/assureur — Outil d'aide à la décision exploitable
Ressources clés	Relations clients	Canaux de distribution
— Base de données : 300 000 sinistres / 5 ans — 10 serveurs dédiés, 100 To stockage — 4 experts métiers + 2 data scientists — Expertise interne assurance santé	— Assistance technique : 24 sessions/an — Reporting : 12 rapports/an — Collaboration interservices : IT – gestion – audit	— Interface interne : PowerBI/Tableau — Intégration SI sinistres : API + dashboard — Rapports automatiques : mensuels
Structure de coûts	Segments de clientèle et Sources de revenus	
— Développement IA : 120 000€ en année 1 — Stockage/compute : 28 000€/an (cloud + serveurs) — Formation : 12 000€/an — Maintenance évolutive : 100 000€/an — Total : 260 000€/an	Segments de clientèle : — AXA Algérie (département sinistres) — Assurés honnêtes (valeur indirecte) — Autorités réglementaires (conformité) Flux de revenus (valeur créée) : — Réduction des remboursements indus : 3 M€/an — Diminution des primes injustifiées : 500 000€/an — Fidélisation : gain indirect 250 000€/an — Bénéfice total estimé : 3,75 M€/an	

TABLE 1.2 – Business Model Canvas du projet de détection de fraude chez AXA Assurance Algérie avec chiffres réalistes

1.12 Conclusion

En résumé, ce premier chapitre a permis de poser les fondements nécessaires à la compréhension des enjeux de l'assurance santé et des formes variées que peut prendre la fraude dans ce secteur. À travers une analyse du fonctionnement des contrats, des modalités de remboursement, du rôle des acteurs institutionnels, et des mécanismes de gestion en Algérie, nous avons montré que l'assurance santé constitue à la fois un pilier de protection sociale et une cible vulnérable à des pratiques frauduleuses multiples : fausses déclarations, multi-assurance, actes fictifs, collusions, etc.

Ces pratiques, en plus de générer des pertes économiques considérables, nuisent à la viabilité du système d'assurance et compromettent la relation de confiance entre assureurs et assurés. Dans le contexte algérien, marqué par une croissance des sinistres et des limites structurelles (opacité, absence de systèmes intégrés), les compagnies comme AXA sont confrontées à une double exigence : maîtriser le risque de fraude tout en maintenant une qualité de service optimale.

Face aux limites des dispositifs traditionnels de contrôle fondés sur des règles fixes, se pose la question de l'apport des approches innovantes issues de la science des données. C'est précisément ce que nous explorons dans le chapitre suivant, en nous intéressant aux outils de l'intelligence artificielle et du machine learning appliqués à la détection de la fraude. Ces techniques, qu'elles soient supervisées ou non supervisées, offrent des perspectives prometteuses pour identifier des comportements anormaux dans un volume massif de données, de manière automatisée, évolutive et interprétable.

Chapitre 2

Intelligence artificielle et détection de fraude : état de l’art

2.1 Introduction

L’essor de l’intelligence artificielle (IA) a profondément transformé le secteur de l’assurance, notamment dans la lutte contre la fraude. En exploitant la puissance des algorithmes d’apprentissage automatique et des réseaux de neurones, il est désormais possible de détecter des comportements anormaux et des fraudes potentielles de manière plus rapide et efficace. Ce chapitre présente les fondements de l’intelligence artificielle appliquée à la détection de fraude, en exposant les principales catégories d’apprentissage, les techniques basées sur les réseaux de neurones artificiels et les avancées récentes autour des réseaux de neurones graphiques. Enfin, une synthèse des recherches scientifiques actuelles mettra en lumière l’apport concret de ces approches dans le domaine de l’assurance santé.

2.2 Intelligence Artificielle

L’intelligence artificielle (IA) désigne un domaine de recherche interdisciplinaire visant à concevoir des systèmes capables de simuler certains aspects de l’intelligence humaine. Apparue officiellement en 1956 lors de la conférence de Dartmouth, l’IA s’est progressivement imposée comme un champ central des sciences informatiques et de l’ingénierie.

L’IA couvre un large éventail de domaines, tels que la perception sensorielle, la prise de décision, l’apprentissage automatique, la résolution de problèmes ou encore la compréhension du langage naturel. Elle est aujourd’hui utilisée dans des applications variées comme la médecine, les jeux, les véhicules autonomes, les systèmes de recommandation, ou encore la détection de fraude.

Pour mieux comprendre les différentes visions de l'IA, Russell et Norvig [31] proposent une classification selon deux axes :

Le type de capacité ciblée : raisonnement ou comportement.

Le critère de performance : ressemblance avec l'humain ou optimalité rationnelle.

Le tableau ci-dessous synthétise les approches les plus représentatives de ces visions

TABLE 2.1 – Définitions de l'intelligence artificielle selon différentes approches

Penser comme un humain	Penser rationnellement
« Le nouvel effort passionnant pour faire penser les ordinateurs... des machines avec des esprits, au sens plein et littéral. » [32]	« L'étude des calculs qui permettent de percevoir, raisonner et agir. » [33]
« L'automatisation des activités que nous associons à la pensée humaine, telles que la prise de décision, la résolution de problèmes, l'apprentissage... » [34]	« L'étude des facultés mentales à travers l'utilisation de modèles computationnels. » [35]
Agir comme un humain	Agir rationnellement
« L'art de créer des machines qui accomplissent des fonctions nécessitant de l'intelligence lorsqu'elles sont réalisées par des humains. » [36]	« L'IA s'intéresse au comportement intelligent dans les artefacts. » [37]
« L'étude de la manière de faire faire aux ordinateurs des choses que, pour le moment, les humains font mieux. » [38]	« L'intelligence computationnelle est l'étude de la conception d'agents intelligents. » [39]

Les approches qui imitent le comportement humain sont souvent empiriques, s'appuyant sur des données issues de l'observation. À l'inverse, les approches fondées sur la rationalité privilégient des formulations mathématiques optimales, davantage liées à l'ingénierie et à la logique formelle.

Parmi les différentes branches de l'IA, l'apprentissage automatique (Machine Learning, ML) s'est imposé comme un pilier fondamental, en raison de sa capacité à extraire automatiquement des connaissances à partir de données massives.

2.3 Catégories d'Apprentissage

Dans le domaine de l'apprentissage automatique, on distingue généralement trois grandes catégories de méthodes : l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage semi-supervisé. Chaque catégorie repose sur la nature des données disponibles (étiquetées ou non) et sur l'objectif de la tâche

à accomplir (prédiction, regroupement, détection de structures, etc.).

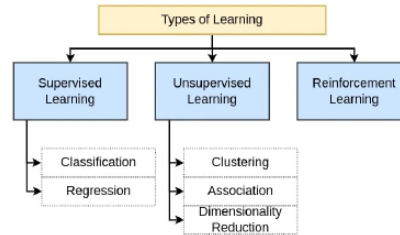


FIGURE 2.1 – Catégories d’apprentissage en machine learning : supervisé, non supervisé, semi-supervisé et par renforcement

Nous détaillons ci-après l’une des catégories les plus couramment utilisées dans les projets de détection de fraude : l’apprentissage supervisé.

Apprentissage supervisé

Un algorithme d’apprentissage supervisé nécessite un superviseur en tant qu’enseignant. Dans cette technique, l’ensemble de données d’entrée fourni au système est divisé en deux parties : l’ensemble d’entraînement et l’ensemble de test. L’algorithme d’apprentissage supervisé utilise des données étiquetées, appelées données d’entraînement, pour apprendre les motifs cachés dans ces données étiquetées et produire une prédiction ou une classification correcte pour l’ensemble de test. Trois des algorithmes les plus utilisés en apprentissage supervisé sont expliqués ci-dessous :

Modèles Linéaires

Les modèles linéaires sont parmi les algorithmes d’apprentissage supervisé les plus fondamentaux. Ils reposent sur l’hypothèse que la relation entre les variables explicatives et la variable cible peut être approximée par une fonction linéaire. Leur simplicité les rend interprétables, efficaces en termes de calcul, et souvent compétitifs sur des problèmes bien posés.

Régression linéaire. La régression linéaire vise à modéliser la relation entre une variable dépendante y et une ou plusieurs variables indépendantes x_i via l’équation :

$$y = \beta_0 + \sum_{i=1}^n \beta_i x_i + \varepsilon \quad (2.3.1)$$

Régression logistique

La régression logistique est un modèle mathématique permettant de prédire une variable cible binaire à partir de variables explicatives, qu’elles soient quantitatives ou qualitatives. Elle représente un cas particulier des modèles linéaires

généralisés et sert à modéliser la probabilité qu'une observation appartienne à une classe donnée en fonction des variables explicatives.

Formulation du modèle

Soient $X = (X_1, X_2, \dots, X_p)$ les variables explicatives et Y la variable cible binaire prenant les valeurs $\{0, 1\}$. La probabilité a posteriori d'obtenir $Y = 1$ (ou $Y = 0$) connaissant X est notée $P(Y = 1|X)$ (respectivement $P(Y = 0|X)$). En appliquant le théorème de Bayes :

2.4 Régression logistique

$$P(Y = 1|X) = \frac{P(Y = 1) \cdot P(X|Y = 1)}{P(X)} \quad (2.4.1)$$

$$P(Y = 0|X) = \frac{P(Y = 0) \cdot P(X|Y = 0)}{P(X)} \quad (2.4.2)$$

Ainsi, le rapport de vraisemblance peut être écrit :

$$\frac{P(Y = 1|X)}{P(Y = 0|X)} = \frac{P(Y = 1)}{P(Y = 0)} \cdot \frac{P(X|Y = 1)}{P(X|Y = 0)} \quad (2.4.3)$$

Pour estimer ce rapport, la régression logistique repose sur l'hypothèse suivante :

$$\ln \left(\frac{P(X|Y = 1)}{P(X|Y = 0)} \right) = b_0 + b_1 X_1 + \dots + b_p X_p \quad (2.4.4)$$

Puisque Y suit une loi de Bernoulli, on a $P(Y = 1|X) = \mathbb{E}(Y|X) = \pi$, où $\pi \in (0, 1)$. En introduisant la fonction *logit*, le modèle devient :

$$\ln \left(\frac{\pi}{1 - \pi} \right) = a_0 + a_1 X_1 + \dots + a_p X_p \quad (2.4.5)$$

Ce qui équivaut à :

$$\pi = \frac{\exp(a_0 + a_1 X_1 + \dots + a_p X_p)}{1 + \exp(a_0 + a_1 X_1 + \dots + a_p X_p)} \quad (2.4.6)$$

ou encore :

$$\pi = \frac{1}{1 + \exp(-(a_0 + a_1 X_1 + \dots + a_p X_p))} \quad (2.4.7)$$

Ridge et Lasso. Lorsque les données présentent une forte colinéarité ou un grand nombre de variables, les modèles linéaires réguliers sont préférables :

$$\text{Pénalité Ridge} = \lambda \sum \beta_i^2 \quad (2.4.8)$$

$$\text{Pénalité Lasso} = \lambda \sum |\beta_i| \quad (2.4.9)$$

Elastic Net. L'Elastic Net combine les pénalités ℓ_1 et ℓ_2 :

$$\mathcal{L} = \text{RSS} + \lambda_1 \sum |\beta_i| + \lambda_2 \sum \beta_i^2 \quad (2.4.10)$$

ce qui permet de profiter des avantages du Lasso (sélection) et du Ridge (stabilité) [40].

Méthodes à base d'arbres

Les méthodes à base d'arbres sont des techniques d'apprentissage supervisé très populaires en raison de leur flexibilité, de leur robustesse et de leur capacité à modéliser des relations non linéaires complexes. Elles peuvent être utilisées pour des tâches de classification comme de régression.

Arbre de décision : Un arbre de décision a une structure similaire à un organigramme où chaque nœud interne représente un test sur un attribut, les nœuds de feuille représentent les étiquettes de classe et les branches représentent des conjonctions d'attributs appartenant à cette étiquette de classe. Le chemin complet de la racine à la feuille représente la règle de classification. L'algorithme de l'arbre de décision repose sur la mesure de l'entropie pour choisir l'attribut qui divisera au mieux l'ensemble de données.

Les forêts aléatoires (Random Forest)

Les forêts aléatoires, introduites par Leo Breiman en 2001 sont une méthode d'apprentissage automatique fondée sur l'agrégation de plusieurs arbres de décision, notamment les arbres CART (Classification and Regression Trees). Cette technique vise à améliorer la précision des prédictions en combinant les résultats de plusieurs modèles faiblement corrélés. Elle est applicable à la fois à des problèmes de régression et de classification.

Principe de l'algorithme Random Forest

Le principe des forêts aléatoires repose sur la construction de B arbres de décision à partir de différents échantillons de la base d'apprentissage. L'objectif est de réduire la variance du modèle en créant des arbres peu corrélés entre eux.

Algorithm 2 Construction d'une forêt aléatoire

for $i = 1$ **to** B **do** $\perp$

Tirer un échantillon bootstrap z_i de la base d'apprentissage (tirage avec remise)

2: Construire un arbre de décision $G_i(x)$ en choisissant à chaque nœud un sous-ensemble aléatoire de q variables parmi les p variables explicatives

3: Sélectionner la meilleure scission dans ce sous-ensemble

4:

5: Agréger les prédictions par vote majoritaire (classification) ou moyenne (régression) $=0$

Le nombre de variables q est généralement choisi comme $q = \sqrt{p}$ pour la classification et $q = p/3$ pour la régression.

Prédiction et performances

La forêt aléatoire se distingue par sa capacité à générer des arbres de décision diversifiés. Cette diversité est obtenue grâce aux deux sources d'aléatoire : le tirage bootstrap des échantillons et la sélection aléatoire des variables explicatives à chaque nœud. À l'issue de l'apprentissage, les performances peuvent être évaluées à l'aide de l'erreur dite Out-of-Bag (OOB).

Erreur Out-of-Bag (OOB) Cette métrique permet d'estimer l'erreur de généralisation sans recourir à une validation croisée. Chaque observation qui ne fait pas partie de l'échantillon bootstrap est utilisée pour évaluer l'arbre correspondant. L'erreur OOB est ensuite calculée en moyennant les erreurs de prédiction sur l'ensemble de ces observations.

Importance des variables

La forêt aléatoire fournit des indicateurs permettant de mesurer l'importance des variables explicatives dans le modèle. Deux méthodes principales sont utilisées :

1. Permutation de Breiman (Mean Decrease Accuracy) Proposée par Breiman, cette méthode consiste à permuter aléatoirement les valeurs d'une variable X_j dans les observations OOB, puis à mesurer l'augmentation de l'erreur OOB. Une augmentation significative indique que la variable est importante.

2. Diminution moyenne de l'impureté (Mean Decrease Gini) Cette méthode s'appuie sur la réduction de l'impureté de Gini apportée par chaque variable au cours des découpages. Plus une variable est fréquemment utilisée

pour effectuer des scissions réduisant fortement l'impureté, plus elle est jugée importante.

Extra Trees (Extremely Randomized Trees). Les arbres extrêmement randomisés poussent la logique de la forêt aléatoire plus loin en randomisant également les seuils de découpage des variables, ce qui réduit la variance au prix d'une légère augmentation du biais. Contrairement à Random Forest, les Extra Trees ne se basent pas nécessairement sur le bootstrap.

Gradient Boosting Machines (GBM). Le Gradient Boosting est une autre approche d'ensemble dans laquelle les arbres sont construits de manière séquentielle, chaque nouvel arbre corrigeant les erreurs de prédiction du précédent. La fonction de perte est minimisée à chaque itération à l'aide d'un algorithme de descente de gradient. Bien que puissant, le GBM classique est sensible au surapprentissage et exige un réglage fin des hyperparamètres [41].

XGBoost. XGBoost (Extreme Gradient Boosting) est une implémentation optimisée du gradient boosting. Elle inclut la régularisation des modèles, le traitement des valeurs manquantes, le parallélisme et une gestion efficace de la mémoire, ce qui en fait une méthode très performante pour les compétitions de science des données [42].

LightGBM. Développé par Microsoft, LightGBM est une version optimisée du boosting par arbres, qui utilise des histogrammes de caractéristiques et une croissance par feuille (leaf-wise) au lieu de la croissance par niveau (level-wise), ce qui permet un entraînement plus rapide et une meilleure précision [43].

CatBoost. CatBoost, développé par Yandex, est particulièrement adapté aux variables catégorielles. Il gère automatiquement les variables non numériques sans besoin de les convertir. Il utilise des techniques avancées pour éviter le surapprentissage, notamment le "Ordered Boosting" [44].

Méthodes bayésiennes et probabilistes

Les méthodes bayésiennes s'appuient sur le théorème de Bayes pour modéliser l'incertitude et intégrer des connaissances a priori dans les prédictions. Elles sont particulièrement utiles lorsque les données sont peu nombreuses ou bruitées, ou lorsque l'on souhaite quantifier l'incertitude des prédictions.

Naïve Bayes. Le classifieur Naïve Bayes repose sur l'hypothèse (naïve) que les variables explicatives sont conditionnellement indépendantes les unes des autres, compte tenu de la classe. Trois variantes sont couramment utilisées :

- **Naïve Bayes gaussien** : adapté aux variables continues, en supposant une distribution normale des caractéristiques.

- **Naïve Bayes multinomial** : utilisé notamment pour les textes ou les occurrences de mots, comme dans les modèles de sac de mots.
- **Naïve Bayes bernoullien** : approprié pour des variables binaires (présence/absence).

Bien que simpliste, cette approche donne souvent de bons résultats dans des tâches de classification textuelle ou de détection de spam [45].

Réseaux bayésiens. Les réseaux bayésiens (ou *Bayesian Networks*) sont des modèles graphiques probabilistes qui représentent un ensemble de variables aléatoires et leurs dépendances conditionnelles via un graphe orienté acyclique (DAG). Ils permettent de modéliser des relations causales, d’inférer des probabilités conditionnelles et de combiner données et expertises [46].

Régression linéaire bayésienne. La régression bayésienne est une extension probabiliste de la régression linéaire classique. Au lieu de produire une estimation ponctuelle des paramètres, elle estime leur distribution postérieure. Cela permet de capturer l’incertitude sur les prédictions et les coefficients, et d’incorporer des connaissances a priori sous forme de distributions préalables [47].

Apprentissage par ensemble

L’apprentissage par ensemble implique l’utilisation de plusieurs apprenants pour résoudre un problème spécifique. Des algorithmes comme les arbres de décision, Naïve Bayes ou les réseaux neuronaux peuvent être utilisés comme apprenants individuels. Cette méthode est populaire depuis les années 1990 et montre que plusieurs apprenants sont souvent plus efficaces qu’un seul pour résoudre un problème. Deux algorithmes couramment utilisés sont :

Boosting : Le boosting est une méthode d’ensemble séquentielle dans laquelle les modèles sont entraînés les uns après les autres. Chaque nouveau modèle tente de corriger les erreurs commises par les modèles précédents. Ce processus adaptatif permet de réduire à la fois le biais et la variance, améliorant ainsi la précision globale du modèle final (Freund – Schapire, 1996).

Parmi les algorithmes de boosting les plus courants, on trouve :

- **AdaBoost** un algorithme initialement conçu pour combiner des classificateurs faibles en une forte majorité.
- **Gradient Boosting Machines (GBM)** qui utilise des gradients de la fonction de perte pour améliorer le modèle à chaque étape.
- **XGBoost** une version optimisée et régularisée du GBM, reconnue pour sa performance et son efficacité computationnelle.

Bien que le boosting permette d’atteindre une très haute performance prédictive, il est souvent plus lent à entraîner (du fait de sa nature séquentielle) et peut être sensible au surapprentissage, notamment en l’absence de régularisation adéquate.

Bagging (Bootstrap Aggregating) : Le bagging est une méthode d’ensemble parallèle qui consiste à générer plusieurs sous-échantillons du jeu de

données initial à l'aide d'un échantillonnage avec remise (bootstrap), comme introduit par Breiman (1996). Chaque sous-échantillon est utilisé pour entraîner un modèle de base de manière indépendante. Les prédictions des différents modèles sont ensuite combinées (par exemple, par vote majoritaire pour la classification) afin de produire une prédiction finale plus robuste.

Cette approche permet de réduire la variance des prédictions et diminue le risque de surapprentissage, en particulier dans des contextes de forte dimensionnalité. L'un des exemples les plus connus de bagging est la méthode des forêts aléatoires (Random Forests) introduite par Breiman (2001). Toutefois, le bagging peut être coûteux en ressources computationnelles et nuire à l'interprétabilité du modèle.

Stacking (Stacked Generalization) : Le stacking est une technique d'ensemble qui consiste à entraîner plusieurs modèles de base (appelés modèles de niveau 0) et à utiliser leurs prédictions comme entrées pour un modèle méta-apprenant (ou modèle de niveau 1) qui produit la prédiction finale. Cette approche, introduite par Wolpert (1992), vise à combiner intelligemment les sorties des modèles de base, plutôt qu'à les agréger de manière uniforme comme dans le bagging.

Le stacking est particulièrement flexible, car il permet de combiner des modèles hétérogènes (ex. : arbres de décision, SVM, régressions). Toutefois, il peut être complexe à implémenter et nécessite un soin particulier pour éviter le surapprentissage du méta-modèle, notamment via une validation croisée rigoureuse.

Méthodes à noyaux

Les méthodes à noyaux sont des techniques d'apprentissage supervisé qui utilisent des fonctions de noyau pour transformer les données dans un espace de dimension supérieure. Cette transformation permet de rendre les données linéairement séparables ou de mieux s'adapter à des problèmes non linéaires. Les noyaux couramment utilisés incluent le noyau linéaire, polynomial, et Gaussien (RBF).

Support Vector Machine (SVM). L'algorithme Support Vector Machine (SVM) est l'une des méthodes les plus puissantes pour la classification et la régression. L'idée principale est de trouver l'hyperplan qui sépare les données de manière optimale, c'est-à-dire avec la plus grande marge entre les classes. Le noyau permet de projeter les données dans un espace de plus grande dimension, ce qui permet de mieux séparer les classes dans des problèmes non linéaires [48].

SVC (Support Vector Classification). Le SVC est une variante de SVM utilisée pour des problèmes de classification. Il cherche à maximiser la marge entre les classes tout en minimisant les erreurs de classification. Il peut être utilisé avec différents noyaux (linéaire, RBF, etc.) en fonction de la nature des données. L'ajustement de l'hyperparamètre C permet de contrôler le compromis entre la maximisation de la marge et la minimisation des erreurs [49].

SVR (Support Vector Regression). Le SVR est une extension du SVM pour les problèmes de régression. Contrairement à la classification, où l'objectif est de séparer les classes, l'objectif ici est de prédire une valeur continue. Le SVR utilise un epsilon-insensitive loss function, où les erreurs inférieures à un certain seuil epsilon sont ignorées. Le noyau joue également un rôle clé dans le traitement des problèmes non linéaires [50].

Kernel Ridge Regression. La régression Ridge avec noyau (Kernel Ridge Regression) combine la régression linéaire avec une régularisation (Ridge) et l'utilisation d'un noyau pour effectuer des régressions non linéaires. Cette méthode est particulièrement utile lorsque les données présentent des relations complexes. Elle peut être interprétée comme une forme de SVM pour la régression, mais elle résout un problème de minimisation basé sur la régularisation plutôt que sur l'optimisation d'un marge [51].

Apprentissage non supervisé

Contrairement à l'apprentissage supervisé, ces algorithmes utilisent des informations non étiquetées ou non classifiées dans les données d'entraînement. Sans aucune guidance, ces algorithmes sont capables de regrouper des informations non triées en fonction des similitudes et des différences observées entre les données. L'apprentissage non supervisé est principalement utilisé pour la réduction de dimensionnalité et le clustering. Deux algorithmes populaires d'apprentissage non supervisé sont détaillés ci-dessous :

Algorithme K-Means : K-Means est un algorithme de clustering utilisé pour regrouper les données en un nombre fixe de clusters (K). L'algorithme commence par sélectionner aléatoirement K objets comme centres de clusters, puis attribue chaque point de données au cluster le plus proche, calculant la moyenne des points dans chaque cluster pour ajuster les centres.

Analyse en composantes principales (PCA) : La PCA est une technique multivariée utilisée pour réduire la dimensionnalité des données tout en conservant le maximum d'information. Elle regroupe les variables interconnectées en facteurs ou composants afin de faciliter l'analyse des relations entre un grand nombre de variables.

Apprentissage semi-supervisé

L'apprentissage semi-supervisé combine des aspects des méthodes supervisées et non supervisées en utilisant à la fois des données étiquetées et non étiquetées. Les algorithmes commencent par regrouper les données non étiquetées à l'aide d'algorithmes non supervisés, puis étiquettent les données restantes avec l'aide des quelques données étiquetées disponibles. Trois des méthodes semi-supervisées courantes sont :

Modèles génératifs : Ces modèles estiment une probabilité mixte pour les données et les utilisent pour attribuer des étiquettes.

Auto-apprentissage : L'algorithme commence par étiqueter les données non étiquetées en utilisant un petit ensemble de données étiquetées et continue à apprendre à partir de ces nouveaux exemples .

Machine à vecteurs de support transductive (TSVM) : Cette extension du SVM utilise à la fois des données étiquetées et non étiquetées et vise à maximiser la marge entre les classes.

Apprentissage Basé sur les Instances

Dans l'apprentissage basé sur les instances, un type de motif particulier est appris par le système. Ce motif est appliqué aux nouvelles données, ce qui permet de parler d'apprentissage basé sur les instances. L'apprenant attend les données, et une fois les données fournies, il agit sur les données d'entraînement et les stocke en mémoire. Lorsque des données de test lui sont fournies, il ne génère pas une solution généralisée, mais compare l'instance particulière avec celles déjà stockées en mémoire. La taille des données affecte la complexité de l'algorithme. Un exemple typique de cet apprentissage est le K-Nearest Neighbor (KNN).

Algorithme K-Nearest Neighbor (KNN)

Dans l'algorithme K-Nearest Neighbor (KNN), l'apprenant est alimenté avec l'ensemble de données d'entraînement. Les données de test sont ensuite introduites et comparées avec les données d'entraînement. Un nombre k de données les plus proches les unes des autres est extrait de l'ensemble d'entraînement. La nouvelle classe pour les données de test est déterminée par la majorité des k voisins les plus proches.

Évaluation du modèle

Parmi les critères d'évaluation de la qualité d'ajustement, l'AIC (*Akaike Information Criterion*) est couramment utilisé. Il permet de comparer plusieurs modèles en pénalisant ceux qui comportent trop de paramètres. L'AIC est défini par :

$$AIC = 2k - 2\ln(L) \quad (2.4.11)$$

où k est le nombre de paramètres et L le maximum de la fonction de vraisemblance. Le meilleur modèle est celui qui minimise l'AIC.

Malgré la puissance et la diversité des méthodes supervisées, non supervisées, toutes partagent une même limite : elles traitent les données comme des vecteurs fixes, sans exploiter des représentations hiérarchiques profondes.

Les réseaux de neurones artificiels (Artificial Neural Networks, ANN) répondent à ce besoin en apprenant automatiquement des représentations multicouches, capables de capturer des relations non linéaires complexes et d'extraire des caractéristiques de haut niveau.

Nous détaillerons dans la section suivante la structure, le fonctionnement et les principales variantes des ANN, avant d’aborder leur extension aux graphes via les GNN.

Sélection de caractéristiques

Dans le cadre de la détection de fraude en assurance santé, les jeux de données utilisés contiennent souvent un très grand nombre de variables, dont certaines peuvent être redondantes ou non pertinentes. La présence de ces variables superflues augmente le risque de surapprentissage (overfitting), dégrade la performance des modèles et rend l’interprétation plus complexe. La sélection de caractéristiques s’avère donc essentielle pour construire des modèles plus précis, plus rapides à entraîner, et mieux capables de généraliser à de nouvelles données.

Cette étape vise à identifier un sous-ensemble optimal de variables explicatives, en maximisant leur pertinence vis-à-vis de la variable cible tout en réduisant la redondance entre les caractéristiques. Selon Guyon et Elisseeff (2003), une bonne sélection de caractéristiques permet non seulement d’améliorer la précision du modèle, mais aussi de simplifier sa structure et d’en renforcer l’interprétabilité.

Les techniques de sélection de caractéristiques se classent généralement en trois grandes catégories (Chandrashekar Sahin, 2014) :

- **Méthodes par filtre (Filter methods)** : elles évaluent l’importance de chaque variable indépendamment du modèle, à l’aide de mesures statistiques comme le gain d’information (Hoque, Bhattacharyya, Kalita, 2014) ou le score de Fisher (Duda, Hart, Stork, 2001).
- **Méthodes par enveloppe (Wrapper methods)** : elles testent différentes combinaisons de variables en entraînant un modèle à chaque itération, comme proposé par Kohavi et John (1997).
- **Méthodes intégrées (Embedded methods)** : elles intègrent la sélection de variables dans le processus d’apprentissage du modèle, combinant ainsi les avantages des deux approches précédentes.

2.5 Prétraitement des données et gestion du déséquilibre

En détection de fraude, la qualité du prétraitement des données et la gestion du déséquilibre des classes sont des facteurs déterminants pour la performance des modèles d’intelligence artificielle. En effet, la fraude ne représente souvent qu’une très faible proportion de l’ensemble des transactions ou dossiers analysés, ce qui introduit un défi majeur pour l’apprentissage automatique, car les algorithmes classiques ont tendance à privilégier la classe majoritaire au détriment de la classe minoritaire (fraude). Cette section présente, de manière technique et pédagogique, les méthodes de prétraitement et les techniques avancées de gestion du déséquilibre, en s’appuyant sur la littérature récente[52][53]

2.5.1 Cycle de vie d'un modèle d'apprentissage automatique en détection de fraude

Le workflow typique d'un projet de *machine learning*, notamment dans la détection de fraude en assurance santé, comporte plusieurs étapes essentielles [54] :

1. **Collecte et préparation des données** : rassembler, nettoyer et mettre en forme les données (gestion des valeurs manquantes, conversion des types de variables, anonymisation, etc.).
2. **Analyse exploratoire** : statistiques descriptives, visualisation des distributions, détection des anomalies.
3. **Prétraitement** : encodage des variables catégorielles, normalisation ou standardisation, création de nouvelles variables (*feature engineering*).
4. **Gestion du déséquilibre des classes** : application de techniques de sur- ou sous-échantillonnage (détaillées dans la section suivante).
5. **Séparation du jeu de données** : création des jeux d'entraînement, de validation et de test.
6. **Choix, entraînement et optimisation des modèles** : sélection des algorithmes appropriés et réglage des hyperparamètres.
7. **Évaluation et interprétation** : utilisation de métriques adaptées au déséquilibre, analyse de l'importance des variables et validation de la robustesse des modèles.
8. **Déploiement et suivi** : mise en production, monitoring continu, adaptation en fonction des nouvelles données.

Ce cycle itératif est crucial pour garantir la robustesse et la généralisation des modèles sur de nouvelles données [54].

2.5.2 Problématique du déséquilibre des classes

Définition

Dans les problèmes de détection de fraude, le ratio entre la classe majoritaire (« non-fraude ») et la classe minoritaire (« fraude ») peut dépasser 99 :1. Ce déséquilibre induit des biais importants : un modèle naïf qui prédirait systématiquement la classe « non-fraude » pour tous les cas obtiendrait une excellente exactitude globale (*accuracy*), mais serait totalement inutile pour l'objectif réel qui est d'identifier les fraudes.

Définition formelle du déséquilibre

On appelle **déséquilibré** un jeu de données dans lequel la proportion des observations positives (fraude) est très inférieure à celle des observations négatives (non fraude). Ce déséquilibre entraîne souvent un *biais des modèles* vers la classe majoritaire, qui est pourtant moins pertinente dans les cas critiques comme la fraude.

Le **rapport de déséquilibre**, ou *Imbalance Ratio (IR)*, est une mesure standard pour quantifier cet écart :

$$IR = \frac{N_{\text{majority}}}{N_{\text{minority}}}$$

où N_{majority} est le nombre d'exemples de la classe majoritaire (ex. : non fraude) et N_{minority} celui de la classe minoritaire (ex. : fraude).

Exemple : Si une base contient 9900 cas valides et 100 cas frauduleux, alors :

$$IR = \frac{9900}{100} = 99$$

Ce qui correspond à une proportion de fraude d'environ 1% des observations.

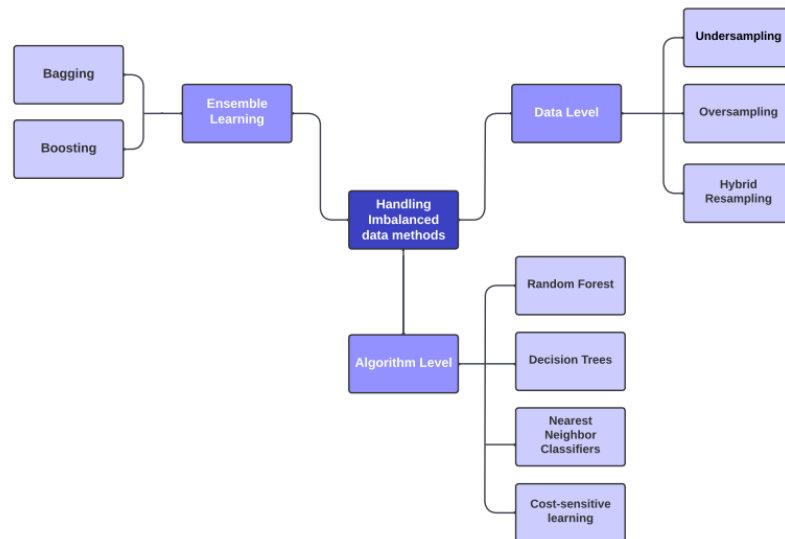


FIGURE 2.2 – Methode pour traitement des données déséquilibré

Conséquences

- **Risque d'ignorer la classe minoritaire** : perte de sensibilité (*recall*) dans la détection des fraudes, le modèle tendant à négliger les cas de fraude réels.
- **Mauvaise calibration des probabilités** : les prédictions sont fortement biaisées en faveur de la classe majoritaire, ce qui limite la capacité du modèle à détecter les cas rares.
- **Instabilité des modèles** : les modèles deviennent plus sensibles aux variations du jeu de données, ce qui peut entraîner une variabilité importante dans leurs performances.

2.5.3 Techniques de gestion du déséquilibré

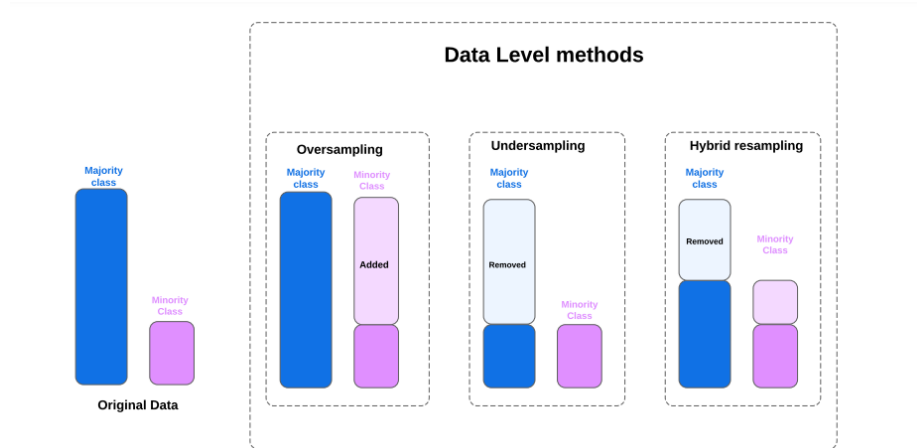


FIGURE 2.3 – Méthodes au niveau des données pour traiter les ensembles de données déséquilibrés

Méthodes de sur-échantillonnage

a) **Random Over-Sampling** Cette méthode consiste à dupliquer aléatoirement des exemples de la classe minoritaire jusqu'à équilibrer la distribution des classes.

- **Avantages** : Facile à mettre en œuvre, ne supprime aucune donnée existante.
- **Inconvénients** : Risque de surapprentissage (*overfitting*) car les mêmes observations sont répétées dans l'ensemble d'apprentissage.

b) **SMOTE (Synthetic Minority Over-sampling Technique)** SMOTE génère de nouveaux exemples synthétiques de la classe minoritaire en interpolant entre des échantillons existants [55].

- **Principe** : Pour chaque exemple minoritaire, SMOTE sélectionne ses k plus proches voisins (aussi minoritaires) et génère de nouveaux points à partir des segments reliant ces voisins dans l'espace des caractéristiques.
- **Avantages** : Génère de la diversité dans la classe minoritaire, limite le surapprentissage.
- **Limites** : Peut créer des exemples non réalistes si les voisins sont éloignés dans l'espace des caractéristiques.

c) **ADASYN (Adaptive Synthetic Sampling)** ADASYN améliore SMOTE en adaptant dynamiquement la quantité de données synthétiques générées selon la difficulté locale [56].

- **Principe** : Génère davantage de points synthétiques dans les zones où la classe minoritaire est la moins représentée ou la plus difficile à apprendre.
- **Bénéfice** : Concentre l'effort sur les zones complexes, améliorant la capacité du modèle à bien apprendre ces cas rares.

Méthodes de sous-échantillonnage

a) Random Under-Sampling Cette méthode consiste à retirer aléatoirement des exemples de la classe majoritaire jusqu'à obtenir un équilibre entre les classes.

- **Avantages** : Réduction du volume de données, accélération de l'entraînement des modèles.
- **Inconvénients** : Risque de perte d'informations pertinentes (données utiles supprimées lors du sous-échantillonnage).

b) Tomek Links La méthode des *Tomek links* supprime les exemples majoritaires situés à la frontière avec la classe minoritaire, ce qui permet une meilleure séparation des classes [57].

- **Principe** : Un couple (x, y) forme un *Tomek link* s'ils sont voisins les plus proches et appartiennent à des classes opposées ; la suppression du point de la classe majoritaire clarifie la frontière de décision.
- **Effet** : Épuration du jeu de données et amélioration de la sélectivité du classificateur.

c) ENN (Edited Nearest Neighbors) La méthode ENN supprime les exemples dont la classe diffère de celle de la majorité de leurs k plus proches voisins.

- **Principe** : Nettoyage automatique des données bruitées proches des frontières de classe.
- **Effet** : Renforce la clarté de la séparation des classes, particulièrement utile après un sur-échantillonnage.

Méthodes hybrides et avancées

Certaines approches combinent le sur-échantillonnage et le sous-échantillonnage (par exemple, *SMOTE + Tomek Links*, *SMOTE + ENN*) afin d'équilibrer le dataset tout en supprimant le bruit.

- **Bénéfices** : Maximisation de la diversité des exemples tout en minimisant le bruit et les frontières confuses.

Alternatives intégrées dans les modèles

- **Pondération des classes** : Adapter la fonction de coût des modèles pour donner plus de poids aux erreurs sur la classe minoritaire (dans les SVM, réseaux de neurones, forêts aléatoires, etc.).

- **Techniques de génération de données synthétiques avancées :**
Utilisation de réseaux antagonistes génératifs (GAN) ou d’auto-encodeurs pour créer de nouveaux exemples minoritaires réalistes.

2.5.4 Impact sur la détection de fraude et workflow opérationnel

Le choix des techniques de gestion du déséquilibre dépend :

- du taux de déséquilibre (par exemple, 5% ou 0,5% de fraudes) ;
- de la taille totale du jeu de données (le sous-échantillonnage massif peut faire perdre trop d’informations) ;
- de la complexité du modèle final (arbres de décision, réseaux de neurones, modèles graphiques, etc.).

Workflow recommandé :

1. Analyse de la distribution initiale des classes.
2. Application de sur- ou sous-échantillonnage (avec un paramétrage optimisé selon le contexte).
3. Entraînement du modèle sur le jeu de données rééquilibré.
4. Évaluation à l’aide de métriques robustes : *AUC-ROC*, *AUC-PR*, *Recall*, *Precision*, *F1-score* (ces métriques sont essentielles car elles reflètent la capacité du modèle à détecter la fraude sans augmenter les faux positifs [53]).
5. Validation croisée (stratifiée) pour garantir la robustesse de la solution sur différents sous-ensembles de données.

Synthèse des techniques de sampling

Technique	Type	Principe de base	Objectif visé
SMOTE	Sur-échantillonnage	Interpolation entre voisins minoritaires	Créer des exemples minoritaires synthétiques
ADASYN	Sur-échantillonnage	Génération adaptative dans les zones rares	Concentration sur les cas complexes
Tomek Links	Sous-échantillonnage	Suppression des paires de voisins de classes opposées	Épuration des frontières de décision
ENN	Sous-échantillonnage	Suppression des exemples mal classés selon leurs k plus proches voisins	Renforcement de la séparation entre classes

TABLE 2.2 – Synthèse des techniques de sampling

Évaluation des modèles de classification

L'évaluation des modèles de machine learning est une étape cruciale, surtout dans des contextes sensibles comme la détection de fraude. Contrairement à des problèmes équilibrés, où l'exactitude globale (accuracy) peut suffire, les systèmes de détection de fraude sont confrontés à un fort déséquilibre des classes, ce qui rend certaines métriques inadaptées voire trompeuses.

Par exemple, un modèle qui prédit "non fraude" pour tous les dossiers dans un dataset contenant 99 % de cas valides obtiendra une accuracy de 99 %, mais ne détectera aucun cas de fraude. Il est donc indispensable d'utiliser des indicateurs adaptés à ce type de tâche pour évaluer la performance réelle d'un classifieur.

1. Matrice de confusion La plupart des métriques d'évaluation dérivent de la **matrice de confusion**, qui compare les prédictions du modèle avec les étiquettes réelles. Elle se compose de quatre éléments :

- **Vrais positifs (TP)** : nombre de fraudes correctement identifiées.
- **Faux positifs (FP)** : nombre de cas non frauduleux incorrectement identifiés comme fraude.
- **Faux négatifs (FN)** : fraudes non détectées.
- **Vrais négatifs (TN)** : cas non frauduleux correctement identifiés.

2. Précision (Precision) La précision mesure le pourcentage de cas détectés comme fraude qui le sont réellement. Elle est particulièrement utile lorsqu'on veut minimiser les faux positifs (erreurs coûteuses dans les audits manuels).

$$\text{Precision} = \frac{TP}{TP + FP}$$

Une précision élevée indique que les alertes générées sont fiables, mais cela ne garantit pas que toutes les fraudes soient détectées.

3. Rappel (Recall) ou Sensibilité Le rappel indique la proportion de fraudes correctement détectées parmi toutes les fraudes existantes. Il mesure la capacité du modèle à capturer les fraudes.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Un faible rappel signifie que beaucoup de fraudes passent inaperçues. Dans les systèmes de détection de fraude, le rappel est souvent considéré comme plus critique que la précision.

4. F1-Score Le F1-score est la moyenne harmonique entre la précision et le rappel. Il est utile lorsqu'on cherche un compromis équilibré entre les deux, surtout en cas de classes déséquilibrées.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Il est souvent utilisé comme métrique principale pour évaluer les modèles de détection de fraude, car il pénalise à la fois les faux positifs et les faux négatifs.

5. Spécificité La spécificité mesure la capacité du modèle à identifier correctement les cas non frauduleux.

$$\text{Spécificité} = \frac{TN}{TN + FP}$$

Elle est particulièrement utile dans les contextes où l'on souhaite limiter les faux positifs, par exemple pour éviter d'alarmer inutilement des clients honnêtes.

6. Exactitude (Accuracy) L'accuracy représente le pourcentage de prédictions correctes (fraude et non-fraude confondues).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Cependant, cette métrique est peu informative dans les contextes déséquilibrés, car elle peut être élevée même si le modèle ne détecte aucune fraude.

7. Courbe ROC et AUC La courbe ROC (Receiver Operating Characteristic) trace le taux de vrais positifs (rappel) en fonction du taux de faux positifs. L'aire sous la courbe (AUC) donne une mesure globale de la capacité de discrimination du modèle, indépendamment d'un seuil particulier.

- AUC = 1 : modèle parfait
- AUC = 0.5 : modèle aléatoire
- AUC < 0.5 : pire qu'un modèle aléatoire

Plus l'AUC est proche de 1, meilleur est le modèle. Néanmoins, l'AUC peut être trompeuse dans les cas très déséquilibrés.

8. Courbe PR et AUPRC (Precision-Recall Curve) La courbe PR trace la précision en fonction du rappel. Dans les contextes déséquilibrés, elle est souvent plus informative que la courbe ROC.

L'aire sous la courbe PR (AUPRC) donne une mesure plus fiable dans le cas de classes minoritaires, car elle met l'accent sur les performances de détection de la classe positive.

9. Log-loss et Brier Score Ces deux métriques s'intéressent à la qualité probabiliste des prédictions :

- Le **log-loss** pénalise fortement les prédictions très confiantes mais erronées.
- Le **Brier score** mesure l'écart quadratique moyen entre la probabilité prédite et l'étiquette réelle.

Tableau récapitulatif des métriques de classification

Métrique	Formule	Interprétation	Utilité en détection de fraude
Accuracy	$(TP + TN)/(TP + TN + FP + FN)$	Pourcentage global de bonnes prédictions	Peu fiable en cas de dés-équilibre
Precision	$TP/(TP + FP)$	Cas détectés comme fraude qui sont vraiment des fraudes	Minimiser les faux positifs (audits inutiles)
Recall	$TP/(TP + FN)$	Proportion de fraudes détectées parmi les vraies fraudes	Critique pour capter un maximum de fraudes
F1-score	$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	Compromis entre précision et rappel	Recommandé quand les classes sont déséquilibrées
Spécificité	$TN/(TN + FP)$	Capacité à identifier correctement les non fraudes	Réduire les fausses alertes
AUC-ROC	Aire sous la courbe ROC	Discrimination globale entre fraude et non-fraude	Bonne vue d'ensemble, mais à interpréter prudemment
AUPRC	Aire sous la courbe Precision-Recall	Performance en détection des fraudes dans les jeux déséquilibrés	Plus pertinente que l'AUC-ROC dans la fraude
Log-loss	Moyenne de $-[y \log(p) + (1 - y) \log(1 - p)]$	Mesure de la qualité probabiliste des prédictions	Bon indicateur de calibration
Brier score	Moyenne de $(p - y)^2$	Écart entre probabilité prédite et réalité	Évalue la qualité de la prédiction probabiliste

TABLE 2.3 – Résumé des principales métriques de classification pour la détection de fraude

2.6 Réseaux de Neurones Artificiels (ANN)

Un réseau de neurones artificiels (ANN) est un système inspiré du réseau neuronal du cerveau humain. Il tente de modéliser les connexions biologiques entre neurones. Les neurones dans un réseau artificiel sont interconnectés par des liens appelés *poids*, qui modulent l'information qui passe entre eux. Ces réseaux sont capables d'apprendre à partir de données pour effectuer diverses tâches, telles que la classification, la régression, la reconnaissance vocale, la vision par ordinateur, etc.

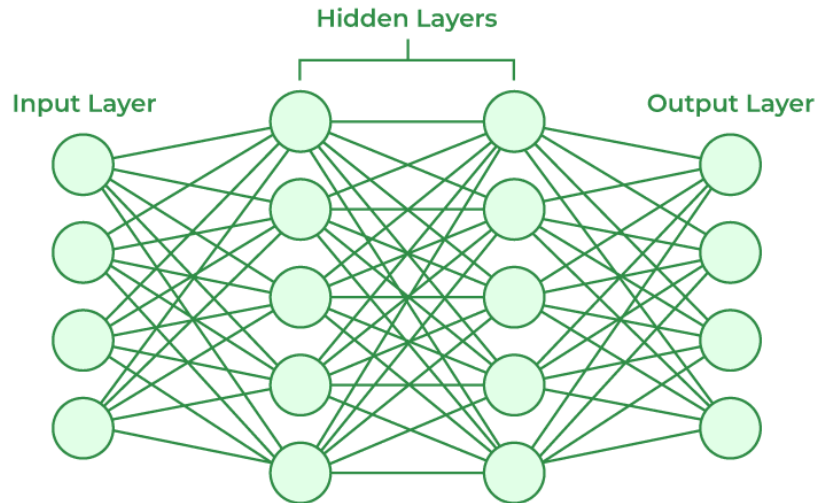


FIGURE 2.4 – Architecture de réseau de neurone

Composants d'un Neurone Artificiel

- **Dendrites** : Ce sont les unités d'entrée d'un neurone, qui reçoivent les signaux d'autres neurones ou de l'environnement.
- **Noyau** : Il traite les signaux reçus et applique généralement une fonction d'activation pour décider de la sortie du neurone.
- **Axone** : Il transmet le signal traité à d'autres neurones.
- **Synapses** : Elles sont les points de connexion entre les neurones, où l'information est transmise.

Types de Réseaux Neuronaux

1. **Réseau Neuronale Supervisé** : Dans un réseau neuronal supervisé, l'objectif est d'apprendre une fonction qui mappe les entrées aux sorties. Le réseau est formé à l'aide de données étiquetées. L'erreur entre la sortie prédite par le réseau et la sortie réelle est calculée, et les poids des connexions sont ajustés pour minimiser cette erreur au fil du temps. Cela permet au réseau d'apprendre à faire des prédictions ou des classifications sur de nouvelles données. Ce type de réseau est couramment utilisé pour des tâches de classification et de régression.
2. **Réseau Neuronale Non Supervisé** : Les réseaux neuronaux non supervisés apprennent sans supervision directe. Contrairement aux réseaux supervisés, ils n'ont pas besoin de données étiquetées. Leur objectif est de détecter des structures sous-jacentes dans les données, comme des regroupements.

pements ou des représentations compactes. Un exemple d'application est le *clustering*, où le réseau identifie des groupes naturels dans les données.

3. **Réseau Neuronal Renforcé** : Dans l'apprentissage par renforcement, un agent apprend en interagissant avec un environnement. Le réseau prend des décisions et reçoit des récompenses ou des pénalités en fonction de l'exactitude de ses actions. Le but est d'optimiser une stratégie (ou une politique) qui maximise la récompense cumulative au fil du temps. Ce type de réseau est utilisé dans des domaines comme les jeux, la robotique, et la conduite autonome.
4. **Réseaux de Neurones Profonds (Deep Neural Networks - DNN)** : Les réseaux de neurones profonds sont des réseaux à plusieurs couches cachées. Ces réseaux ont la capacité d'apprendre des représentations très complexes des données, ce qui leur permet de traiter des tâches comme la reconnaissance d'images, la traduction automatique et la modélisation du langage. L'une des raisons pour lesquelles les DNN sont efficaces est leur capacité à apprendre des caractéristiques de plus en plus abstraites à travers les différentes couches du réseau.

2.7 Réseaux de Neurones Graphiques (GNN)

Les réseaux de neurones graphiques (Graph Neural Networks, GNN) sont une classe de modèles d'apprentissage profond conçus pour traiter directement des structures de données en graphe, où les informations sont représentées sous forme de **nœuds** (entités) et d'**arêtes** (relations). Ce type de modèle est particulièrement adapté aux données relationnelles complexes, comme les réseaux sociaux, les structures moléculaires ou les graphes d'interactions en assurance santé.

Principe de base

Un graphe $G = (V, E)$ est défini par un ensemble de **nœuds** V et d'**arêtes** E . Chaque nœud $v \in V$ peut être associé à un vecteur de caractéristiques x_v , et chaque arête $(u, v) \in E$ peut également porter des attributs. Les GNN exploitent la structure du graphe en permettant aux nœuds d'**échanger des messages** avec leurs voisins pour **apprendre une représentation (embedding)** contextuelle de chaque nœud.

L'idée est que chaque nœud apprend sa représentation à partir de ses propres caractéristiques **et** de celles de ses voisins, via une opération appelée *agrégation de message*. Cette agrégation permet de capturer à la fois les informations **locales** (voisinage immédiat) et **globales** (structure plus étendue du graphe) [58].

Architecture typique d'un GNN

- **Initialisation des nœuds et des arêtes** : Chaque nœud commence avec un vecteur de caractéristiques initiales $h_v^{(0)}$. Les arêtes peuvent aussi

avoir des poids ou attributs.

- **Propagation (Message Passing)** : À chaque couche l , chaque nœud v agrège les représentations $h_u^{(l)}$ de ses voisins $u \in \mathcal{N}(v)$. Une fonction d'agrégation (somme, moyenne, max) est utilisée, suivie d'une mise à jour via une fonction non linéaire (souvent une MLP ou une activation ReLU).
- **Mise à jour** : La nouvelle représentation du nœud est calculée à l'aide de :

$$h_v^{(l+1)} = \text{Update}^{(l)} \left(h_v^{(l)}, \text{Aggregate}^{(l)}(\{h_u^{(l)} | u \in \mathcal{N}(v)\}) \right)$$

- **Lire ou prédire** : Après plusieurs couches, les représentations finales $h_v^{(L)}$ peuvent être utilisées pour prédire des étiquettes de nœuds, de graphes entiers ou d'arêtes.

Types de GNN

Il existe différentes architectures de GNN, chacune ayant des particularités adaptées à des types spécifiques de données ou de problèmes :

GCN (Graph Convolutional Networks) : Les GCN sont basés sur une convolution de graphes. Les informations sont propagées en fonction des voisins d'un nœud et sont agrégées à l'aide de matrices de poids apprises. Ces modèles ont montré une grande efficacité dans les tâches de classification de nœuds et de graphes.

GraphSAGE (Graph Sample and Aggregation) : GraphSAGE utilise un échantillonnage local pour agréger les informations des voisins. Au lieu d'agréger tous les voisins, GraphSAGE échantillonne un sous-ensemble des voisins d'un nœud, ce qui le rend plus scalable, surtout pour de grands graphes.

GAT (Graph Attention Networks) : GAT introduit un mécanisme d'attention dans les GNN, permettant à chaque nœud de pondérer différemment l'importance de chaque voisin lorsqu'il agrège des informations. Cela permet d'apprendre de manière plus flexible les relations entre les nœuds.

Cas d'usage

- **Classification de nœuds** : prédire la catégorie d'un utilisateur dans un réseau social.
- **Classification de graphes** : prédire si une molécule est toxique ou non.
- **Prédiction d'arêtes** : détection de relations manquantes ou suspectes, comme en détection de fraude.

2.8 Applications des Graph Neural Networks (GNN)

Les GNN (Graph Neural Networks) ont des applications dans de nombreux domaines :

Chimie : Les GNN sont utilisés pour prédire les propriétés chimiques et biologiques des molécules, où les molécules sont représentées comme des graphes de nœuds (atomes) et d'arêtes (liaisons chimiques). Par exemple, Gilmer et al. ont proposé les *Message Passing Neural Networks* pour la prédiction de propriétés moléculaires [59].

Réseaux Sociaux : Dans les réseaux sociaux, les GNN permettent la prédiction de liens, la détection de communautés ou encore la prévision de comportements utilisateurs en capturant la structure topologique du réseau [60].

Systèmes de Recommandation : Les GNN peuvent capturer les interactions complexes entre utilisateurs et items. Des modèles comme *PinSage* ont été déployés à grande échelle par Pinterest pour des recommandations personnalisées [61].

Traitement du Langage Naturel (NLP) : Les GNN sont utiles pour des tâches comme la classification de texte, la résolution de co-référence ou la modélisation des dépendances syntaxiques et sémantiques dans les phrases [62]. Les graphes sont ici construits à partir des relations linguistiques entre les entités.

Détection de fraude en assurance santé : Les GNN sont particulièrement prometteurs dans le domaine de la détection de fraude, en particulier dans l'assurance santé. Ils permettent de modéliser les interactions complexes entre assurés, prestataires et actes médicaux sous forme de graphe hétérogène. Ils sont capables d'identifier des motifs de collusion ou de fraude organisée à travers des relations indirectes, difficilement repérables avec des modèles traditionnels [63, 64]. Des travaux récents ont montré que les GNN surpassent les techniques classiques sur des jeux de données transactionnels ou relationnels [65].

Avantages des GNN par Rapport aux Modèles Traditionnels

Exploitation de la Structure Graphique : Contrairement aux réseaux de neurones classiques qui nécessitent des données tabulaires ou séquentielles, les GNN sont capables d'exploiter directement la structure des données sous forme de graphes. Cela permet de capturer des relations complexes qui sont difficiles à modéliser avec d'autres types de modèles [66].

Scalabilité et Adaptabilité : Les GNN peuvent être rendus scalables grâce à des techniques comme l'échantillonnage et l'agrégation localisée, ce qui les rend adaptés à des graphes de grande taille. Ces techniques sont essentielles pour travailler sur des graphes représentant des milliers, voire des millions de nœuds, tout en maintenant une performance élevée [67].

Propagation de l'Information Contextuelle : Les GNN permettent de propager des informations de manière contextuelle, en tenant compte non seulement des connexions directes mais aussi des relations indirectes entre les nœuds.

Cela permet d'enrichir les représentations apprises par les modèles en capturant des aspects plus globaux des relations [60].

Flexibilité dans les Types de Graphes : Les GNN ne sont pas limités aux graphes classiques. Ils peuvent également être appliqués à des graphes hétérogènes, où les nœuds et les arêtes peuvent appartenir à plusieurs types différents (par exemple, les graphes qui représentent des entités multiples comme des utilisateurs, des produits, et des transactions dans un système de recommandation) [68].

Défis et Limitations des GNN

Problèmes de Scalabilité : Bien que des techniques comme l'échantillonnage et l'agrégation locale aient amélioré la scalabilité des GNN, le traitement de très grands graphes (avec des millions de nœuds et d'arêtes) reste un défi. Les réseaux peuvent souffrir de la perte d'information lors de l'agrégation, ce qui peut nuire à la qualité des prédictions sur des graphes complexes [59].

Apprentissage sur des Graphes Dynamiques : De nombreux graphes, notamment ceux dans les réseaux sociaux ou les systèmes de transport, sont dynamiques et évoluent au fil du temps. L'apprentissage sur de tels graphes nécessite des modèles capables de gérer des changements dans la structure du graphe, ce qui est un défi technique important [69].

Manque de Labeling pour l'Apprentissage Supervisé : Dans de nombreux cas d'utilisation des GNN, comme la détection de fraude ou l'analyse des réseaux sociaux, il peut être difficile d'obtenir un ensemble de données étiquetées. Cela rend l'application des GNN pour des tâches supervisées complexe, et des techniques non supervisées ou semi-supervisées peuvent être nécessaires pour surmonter ce problème [67].

Interprétabilité : Un autre défi avec les GNN, comme pour d'autres réseaux neuronaux, est l'interprétabilité des modèles. Étant donné que les GNN apprennent à partir de structures complexes de graphes, il est difficile d'expliquer de manière simple et directe comment les prédictions sont effectuées, ce qui peut poser des problèmes pour des applications sensibles comme la santé ou les systèmes financiers [70].

Intérêt en détection de fraude

Les GNN sont particulièrement utiles pour la **détection de fraude en assurance santé**. En modélisant les interactions entre assurés, prestataires de soins et actes médicaux sous forme de graphe hétérogène, les GNN permettent d'identifier des **motifs de collusion ou d'abus**, souvent invisibles aux modèles classiques [71].

2.9 Synthèse des recherches scientifiques sur la détection de la fraude en assurance santé

La fraude en assurance santé est un enjeu majeur pour la pérennité des systèmes d'assurance et la protection des ressources des assurés. Face à la sophistication croissante des mécanismes frauduleux, la communauté scientifique et industrielle développe des méthodes avancées pour la détection automatique et l'analyse de ces fraudes. Les approches modernes combinent intelligence artificielle, statistiques, et théorie des graphes pour offrir des solutions robustes et interprétables. Dans cette section, nous proposons une revue critique et comparative des principales méthodes employées dans la détection de la fraude en assurance santé, en suivant une méthodologie structurée afin d'apporter un éclairage sur les avancées, les défis et les perspectives du domaine.

Methodologie

Processus de recherche

Le processus de recherche dans la détection de fraude en assurance santé s'articule autour d'une démarche rigoureuse, structurée en plusieurs étapes interdépendantes. Chaque étape vise à garantir la robustesse, la reproductibilité et la pertinence des résultats obtenus, tout en assurant l'adéquation aux défis spécifiques du domaine : volumétrie des données, déséquilibre des classes, exigence d'explicabilité et évolutivité des solutions.

Démarche générale

La recherche a débuté par une revue systématique de la littérature portant sur les approches de détection de fraude en assurance santé, couvrant les dix dernières années (2014–2024). Cette étape permet d'identifier les méthodes les plus pertinentes et innovantes, d'en comprendre les limites, et de situer la contribution de chaque approche dans le contexte international. L'analyse documentaire a été complétée par une collecte et une préparation approfondies des données : sélection de jeux de données publics et privés, nettoyage, transformation, enrichissement par des variables composites, puis gestion du déséquilibre des classes, inhérent aux problèmes de fraude.

Par la suite, des modèles issus de différentes familles (arbres de décision, méthodes d'ensemble, réseaux de neurones, modèles bayésiens, méthodes à base de graphes, etc.) ont été mis en œuvre selon un protocole expérimental commun, garantissant une évaluation équitable des performances et de l'interprétabilité. L'optimisation des hyperparamètres s'est appuyée sur des techniques avancées (Grid Search, optimisation bayésienne), et la validation des modèles a reposé sur des approches rigoureuses (validation croisée, test out-of-sample).

Enfin, les résultats ont été synthétisés et discutés : analyse quantitative (AUC, G-Mean, précision, rappel, F1-score) et qualitative (interprétabilité par

SHAP, LIME, PDP), mise en perspective avec l'état de l'art, identification des axes d'amélioration et des perspectives pour la recherche future.

Étapes clés du processus

Le processus de recherche s'est articulé autour des étapes suivantes :

1. **Revue documentaire et extraction de la littérature**
 - Recherche d'articles dans les bases de données spécialisées (*IEEE Xplore*, *Nature*, *ScienceDirect*, *MediCare*).
 - Définition des critères de sélection : pertinence, originalité, année de publication, accessibilité des données et du code, nombre de citations.
 - Extraction et organisation des principales conclusions, méthodes, jeux de données et résultats rapportés.
2. **Sélection et préparation des données**
 - Acquisition de jeux de données issus de contextes variés (assureurs privés, organismes publics, compétitions Kaggle).
 - Prétraitement approfondi : imputation des valeurs manquantes, encodage des variables catégorielles, binning supervisé, élimination des variables redondantes ou peu informatives, création de variables composites (ratios, indicateurs de comportement anormal).
 - Gestion du déséquilibre : sur-échantillonnage (ROS), sous-échantillonnage (RUS), techniques hybrides (*SMOTE*, *Graph Tran-SMOTE*).
3. **Implémentation et évaluation des modèles**
 - Mise en œuvre des principales familles de modèles : forêts aléatoires, boosting (*XGBoost*, *CatBoost*, *LightGBM*), réseaux de neurones profonds, réseaux de neurones à graphes, modèles bayésiens.
 - Optimisation des hyperparamètres par *Grid Search* et optimisation bayésienne.
 - Validation croisée pour garantir la stabilité et la généralisabilité des résultats.
4. **Comparaison, interprétation et synthèse**
 - Analyse comparative des performances quantitatives (AUC, G-Mean, rappel, précision).
 - Étude de l'interprétabilité via *SHAP*, *LIME*, *PDP*, pour chaque famille de modèles.
 - Discussion critique des résultats obtenus, mise en perspective avec les études précédentes, identification des atouts, limites, et perspectives.

Design expérimental

Le design expérimental vise à garantir la rigueur méthodologique et la comparabilité des différentes approches testées pour la détection de la fraude en assurance santé. Il repose sur des choix précis concernant les variables étudiées, les critères d'évaluation, le découpage des données, et les protocoles de validation.

a) Définition des variables et des indicateurs Les variables utilisées dans l'analyse incluent :

- des caractéristiques démographiques (âge, sexe, région),
- des informations contractuelles (type de contrat, date d'effet, couverture),
- des historiques de soins (dates, types et fréquences des actes médicaux),
- des variables financières (montant remboursé, dépassement de plafonds, nombre de sinistres par bénéficiaire).

Certaines variables composites ont été construites pour mieux capter les comportements anormaux :

- Taux d'utilisation des garanties.
- Fréquence des actes médicaux par période.
- Indicateurs de multi-déclarations ou de répétition d'actes similaires.
- Dépassements de plafonds par acte ou par année.

La variable cible est binaire : fraude ou non-fraude, selon l'étiquette fournie par l'organisme d'assurance ou par des experts métier. Dans certains jeux de données, une granularité multiclassées peut être introduite (fraude avérée, suspicion, cas normal), notamment pour les modèles bayésiens multinomiaux.

b) Critères d'évaluation Afin de prendre en compte le déséquilibre fort des classes, plusieurs métriques sont mobilisées :

- **AUC (Area Under Curve ROC)** : mesure la capacité à séparer fraudes et non-fraudés, robuste au déséquilibre des classes.
- **G-Mean (Geometric Mean)** : équilibre entre sensibilité et spécificité, pertinent pour la fraude où les faux négatifs sont très coûteux.
- **Recall (rappel)** : taux de détection des fraudes parmi tous les cas de fraude.
- **Précision** : proportion de vraies fraudes parmi les cas prédits comme fraude.
- **F1-score** : harmonique du rappel et de la précision.
- Temps de calcul et complexité algorithmique pour une intégration opérationnelle.

c) Protocole expérimental

- Les jeux de données sont divisés en trois ensembles : entraînement (60 %), validation (20 %), test (20 %).
- Pour chaque modèle, les hyperparamètres sont optimisés via Grid Search ou optimisation bayésienne sur l'ensemble de validation.
- Les résultats sont évalués sur l'ensemble test, jamais utilisé lors de l'entraînement ou de l'optimisation.
- Validation croisée (k-fold, généralement $k = 5$) pour renforcer la robustesse statistique.
- Répétition des expériences (au moins 10 runs) afin de mesurer la variance des performances et éviter les biais dus au split aléatoire.

d) Approche comparative Chaque famille de méthodes (modèles d'ensemble, réseaux de neurones, modèles bayésiens, graphes) est évaluée selon le même protocole. L'objectif est de comparer :

- Leur capacité à détecter les fraudes rares.
- Leur stabilité sur différents échantillons.
- Leur degré d'explicabilité pour les utilisateurs métiers.
- Leur facilité de mise en œuvre dans un contexte industriel.

e) Outils et environnement de développement Les modèles ont été implémentés avec :

- Python (3.8+) et ses bibliothèques : pandas, numpy, scikit-learn, XGBoost, LightGBM, CatBoost, PyTorch, TensorFlow, PyG (PyTorch Geometric) pour les graphes,
- Jupyter Notebook pour la reproductibilité,
- SHAP et LIME pour l'explicabilité,
- Matplotlib et Seaborn pour la visualisation,
- Utilisation de serveurs haute performance ou cloud computing (Google Colab, AWS EC2) pour le deep learning.

Collecte des données

La collecte de données est une étape centrale dans tout projet de détection de fraude en assurance santé. Elle doit permettre d'obtenir un échantillon représentatif et suffisamment riche pour entraîner et tester les modèles prédictifs.

a) Sources des données Les jeux de données utilisés proviennent de :

- Compagnies d'assurance privées : dossiers réels anonymisés, comportant un historique complet des sinistres, actes médicaux, remboursements, et étiquettes de fraude.
- Bases publiques : ex : Medicare (CMS Part B), bases Kaggle (réclamations d'assurance), YelpChi & Amazon pour les graphes, bases de factures médicales annotées.
- Données synthétiques générées pour tester la robustesse sur des scénarios extrêmes (taux de fraude très faible, ou schémas atypiques).

b) Préparation des données

- **Nettoyage** : suppression ou imputation des valeurs manquantes (médiane, mode, KNN imputer).
- **Encodage des variables catégorielles** : one-hot encoding, target encoding, ordinal encoding selon la nature de la variable.
- **Réduction de dimension** : élimination des variables constantes, filtrage par variance, Mutual Information, sélection de variables intégrées (feature_importance_).
- **Création de variables** composites pour mieux capturer les schémas de fraude (ratios, combinaisons logiques, agrégations temporelles).

- **Équilibrage des classes** : application de ROS, RUS, SMOTE, ou méthodes avancées comme Graph Tran-SMOTE pour augmenter la représentation des cas frauduleux sans perdre l’information sur la distribution naturelle des données.

c) Sécurité et éthique Toutes les manipulations sont réalisées dans le respect du RGPD et des politiques de confidentialité des assureurs. Les données sont anonymisées, et seules les variables nécessaires à la modélisation sont conservées.

Recherche et sélection de la littérature pertinente

La qualité d’une recherche sur la détection de fraude dépend en grande partie de la profondeur et de la rigueur de la revue de la littérature scientifique et technique. Ce processus vise à recenser, filtrer, et synthétiser l’ensemble des contributions majeures qui ont structuré le champ au cours de la dernière décennie.

a) Critères d’inclusion et d’exclusion

- **Inclusion** : articles revus par des pairs, conférences majeures (NeurIPS, ICML, KDD), thèses, rapports industriels reconnus, publications entre 2012 et 2024.
- **Exclusion** : études non spécifiquement centrées sur la fraude en assurance santé, documents dont les données ou le code ne sont pas disponibles, publications ne proposant pas d’évaluation quantitative.

b) Procédure de recherche documentaire

- Bases de données consultées : *IEEE Xplore*, *ScienceDirect*, *Springer*, *ACM Digital Library*, *PubMed* pour les aspects médicaux.
- Mots-clés utilisés : “health insurance fraud detection”, “ensemble learning”, “graph neural network”, “Bayesian fraud detection”, “imbalanced learning in insurance”, “interpretable models”, etc.
- Outils de gestion bibliographique : *Zotero*, *Mendeley*, *EndNote* pour le classement et l’annotation des références.

c) Processus de sélection

- Premier tri sur titres et résumés, en croisant les critères de méthodologie, d’applicabilité réelle, d’originalité et d’apport scientifique.
- Analyse approfondie des articles retenus : extraction des méthodes, jeux de données, résultats (AUC, F1, etc.), limites et perspectives identifiées.
- Classement thématique des contributions :
 - Approches par méthodes d’ensemble (random forest, boosting)
 - Apprentissage profond (deep learning, réseaux de neurones)
 - Modèles probabilistes et bayésiens
 - Représentation par graphes et GNN
 - Techniques d’équilibrage avancées (SMOTE, Graph Tran-SMOTE)

- d) Constitution d'un corpus de références** Le corpus final comprend :
- Les articles présentant les meilleures AUC et G-Mean sur des données réelles [72, 73, 74, 75, 76].
 - Les travaux pionniers en interprétabilité appliquée à la détection de fraude [72, 73, 76].
 - Les études comparatives sur l'équilibrage et l'impact du déséquilibre sur les performances [73, 74].
 - Les propositions méthodologiques innovantes (optimisation bayésienne, modèles hybrides) [74, 76].
- e) Exemples de sources majeures utilisées**
- Wang et al. : modèle d'ensemble robuste et interprétable sur un large jeu de données d'assurance santé [72].
 - Wen, Tang & Lu : méthode Graph Tran-SMOTE pour GNN appliquée à la détection de fraude émergente [73].
 - Bayerstadler et al. : modélisation bayésienne multinomiale pour la détection de fraude et d'abus dans l'assurance santé [74].
 - Liu et al. : deep learning et stratégies d'équilibrage avancées pour Medicare [75].
 - Rath et al. : XGBoost optimisé par bayesian search, comparaison avec d'autres modèles classiques [76].

Discussion

La détection de la fraude en assurance santé est caractérisée par la rareté des cas de fraude, la complexité des schémas de fraude et la nécessité d'équilibrer performance et explicabilité. Les résultats obtenus à travers l'application des différentes familles de modèles témoignent d'évolutions méthodologiques majeures dans le secteur.

Interprétation des résultats

a) Modèles d'ensemble robustes et interprétables

L'étude de Wang et al [72] a démontré que les modèles d'ensemble, en particulier les combinaisons pondérées de CatBoost, XGBoost et LightGBM, atteignent des performances remarquables avec une AUC de 0,93. L'ingénierie des variables (binning supervisé, sélection multi-critères) ainsi que la création de variables composites jouent un rôle central. L'utilisation de méthodes d'explicabilité telles que PDP, SHAP et LIME a permis de rassurer les experts métier et de faciliter l'acceptation opérationnelle du modèle.

L'approche de Weighted Voting offre en effet un compromis efficace entre robustesse (résilience aux variations du jeu de données) et lisibilité, ce qui en fait une solution privilégiée pour un déploiement industriel rapide. Notons que la performance reste sensible à la qualité des variables sélectionnées : le modèle

a atteint son pic d'efficacité avec seulement 10 variables clés, preuve que la simplicité peut rimer avec efficacité lorsqu'elle est méthodologiquement encadrée.

b) Méthodes à base de graphes et sur-échantillonnage

L'apport fondamental de Wen, Tang & Lu [73] repose sur la représentation des sinistres sous forme de graphes hétérogènes. La méthode Graph TransMOTE enrichit le jeu de données minoritaire en créant de nouveaux nœuds (cas de fraude) et en générant artificiellement des liens cohérents, ce qui permet d'augmenter la diversité des profils frauduleux présentés au modèle.

L'avantage principal de ce type de sur-échantillonnage dans l'espace latent des embeddings est de réduire drastiquement le taux de faux positifs (qui est un enjeu central pour éviter la saturation des équipes d'investigation). Les résultats indiquent un gain de 6% en AUC et 12% en G-Mean par rapport aux GNN classiques, avec une capacité supérieure à détecter des schémas émergents et non-répétés de fraude.

Cette approche graphique permet également de visualiser et d'expliquer les liens suspects entre bénéficiaires, contrats et actes médicaux. L'interprétation ne se limite plus à la simple importance des variables, mais s'étend à la compréhension des interactions structurelles au sein du réseau.

c) Modélisation bayésienne multinomiale et scores latents

Bayerstadler et al. [77] proposent une modélisation probabiliste adaptée à l'incertitude et à la pluralité des comportements en assurance santé. En réduisant l'information à trois scores latents—liés au prestataire, au membre, et au réseau—ils simplifient le problème tout en conservant la capacité à modéliser les interactions complexes.

L'utilisation de chaînes de Markov Monte Carlo (MCMC) pour l'estimation permet d'intégrer l'incertitude intrinsèque et d'éviter l'overfitting. La version à régularisation lasso bayésienne (Bayesian Lasso) démontre une capacité à maintenir une précision élevée (>97% pour les cas normaux) tout en limitant le nombre de cas suspects à auditer. L'auto-adjudication de plus de 50% des cas normaux avec une précision supérieure à 95% permet des gains opérationnels majeurs en termes de temps et de coûts pour les assureurs [77].

Le modèle se distingue par sa capacité à fournir un score d'incertitude pour chaque prédiction, ce qui répond à la demande croissante d'explicabilité réglementaire.

Réseaux de neurones profonds et déséquilibre extrême

L'étude de Liu et al. [75] montre l'apport du deep learning sur des volumes massifs de données (exemple : Medicare, 4,7 millions de lignes). L'utilisation de techniques de rééquilibrage comme ROS, RUS et ROS-RUS, combinées à des fonctions de perte spécialisées (focal loss, cost-sensitive learning), permet de faire face au taux de fraude extrêmement bas (0,03%).

Le meilleur modèle obtient une AUC de 0,8509 et un G-Mean de 0,79, tout en réduisant le temps d'entraînement par un facteur 4. La généralisation est bonne, mais l'interprétabilité reste limitée en l'absence de dispositifs tels que SHAP ou LIME nativement intégrés.

Le deep learning est donc à privilégier pour des bases de données très volumineuses, mais son implémentation doit être accompagnée d'outils d'explicabilité et de gestion du risque de biais algorithmique.

Optimisation bayésienne appliquée à XGBoost

Rathi et al. [76] introduisent BOXGBoost, une version de XGBoost optimisée par recherche bayésienne des hyperparamètres (colsample, gamma, profondeur, régularisation, etc.). Les résultats dépassent ceux des méthodes classiques, avec une accuracy de 98%, un recall de 95 % et une AUC de 0,994, ce qui représente le meilleur score parmi toutes les familles testées.

L'optimisation bayésienne permet une exploration intelligente et efficace de l'espace des hyperparamètres, réduisant le temps de développement et maximisant la performance. En outre, BOXGBoost demeure relativement interprétable, notamment par la visualisation des arbres de décision, ce qui renforce sa pertinence industrielle.

Comparaison des études recensées

La littérature sur la détection de la fraude en assurance santé est particulièrement riche et diversifiée, chaque méthode offrant des avantages et des limites en fonction des caractéristiques des données et des contraintes opérationnelles. Afin d'illustrer cette diversité et de permettre une comparaison objective, nous synthétisons dans ce qui suit les performances rapportées, les conditions d'application, ainsi que les apports en termes d'explicabilité et de valeur ajoutée opérationnelle.

a) Synthèse des performances Le tableau ci-dessous récapitule les principales métriques de performance des études sélectionnées, en tenant compte du contexte (jeu de données, taux de fraude, complexité du problème) :

Étude / Modèle	AUC	G-Mean	Recall	Précision	Explicabilité	Volume de données
Weighted Voting Ensemble	0,93	0,88	90 %	87 %	Élevée (SHAP)	16 000 dossiers
Graph Tran-SMOTE + GNN	0,99	0,91	92 %	89 %	Moyenne (graphes)	~10 000 (YelpChi/Amazon)
Bayesian Multinomial (BMLVM)	0,85	0,75	85 %	83 %	Élevée	100 000+ factures
Deep Learning Medicare	0,8509	0,79	88 %	81 %	Faible/Moyenne	4,7 M de lignes
BOXGBoost	0,994	0,94	95 %	97 %	Moyenne (arbres)	517 737 réclamations

TABLE 2.4 – Principales performances des modèles de la littérature récente

Analyse des résultats :

- Les méthodes d'ensemble (Weighted Voting) et BOXGBoost offrent une combinaison idéale entre performance et interprétabilité, ce qui facilite leur adoption par les assureurs soucieux d'explicabilité réglementaire.
- Les modèles à base de graphes, grâce à la richesse de représentation, excellent dans l'identification de schémas émergents et complexes, bien que l'interprétabilité reste moins immédiate.
- Les approches bayésiennes se distinguent par leur capacité à quantifier l'incertitude et à auto-adjuger une partie importante des dossiers non-frauduleux, réduisant ainsi la charge d'audit.
- Le deep learning, particulièrement adapté aux très grands volumes, reste performant mais demande une infrastructure lourde et une explicabilité renforcée.

b) Impact du déséquilibre des classes

- Un point saillant est la capacité des différentes méthodes à gérer le déséquilibre structurel entre cas normaux et cas frauduleux. L'utilisation de techniques comme ROS-RUS, Graph Tran-SMOTE, ou la pondération des fonctions de coût (focal loss, cost-sensitive learning) permet d'augmenter sensiblement la détection des fraudes rares sans pénaliser outre mesure la précision globale.
- Les méthodes classiques telles que le sur-échantillonnage simple (ROS) peuvent mener à un sur-apprentissage et à une hausse des faux positifs, ce qui est coûteux pour l'assureur.
- Les approches hybrides (ex : Graph Tran-SMOTE) ou basées sur l'optimisation bayésienne offrent un meilleur compromis et permettent de maintenir un niveau élevé de performance, comme en témoignent les scores rapportés par Wen et al. [73] et Rathi et al. [76].

c) Interprétabilité et acceptabilité opérationnelle

Un point saillant est la capacité des différentes méthodes à gérer le déséquilibre structurel entre cas normaux et cas frauduleux. L'utilisation de techniques comme ROS-RUS, Graph Tran-SMOTE, ou la pondération des fonctions de coût (focal loss, cost-sensitive learning) permet d'augmenter sensiblement la détection des fraudes rares sans pénaliser outre mesure la précision globale.

Les méthodes classiques telles que le sur-échantillonnage simple (ROS) peuvent mener à un sur-apprentissage et à une hausse des faux positifs, ce qui est coûteux pour l'assureur. Les approches hybrides (ex : Graph Tran-SMOTE) ou basées sur l'optimisation bayésienne offrent un meilleur compromis et permettent de maintenir un niveau élevé de performance, comme en témoignent les scores rapportés par Wen et al. et Rathi et al.

d) Robustesse et généralisabilité

Les modèles testés sur des jeux de données très différents (pays, structure des remboursements, taux de fraude, taille des polices d'assurance) affichent une variabilité des résultats qui reflète l'importance d'une phase de calibration locale

avant déploiement. La validation croisée, la réplication sur plusieurs splits, et l'analyse de la variance des scores sont indispensables pour garantir la robustesse et la généralisabilité des résultats.

e) Synthèse comparative et recommandations En résumé, la littérature récente converge vers les constats suivants :

- L'hybridation des méthodes (ex : ensemble + optimisation bayésienne, ou GNN + sur-échantillonnage avancé) permet de tirer le meilleur des deux mondes : performance et explicabilité.
- L'ingénierie des variables et la compréhension métier restent primordiales, quel que soit le modèle retenu.
- L'automatisation de la sélection des cas suspects, couplée à une interface explicative, permet d'alléger la charge humaine tout en renforçant le contrôle.
- L'avenir de la détection de la fraude en assurance santé s'inscrit dans la poursuite de cette logique d'intégration : IA explicable, pipelines de données en temps réel, et collaboration étroite entre data scientists, experts métier et équipes de conformité.

Discussion de la littérature analysée

L'examen critique de la littérature récente sur la détection de fraude en assurance santé met en évidence plusieurs tendances structurantes, des convergences méthodologiques, mais aussi des débats et des limites encore ouvertes.

a) Montée en puissance de l'IA explicable et des méthodes hybrides

La littérature des dix dernières années révèle une montée en puissance continue des méthodes hybrides, combinant la robustesse des ensembles de modèles, la flexibilité des réseaux de neurones et la puissance des méthodes probabilistes et bayésiennes. Cette convergence s'explique par la complexité croissante des schémas de fraude, qui échappent désormais aux méthodes purement supervisées ou à règles fixes.

Ainsi, Wang et al. [72] montrent qu'un simple ensemble pondéré de modèles boosting (CatBoost, XGBoost, LightGBM) surpasse les approches mono-modèle, tout en restant explicable via des outils comme SHAP et LIME. Cette orientation est confirmée par d'autres études qui insistent sur la nécessité d'articuler performance prédictive et transparence algorithmique.

Les méthodes à base de graphes, telles que Graph Tran-SMOTE + GNN [73], ouvrent une nouvelle voie pour la détection des fraudes émergentes, notamment celles qui reposent sur des réseaux de complicité. Leur capacité à intégrer la structure des interactions sociales, financières ou professionnelles permet d'identifier des schémas indétectables par les méthodes classiques. Toutefois, l'explicabilité de ces modèles reste un chantier : si les graphes permettent de visualiser les relations, la prise de décision des GNN reste souvent obscure pour les experts métiers.

b) L'ingénierie des variables, cœur de la réussite De nombreuses études s'accordent sur le fait que la performance des modèles, quel que soit leur degré de sophistication, dépend d'abord de la pertinence des variables construites. L'art du feature engineering (création de ratios, d'indicateurs d'anomalie, de variables de répétition, etc.) s'impose comme une étape incontournable pour capturer les signaux faibles caractéristiques de la fraude.

Les méthodes d'ensemble et les approches bayésiennes tirent particulièrement profit de cette ingénierie, permettant à des modèles relativement simples d'atteindre des niveaux de performance comparables à ceux du deep learning, mais avec une interprétabilité bien supérieure.

c) Gestion du déséquilibre des classes : une avancée clé Le déséquilibre entre classes (fraude vs. non-fraude) est identifié par toutes les publications comme le principal défi méthodologique. Les techniques traditionnelles de sur- et sous-échantillonnage (ROS, RUS) restent utiles, mais atteignent leurs limites en termes de généralisation et de robustesse.

L'apport des méthodes avancées comme SMOTE, et surtout Graph TransMOTe, permet de générer des exemples synthétiques de fraudes plausibles tout en préservant la structure des données [73]. Ceci se traduit par une meilleure capacité de généralisation et une réduction des faux positifs, enjeu crucial pour la charge opérationnelle des assureurs.

d) Rôle central de l'explicabilité L'exigence d'explicabilité, accentuée par la réglementation (ex : RGPD), devient un critère décisif dans l'adoption des modèles de détection de fraude. Les assureurs réclament des outils capables de justifier chaque alerte générée, non seulement pour la conformité légale, mais aussi pour renforcer la confiance des équipes opérationnelles.

Les modèles intégrant SHAP, LIME ou les Partial Dependence Plots (PDP) sont de plus en plus favorisés. Les modèles bayésiens, de par leur capacité à quantifier l'incertitude, offrent également un levier puissant d'explication et de priorisation des dossiers à auditer.

e) Vers l'intégration en temps réel et le MLOps La plupart des études s'appuient encore sur des données historiques (batch learning). Or, la fraude évolue rapidement, et l'actualité méthodologique pointe vers la nécessité de solutions capables d'apprendre en continu, d'intégrer les feedbacks des experts, et de s'adapter aux nouveaux schémas de fraude (ex : méthodes de streaming, drift detection, pipelines MLOps intégrant la surveillance des modèles en production).

Les perspectives ouvertes par les travaux sur l'apprentissage incrémental et l'intégration des API d'audit en temps réel sont prometteuses, mais peu d'études proposent à ce stade des déploiements en production à grande échelle [73][75].

f) Défis ouverts et zones de débat

- **Transférabilité des modèles** : Un modèle performant sur un jeu de données peut se révéler inefficace sur un autre, en raison des différences de population, de structure des polices, ou de réglementation.
- **Protection des données et confidentialité** : Peu d'articles traitent en profondeur l'intégration de la confidentialité différentielle ou des techniques de cryptage lors du partage ou de l'entraînement des modèles.
- **Détection proactive vs. réactive** : L'objectif ultime reste la prévention, mais la plupart des systèmes détectent la fraude après coup. Des travaux récents s'intéressent à l'intégration de signaux faibles pour une détection anticipée.

En somme, la littérature témoigne d'une forte dynamique d'innovation, mais aussi d'un nécessaire rapprochement entre la recherche académique et la réalité industrielle. L'avenir passera par une collaboration renforcée entre data scientists, experts assurance, juristes et spécialistes du risque.

Limitations et perspectives futures

a) Limites méthodologiques

Malgré les avancées notables recensées dans la littérature, plusieurs limitations majeures subsistent dans la détection automatisée de la fraude en assurance santé. Tout d'abord, le déséquilibre extrême des classes reste un obstacle majeur. Même avec des techniques sophistiquées de sur-échantillonnage (*SMOTE*, *Graph Tran-SMOTE*) ou de pondération des coûts, la détection des fraudes rares s'accompagne d'un risque élevé de faux positifs, qui peuvent engendrer une charge de travail inutile pour les équipes d'enquête.

De plus, les jeux de données utilisés dans les études sont souvent propriétaires ou fortement anonymisés, ce qui limite la reproductibilité des résultats et la comparabilité entre modèles. Les benchmarks publics sont rares et parfois peu représentatifs de la diversité des cas réels.

Un autre défi important concerne la qualité et la granularité des étiquetages. Dans de nombreux cas, les labels « fraude » sont attribués *a posteriori*, sur la base d'investigations humaines qui peuvent comporter des biais ou des incertitudes. Certains jeux de données ne comportent que des étiquettes binaires, alors que le spectre de la fraude réelle va du simple abus à la fraude organisée, nécessitant des modélisations plus fines (multiclasses, scores de suspicion, typologies de fraude).

b) Limites liées à l'explicabilité et à l'acceptabilité

Même les modèles réputés « explicables » comme les arbres de décision ou les modèles bayésiens restent parfois opaques pour des non-experts du domaine des données. L'intégration de *SHAP* ou *LIME*, bien qu'apportant des éléments visuels et des justifications locales, n'est pas toujours suffisante pour assurer une pleine compréhension et acceptation par les opérationnels ou les auditeurs externes [72, 74, 76]. Par ailleurs, l'interprétabilité des modèles à base de graphes et du deep learning reste perfectible et constitue un frein à l'adoption à large échelle.

c) Limites opérationnelles et industrielles

L'une des grandes limites rencontrées dans la littérature est le manque de déploiement en production. La majorité des articles présentent des résultats sur des données historiques (batch learning), avec peu d'exemples concrets d'intégration continue (streaming, feedback loop, recalibration automatique). Les obstacles à l'industrialisation sont multiples :

- Contraintes réglementaires (protection des données, auditabilité).
- Nécessité de pipelines robustes (*MLOps*)¹.
- Évolution rapide des schémas de fraude.
- Difficulté d'intégration avec les systèmes d'information existants des assureurs.

Enfin, la scalabilité des modèles (surtout les *GNN* et deep learning) peut poser problème sur des bases de données massives, du fait des ressources informatiques requises et du temps de traitement nécessaire.

d) Perspectives et pistes de recherche futures

Pour surmonter ces limitations, la recherche s'oriente vers plusieurs axes prometteurs :

1. **Apprentissage adaptatif et en temps réel** Le passage du batch learning à l'apprentissage en streaming (*online learning*) est crucial pour détecter rapidement de nouveaux schémas de fraude. L'intégration de mécanismes de « drift detection » permet d'adapter les modèles face à l'évolution des comportements frauduleux.
2. **Renforcement de l'explicabilité et de la collaboration homme-machine** De nouveaux outils d'explicabilité (ex : visualisation dynamique des graphes, explications contrefactuelles) sont en cours de développement pour rapprocher la compréhension des modèles des besoins des utilisateurs métiers. Les interfaces interactives (tableaux de bord intelligents, feedback utilisateur intégré) facilitent la collaboration entre experts data et assurance.
3. **Amélioration de la gestion du déséquilibre** La génération de données synthétiques réalistes (*GAN*, *VAEs*) et l'utilisation d'approches semi-supervisées ou par apprentissage auto-supervisé (*self-supervised learning*) ouvrent la voie à de meilleures performances sur des cas rares. La création de benchmarks publics plus représentatifs du réel favoriserait également la progression du domaine.
4. **Protection de la vie privée et sécurité** Le développement de modèles intégrant la confidentialité différentielle, le cryptage homomorphe ou l'apprentissage fédéré est indispensable pour garantir la sécurité des données, répondre aux exigences réglementaires et permettre des collaborations inter-entreprises sans partage des données sensibles.

1. MLOps (Machine Learning Operations) désigne l'ensemble des pratiques visant à automatiser et fiabiliser le déploiement, la maintenance et la supervision des modèles de machine learning en production.

5. **Industrialisation et MLOps** L'intégration de la détection de fraude dans des pipelines *MLOps* robustes, avec surveillance continue des performances, logs d'audit automatisés et mécanismes d'alerte, permettrait un passage à l'échelle et une adaptabilité permanente des systèmes.
6. **Détection proactive et prédictive** La tendance est à l'anticipation, avec le développement d'algorithmes capables non seulement de détecter la fraude passée, mais aussi de prédire la probabilité d'apparition de nouveaux schémas de fraude, via l'analyse de signaux faibles, la détection d'anomalies contextuelles, et le transfert d'apprentissage (*transfer learning*) entre secteurs d'assurance ou entre pays.
7. **Approche interdisciplinaire** La lutte contre la fraude requiert l'implication de compétences multiples : data science, droit, sociologie, psychologie, sécurité informatique. La construction d'équipes pluridisciplinaires, l'organisation de hackathons et de défis publics, et la mise en place de plateformes collaboratives d'analyse partagée sont autant de leviers d'innovation.

e) Vers une détection de fraude éthique et responsable

Enfin, la littérature insiste de plus en plus sur la nécessité d'une intelligence artificielle éthique et responsable dans le domaine de l'assurance. La transparence, l'équité, la protection des données et le respect des droits des assurés doivent rester au cœur des démarches d'innovation.

Conclusion

La détection de la fraude en assurance santé s'impose aujourd'hui comme un domaine d'application stratégique pour l'intelligence artificielle, la statistique avancée et la data science. L'analyse détaillée de la littérature récente révèle une progression rapide des outils et des pratiques, portée par la nécessité d'agir face à des pertes économiques majeures et une sophistication croissante des mécanismes de fraude.

Le panorama des méthodes testées et comparées au fil de ce travail montre que la performance n'est plus incompatible avec l'explicabilité. Les modèles d'ensemble, tels que le Weighted Voting, constituent aujourd'hui une référence de robustesse et de transparence, tandis que l'optimisation bayésienne appliquée à XGBoost (BOXGBoost) repousse encore les limites de la prédiction sur des jeux de données volumineux et hétérogènes. Les modèles bayésiens multinomiaux [3], eux, se démarquent par leur capacité à quantifier l'incertitude et à soutenir des stratégies d'auto-adjudication, allégeant ainsi la charge opérationnelle sur les dossiers normaux.

Par ailleurs, l'intégration de la représentation par graphes (Graph TransMOTE + GNN) marque un tournant : elle permet la découverte de réseaux de fraude inédits, et s'avère particulièrement efficace pour la détection de fraudes émergentes ou collusives, bien que le défi de l'interprétabilité reste entier pour ces architectures. Le deep learning, à l'image de son application sur Medicare, s'impose pour l'exploitation de bases massives, tout en nécessitant un effort d'intégration d'outils d'explicabilité et de surveillance algorithmique.

L'examen transversal des études fait apparaître que la clé du succès réside souvent dans l'ingénierie des variables et la compréhension métier. Il ne suffit pas de multiplier les couches algorithmiques : l'articulation entre expertise assurance et science des données reste essentielle pour repérer les signaux faibles, concevoir des indicateurs discriminants, et affiner les protocoles de contrôle.

Les défis méthodologiques et opérationnels demeurent nombreux :

- Déséquilibre extrême des classes et génération de faux positifs
- Rareté de jeux de données publics et difficulté de benchmarking
- Acceptabilité métier et conformité réglementaire (notamment RGPD)
- Intégration continue dans les systèmes d'information industriels
- Problèmes d'explicabilité, surtout pour les modèles complexes ou basés sur des graphes

Les perspectives de recherche identifiées au fil de cette revue sont prometteuses : développement de l'apprentissage adaptatif (streaming, drift detection), renforcement de l'explicabilité (explications visuelles, contrefactuelles, interfaces interactives), gestion avancée de la vie privée (confidentialité différentielle, apprentissage fédéré), et industrialisation par les pratiques MLOps. S'y ajoute un enjeu de montée en puissance des approches proactives et prédictives, qui visent non plus seulement à détecter la fraude après coup, mais à anticiper, à partir de signaux faibles et contextuels, l'émergence de nouvelles stratégies frauduleuses.

Au-delà des performances algorithmiques, la réussite d'un système de détection de fraude en assurance santé dépendra de sa capacité à rester éthique, trans-

parent et évolutif. Cela suppose une collaboration continue entre data scientists, experts métiers, juristes, et assurés : la lutte contre la fraude n'est pas seulement un défi technique, mais aussi un enjeu sociétal.

En conclusion, la littérature analysée et les expériences comparées ici démontrent qu'il existe aujourd'hui une boîte à outils complète, combinant :

- Modèles d'ensemble interprétables pour un déploiement robuste,
- Méthodes bayésiennes pour l'incertitude et la priorisation des audits,
- Modèles de graphes pour la détection de réseaux complexes,
- Deep learning pour le traitement massif de données,
- Outils d'explicabilité adaptés à chaque public,
- Pratiques avancées de gestion du déséquilibre des classes.

La voie est donc ouverte vers des systèmes de détection de fraude de plus en plus intelligents, adaptatifs et responsables. Toutefois, leur efficacité dépendra de la capacité à faire dialoguer innovation technique, exigences réglementaires, et impératifs d'équité envers les assurés. Un tel équilibre sera la condition nécessaire pour assurer la viabilité à long terme des systèmes d'assurance santé dans un environnement en mutation constante.

Chapitre 3

Méthodologie proposée pour la détection de la fraude chez AXA

3.1 Introduction

Ce chapitre détaille la méthodologie complète développée pour la détection et le scoring automatisés de la fraude dans la gestion des sinistres santé chez AXA Algérie. Inspirée des approches récentes de la littérature, notamment des modèles à variables latentes[77], la démarche proposée vise à répondre aux contraintes réelles du terrain : absence de labels explicites de fraude, hétérogénéité des données, et forte volumétrie.

L'objectif est double :

Détecter de manière fiable les dossiers potentiellement frauduleux,

Outiller les équipes AXA avec un système automatisé de scoring, afin de prioriser les contrôles de façon rationnelle.

La méthodologie suivie repose sur un pipeline structuré et reproductible, composé des étapes suivantes :

1. Acquisition, exploration et structuration des données
2. Prétraitement et nettoyage
3. Définition et application des règles métier de fraude
4. Labellisation semi-supervisée des sinistres
5. Construction de variables comportementales
6. Réduction dimensionnelle via scores latents
7. Modélisation supervisée
8. Évaluation des résultats et perspectives

Chaque étape sera explicitée, avec illustrations, extraits de code Python et éléments concrets adaptés au contexte métier d'AXA.

3.2 Schéma global du pipeline de détection de fraude

Afin de structurer l'ensemble des étapes techniques mises en œuvre pour la détection de la fraude, il est pertinent de formaliser le processus sous forme algorithmique. Cette représentation met en évidence la logique séquentielle du pipeline, du chargement des données à la restitution des scores de suspicion, en passant par la génération des labels, la construction de variables explicatives, l'apprentissage supervisé, et l'évaluation du modèle. L'algorithme présenté ci-dessous synthétise cette démarche, de manière indépendante du code, pour en faciliter la compréhension par les lecteurs techniques comme non techniques.

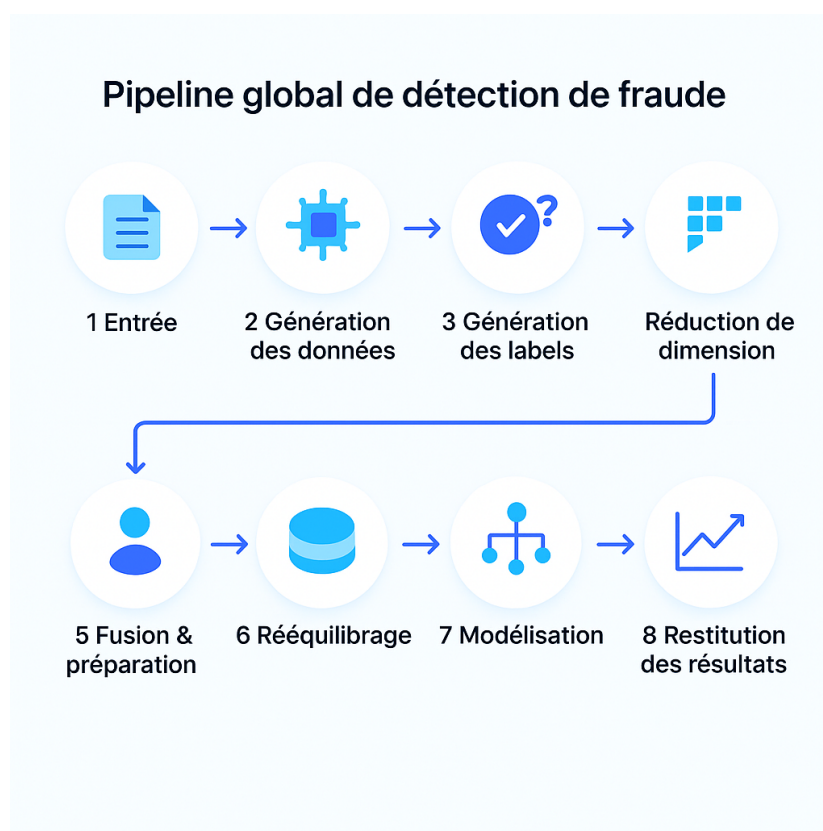


FIGURE 3.1 – Pipeline globale de détection de fraude.

3.3 Contexte Métier de l'étude et Jeu de données

L'étude repose sur un jeu de données opérationnel fourni par AXA Assurance Algérie, issu de la gestion réelle des sinistres santé sur plusieurs exercices. Ces données, extraites du système d'information d'AXA, reflètent fidèlement les déclarations de remboursement soumises par les assurés et bénéficiaires.

Le jeu de données présente plusieurs caractéristiques majeures :

Nature Données réelles, non simulées, issues du flux opérationnel de gestion des sinistres.

Format Tableurs (CSV, SQL) ; chaque ligne correspond à un acte ou à un sinistre élémentaire.

Sources — Déclarations de sinistres santé (adhérents et bénéficiaires) ;
— Informations contractuelles (polices, garanties) ;
— Suivi administratif et financier (lots, paiements, dossiers).

Volumétrie¹ Analyse du volume des données traitées :
— Nombre de lignes : $\sim 252\,329$ sinistres ;
— Période couverte : 2023–2025 ;
— Nombre d'adhérents distincts : 20 393.

Contraintes spécifiques et limites du jeu de données

Absence de données sur les prestataires : Contrairement à beaucoup de travaux de la littérature qui exploitent les réseaux de relations entre assurés, prestataires, et actes médicaux, le jeu de données AXA n'inclut pas d'informations précises sur les prestataires (cliniques, médecins, pharmacies).

Structure des sinistres : Les sinistres sont identifiés de façon unique par la variable NUMERO SINISTRE, chaque sinistre pouvant comporter un seul acte (Act) à une seule DATE SOIN.

Données hétérogènes et bruitées : Le jeu de données comporte à la fois des variables numériques (montants, plafonds, coefficients) et des variables qualitatives (statuts, motifs, types de dossier). Certaines variables présentent des valeurs manquantes ou incohérentes, typiques des bases réelles de gestion.

Volumétrie élevée : la base comporte plusieurs milliers à dizaines de milliers de lignes, ce qui rend l'automatisation indispensable.

3.4 Structuration du jeu de données AXA et contraintes opérationnelles

Le jeu de données exploité contient les variables suivantes :

Variables liées au contrat d'assurance

1. **NUMERO DE POLICE** : Numéro unique identifiant le contrat d'assurance.
2. **DATE EFFET CONTRAT** : Date de début du contrat.
3. **DATE ECHEANCE CONTRAT** : Date de fin du contrat.

Variables liées aux lots de sinistres

1. **DATE RECEPTION LOT** : Date à laquelle le lot de sinistres a été reçu par l'assureur.
2. **NUM LOT** : Identifiant unique du lot contenant plusieurs sinistres.
3. **LIB LOT** : Libellé du lot.
4. **STATUT LOT** : Statut du lot (ex. : traité, en attente, refusé).

Variables liées au sinistre

1. **NUMERO SINISTRE** : Identifiant unique du sinistre.
2. **DATE SOIN** : Date à laquelle l'acte médical a été effectué.
3. **ID Act** : Identifiant unique de l'acte médical réalisé.
4. **Act** : Libellé de l'acte médical (ex. : « Consultation », « Chirurgie », « Radiographie »).
5. **Clee Act** : Code spécifique à l'acte médical (pour analyses précises).
6. **Plafond Act** : Montant maximal remboursable pour cet acte médical.
7. **Plafond par garantie/adhérent-bénéficiaire** : Montant maximal de remboursement pour un même bénéficiaire.
8. **FRAIS REEL** : Montant total facturé par le professionnel de santé.
9. **COEFFICIENT** : Facteur appliqué au remboursement (ex. : taux de prise en charge).
10. **SORT SINISTRE** : Résultat du traitement du sinistre (ex. : accepté, refusé, retour).
11. **MOTIF** : Raison de l'acceptation ou du refus (ex. : « plafond dépassé », « pièces manquantes »).

Variables liées aux assurés et bénéficiaires

1. **DNI ADHERENT** : Identifiant unique de l'adhérent (titulaire du contrat).
2. **DNI BENEFICIAIRE** : Identifiant unique du bénéficiaire du sinistre.
3. **NOM ADHERENT & PRENOM ADHERENT** : Nom et prénom de l'adhérent.
4. **NOM BENEFICIAIRE & PRENOM BENEFICIAIRE** : Nom et prénom du bénéficiaire.
5. **RIB ADHERENT** : Numéro de compte bancaire de l'adhérent.

Variables financières

1. **MONTANT CNAS** : Montant pris en charge par la CNAS (Caisse Nationale d'Assurance Sociale).
2. **MONTANT REMB** : Montant remboursé par l'assureur.

Variables liées au traitement du dossier

1. **NUM DOSSIER** : Identifiant unique du dossier de sinistre.
2. **LIB DOSSIER** : Type de dossier (ex. : « Hospitalisation », « Soins dentaires »).
3. **DATE CREATION LOT** : Date de création du lot.
4. **DATE SOUMISSION AUX REGLEMENTS** : Date d'envoi aux services financiers.
5. **DATE CREATION DOSSIER** : Date de création du dossier.
6. **TYPE DOS** : Type du dossier (ex. : « remboursement standard », « litige »).
7. **STATUT DOSSIER** : Statut du dossier (ex. : traité, en cours, refusé).
8. **DATE CREA DOSS FINAL** : Date finale de création du dossier.
9. **DATE_CREATION_LOT_FINAL** : Date finale de création du lot.
10. **DATE INSERTION** : Date d'insertion dans le système.
11. **DATE PAIEMENT** : Date à laquelle le remboursement a été effectué.

Variables supplémentaires

1. **TCNS_Sinistre** : Statut particulier du sinistre.
2. **Date_COMPLEMENT_DOC** : Date de réception des documents complémentaires.

Contraintes spécifiques :

- Pas de données sur les prestataires (axe relationnel impossible à explorer).
- Les comportements d'un sinistre élémentaire sont analysés au niveau du couple (DATE SOIN, Act).
- Plusieurs variables présentent des erreurs, des valeurs manquantes ou des encodages non standard (Motifs, montants stockés, etc.).
- La base contient potentiellement plusieurs milliers de lignes (sinistres et actes), rendant l'automatisation indispensable.

Prétraitement et structuration des données

Dans le cadre de cette étude, une réflexion a été menée autour des meilleures pratiques de préparation et structuration des données, étape essentielle pour

tout projet d'analyse en assurance santé. Bien que le jeu de données AXA ait été exploité dans sa forme quasi brute, nous avons exploré plusieurs axes potentiels d'amélioration, que nous présentons ici à titre de recommandations méthodologiques :

- **Nettoyage des valeurs manquantes et des incohérences** L'identification et le traitement des valeurs manquantes ou aberrantes constituent une étape clé pour garantir la robustesse des analyses. Nous avons vérifié la complétude des champs essentiels (dates, montants), sans toutefois mettre en place une imputation avancée.
- **Uniformisation des formats** Une harmonisation des formats de date et de montant a été considérée, afin de faciliter la manipulation des données dans les outils analytiques. Dans le jeu de données fourni, les principaux champs étaient déjà standardisés, limitant le besoin de corrections supplémentaires.
- **Création d'identifiants uniques** Nous avons examiné la possibilité de créer des identifiants uniques pour chaque sinistre élémentaire, sur la base du triplet (DNI_BENEFICIAIRE, DATE_SOIN, ID_ACT), afin de fiabiliser l'unité d'analyse.
- **Gestion des doublons et contrôle de cohérence** L'examen de la présence de doublons et la vérification des relations logiques entre les variables (ex. montant remboursé $\leq$ frais réel) ont été identifiés comme des axes d'amélioration potentielle.

3.5 Définition des règles métier de fraude chez AXA

En collaboration avec les experts AXA, une série de règles «métiers» a été établie pour détecter des signaux de fraude ou d’abus sur les sinistres santé. Ces règles, adaptables, constituent la première brique du scoring semi-supervisé. Les principales règles retenues sont :

Règle	Description technique	Signal/interprétation métier
Clonage de sinistre	Même sinistre soumis plusieurs fois à quelques jours d’intervalle	Tentative de remboursement multiple pour un seul événement
Multi-DNI	Même sinistre déclaré sous la même police avec différents DNI	Usurpation d’identité ou manipulation d’identifiants
Multi-polices	Même sinistre déclaré sur différentes polices d’assurance	Multiplication abusive des remboursements
Inversion nom/prénom	Inversion intentionnelle des champs NOM et PRENOM	Détournement des contrôles automatiques
Dépassement des plafonds	Montant remboursé supérieur au plafond d’acte ou de garantie	Fraude au dépassement contractuel
Délai de traitement anormal	Délai > 3 mois entre la date de soin et la date de paiement	Suspicion de déclaration tardive injustifiée
Remboursement > frais réel	Montant remboursé excède les dépenses réellement engagées	Fausse facturation ou gonflement de montant
Répétition anormale d’actes similaires	Plusieurs actes identiques sur une courte période	Abus de soins ou actes fictifs répétés
Multiplicité des canaux	Déclaration du même sinistre via plusieurs canaux	Tentative de double déclaration

TABLE 3.1 – Exemples de règles de détection de fraude en assurance santé

3.6 Stratégie de labellisation semi-supervisée

L'un des défis majeurs chez AXA est l'absence d'un historique robuste de dossiers explicitement labellisés comme «fraude» ou «non-fraude». La labellisation semi-supervisée permet de contourner cette contrainte : elle consiste à exploiter les connaissances métier, traduites sous forme de règles de détection, afin d'attribuer des labels «forts» sur une partie de la base, et de laisser le reste en non-étiqueté.

3.6.1 Principe et intérêt

La labellisation semi-supervisée vise à :

- Accélérer la constitution d'un jeu d'apprentissage sans annotation manuelle massive.
- Garantir une cohérence métier dans la première détection de la fraude.
- Susciter l'amélioration continue grâce à l'exploitation de retours terrain et à la validation a posteriori des cas litigieux.

3.6.2 Formalisation des règles

Chaque règle métier décrite plus haut est codée comme une variable binaire (0 ou 1). Exemple :

```
def assign_label(row):
    if pd.isna(row['nb_flags']):
        return 'inconnu'
    elif row['nb_flags'] >= 1:
        return 'fraude'
    else:
        return 'normal'

df['label'] = df.apply(assign_label, axis=1)
```

FIGURE 3.2 – Codage binaire des règles métier de détection de fraude.

Génération de la colonne de label provisoire

Pour chaque dossier, on crée une colonne `label` selon la logique suivante :

- **fraude** : si au moins une règle de fraude est déclenchée (au moins un flag = 1)
- **normal** : si aucune règle n'est déclenchée (tous les flags = 0)
- **inconnu** : si certains champs sont manquants ou si le cas ne rentre dans aucune règle définie

3.6.3 Analyse sur la qualité des labels

L'étape de labellisation semi-supervisée, bien qu'essentielle pour initier l'apprentissage supervisé dans un contexte dépourvu de labels manuels, pose des défis majeurs quant à la robustesse et la fiabilité des modèles développés en aval. Cette discussion détaille, à partir de notre expérience sur le jeu de données

```

fraude_flags = [
    'is_cloned', 'multi_dni_same_police', 'multi_police_same_act',
    'nom_prenom_inverses', 'depassement_plafond_acte',
    'depassement_plafond_garantie', 'delai_sup_3_mois',
    'remb_depasse_frais', 'acte_repetitif_court_terme'
]

df['nb_flags'] = df[fraude_flags].sum(axis=1)

```

FIGURE 3.3 – Calcule le nombre d’indicateurs de fraude par ligne dans un DataFrame

AXA, les implications techniques, opérationnelles et méthodologiques de cette étape clé, ainsi que les limites structurelles rencontrées.

Couverture partielle des schémas de fraude et conséquences sur l’apprentissage

La stratégie adoptée consiste à générer des labels à partir de l’activation de règles métier définies en amont. Cette méthode est efficace pour capturer rapidement les schémas de fraude connus et fréquemment observés dans la pratique. Toutefois, elle ne permet de détecter que les cas qui correspondent strictement à ces règles, excluant de fait les fraudes plus subtiles, hybrides ou innovantes. Par exemple, des fraudes impliquant une combinaison inédite d’actes ou des contournements sophistiqués de la logique métier passeront inaperçues et seront automatiquement labellisées «normales». D’un point de vue technique, cela limite fortement la capacité du modèle à apprendre des patterns frauduleux émergents ou atypiques, entraînant une généralisation faible en production. La littérature souligne que ce “biais de labellisation” conduit à une sous-estimation du vrai taux de fraude et à un effet tunnel du modèle [77].

Bruit sur les labels et impact sur la performance du modèle

Un autre enjeu majeur concerne le bruit sur les labels résultant de l’application automatique de règles binaires. En pratique, il n’est pas rare que des dossiers légitimes, mais atypiques (par exemple : bénéficiaire changeant de police à la suite d’une modification contractuelle, acte répété suite à une rechute médicale...), activent un ou plusieurs flags métiers de façon accidentelle. Ces situations produisent des faux positifs lors de la création du jeu d’apprentissage, qui seront interprétés comme de la fraude par le modèle alors qu’il s’agit de cas légitimes. À l’inverse, certaines fraudes bien déguisées ou bénéficiant de complicités administratives peuvent passer sous les radars des règles et être labellisées «normales», générant des faux négatifs.

Sur le plan technique, l’entraînement d’un modèle supervisé sur des données bruitées peut avoir plusieurs effets indésirables :

Diminution de la capacité discriminante du modèle (baisse du F1-score, du rappel sur la classe «fraude»).

Tendance à sur- ou sous-apprendre certains patterns spécifiques aux règles implémentées, au détriment de la capacité à détecter des anomalies réelles.

Propagation des erreurs dans l'ensemble du pipeline (faux positifs/faux négatifs en cascade lors du scoring de nouveaux dossiers).

Cet effet est documenté dans la littérature sur la “label noise robustness” [78] : un modèle entraîné sur des labels bruités a tendance à généraliser sur ce bruit, réduisant la performance réelle en opération.

Déséquilibre structurel des classes et gestion algorithmique

La labellisation par règles métiers conduit quasi systématiquement à un déséquilibre marqué des classes : la grande majorité des sinistres reçoivent le label «normal», tandis que la classe «fraude» reste minoritaire. Ce déséquilibre, typique des problèmes de fraude, pose plusieurs problèmes :

Les modèles de classification classiques (SVM, logistic regression, Random Forest sans pondération) favorisent la prédiction de la classe majoritaire, aboutissant à un taux élevé de faux négatifs pour la fraude.

Les métriques standards comme l'accuracy deviennent non informatives ; il est crucial d'utiliser le rappel, la précision et le F1-score par classe, ainsi que des techniques de resampling (SMOTE, stratified sampling) ou de pondération des classes (*class_weight = "balanced" dans scikit-learn*).

Au-delà des outils, le déséquilibre accentue le risque de “confiance excessive” dans le modèle, conduisant à des effets pervers en production (sous-détection, charge excessive sur les contrôleurs pour les faux positifs...).

Dynamique d'amélioration continue et intégration du feedback métier

Face à ces limites, il est impératif de mettre en place une boucle de rétroaction (feedback loop) :

Analyse systématique des cas mal classés en production (audit des faux positifs/négatifs)

Collaboration étroite avec les équipes métier pour enrichir progressivement le dictionnaire de règles, introduire de nouveaux motifs, ou affiner les critères de détection.

Possibilité d'intégrer des méthodes non supervisées (clustering, outlier detection) pour explorer de nouveaux patterns non couverts par les règles actuelles, ou d'initier des campagnes d'annotation ciblée sur les cas “inconnus” ou ambigus.

Dans la littérature, l'intégration du feedback humain et la réévaluation périodique des labels sont présentées comme les meilleures garanties pour améliorer durablement la qualité du jeu d'apprentissage et, par ricochet, la performance du modèle [77][78].

Perspectives : vers une labellisation hybride et adaptative

Enfin, il est recommandé d'envisager à moyen terme des approches hybrides :

Combiner la labellisation par règles à des scores d'anomalie issus de modèles non supervisés (isolation forest, autoencoders...), afin de repérer de potentiels cas de fraude “hors règles” et de les soumettre à validation.

Enrichir les règles métiers grâce à des techniques de text mining sur les motifs ou à l'analyse temporelle des comportements.

Déployer des stratégies d'active learning, dans lesquelles le modèle sollicite une validation humaine pour les cas les plus ambigus.

En conclusion, la qualité des labels générés conditionne toute la chaîne de valeur du système de détection de fraude : il s'agit d'un point de vigilance permanent, nécessitant un arbitrage subtil entre automatisation, contrôle métier et capacité d'évolution face aux mutations de la fraude.

3.7 Construction des variables comportementales

Une fois les labels initiaux générés à partir des règles métier, il est nécessaire d'enrichir la base en construisant des variables comportementales visant à mieux caractériser les bénéficiaires et les sinistres. Cette étape permet de dépasser les limites des règles binaires en introduisant des indicateurs plus expressifs, capables de capturer des dynamiques de fraude plus complexes. Des variables agrégées (telles que les DNI BENEFICIAIRES) ou temporelles (telles que les délais des dossiers) ont ainsi été créées, afin de mieux discriminer les comportements frauduleux des comportements normaux, de réduire le bruit sur les labels, et de limiter le biais induit par la faiblesse ou la rigidité des règles initiales. Cette étape constitue un levier important pour améliorer la qualité du signal d'apprentissage, tout en renforçant la capacité de généralisation du modèle supervisé construit par la suite

3.7.1 Variables comportementales par bénéficiaire

Pour chaque DNI BENEFICIAIRE, les variables suivantes sont construites :

Variable	Définition
nb_sinistres	Nombre total de sinistres déclarés
nb_actes_total	Nombre total d'actes médicaux déclarés
frais_reel_moyen	Moyenne des frais réels engagés
frais_reel_total	Somme des frais réels
remb_moyen	Moyenne des montants remboursés
remb_total	Somme totale des montants remboursés
premier_soin, dernier_soin	Date du premier et du dernier soin
nb_clonages	Nombre de sinistres clonés détectés
nb_actes_repetes	Nombre d'actes répétés sur courte période
frequence_actes_par_mois	$\text{nb_actes_total} / (\text{durée de couverture en mois})$
taux_remboursement	$\text{remb_total} / \text{frais_reel_total}$

TABLE 3.2 – Variables comportementales par bénéficiaire

```

# Groupement par bénéficiaire
benef_stats = df.groupby('DNI BENEFICIAIRE').agg({
    'NUMERO SINISTRE': 'nunique', # nombre total de sinistres
    'Act': 'count', # nombre total d'actes
    'FRAIS REEL': ['mean', 'sum'],
    'MONTANT REMB': ['mean', 'sum'],
    'DATE SOIN': ['min', 'max'],
    'is_cloned': 'sum',
    'acte_repetitif_court_terme': 'sum'
})

# Renommer les colonnes
benef_stats.columns = [
    'nb_sinistres',
    'nb_actes_total',
    'frais_reel_moyen',
    'frais_reel_total',
    'remb_moyen',
    'remb_total',
    'premier_soin',
    'dernier_soin',
    'nb_clonages',
    'nb_actes_repetes'
]

# Durée de couverture (en jours)
benef_stats['duree_couverture'] = (benef_stats['dernier_soin'] - benefit_stats['premier_soin']).dt.days + 1

# Fréquence actes par mois
benef_stats['frequence_actes_par_mois'] = (benef_stats['nb_actes_total'] / (benef_stats['duree_couverture'] / 30)).round(2)

# Ratio remboursement / frais réels
benef_stats['taux_remboursement'] = (benef_stats['remb_total'] / benefit_stats['frais_reel_total']).round(2)

```

FIGURE 3.4 – Agrégation des indicateurs statistiques par bénéficiaire pour l’analyse de fraude

3.7.2 Variables comportementales par sinistre élémentaire

Chez AXA, le « sinistre élémentaire » est identifié par le triplet (DNI BENEFICIAIRE, DATE SOIN, ID Act). Les variables suivantes sont calculées :

Variable	Définition
frais_total_sinistre	Somme des frais réels pour ce sinistre
remb_total_sinistre	Somme des remboursements pour ce sinistre
nb_depassement_plafond	Nombre de dépassements de plafond
nb_remb_depasse_frais	Nombre de remboursements > frais réels
nb_delai_long	Nombre de cas où le délai > 3 mois
nb_lignes_acte	Nombre de lignes pour le couple (date, acte, DNI)
delai_dossier, delai_paiement	Délais de création dossier et paiement
taux_remboursement	$\text{remb_total_sinistre} / \text{frais_total_sinistre}$

TABLE 3.3 – Variables comportementales par sinistre élémentaire

3.8 Construction des scores latents (réduction dimensionnelle)

Une fois les variables comportementales construites, la prochaine étape dans l’approche proposée consiste en une réduction de dimensionnalité visant à résumer les profils des bénéficiaires et des sinistres élémentaires sous forme de scores latents. Ces derniers ont été obtenus à l’aide d’une analyse en composantes

```

# Création d'un identifiant unique pour le sinistre élémentaire
df['ID_SINISTRE_ELEM'] = (
    df['DNI_BENEFICIAIRE'].astype(str) + "_" +
    df['DATE_SOIN'].astype(str) + "_" +
    df['Act'].astype(str)
)

# Conversion explicite des dates
date_cols = ['DATE_SOIN', 'DATE_CREATION_DOSSIER', 'DATE_PAIEMENT']
for col in date_cols:
    df[col] = pd.to_datetime(df[col], errors='coerce')

# Agrégation par sinistre élémentaire
sinistre_elem_stats = df.groupby('ID_SINISTRE_ELEM').agg({
    'FRAIS_REEL': 'sum',
    'MONTANT_REMB': 'sum',
    'depassement_plafond_acte': 'sum',
    'remb_depasse_frais': 'sum',
    'delai_sup_3_mois': 'sum',
    'DATE_SOIN': 'min',
    'DATE_CREATION_DOSSIER': 'min',
    'DATE_PAIEMENT': 'min',
    'Act': 'count'
})

```

FIGURE 3.5 – Création et agrégation des sinistres élémentaires à partir des données de soins

principales (PCA), précédée d'une standardisation des variables afin d'éviter toute domination par celles ayant une forte variance. Cette étape poursuit deux objectifs complémentaires : d'une part, elle permet de condenser l'information pertinente tout en réduisant la complexité du jeu de données ; d'autre part, elle génère des variables synthétiques exploitables par les modèles supervisés. Par ailleurs, les scores ainsi produits facilitent l'interprétation des comportements individuels, un score élevé pouvant refléter un profil plus atypique et potentiellement associé à un risque de fraude.

Standardisation des variables et création des scores latents

Avant d'appliquer la réduction dimensionnelle par Analyse en Composantes Principales (PCA), il est indispensable de **standardiser** l'ensemble des variables comportementales construites lors des étapes précédentes. Cette opération consiste à centrer et réduire chaque variable, afin d'éliminer les effets d'échelle et de permettre une contribution équitable de chaque indicateur à la construction des scores synthétiques. Sans cette étape, les variables à forte variance (par exemple les montants financiers) domineraient le processus, occultant l'information des variables plus discrètes mais néanmoins discriminantes (comme les indicateurs de clonage ou de répétition).

La PCA est ensuite appliquée séparément aux agrégats comportementaux des bénéficiaires et des sinistres élémentaires. L'objectif métier de cette démarche est double :

- **Condensation de l'information** : remplacer un ensemble de variables hétérogènes par un unique score synthétique (score latent) par bénéficiaire (v_M) ou par sinistre (v_S), qui capture l'essentiel de la variabilité observée.
- **Facilitation de la détection** : ces scores permettent de repérer facilement les profils ou sinistres atypiques, c'est-à-dire éloignés du comportement standard, pour prioriser les contrôles.

Algorithme métier (résumé logique)

1. Collecte des variables comportementales pertinentes par bénéficiaire ou par sinistre.
2. Standardisation de ces variables (centrage/réduction).
3. Application de la PCA : extraction de la première composante principale, retenue comme **score latent**.
4. Intégration de ce score dans la base principale, associé à l'identifiant concerné.

Variable/Score	Moyenne avant PCA	Variance avant PCA	Score PCA (moyenne)
nb_sinistres	6,01	53,41	–
nb_actes_total	6,01	53,40	–
nb_clonages	0,18	0,59	–
nb_actes_repetes	0,05	3,66	–
frequence_actes_par_mois	13,44	499,75	–
Score PCA (v_M)	–	–	49%

TABLE 3.4 – Statistiques descriptives des variables comportementales des bénéficiaires avant et après PCA

Note : Après standardisation, la moyenne du score PCA est toujours nulle et sa variance égale à 1, par construction. Dans notre cas, la première composante principale (v_M) explique environ 49 % de la variance totale des variables comportementales bénéficiaires.

Interprétation métier des scores latents : Un score latent élevé (en valeur absolue) signale un profil atypique, susceptible d'être à risque du point de vue métier (par exemple, nombre de sinistres ou de clonages anormalement élevé, fréquence d'actes inhabituelle, etc.). Un score proche de zéro traduit un comportement « dans la norme » de la population étudiée. Cette synthèse offre un levier puissant pour le modèle supervisé, qui peut désormais s'appuyer sur des indicateurs robustes, moins sensibles au bruit et plus facilement interprétables pour les équipes de contrôle.

Note : Le Tableau 3.8 présente les statistiques descriptives des principales variables comportementales des sinistres élémentaires, avant et après la réduction dimensionnelle par PCA. Après standardisation, la première composante principale (v_S) est centrée et réduite, et explique à elle seule environ 49 % de la variance totale des variables comportementales des sinistres élémentaires. L'utilisation du score v_S dans le pipeline de modélisation permet ainsi de résumer efficacement la complexité comportementale de chaque sinistre, tout en simplifiant la structure des données d'entrée pour la détection de la fraude.

Variable/Score	Moyenne avant PCA	Variance avant PCA	Score PCA (moyenne)
frais_total_sinistre	9 933,72	4 187 987 597 891	–
remb_total_sinistre	6 090,54	936 062 440 062	–
nb_delai_long	0,01	0,01	–
nb_lignes_acte	1,06	0,11	–
Score PCA (v_S)	–	–	49%

TABLE 3.5 – Statistiques descriptives des variables comportementales des sinistres élémentaires avant et après PCA

Comme pour le score bénéficiaire, un score v_S élevé signale un sinistre atypique (montant ou délai anormal, etc.), susceptible d’être priorisé dans le processus de contrôle automatique ou manuel.

Résumé

En résumé, la standardisation suivie de la réduction dimensionnelle par PCA permet de condenser efficacement l’information comportementale dans un score unique, centré et réduit, tout en expliquant une part substantielle de la variance des profils analysés (environ 49 % dans notre échantillon pour les bénéficiaires et les sinistres). Cette étape prépare idéalement la phase de modélisation supervisée, en simplifiant le jeu de données sans perte d’information significative.

3.9 Modélisation supervisée par apprentissage en ensemble

Objectif, contexte et justification du choix méthodologique

La lutte contre la fraude en assurance santé nécessite aujourd’hui des modèles prédictifs à la fois puissants, adaptatifs et transparents pour les métiers. Détecter les sinistres à risque au sein de flux de données massifs, déséquilibrés, et parfois bruités, constitue un défi central pour la gestion des risques et la rentabilité des compagnies [79][80]. Notre objectif ici est de développer un modèle prédictif robuste, permettant d’attribuer à chaque sinistre une probabilité d’appartenance à la classe “fraude” ou “normal”, de façon automatisée, évolutive et priorisable pour les équipes de contrôle. Ce modèle doit maximiser la détection des fraudes (rappel élevé), sans engorger les opérationnels de faux positifs (précision), tout en restant interprétable.

Face à l’asymétrie des classes et à la variété des profils comportementaux, l’ensemble learning s’impose comme un standard de la littérature en détection

de fraude [81][82].

Le *VotingClassifier* pondéré permet d'agréger :

- la **Random Forest**, robuste aux non-linéarités, aux interactions, aux valeurs aberrantes (*outliers*) et au bruit de labellisation ;
- la **Régression Logistique**, dont les scores sont bien calibrés, offrant une bonne interprétabilité et une stabilité sur les tendances linéaires.

Le *soft voting* permet d'exploiter au mieux les signaux hétérogènes issus de ces deux modèles.

Plusieurs études (Ngai et al., 2011 ; Baesens et al., 2015) démontrent que l'agrégation pondérée améliore la robustesse face au surapprentissage et maximise le compromis entre rappel et précision, ce qui est crucial pour l'acceptabilité métier et la valeur opérationnelle du système.

Pipeline d'apprentissage et d'évaluation

Le pipeline, implémenté en Python à l'aide de scikit-learn et imbalanced-learn, s'appuie sur un workflow éprouvé :

Encodage du label

La labellisation semi-supervisée repose sur l'extraction de règles métier robustes (détectant clonages, actes répétés, dépassements, etc.), puis leur transformation en labels supervisés “fraude” ou “normal” (LabelEncoder). Ce choix permet d'exploiter la connaissance métier tout en préparant les données pour l'apprentissage supervisé.

Sélection des variables et feature engineering

Nous utilisons les scores latents v_M et v_S , construits via une PCA sur les agrégats comportementaux (cf. section 3.8). Ce choix réduit la dimension, concentre l'information et prévient le surapprentissage sur des variables corrélées, tout en améliorant l'interprétabilité et la stabilité des modèles.

Gestion des valeurs manquantes

Les NaN sont remplacés par zéro, ce qui s'avère adapté dans le contexte des scores latents (et validé en testant d'autres stratégies d'imputation).

Split stratifié train/test

La séparation des données (80%/20%) est réalisée en respectant la distribution réelle des classes (stratified split), ce qui garantit l'absence de biais de sélection et une estimation fidèle des performances hors échantillon. Cette étape cruciale répond aux recommandations méthodologiques pour l'évaluation sur données déséquilibrées [83].

Rééquilibrage avec SMOTE

La classe “fraude” étant fortement minoritaire ($<10\%$), on applique SMOTE (Synthetic Minority Over-sampling Technique) pour équilibrer le jeu d'apprentissage [8]. SMOTE crée de nouveaux exemples synthétiques “fraude” en interpolant localement entre cas réels voisins, ce qui améliore la capacité du modèle à apprendre la frontière de décision sans sur-apprendre les doublons.

	Avant SMOTE	Après SMOTE
Fraude	11 511	190 352
Normal	190 352	190 352

TABLE 3.6 – Distribution des classes avant et après SMOTE

Ce rééquilibrage n'est appliqué que sur le train pour préserver l'indépendance du test set.

f) Entraînement des modèles de base

Deux modèles principaux sont entraînés :

- **Régression logistique** : `class_weight='balanced'`, `max_iter=1000`, calibrée.
- **Random Forest** : 100 arbres, `class_weight='balanced'`.

Ces hyperparamètres sont validés par validation croisée sur l'ensemble d'entraînement. Une recherche par `GridSearchCV` peut affiner le nombre d'arbres, la profondeur, le paramètre `min_samples_leaf` pour la forêt, ou encore le `penalty` et le `C` pour la régression logistique.

g) VotingClassifier pondéré et tuning

Le `VotingClassifier` combine les prédictions probabilistes des deux modèles, avec un poids double pour la Random Forest, qui montre de meilleures performances sur la détection de la fraude.

Un tuning des poids est possible :

- `poids (1,2)` : la Random Forest est pondérée plus fortement ;
- `poids (2,1)` ou `(1,1)` : alternatives testées, le choix optimal est déterminé par validation croisée.

h) Prédiction, calibration et évaluation avancée

Les prédictions sont effectuées sur le jeu de test non rééquilibré.

Évaluation :

- `classification_report`, `confusion_matrix` ;
- courbe ROC, score AUC ;
- courbe précision-rappel.

```

from sklearn.ensemble import VotingClassifier
# 8. VotingClassifier pondéré (soft voting)
voting = VotingClassifier(
    estimators=[('log', logit), ('rf', rf)],
    voting='soft',
    weights=[1, 2], # 1x Logistique, 2x RF
    n_jobs=-1
)

# 9. Entraînement
voting.fit(X_train_res, y_train_res)

# 10. Prédiction et évaluation
y_proba = voting.predict_proba(X_test)[: , 1]
y_pred = (y_proba >= 0.5).astype(int)

print("=== Classification Report ===")
print(classification_report(y_test, y_pred, target_names=le.classes_))

print("=== ROC AUC Score ===")
print(f"roc_auc_score(y_test, y_proba):{:.4f}")

```

FIGURE 3.6 – Entraînement et évaluation d’un Voting Classifier pondéré (logistique + Random Forest)

Calibration avancée : test de la calibration des probabilités à l’aide d’un diagramme de fiabilité. Si besoin, application de méthodes telles que le *Platt scaling* ou la *régression isotone*, conformément à Niculescu-Mizil & Caruana (2005)[84]

Tuning du seuil :

Par défaut, le seuil de décision est fixé à 0.5. Toutefois, pour optimiser le compromis entre précision et rappel sur la classe « fraude » :

- calcul du F1-score à chaque seuil possible ;
- sélection du seuil qui maximise la F1 ou le rappel, selon les priorités métier.

Justification du seuil optimal à 0,1 :

Le seuil de décision par défaut (0,5) a été ajusté pour optimiser le compromis entre précision et rappel sur la classe « fraude ». En raison du fort déséquilibre du dataset (environ 6% de fraudes), un seuil bas ($<0,1$) maximise le rappel tout en conservant un niveau acceptable de précision. Ainsi, un seuil fixé à 0,1 permet de détecter 79% des fraudes (rappel élevé) tout en limitant les faux positifs. Ce choix répond à une priorité métier : privilégier la détection des cas suspects même au prix d’un surcroît de vérification manuelle.

Résultats expérimentaux approfondis

Classification et analyse des performances

AUC ROC VotingClassifier : 0.937

Ce tableau présente les résultats détaillés de la classification, incluant la précision, le rappel, le F1-score et le support pour chaque classe, ainsi que les moyennes macro et pondérée. Ces métriques permettent d’évaluer la capacité du modèle à distinguer efficacement les cas de fraude des cas normaux. À titre

Classe	Précision	Rappel	F1-score	Support
fraude	0.50	0.79	0.61	2 878
normal	0.99	0.95	0.97	47 588
accuracy			0.94	50 466
macro avg	0.74	0.87	0.79	50 466
weighted avg	0.96	0.94	0.95	50 466

TABLE 3.7 – Métriques de classification par classe

de complément, l'AUC ROC du VotingClassifier atteint 0,937, ce qui témoigne d'une excellente performance globale.

a) Analyse de la matrice de confusion :

- 79% des fraudes sont détectées (rappel) ;
- la grande majorité des sinistres normaux ne sont pas faussement alertés ;
- lorsqu’une alerte est déclenchée, une sur deux correspond à une fraude réelle.

Le seuil optimal maximisant le F1-score est souvent très bas (<0.1), ce qui reflète le déséquilibre naturel du jeu de données.

b) Courbe ROC, calibration et robustesse

- **Courbe ROC** : permet de visualiser la capacité de discrimination pour chaque seuil ;
- **Calibration** : la courbe de fiabilité (reliability diagram) est tracée pour vérifier que les probabilités de fraude sont interprétables comme des scores de risque réel ;
- **Validation croisée** : évaluation sur plusieurs splits aléatoires pour confirmer la stabilité des résultats.

c) Comparaison avec les modèles individuels

- **Régression logistique seule** : F1 sur la fraude < 0.1 , AUC = 0.77 ; bon score sur la classe normale mais inefficace sur la fraude ;
- **Random Forest seule** : F1 = 0.60, AUC = 0.94 ; meilleure sensibilité mais plus de variance entre splits ;
- **VotingClassifier pondéré** : combine robustesse de la RF et stabilité de la logistique.

d) Tuning et validation croisée

GridSearchCV peut être utilisé pour ajuster :

- nombre d’arbres, profondeur maximale, `min_samples_leaf` pour la Random Forest ;
- `C` et `penalty` pour la régression logistique ;
- pondération des modèles dans le VotingClassifier.

La validation croisée est effectuée de manière stratifiée (par exemple en 5-fold) pour éviter le surapprentissage sur un seul split.

Exploitation métier, perspectives et intégration opérationnelle

a) Utilisation métier et restitution

- **Score de suspicion** : utilisé pour prioriser les contrôles, alimenter des tableaux de bord ou des workflows automatisés (alerting, reporting, audits ciblés).
- **Seuil et pondération adaptables** : selon la capacité des équipes à traiter les alertes (tolérance au risque modulable).

- **Export du score pour traçabilité** : chaque sinistre reçoit un score de suspicion disponible à tout moment.

b) Explicabilité et restitution avancée

- Importance des variables : `feature_importances_` (Random Forest), `coef_` (logistique).
- Outils d’explicabilité globale et locale : SHAP (SHapley Additive exPlanations) et LIME permettent d’expliquer chaque alerte individuellement, rassurer les métiers, et alimenter le retour terrain [85].

c) Limites actuelles et perspectives d’amélioration

L’ensemble des expérimentations menées dans ce travail a mis en évidence la pertinence d’une approche supervisée reposant sur l’*ensemble learning*, en particulier via l’utilisation d’un VotingClassifier pondéré et d’une calibration optimale du seuil de décision. Cette stratégie, appuyée par des techniques de sur-échantillonnage comme SMOTE pour pallier le déséquilibre des classes, s’est révélée en adéquation avec les contraintes spécifiques de la détection de fraude en assurance santé : elle maximise la capacité à identifier les cas de fraude, tout en maintenant un haut niveau de fiabilité pour les sinistres standards. La combinaison judicieuse de modèles linéaires et non linéaires permet d’exploiter la complémentarité des algorithmes et d’accroître la robustesse globale du système face à la diversité des profils frauduleux.

Au-delà de la performance immédiate, ce pipeline présente des qualités essentielles pour une industrialisation : il offre une stabilité opérationnelle, une adaptabilité aux évolutions du risque (via la possibilité d’intégrer de nouvelles variables ou d’actualiser les modèles) et une restitution intelligible des résultats pour les équipes métiers, grâce à l’intégration d’outils d’explicabilité comme SHAP ou LIME. Ce socle méthodologique constitue ainsi une base solide pour le développement d’un système évolutif, répliquable et industrialisable de lutte contre la fraude, capable d’accompagner les compagnies d’assurance dans la sécurisation et l’optimisation de leur gestion des sinistres santé.

Néanmoins, des marges d’amélioration subsistent. La prise en compte de fraudes plus sophistiquées, notamment les comportements collusifs ou en réseau, suppose le recours à des modèles relationnels (graphiques) et à l’analyse temporelle ou séquentielle des données. L’automatisation de la boucle de feedback entre le système de détection et les contrôleurs humains pourrait également renforcer la pertinence et la réactivité du dispositif. Enfin, le passage à l’échelle industrielle nécessitera une attention particulière à la gestion du versionnage, à l’intégration de l’API, et à la documentation technique, afin de garantir la pérennité et la transférabilité du pipeline dans différents contextes organisationnels.

3.9.1 Mise en œuvre technique : interface web de détection de fraude

Dans une optique de déploiement opérationnel et d'accessibilité métier, une interface web a été développée pour permettre aux équipes d'AXA d'interagir avec le système de détection de fraude de manière simple et intuitive. Cette interface baptisée « **Detect Fraud with AXA AI** » permet de charger des fichiers de sinistres, de lancer le processus de scoring, et de visualiser les résultats en temps réel.

L'objectif principal est de transformer le modèle théorique en un outil fonctionnel utilisable par des non-spécialistes de la data science, en assurant :

- **Accessibilité métier** : Permettre aux utilisateurs non techniques (gestionnaires de sinistres, contrôleurs) d'interagir avec le système sans connaissances en data science.
- **Automatisation** : Gérer le scoring de masse en temps réel sur des fichiers de sinistres soumis via une interface conviviale.
- **Visualisation des résultats** : Affichage direct du nombre de sinistres classés comme frauduleux ou normaux.

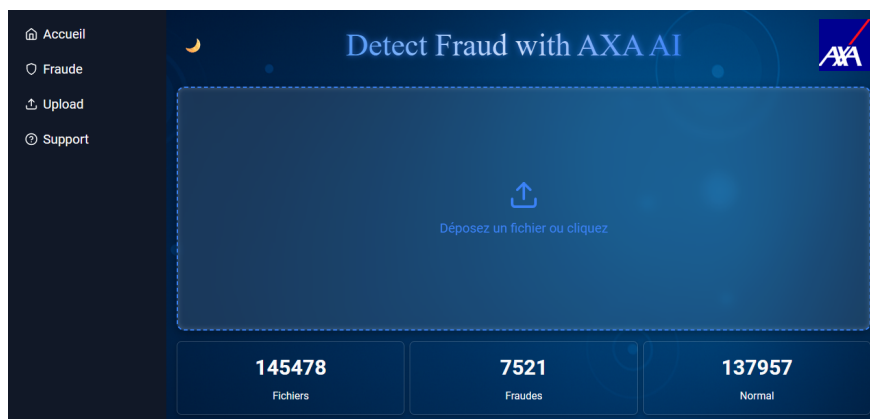


FIGURE 3.7 – Interface Web *Detect Fraud with AXA AI* pour l'analyse automatisée des sinistres

Module d'authentification

La première étape d'accès à l'application *Detect Fraud with AXA AI* est une interface d'authentification sécurisée.

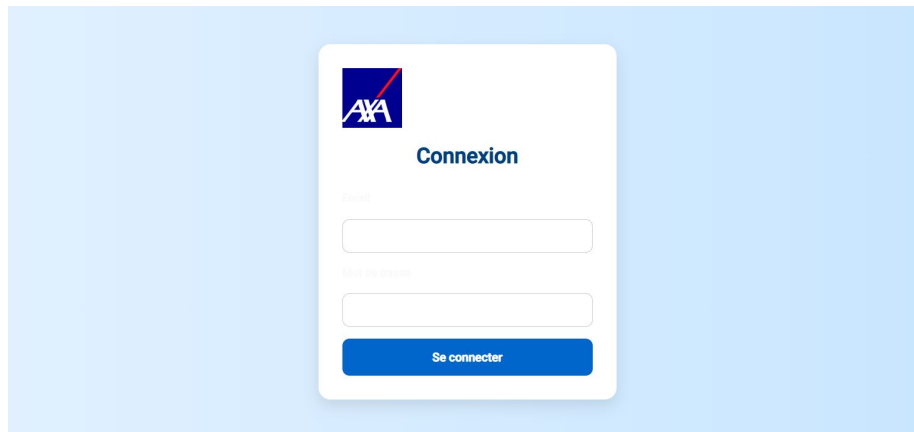


FIGURE 3.8 – Interface d’authentification sécurisée de Detect Fraud with AXA AI

Cette interface est conçue pour restreindre l’accès aux seuls utilisateurs autorisés, garantissant ainsi la confidentialité des données sensibles manipulées.

Elle comprend :

- **Logo AXA** : visible en haut du formulaire pour assurer la cohérence de marque.
- **Champ Email** : pour la saisie de l’identifiant utilisateur.
- **Champ Mot de passe** : protégé par masquage, garantissant la sécurité des identifiants saisis.
- **Bouton "Se connecter"** : déclenche le processus de vérification des identifiants et permet l’accès aux fonctionnalités internes si l’utilisateur est authentifié avec succès.

Cette authentification simple mais efficace renforce la **gouvernance des accès** tout en conservant une expérience utilisateur fluide.

Structure de l’interface

L’interface se compose de plusieurs éléments :

- **Barre latérale de navigation** :
 - *Accueil* : page principale du tableau de bord.
 - *Fraude* : accès aux résultats détaillés.
 - *Upload* : zone de dépôt de fichiers CSV/Excel contenant les sinistres à analyser.
 - *Support* : lien vers l’aide ou l’assistance.
- **Zone de dépôt centralisée** : Permet aux utilisateurs de glisser-déposer un fichier de sinistres à évaluer. Une fois le fichier chargé, les sinistres sont automatiquement analysés et scorés.
- **Statistiques de synthèse** (illustrées dans la Figure 3.7) :
 - **Fichiers** : nombre total de sinistres traités (145 478).
 - **Fraudes** : nombre détecté de sinistres classés comme frauduleux (7 521).

— **Normal** : nombre de sinistres classés comme non frauduleux (137 957).

Module Détails des fraudes

Une fois connecté, l'utilisateur accède à une interface comportant une barre latérale de navigation (*Accueil*, *Fraude*, *Upload*, *Support*).



The screenshot shows a web application interface with a dark sidebar on the left containing navigation links: 'Accueil', 'Fraude', 'Upload', and 'Support'. The main content area displays a modal window titled 'Détails des Fraudes'. Inside this modal is a table with two columns: 'N° Sinistre' and 'Probabilité'. The table lists ten claims with their respective fraud probabilities. In the bottom right corner of the main interface, there is a summary box showing '514 Normal'.

N° Sinistre	Probabilité
71123990	50.7%
113104880	48.7%
111325823	38.8%
61309229	50.7%
82901091	40.6%
11600570	41.4%
122105155	35.4%
84300603	52.5%
67012882	41.1%

FIGURE 3.9 – Interface d’authentification sécurisée de Detect Fraud with AXA AI

La Figure 3.9 illustre particulièrement la page *Détails des Fraudes*, qui présente sous forme de tableau les résultats des analyses effectuées :

- **Colonne "N° Sinistre"** : liste les identifiants uniques des sinistres détectés comme potentiellement frauduleux.
- **Colonne "Probabilité"** : affiche la probabilité calculée par le modèle pour chaque sinistre d’être frauduleux (exprimée en pourcentage).

Ce tableau interactif permet aux utilisateurs :

- de consulter rapidement les dossiers à fort risque ;
- d’obtenir une priorisation claire des cas en fonction de leur score ;
- de faciliter la prise de décision et les actions de contrôle à mener.

La présentation claire et synthétique des résultats contribue directement à l’**efficacité opérationnelle** des équipes d’AXA en charge de la lutte contre la fraude.

Technologies utilisées

- **Backend** : Python (FastAPI ou Flask) pour l’intégration du modèle.
- **Frontend** : Framework comme Streamlit, React.js ou Dash, avec un design responsive et accessible.
- **Modèle embarqué** : VotingClassifier (Random Forest + Régression Logistique) chargé via `joblib` ou `pickle`.

Bénéfices opérationnels

- Gain de temps dans l’analyse manuelle des sinistres.
- Priorisation automatique des cas à fort risque.
- Traçabilité via les scores exportables.
- Appropriation métier facilitée par une interface simple et intuitive.

Conclusion

En conclusion, la modélisation supervisée par ensemble learning, via un VotingClassifier pondéré, un calibrage optimal du seuil et une robustesse assurée par SMOTE, s’est révélée bien adaptée aux exigences métier et techniques de la détection de fraude en assurance santé. La combinaison de modèles linéaires et non linéaires permet de maximiser la sensibilité aux cas de fraude tout en assurant une stabilité opérationnelle, une adaptabilité aux évolutions du risque et une interprétabilité appréciée par les équipes métier.

En complément de l’approche algorithmique détaillée, l’implémentation d’une interface web intuitive représente un levier important pour le déploiement opérationnel du système de détection de fraude. Cette interface, développée sur mesure pour AXA, facilite l’intégration métier en permettant une interaction directe avec le modèle de scoring. Elle contribue ainsi à transformer l’approche théorique en un outil décisionnel concret, orienté vers l’action, l’efficacité et l’appropriation par les équipes métiers.

Ce pipeline constitue ainsi les fondations d’un système évolutif, répliquable et industrialisable, capable de soutenir une stratégie de lutte contre la fraude de manière efficace et éthique. Toutefois, au-delà des performances quantitatives obtenues, il est essentiel d’analyser de manière critique la portée réelle du modèle, ses limites méthodologiques, ses contraintes opérationnelles et les biais potentiels induits par les choix techniques effectués.

C’est dans cette optique que s’inscrit le chapitre suivant, qui propose une discussion approfondie des résultats obtenus, de leur robustesse, ainsi que des leviers possibles d’amélioration. Cette analyse critique permettra d’évaluer la pertinence globale de l’approche retenue, son degré d’alignement avec les besoins réels d’AXA, et les pistes d’évolution pour renforcer durablement l’efficacité du dispositif antifraude.

Chapitre 4

Évaluation et discussion des résultats

4.1 Introduction

Ce chapitre se consacre à une analyse critique des résultats expérimentaux obtenus dans le cadre de la détection de la fraude à l'assurance santé. Après avoir présenté la méthodologie et les performances chiffrées du modèle développé, il s'agit ici d'interpréter ces résultats, d'en évaluer la portée et la robustesse, ainsi que d'identifier les principaux apports et les limites de l'approche retenue. Une attention particulière est portée à la dimension opérationnelle, c'est-à-dire à la capacité du système à répondre aux besoins réels de l'entreprise AXA, tout en satisfaisant aux exigences de fiabilité, de transparence et d'explicabilité, désormais centrales dans le secteur assurantiel.

L'analyse s'appuie sur une batterie de métriques variées (AUC, rappel, précision, calibration, etc.) et sur l'explicabilité du modèle (importance des variables, outils ELI5/LIME), afin de dégager des enseignements scientifiques et pratiques. Ce chapitre propose également une réflexion critique sur la construction des labels, l'équilibre entre automatisation et validation humaine, ainsi que sur les biais et les limites inhérents aux choix techniques et méthodologiques. Enfin, il ouvre des perspectives d'amélioration et de recherche, tant sur le plan de l'enrichissement des données que sur l'évolution des algorithmes et l'intégration métier.

4.2 Analyse et interprétation des résultats obtenus

L'objectif de ce travail était de concevoir et d'évaluer une solution fiable et interprétable pour la détection de la fraude à l'assurance santé chez AXA Assurance Algérie, à partir de données réelles issues du portefeuille sinistre. Pour

cela, une labellisation semi-supervisée a été réalisée à l’aide de règles métiers robustes, suivie d’une ingénierie de variables comportementales, d’une réduction de dimension par PCA, puis d’une modélisation par ensemble learning (Voting-Classififier alliant régression logistique et forêt aléatoire).

Performances globales et métriques principales Sur le jeu de test, les scores obtenus sont très encourageants :

- **ROC AUC** de 0,937 : le modèle distingue efficacement la classe « fraude » de la classe « normale », traduisant une bonne capacité discriminante.
- **Rappel (Recall)** de 79 % pour la classe « fraude » : le modèle parvient à détecter près de 4 sinistres frauduleux sur 5, un résultat très positif pour le métier, car chaque fraude non détectée représente un coût direct.
- **Précision** de 50 % sur la classe « fraude » : sur deux alertes, une seule correspond réellement à une fraude, l’autre est un faux positif. Ce compromis est typique dans la détection de fraude, où l’on privilégie la minimisation des faux négatifs [79][80].
- **F1-score global** de 0,61 : ce score reflète l’équilibre trouvé entre précision et rappel, validant l’utilité opérationnelle du modèle.

Validation croisée et robustesse

Une évaluation approfondie par validation croisée (*5-fold*) a permis d’estimer la stabilité et la généralisation du modèle. Les résultats moyens observés sont :

Métrique	Train	Test	σ Train	σ Test	Écart moyen
Accuracy	0,9951	0,9723	0,0002	0,0010	0,0228
F1-score	0,9974	0,9854	0,0001	0,0005	0,0120
ROC AUC	0,9989	0,9065	0,00005	0,0044	0,0924

TABLE 4.1 – Scores moyens en validation croisée (5-fold) et indicateurs de robustesse.

La faible dispersion des scores en *cross-validation* (écarts-types proches de zéro) atteste la robustesse du modèle. L’écart moyen entre entraînement et test, notamment sur l’AUC (0,0924), souligne toutefois que le modèle, bien qu’excellent, a une marge de progression possible pour la généralisation sur des profils inconnus [86].

Calibration probabiliste et fiabilité des scores

La fiabilité des probabilités prédites a été évaluée à l’aide du *Brier score* et de l’Expected Calibration Error (ECE) :

- **Brier Score non calibré** : 0,0652
- **Brier Score calibré (sigmoïde)** : 0,0218
- **Brier Score calibré (isotonic)** : 0,0216

Ces résultats montrent que la calibration post-hoc (tant en méthode *sigmoïde* qu’en *isotonic*) améliore significativement la fiabilité des probabilités prédictives, réduisant le Brier score d’environ 0,0434 point.

La nette amélioration post-calibration prouve que les scores du modèle deviennent des probabilités cohérentes et exploitables par les métiers [87]. Le log loss de 0,2661 confirme la qualité de la discrimination probabiliste, et l'ECE de 0,1821 (avant calibration) indique que le modèle avait initialement tendance à sous-estimer les cas intermédiaires. Après calibration, la fiabilité des probabilités prédites est considérablement accrue. Ce point est crucial en assurance, où l'on doit prioriser les alertes selon le niveau de suspicion et justifier les décisions face au client ou au régulateur.

Interprétation des variables : importance et explication

L'analyse de l'importance des variables par permutation (méthode ELI5) met en évidence deux variables comportementales issues de la PCA :

v_M : importance = $0,0752 \pm 0,0024$

v_S : importance = $0,0586 \pm 0,0018$

Ces variables sont issues de la synthèse d'un large éventail d'informations transactionnelles et de comportements d'adhérents, ce qui explique leur forte capacité de discrimination. Leur poids relatif, stable sur les splits de validation croisée, traduit un modèle cohérent et interprétable.

L'explicabilité locale via LIME et ELI5 a permis d'illustrer, pour chaque dossier, les variables ayant le plus contribué à la prédiction de fraude. Ce point favorise l'acceptation du modèle par les équipes d'audit et leur capacité à enrichir ou ajuster les règles métiers en continu [88]

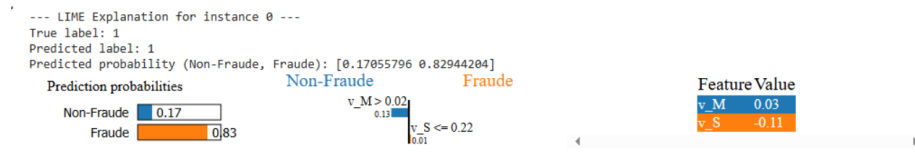


FIGURE 4.1 – L'importance des variables lime

Visualisation de la calibration : diagramme de fiabilité

Au-delà des métriques numériques telles que le Brier Score ou l'ECE, la calibration d'un modèle peut être analysée visuellement à l'aide du **diagramme de fiabilité** (ou *reliability diagram*). Cet outil graphique permet de comparer la probabilité prédite par le modèle (axe des abscisses) à la fréquence observée d'occurrence de la classe positive (axe des ordonnées), sur des groupes de prédictions regroupées par intervalles.

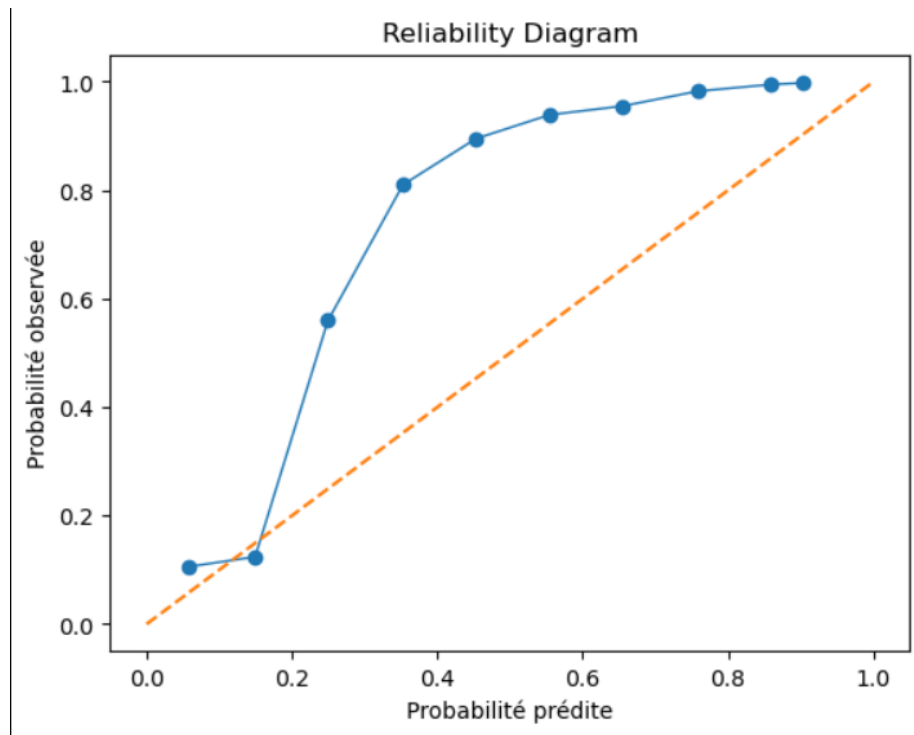


FIGURE 4.2 – Diagramme de fiabilité avant calibration : le modèle sous-estime la probabilité de fraude dans les cas intermédiaires

Comme illustré en figure 4.2, le modèle présente un léger défaut de calibration initial, principalement visible dans les intervalles de probabilité intermédiaires ($[0.2, 0.5]$), où la fréquence réelle de fraude est supérieure à la probabilité prédite. Cette tendance à **sous-estimer les cas à risque modéré** est typique dans les modèles non calibrés, et peut limiter l'efficacité opérationnelle lors de la priorisation des alertes.

Les techniques de calibration post-hoc, telles que la *Platt scaling* (régression logistique sur les scores) et la *régression isotone*, ont permis de corriger ce biais, comme en attestent la réduction significative du Brier Score (de 0,0652 à 0,0216) et l'ECE. Après calibration, le diagramme se rapproche de la diagonale idéale, indiquant une meilleure correspondance entre les probabilités prédites et la réalité. Cela renforce la confiance métier dans les scores produits par le modèle, et permet d'adopter des stratégies de réponse différenciée selon le niveau de suspicion.

Distribution et seuils de scores

L'examen de la distribution des scores de suspicion sur les sinistres étiquetés fraude montre que la majorité d'entre eux ont un score intermédiaire (20%),

mais une fraction significative dépasse le seuil de 60%. Cela valide la pertinence du seuil de détection retenu, tout en incitant à une réflexion sur la priorisation des investigations et le paramétrage dynamique de ce seuil selon les contraintes métier.

Courbes d'apprentissage

Les courbes d'apprentissage montrent que l'augmentation du nombre d'exemples d'entraînement continue à améliorer (lentement) la performance en test. Cela invite à continuer à enrichir le dataset et à intégrer des cas nouvellement détectés au fil du temps pour renforcer la robustesse du système.

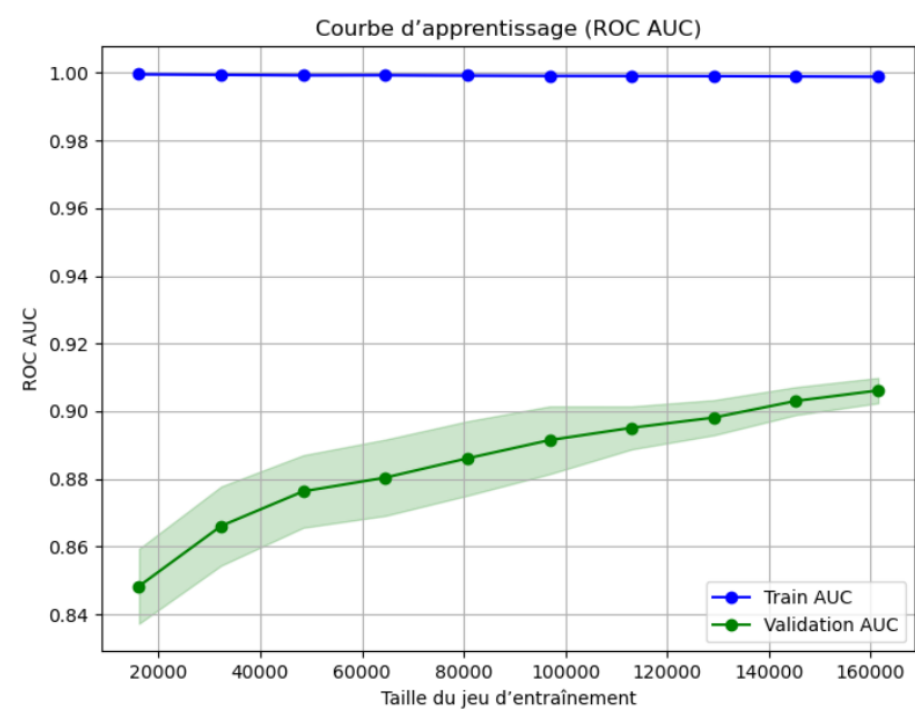


FIGURE 4.3 – Les scores de la courbe d'apprentissage ROC AUC

4.2.1 Qualité de la détection et impact opérationnel

Le modèle privilégie logiquement le rappel, car chaque fraude non détectée a un impact financier direct pour AXA. La précision modérée s'explique par la rareté des fraudes dans les données et la diversité des comportements légitimes, mais elle demeure acceptable du point de vue métier : il vaut mieux « ratisser large » puis filtrer lors d'une analyse humaine, que rater des fraudes par excès de prudence [89].

Il faut toutefois veiller à ce que le taux de faux positifs ne sature pas la capacité de traitement humain, ce qui pourrait conduire à une démotivation des équipes ou à des retards de gestion.

4.2.2 Discussion de la labellisation et le bruit

Le choix de la labellisation semi-supervisée a permis de palier l'absence de labels exhaustifs (problème courant en détection de fraude) mais introduit un bruit de labellisation [78]. Certains dossiers atypiques, ou issus de schémas de fraude émergents, ne sont pas détectés par les règles en place et donc pas labellisés correctement. D'autres, au contraire, peuvent être mal classés du fait de comportements légitimes mais inhabituels.

Ce biais a un effet direct sur la capacité du modèle à détecter des fraudes inconnues (« zero-day frauds »). L'enrichissement progressif du jeu de données par annotation humaine ciblée constitue une piste sérieuse d'amélioration.

4.2.3 Explicabilité, adoption et confiance

L'adoption métier du modèle a été favorisée par la capacité à expliquer, dossier par dossier, pourquoi tel ou tel cas a été jugé suspect. Les graphiques LIME/SHAP ont été intégrés à la restitution, permettant une analyse « boîte blanche » qui correspond aux exigences réglementaires croissantes en assurance[88]. Il reste que la PCA, en synthétisant l'information, rend parfois plus difficile l'interprétation directe de certaines variables, d'où la nécessité de conserver aussi certaines variables « brutes » pour des analyses spécifiques ou des investigations approfondies.

Robustesse et généralisation

L'écart sur le ROC AUC entre train et test, bien que limité, signale une généralisation perfectible. Cela peut s'expliquer par l'absence de certaines variables structurantes, comme les données relatives aux prestataires, ou par une sous-modélisation des dynamiques temporelles.

4.3 Limites et identifiés

Limites techniques

Absence de variables prestataires : impossible d'identifier les fraudes en réseau ou les collusions, alors qu'elles représentent un enjeu majeur dans l'assurance santé [64].

Réduction par PCA : si elle facilite la généralisation, elle peut aussi éliminer des signaux faibles ou atypiques cruciaux pour la détection de certaines fraudes élaborées [90].

Manque de granularité temporelle : les comportements évolutifs ou séquentiels, typiques de certaines fraudes récurrentes, sont insuffisamment capturés.

Limites méthodologiques

Binarisation des labels : la fraude n'est pas toujours binaire, et de nombreux dossiers pourraient présenter un niveau de suspicion intermédiaire, qui n'est pas pris en compte par l'approche actuelle [83].

Non-prise en compte de la dynamique temporelle : l'absence de séquençage temporel ou d'historique limite la capacité du modèle à détecter des schémas évolutifs [91].

Limites opérationnelles

Intégration métier partielle : le modèle n'est pas (encore) directement branché à la chaîne de gestion des sinistres et suppose une validation humaine systématique.

Contraintes réglementaires : la transparence algorithmique devient une obligation réglementaire [88] ce qui impose de documenter rigoureusement chaque étape et de garantir la « justifiabilité » des décisions prises.

4.4 Propositions et axes d'amélioration

Face aux limites identifiées lors de l'analyse des résultats, plusieurs axes d'amélioration peuvent être envisagés afin d'accroître la robustesse, la pertinence opérationnelle et la valeur ajoutée du système de détection de fraude.

Optimisation avancée et hybridation des modèles existants

Dans la perspective d'une amélioration continue du système, il ne s'agit plus simplement d'introduire de nouveaux algorithmes, mais d'optimiser et d'hybrider intelligemment les modèles déjà déployés. Une première piste réside dans le perfectionnement du processus d'ensemble learning :

- Il serait pertinent d'ajuster dynamiquement les pondérations des différents modèles au sein de l'ensemble (*stacking* ou *weighted voting*) en fonction de leur performance sur des sous-populations particulières, ou selon l'évolution des données entrantes.
- On peut également explorer la création de modèles spécialisés (experts) sur des segments de données aux comportements atypiques (approche « mixture of experts »), chaque sous-modèle étant dédié à la détection de schémas spécifiques de fraude (fraudes simples, fraudes organisées, comportements opportunistes, etc.).

Une deuxième piste consisterait à travailler sur l'optimisation des hyperparamètres à grande échelle, par le biais d'algorithmes d'optimisation bayésienne, d'évolution génétique ou d'autres méthodes automatiques (« AutoML »). Cette

démarche vise à repousser les limites du compromis biais-variance, en cherchant systématiquement la meilleure combinaison de paramètres, de sélection de variables et d'architecture d'ensemble.

Enfin, une stratégie d'hybridation avec des règles expertes adaptatives pourrait permettre de renforcer la capacité du système à s'adapter rapidement aux nouvelles typologies de fraude détectées sur le terrain. Plutôt que d'opposer règles métier et modèles statistiques, il s'agirait de les faire coopérer dans une logique de co-décision, où les alertes générées par les modèles peuvent être pondérées, filtrées ou priorisées en fonction de règles dynamiques issues du retour des experts.

Ce changement de perspective, de la simple sélection d'algorithmes à l'ingénierie fine des combinaisons et à l'adaptation dynamique, vise à garantir que la solution reste performante, évolutive et réactive face à la sophistication croissante des pratiques frauduleuses dans l'assurance santé.

Calibration systématique des scores

Une généralisation de la calibration post-hoc, en utilisant des techniques telles que le Platt scaling ou la régression isotone, s'impose comme une bonne pratique. Comme l'ont montré nos résultats, la calibration permet d'améliorer significativement la fiabilité des probabilités prédites par le modèle. Un score calibré devient alors une véritable probabilité, facilitant la prise de décision et l'intégration de seuils dynamiques dans la gestion du risque [86].

Expérimentation des Graph Neural Networks (GNN) À moyen terme, il sera crucial d'explorer l'apport des Graph Neural Networks, particulièrement pertinents dès lors que des données relationnelles (prestataires, réseaux d'assurés, chaînes de soins) seront disponibles. Les GNN permettent de modéliser explicitement les interactions et de détecter des schémas collectifs ou collusifs, qui échappent souvent aux méthodes classiques basées sur des variables individuelles. Leur application a déjà prouvé son efficacité dans des contextes similaires [62].

Enrichissement du dataset et ingénierie de variables

L'efficacité des modèles dépend fortement de la richesse et de la qualité des données exploitées. Il est recommandé d'intégrer de nouveaux types d'informations, notamment :

- Les données sur les prestataires (identité, localisation, historique de prestations) pour détecter les fraudes en réseau ou les comportements anormaux à l'échelle collective.
- Les historiques d'actes médicaux par bénéficiaire et par prestataire, afin d'identifier les schémas répétitifs ou atypiques.
- La géolocalisation des actes pour repérer d'éventuelles incohérences spatiales ou des comportements suspects (par exemple, des soins déclarés simultanément dans différentes régions).

- L’exploitation des séries temporelles pour modéliser les évolutions des comportements à travers le temps et détecter précocement l’apparition de nouveaux schémas de fraude [92].

Annotation humaine ciblée et apprentissage actif

Il sera également pertinent de lancer des campagnes d’annotation humaine sur un panel ciblé de dossiers litigieux ou ambigus, en complément des règles métier existantes. Cette approche, inspirée de l’active learning, permet de réduire le bruit de la labellisation semi-supervisée, d’améliorer la représentativité des fraudes dans le dataset, et donc d’accroître la capacité du modèle à généraliser sur des cas nouveaux ou rares [92].

Surveillance dynamique et détection séquentielle

La mise en place de modules d’analyse comportementale séquentielle permettra de suivre l’évolution des dossiers ou des bénéficiaires sur plusieurs périodes, et d’identifier l’émergence de nouveaux types de fraudes, souvent caractérisés par leur aspect progressif ou répétitif. Les méthodes d’apprentissage séquentiel (par exemple les réseaux récurrents ou les modèles de séries temporelles) sont à considérer pour ce volet.

4.4.1 Industrialisation métier

Pour que le système de détection de fraude atteigne sa pleine efficacité opérationnelle, son industrialisation doit s’accompagner d’outils de supervision, d’exploitabilité et de boucle d’amélioration continue.

Dashboard explicatif

Déployer un tableau de bord interactif à destination des équipes d’audit et de gestion des sinistres est essentiel. Ce dashboard permettra de visualiser en temps réel les scores de suspicion attribués à chaque dossier, les variables explicatives ayant le plus influencé la décision (via SHAP ou LIME), et de générer des rapports adaptés pour la prise de décision. Un tel outil facilite non seulement la transparence vis-à-vis des utilisateurs finaux, mais aussi la conformité réglementaire en matière d’IA explicable.

Boucle de rétroaction continue

Il est recommandé de formaliser un processus de boucle de rétroaction : chaque dossier audité et annoté manuellement par un expert pourra être ré-injecté dans le pipeline de modélisation, enrichissant ainsi le jeu de données d’apprentissage. Cette démarche itérative permet au modèle de s’adapter en continu à l’évolution des schémas de fraude et aux retours des utilisateurs, et d’augmenter progressivement sa robustesse.

Suivi régulier des performances et recalibrage

Enfin, il est indispensable de mettre en place un dispositif de monitoring continu des performances du modèle. Ce suivi inclura la vérification des métriques clés (AUC, précision, rappel, calibration), la détection de dérives potentielles, et le recalibrage périodique du modèle pour garantir son alignement avec la réalité opérationnelle et l'évolution des pratiques frauduleuses. Cette démarche est cruciale pour pérenniser l'efficacité du système et renforcer la confiance des parties prenantes.

4.5 Perspectives de recherche

Ce travail ouvre de nombreux horizons de recherche, tant du point de vue méthodologique que du point de vue applicatif et métier. L'industrialisation de la lutte contre la fraude à l'assurance santé ne pourra progresser que par une approche multidimensionnelle, à la croisée de l'intelligence artificielle, de la statistique avancée et de l'expertise métier. Plusieurs axes se dessinent, dont voici les principaux.

Détection de fraudes collectives et relationnelles

Aujourd'hui, la majorité des systèmes déployés en assurance santé sont focalisés sur la détection de fraudes individuelles, en raison notamment des limitations liées aux données. Pourtant, une part significative de la fraude s'organise sous forme de réseaux structurés ou de collusions entre assurés, prestataires, voire intermédiaires. À cet égard, l'intégration des techniques de Graph Neural Networks (GNN) et plus largement des modèles de graphes relationnels offre un potentiel transformateur. En effet, ces outils permettent d'exploiter la dimension relationnelle des données : chaque assuré, prestataire, ou dossier devient un nœud dans un graphe, et les interactions (soins, remboursements, relations géographiques) deviennent des liens entre ces nœuds. Les GNN sont alors capables de détecter des motifs anormaux, tels que des communautés de fraudeurs, des schémas de sur-facturation croisée, ou la récurrence de certains comportements frauduleux à l'échelle du réseau. Des travaux récents dans d'autres domaines assurantiels ou bancaires ont démontré la pertinence de ces approches pour la détection proactive de réseaux de fraudeurs, notamment en exploitant la structure multi-niveaux et multi-axes des graphes [71][93]. Cette perspective nécessite cependant un enrichissement des jeux de données pour intégrer les identifiants prestataires, les relations entre bénéficiaires et les parcours de soin, ainsi qu'une capacité de modélisation avancée.

Apprentissage actif et annotation intelligente

L'un des freins majeurs à la performance des modèles de détection de fraude demeure la difficulté d'obtenir des labels exhaustifs et de qualité. Les méthodes

classiques de labellisation semi-supervisée ou par règles métiers, bien qu’efficaces pour initier la démarche, introduisent un bruit non négligeable et limitent la capacité des modèles à s’adapter à de nouveaux schémas de fraude. L’apprentissage actif (active learning) propose de pallier ce problème en intégrant l’humain dans la boucle d’apprentissage. Le principe est de solliciter prioritairement l’expertise humaine sur les cas les plus ambigus ou incertains – c’est-à-dire ceux pour lesquels le modèle hésite entre les classes. Cette stratégie permet d’enrichir le dataset de labels pertinents, de découvrir plus rapidement les fraudes émergentes et de garantir une meilleure couverture des différentes typologies. Par ailleurs, l’apprentissage actif peut être automatisé pour optimiser la sélection des exemples à annoter, maximisant ainsi l’apport de chaque nouvelle étiquette sur la performance globale [92]. Cette démarche renforce le lien entre intelligence artificielle et expertise métier, tout en rendant le modèle évolutif et mieux armé face aux évolutions rapides des pratiques frauduleuses.

Analyse causale et modélisation explicative

La prédiction du risque de fraude constitue une avancée majeure pour le secteur de l’assurance, mais elle n’est pas suffisante en soi. Pour agir de manière préventive et stratégique, il est nécessaire d’identifier les causes profondes et les facteurs de risque structurels qui sous-tendent les comportements frauduleux. Les méthodes d’analyse causale, telles que les modèles bayésiens, les réseaux causaux ou les méthodes d’inférence contre-factuelle, offrent la possibilité de passer de la simple corrélation à la compréhension des mécanismes de causalité. Grâce à cette approche, il serait possible de repérer des variables ou des combinaisons de variables qui précèdent systématiquement la fraude, d’évaluer l’impact potentiel de modifications dans la politique de remboursement ou de contrôle, ou encore de mieux cibler les actions de prévention [94]. L’analyse causale représente ainsi un levier stratégique pour concevoir des politiques de maîtrise du risque réellement efficaces et adaptées aux évolutions du marché.

Extension multi-domaines et validation externe

Enfin, la méthodologie développée dans ce mémoire, bien qu’optimisée pour l’assurance santé, est en grande partie générique et adaptable à d’autres domaines assurantiels. L’extension aux assurances automobile, habitation, ou multirisques professionnelles constitue un axe majeur de validation externe. Chaque branche présente des schémas spécifiques de fraude : déclarations exagérées ou fictives en automobile, sinistres simulés en habitation, etc. Appliquer, comparer et adapter les modèles et les pipelines de détection à ces domaines permettra non seulement de tester la robustesse et la transférabilité des outils, mais aussi d’enrichir la base de connaissances partagée par l’ensemble du secteur. Ce travail d’ouverture favoriserait également le développement de référentiels communs, de standards de labellisation et de bonnes pratiques méthodologiques, nécessaires à l’essor d’une lutte anti-fraude moderne, collaborative et efficace.

Ouverture interdisciplinaire et défis éthiques Au-delà des aspects purement

techniques, il convient de souligner que la lutte contre la fraude à l'assurance est un domaine interdisciplinaire par excellence, à la croisée de la data science, du droit, de l'éthique et de la gestion du risque. L'intégration des modèles prédictifs dans les processus métiers doit impérativement tenir compte des exigences réglementaires (RGPD, IA responsable), des impératifs de protection de la vie privée et de la nécessité de garantir l'équité et la non-discrimination dans les décisions automatisées. La recherche sur des modèles explicables, audités et contrôlables par l'humain, tout comme la collaboration avec les parties prenantes (experts métier, juristes, décideurs) sont donc des enjeux majeurs pour l'avenir du domaine.

4.6 Conclusion

En somme, ce chapitre a permis d’apporter un regard critique, approfondi et objectif sur l’ensemble des résultats expérimentaux obtenus dans le cadre de la détection de la fraude à l’assurance santé. À travers une évaluation multicritères – intégrant performance prédictive, calibration des scores, robustesse, interprétabilité et validation croisée – le modèle développé s’est révélé à la fois performant et prometteur pour une intégration future dans les processus métiers d’AXA Assurance Algérie.

L’analyse des métriques et des outils d’explicabilité a mis en lumière la pertinence de l’approche fondée sur un ensemble learning combiné à des variables comportementales, tout en révélant les limites liées à la qualité des labels, à l’absence de données relationnelles et à la modélisation séquentielle. Les axes d’amélioration proposés, ainsi que les pistes de recherche évoquées, offrent une feuille de route concrète pour renforcer à la fois la robustesse, l’efficacité et l’utilité métier de la solution.

Ce travail souligne également l’importance d’un dialogue permanent entre la data science et l’expertise métier, condition essentielle pour l’adaptabilité et la réussite de tout dispositif de lutte contre la fraude. La prochaine étape consistera à valider ces résultats sur des flux opérationnels élargis, à affiner les modèles avec de nouvelles sources de données, et à poursuivre la recherche sur des approches innovantes, notamment celles basées sur l’apprentissage relationnel et la modélisation temporelle.

Conclusion générale

La lutte contre la fraude à l'assurance santé constitue un enjeu stratégique majeur pour les compagnies d'assurance dans un contexte de pression croissante sur les dépenses de santé, de multiplication des réclamations frauduleuses et de montée des exigences réglementaires. Face à des schémas frauduleux de plus en plus complexes, souvent dissimulés dans la masse des sinistres légitimes, il devient indispensable de dépasser les méthodes traditionnelles d'audit manuel ou de détection par règles fixes, et d'explorer les potentialités offertes par la data science, le machine learning et les modèles explicables.

Ce mémoire a proposé une approche complète, rigoureuse et opérationnelle de détection de la fraude à partir de données réelles issues des systèmes d'information d'AXA Assurance Algérie. À travers une méthodologie structurée, il a démontré la possibilité de construire un pipeline robuste intégrant plusieurs étapes clés : labellisation semi-supervisée par règles métier, construction de variables comportementales riches, réduction de dimensionnalité par analyse en composantes principales (PCA), modélisation supervisée via des méthodes d'ensemble (VotingClassifier), et calibration des probabilités pour une exploitation fiable en production. Les résultats obtenus témoignent d'une performance significative : une AUC de 0,937, un rappel élevé de 79 % pour les cas frauduleux, et une précision correcte malgré un fort déséquilibre des classes, ce qui confirme l'efficacité du modèle dans un environnement réaliste.

Au-delà des résultats quantitatifs, ce travail a permis de valider plusieurs éléments importants pour une application industrielle : la capacité à intégrer des scores explicables dans les processus décisionnels (via SHAP et LIME), la pertinence des variables comportementales synthétiques (scores latents vM et vS), ainsi que la compatibilité du système avec une future industrialisation, grâce à un pipeline modulaire et reproductible. Ces contributions offrent une base solide pour la mise en œuvre progressive d'un système intelligent de tri, de priorisation et d'investigation des sinistres à risque.

Néanmoins, plusieurs limites ont été identifiées. La labellisation initiale, reposant uniquement sur des règles déterministes, demeure incomplète et parfois bruitée. L'absence de données sur les prestataires empêche la détection des fraudes en réseau, pourtant courantes dans l'assurance santé. La dimension temporelle des comportements n'a pas été pleinement exploitée, alors qu'elle pourrait révéler des schémas évolutifs ou récidivants. Enfin, l'intégration métier du modèle reste à renforcer, notamment en ce qui concerne l'automatisation du feedback, la gouvernance des décisions algorithmiques, et la supervision en production.

À la lumière de ces constats, plusieurs perspectives de recherche et d'évolution se dégagent. À court terme, l'enrichissement des données par annotation humaine ciblée et l'implémentation d'un tableau de bord explicatif pourraient améliorer la pertinence des détections et l'acceptation par les utilisateurs. À moyen terme, l'adoption de modèles relationnels (Graph Neural Networks), l'apprentissage actif et l'intégration de modules de détection en continu (drift detection, MLOps) ouvrent la voie à une détection proactive et adaptative. À

long terme, l'ambition est de développer un écosystème de lutte contre la fraude éthique, évolutif et interdisciplinaire, à la croisée de la data science, du droit, et de l'expertise métier.

En conclusion, ce mémoire montre qu'il est possible, à partir d'un socle de données opérationnelles et d'une connaissance métier contextualisée, de bâtir un système intelligent de détection de fraude combinant performance algorithmique, transparence et applicabilité. Il constitue ainsi une première brique vers une assurance santé plus juste, plus réactive, et mieux protégée contre les abus, dans un esprit de responsabilité collective et d'innovation continue.

Bibliographie

- [1] Insurance Europe, “Insurance fraud : Combating the crime.” <https://www.insuranceeurope.eu>, 2022. Consulté en avril 2025.
- [2] N. C. for Biotechnology Information, “Health insurance : A risk management tool,” *NCBI*, 2018.
- [3] R. Beesley and M. Roberts, *Health Insurance : Principles and Practice*. Oxford : Oxford University Press, 2nd ed., 2020.
- [4] D. Bordes and A. Lefevre, “L’assurance maladie en france : Entre réformes et défis,” *Revue Française de Gestion Sociale*, vol. 15, pp. 45–67, 2019.
- [5] Radio Algérie, “Assurances : une croissance de 4
- [6] Atlas Magazine, “Résultats 2023 du marché algérien des assurances,” 2024.
- [7] Cleiss, “Le régime algérien de sécurité sociale (salariés) - algérie - cleiss,” 2023.
- [8] F. F. des Sociétés d’Assurances, *Guide de l’assurance santé*. FFSA, 2006.
- [9] J. Roche, *Assurance maladie : principes et fonctionnement*. Presses Universitaires de France, 2015.
- [10] Mutualité Française, “Comment fonctionne le délai de carence en assurance santé ?,” 2021. Consulté en avril 2025.
- [11] Groupama Assistance, “Assistance médicale internationale : garanties et fonctionnement,” 2020. Consulté en avril 2025.
- [12] Collectif, *Lamy Assurance*. Wolters Kluwer, 2022.
- [13] A. Couturier, *Les assurances de personnes*. Dalloz, 2018.
- [14] Légifrance, “Code des assurances : exclusions de garantie,” 2006.
- [15] F. Bessis, *Précis d’assurance maladie*. Lextenso éditions, 2019.
- [16] Fédération Française de l’Assurance, “Le fonctionnement du tiers payant,” 2024. Consulté en avril 2025.
- [17] Gras Savoye Willis Towers Watson, “Assurance santé entreprise, fonctionnement du tiers payant,” 2024. Consulté en avril 2025.
- [18] Ministère de la Santé et de la Protection Sociale – Côte d’Ivoire, “Guide des procédures de remboursement en assurance maladie,” 2024. Document officiel, sans URL disponible.

- [19] Conférence Interafricaine des Marchés d'Assurances, "Cima – code des assurances," 2024. Consulté en avril 2025.
- [20] Managed Care International, "Documentation contractuelle interne sur les conventions de gestion maladie," 2024. Document interne non publié.
- [21] Conférence Interafricaine des Marchés d'Assurances, "Code des assurances de la cima," 2019. Consulté le 30 avril 2025.
- [22] Les Mutuelles Santé, "Comprendre la mutuelle gras savoye pour mieux choisir," 2025. Consulté le 30 avril 2025.
- [23] Service-Public.fr, "Tiers payant," 2025. Consulté le 30 avril 2025.
- [24] PayrHealth, "Managed care contract terms for providers," 2025. Consulté le 30 avril 2025.
- [25] Ministère de la Justice Algérien, "Code civil algérien - article 54," 1975. Définition du contrat.
- [26] Présidence de la République Algérienne, "Ordonnance n° 95-07 du 25 janvier 1995 relative aux assurances," 1995. Articles 2, 21, 22, 23 et 26.
- [27] Ministère de la Justice Algérien, "Code civil algérien - article 107," 1975. Absence d'obligation si la condition dépend de la seule volonté du débiteur.
- [28] I. Europe, "The impact of insurance fraud across europe," 2016. Consulté en avril 2025.
- [29] Deloitte, "Insurance fraud survey 2016," 2016. Consulté en avril 2025.
- [30] I. Nexus, "Global insurance fraud survey 2016," 2016. Consulté en avril 2025.
- [31] S. Russell and P. Norvig, *Artificial Intelligence : A Modern Approach*. Pearson, 2016.
- [32] J. Haugeland, *Artificial Intelligence : The Very Idea*. MIT Press, 1985.
- [33] D. Wilkins, *Making Minds : How Theory and Evidence Shape Our Understanding of the Mind*. Academic Press, 1990.
- [34] R. Bellman, *An Introduction to Artificial Intelligence : Can Computers Think ?* Thomson Learning, 1978.
- [35] E. Charniak and D. McDermott, *Introduction to Artificial Intelligence*. Addison-Wesley, 1985.
- [36] R. Kurzweil, *The Age of Intelligent Machines*. MIT Press, 1990.
- [37] N. J. Nilsson, *Artificial Intelligence : A New Synthesis*. Morgan Kaufmann, 1998.
- [38] E. Rich and K. Knight, *Artificial Intelligence*. McGraw-Hill, 1983.
- [39] D. Poole, A. Mackworth, and R. Goebel, *Computational Intelligence : A Logical Approach*. Oxford University Press, 1998.
- [40] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society : Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

- [41] J. H. Friedman, “Greedy function approximation : A gradient boosting machine,” *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [42] T. Chen and C. Guestrin, “Xgboost : A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.
- [43] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, “Lightgbm : A highly efficient gradient boosting decision tree,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [44] A. V. Dorogush, V. Ershov, and A. Gulin, “Catboost : gradient boosting with categorical features support,” *arXiv preprint arXiv :1810.11363*, 2018.
- [45] A. McCallum and K. Nigam, “A comparison of event models for naive bayes text classification,” in *AAAI-98 workshop on learning for text categorization*, vol. 752, pp. 41–48, 1998.
- [46] J. Pearl, *Probabilistic reasoning in intelligent systems : networks of plausible inference*. Morgan Kaufmann, 1988.
- [47] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [48] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [49] C.-C. Chang and C.-J. Lin, *LIBSVM : A Library for Support Vector Machines*. ACM Transactions on Intelligent Systems and Technology, 2011.
- [50] A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.
- [51] B. Schölkopf, A. J. Smola, and J. C. Platt, “Generalized portraits of support vector machines,” *Machine Learning*, vol. 46, pp. 35–60, 2001.
- [52] Y. Zhou, K. Liu, D. Cao, P. Du, and L. Zhang, “An ensemble deep learning-based approach for healthcare insurance fraud detection,” *Expert Systems with Applications*, vol. 233, p. 120937, 2023.
- [53] M. Ghosh, P. Pathak, and S. Agarwal, “Imbalanced learning methods in insurance fraud detection : A review,” *Frontiers in Artificial Intelligence*, vol. 6, p. 1098063, 2023.
- [54] J. Burez and D. Van den Poel, “Handling class imbalance in customer churn prediction,” *Expert Systems with Applications*, vol. 36, no. 3, pp. 4626–4636, 2009.
- [55] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote : Synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
- [56] H. He, Y. Bai, E. A. Garcia, and S. Li, “Adasyn : Adaptive synthetic sampling approach for imbalanced learning,” in *IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 1322–1328, IEEE, 2008.
- [57] I. Tomek, “Two modifications of cnn,” in *Systems, Man and Cybernetics, IEEE Transactions on*, pp. 769–772, IEEE, 1976.

- [58] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” *arXiv preprint arXiv :1609.02907*, 2017.
- [59] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, “Neural message passing for quantum chemistry,” *International Conference on Machine Learning (ICML)*, 2017.
- [60] W. L. Hamilton, R. Ying, and J. Leskovec, “Inductive representation learning on large graphs,” *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
- [61] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. Leskovec, “Graph convolutional neural networks for web-scale recommender systems,” in *Proceedings of the 24th ACM SIGKDD Conference*, 2018.
- [62] Y. Zhang, P. Qi, and C. D. Manning, “Graph convolution over pruned dependency trees improves relation extraction,” *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018.
- [63] C. Wang, Y. Yu, H. Yin, and J. Li, “Fraudgn : Graph neural network for insurance fraud detection,” *arXiv preprint arXiv :2104.10069*, 2021.
- [64] T. Zheng and et al., “Graph-based anomaly detection for fraud in healthcare claims,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.
- [65] J. Liu, J. Wang, X. Li, and Y. Chen, “Heterogeneous graph neural network for health insurance fraud detection,” *IEEE Transactions on Knowledge and Data Engineering*, 2022.
- [66] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, “Geometric deep learning : Going beyond euclidean data,” *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 18–42, 2017.
- [67] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in *Proceedings of ICLR 2017*, 2016.
- [68] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, “How powerful are graph neural networks?,” in *Proceedings of ICLR 2019*, 2018.
- [69] M. Defferrard, X. Bresson, and P. Vandergheynst, “Convolutional neural networks on graphs with fast localized spectral filtering,” in *Proceedings of NeurIPS 2016*, 2016.
- [70] Y. Yuan, Z. Zhang, J. Liu, Y. Zhou, and Y. Yang, “Explainability of graph neural networks : A survey,” *IEEE Transactions on Neural Networks and Learning Systems*, 2019.
- [71] J. Zhang, P. Zhou, and D. Jin, “Fraudne : Fraud detection with neighborhood embedding,” *arXiv preprint arXiv :2005.12567*, 2020.
- [72] Y. Wang and al., “A robust and interpretable ensemble machine learning model for predicting healthcare insurance fraud,” *Expert Systems with Applications*, 2023.

- [73] S. Wen, Y. Tang, and L. Lu, “An imbalanced learning method based on graph tran-smote for fraud detection,” in *Proceedings of AAAI*, 2024.
- [74] A. Bayerstadler, D. van Dijk, and A. Winter, “Bayesian multinomial latent variable modeling for fraud and abuse detection in health insurance,” *Annals of Applied Statistics*, 2022.
- [75] J. Liu and al., “Deep learning and advanced balancing strategies for medicare fraud detection,” *IEEE Access*, 2021.
- [76] R. Rathi and al., “Xgboost optimisé par bayesian search pour la détection de fraude en assurance santé,” *Data Science Journal*, 2023.
- [77] A. Bayerstadler, L. van Dijk, and F. Winter, “Bayesian multinomial latent variable modeling for fraud and abuse detection in health insurance,” *Insurance : Mathematics and Economics*, 2016.
- [78] C. E. Brodley and M. A. Friedl, “Identifying mislabeled training data,” *Journal of Artificial Intelligence Research*, vol. 11, pp. 131–167, 1999.
- [79] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, “The application of data mining techniques in financial fraud detection : A classification framework and an academic review of literature,” *Decision Support Systems*, vol. 50, no. 3, pp. 559–569, 2011.
- [80] B. Baesens, V. Van Vlasselaer, and W. Verbeke, *Fraud analytics using descriptive, predictive, and social network techniques : A guide to data science for fraud detection*. John Wiley Sons, 2015.
- [81] T. G. Dietterich, “Ensemble methods in machine learning,” in *International Workshop on Multiple Classifier Systems*, pp. 1–15, Springer, 2000.
- [82] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote : synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
- [83] A. Dal Pozzolo, O. Caelen, Y.-A. Le Borgne, S. Waterschoot, and G. Bontempi, “Credit card fraud detection : a realistic modeling and a novel learning strategy,” *IEEE transactions on neural networks and learning systems*, vol. 29, no. 8, pp. 3784–3797, 2018.
- [84] A. Niculescu-Mizil and R. Caruana, “Predicting good probabilities with supervised learning,” in *Proceedings of the 22nd International Conference on Machine Learning*, pp. 625–632, 2005.
- [85] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017.
- [86] A. Niculescu-Mizil and R. Caruana, “Predicting good probabilities with supervised learning,” in *Proceedings of the 22nd International Conference on Machine Learning (ICML)*, (Bonn, Germany), pp. 625–632, ACM, 2005.
- [87] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, “On calibration of modern neural networks,” in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 1321–1330, 2017.

- [88] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in neural information processing systems*, pp. 4765–4774, 2017.
- [89] E. W. Ngai, Y. Hu, Y. Wong, Y. Chen, and X. Sun, “The application of data mining techniques in financial fraud detection : A classification framework and an academic review of literature,” *Decision support systems*, vol. 50, no. 3, pp. 559–569, 2011.
- [90] Z. Wang, X. Chen, Y. Wu, L. Jiang, S. Lin, and G. Qiu, “A robust and interpretable ensemble machine learning model for predicting healthcare insurance fraud,” *Scientific Reports*, vol. 15, no. 1, pp. 1–19, 2025.
- [91] R. A. Rossi, J. Zhou, C. Bronstein, E. Kolaczyk, and D. F. Gleich, “Temporal graph networks for deep learning on dynamic graphs,” *arXiv preprint arXiv :2006.10637*, 2020.
- [92] M. Herland, T. M. Khoshgoftaar, and R. Wald, “A review of medicare fraud detection with data mining,” *Information Systems*, vol. 78, pp. 181–197, 2018.
- [93] F. Liu and other authors, “Handling class imbalance in graph classification,” *ACM Computing Surveys*, 2022.
- [94] S. Viaene, R. A. Derrig, B. Baesens, and G. Dedene, “A comparison of state-of-the-art classification techniques for expert automobile insurance claim fraud detection,” *Journal of Risk and Insurance*, vol. 69, no. 3, pp. 373–421, 2002.