

Mémoire présenté le :
pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA
et l'admission à l'Institut des Actuaires

Par : Eliot NEHME

Titre : Mémoire de zinzin lol rendu avant 2030

Confidentialité : ☒ NON ☐ (Durée : ☐ 1 an ☐ 2 ans)

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

*Membres présents du jury de Signature
l'Institut des Actuaires*

.....

.....

.....

*Membres présents du jury de
l'ISFA*

.....

.....

.....

Entreprise :

Nom :

Signature :

*Directeur de mémoire en entre-
prise :*

Nom :

Signature :

Invité :

Nom :

Signature :

***Autorisation de publication et
de mise en ligne sur un site de
diffusion de documents actua-
riels (après expiration de l'éventuel
délai de confidentialité)***

Signature du responsable entreprise

Signature du candidat

Table des matières

Résumé	iv
Abstract	v
Remerciements	vi
Synthèse	vii
Synthesis	viii
Introduction	1
1 Enjeux économiques et actuariels de la fraude	2
1.1 La fraude en assurance non-vie (auto/MRH)	3
1.1.1 Définitions, typologies (fraude opportuniste vs organisée).	3
1.2 Les sanctions et le cadre réglementaire	3
1.2.1 Souscription	4
1.2.2 Sinistre	4
1.3 Le dilemme de la rentabilité	4
1.3.1 L'ampleur du marché de la fraude	5
1.3.2 Analyse du coût : Détecté vs Estimé	5
1.3.3 Analyse du coût de l'investigation (experts, avocats) par rapport au montant de l'enjeu. Le seuil de déclenchement d'une enquête	5
1.4 Impact sur la tarification	5
1.4.1 Biais dans les statistiques de sinistralité (fréquence et coût moyen).	6
1.4.2 Conséquence sur la prime pure : comment la fraude fait payer les assurés honnêtes.	6
2 État de l'Art et Cadre Théorique	7
2.1 Revue de littérature	8
2.1.1 Ce qui se fait à l'Institut des Actuaires (méthodes traditionnelles vs IA).	8
2.2 Les algorithmes d'IA pour la fraude	8
2.2.1 Apprentissage supervisé	9
2.2.2 Apprentissage non supervisé	9
2.2.3 Deep Learning & Computer Vision	9

2.3	Analyse de réseaux (Graph Mining)	9
2.3.1	Pour détecter les bandes organisées	10
3	Stratégie de Données et Prétraitement	11
3.1	Sourcing et acquisition des bases de données	12
3.1.1	Piste A : Collaboration avec la Direction IARD	12
3.1.2	Piste B : Recherche de bases de données externes (Kaggle)	12
3.1.3	Piste C : Open Data et génération de données synthétiques	12
3.2	Préparation et enrichissement des données	12
3.2.1	Traitement des données structurées	13
3.2.2	Extraction de caractéristiques non-structurées	13
3.3	Gestion du déséquilibre des classes et falsification	13
3.3.1	La problématique de la rareté des cas de fraude	14
3.3.2	Techniques de ré-échantillonnage et GANs	14
4	Mise en œuvre technique et détection par l'IA	15
4.1	Détection de la fraude opportuniste (Auto et MRH)	16
4.1.1	L'IA au service de l'expertise visuelle (YOLO)	16
4.1.2	Analyse de la falsification documentaire	16
4.2	Détection de la fraude en bande organisée	16
4.2.1	Identification de patterns de coûts identiques	17
4.2.2	Analyse spatiale et comportementale	17
4.3	Évaluation de la performance des modèles	17
4.3.1	Métriques de précision et de rappel	18
4.3.2	Interprétabilité des résultats (XAI)	18
5	Impacts Actuariels, Rentabilité et Perspectives	19
5.1	Analyse de la rentabilité économique (Business Case)	20
5.1.1	Le dilemme du coût de l'investigation	20
5.1.2	Réduction des frais de gestion	20
5.2	Conséquences sur la tarification et le provisionnement	20
5.2.1	Biais statistiques induits par la fraude	21
5.2.2	Impact sur la prime pure et l'équité	21
5.3	Conclusion et ouverture	21
5.3.1	Synthèse des travaux	22
5.3.2	Limites éthiques et réglementaires (RGPD)	22
5.3.3	Perspectives sur l'évolution de la fraude (Deepfakes)	22
6	Conclusion	23
6.1	Résumé des résultats	24
6.1.1	Synthèse des principaux résultats obtenus	24
6.2	ffhfhf	24
	Annexes	26



Bibliographie

28

Résumé

Petit résumé en français de mon mémoire ?

Abstract

It's a brief sum up in english !

Remerciements

Merci Gaylord

Synthèse

Un long résumé en français de mon mémoire

Synthesis

Un long résumé en anglais de mon mémoire hehe

Introduction

Ce mémoire est vraiment trop bien hhihihi Je complète mon intro car je suis bg et incroyablement intelligent ! Tu complètes dkdjff

Chapitre 1

Enjeux économiques et actuariels de la fraude

1.1 La fraude en assurance non-vie (auto/MRH)

1.1.1 Définitions, typologies (fraude opportuniste vs organisée).

La fraude en assurance non-vie est un acte intentionnel visant à tromper une compagnie d'assurance dans le but d'obtenir un avantage financier, de la part d'une personne morale ou physique.

1.1.1.1 La fraude à la souscription

L'assurance peut être victime de fraude dès la phase de souscription du contrat. Cela peut inclure la fourniture d'informations fausses ou incomplètes sur le demandeur d'assurance, telles que l'omission de déclarer des antécédents médicaux, des informations sur le véhicule ou le lieu de résidence. Ces fraudes visent à obtenir des primes plus basses ou à garantir la couverture d'un risque qui serait autrement exclu.

1.1.1.2 La fraude lors de la survenance d'un sinistre

La fraude la plus courante en assurance non-vie se produit lors de la déclaration d'un sinistre. Cela peut inclure la surestimation des dommages, la déclaration de sinistres fictifs, la falsification de preuves (comme des photos ou des rapports de police), ou tout simplement inventer des sinistres artificiels. Ces fraudes visent à obtenir des indemnités plus élevées que celles auxquelles l'assuré a droit, ou bien créer une indemnité qui n'aurait tout simplement pas dû exister.

1.1.1.3 La fraude liée aux conditions générales

La fraude peut également se produire en exploitant les conditions générales des contrats d'assurance. Par exemple, un assuré peut tenter de contourner les exclusions de couverture en manipulant les circonstances d'un sinistre ou en exploitant des ambiguïtés dans le contrat. La fraude documentaire (faux certificats, fausses factures, photos retouchées) est en explosion et représente désormais 39% (ALFA) des cas de fraude détectés en IARD. Cela prouve que les méthodes classiques ne suffisent plus et qu'il faut analyser les documents eux-mêmes.

1.2 Les sanctions et le cadre réglementaire

1.2.1 Souscription

Le code des assurances article L113-8 stipule que toute fausse déclaration intentionnelle de la part de l'assuré lors de la souscription d'un contrat d'assurance entraîne la nullité du contrat. En cas de découverte de la fraude, l'assureur peut résilier le contrat et refuser toute indemnisation en cas de sinistre. De plus, les primes acquittées restent acquises à l'assureur en guise de dommages et intérêts.

1.2.2 Sinistre

Selon l'article L113-9 du code des assurances, en cas de fraude lors de la déclaration d'un sinistre, l'assureur peut refuser d'indemniser l'assuré. De plus, si la fraude est découverte après le paiement de l'indemnité, l'assureur peut exiger le remboursement des sommes versées. Des sanctions pénales peuvent également être appliquées, incluant des amendes et des peines d'emprisonnement, en fonction de la gravité de la fraude.

1.3 Le dilemme de la rentabilité

1.3.1 L'ampleur du marché de la fraude

Le marché de l'assurance est si important en France qu'il n'est pas possible de tout vérifier à la main. En 2023, les assureurs ont versé 56,4 milliards d'euros de prestations (sinistres et indemnités). Selon le rapport 2023 de l'ALFA (Association de Lutte contre la Fraude en Assurance), la fraude en assurance non-vie représente environ entre 5% à 10% des sinistres déclarés, soit une perte estimée à plusieurs milliards d'euros chaque année pour les compagnies d'assurance. Cette fraude se manifeste sous diverses formes, allant de la surestimation des dommages à la déclaration de sinistres fictifs. Cependant, la fraude détectée que 902 millions d'euros en 2023 (dont 656 millions rien que pour l'IARD Auto/MRH). Cela signifie qu'une grande partie de la fraude reste non détectée, ce qui représente un défi majeur pour les assureurs.

1.3.2 Analyse du coût : Détecté vs Estimé

En 2023, le montant total des indemnisations versées par les assureurs s'élevait à 49 Milliards d'euros. En comparaison, la fraude détectée en IARD avoisine les 656 millions d'euros (Source : ALFA). Ce différentiel met en lumière le "chiffre noir" de la fraude : alors que la fraude détectée représente environ 1,3% des prestations, les estimations actuarielles situent la fraude réelle entre 3% et 5%. L'enjeu n'est donc pas seulement de stopper l'hémorragie financière, mais de préserver la marge technique dans un contexte où la charge des sinistres climatiques a déjà coûté 6,5 Md€ en 2023.

1.3.3 Analyse du coût de l'investigation (experts, avocats) par rapport au montant de l'enjeu. Le seuil de déclenchement d'une enquête

1.4 Impact sur la tarification

1.4.1 Biais dans les statistiques de sinistralité (fréquence et coût moyen).

L'absence de détection de la fraude introduit un biais systématique dans les bases de données utilisées pour la tarification. La prime pure (π) est modélisée par l'équation :

$$\pi = E[N] \times E[C] \quad (1.1)$$

Où N est la fréquence et C le coût moyen.

- **Impact sur la Fréquence (N)** : La fraude opportuniste (sinistres imaginaires) augmente artificiellement la fréquence des sinistres dans le portefeuille.
- **Impact sur le Coût Moyen (C)** : La majoration de dommages (ajout d'éléments sur une facture) gonfle la sévérité moyenne.

Si l'actuaire calibre ses modèles (GLM, XGBoost) sur ces données polluées, il surestime le risque réel du portefeuille sain.

1.4.2 Conséquence sur la prime pure : comment la fraude fait payer les assurés honnêtes.

Chapitre 2

État de l'Art et Cadre Théorique

2.1 Revue de littérature

2.1.1 Ce qui se fait à l'Institut des Actuaire (méthodes traditionnelles vs IA).

2.2 Les algorithmes d'IA pour la fraude

2.2.1 Apprentissage supervisé

Détecter la fraude connue (Random Forest XGBoost).

2.2.2 Apprentissage non supervisé

Détecter des anomalies / comportements atypiques (Isolation Forest, Clustering).

2.2.3 Deep Learning & Computer Vision

Utilisation de YOLO pour l'analyse d'images de chocs et GANs pour la gestion des données déséquilibrées.

2.3 Analyse de réseaux (Graph Mining)

2.3.1 Pour détecter les bandes organisées

Liens entre experts, garages, et assurés

Chapitre 3

Stratégie de Données et Prétraitement

3.1 Sourcing et acquisition des bases de données

3.1.1 Piste A : Collaboration avec la Direction IARD

Exploitation des données réelles de l'assureur et sollicitation du réseau du directeur pour obtenir des bases qualifiées.

3.1.2 Piste B : Recherche de bases de données externes (Kaggle)

Utilisation de jeux de données issus de compétitions de data science spécialisées en détection d'anomalies.

3.1.3 Piste C : Open Data et génération de données synthétiques

Recours à des bases anonymisées disponibles en ligne ou création de données via des techniques de falsification (expérience du TER).

3.2 Préparation et enrichissement des données

3.2.1 Traitement des données structurées

Nettoyage des bases de sinistres classiques (âge du conducteur, lieu, montant, type de garantie).

3.2.2 Extraction de caractéristiques non-structurées

Utilisation du NLP pour les rapports d'experts et des Transformers pour l'analyse des pièces justificatives.

3.3 Gestion du déséquilibre des classes et falsification

3.3.1 La problématique de la rareté des cas de fraude

Analyse statistique du faible taux de fraude dans un portefeuille sain.

3.3.2 Techniques de ré-échantillonnage et GANs

Utilisation de réseaux antagonistes génératifs (GANs) pour créer des exemples de fraude réalistes à partir de ton projet sur les faux documents (CNI, factures).

Chapitre 4

Mise en œuvre technique et détection par l'IA

4.1 Détection de la fraude opportuniste (Auto et MRH)

4.1.1 L'IA au service de l'expertise visuelle (YOLO)

Remplacer l'œil de l'expert pour détecter les incohérences : identifier si une griffure latérale est compatible avec un choc frontal sur un poteau.

4.1.2 Analyse de la falsification documentaire

Détection de retouches sur les photos d'accidents ou sur les documents d'identité.

4.2 Détection de la fraude en bande organisée

4.2.1 Identification de patterns de coûts identiques

Repérer les sinistres ayant exactement le même coût moyen, signe potentiel d'une fraude industrielle.

4.2.2 Analyse spatiale et comportementale

Détection de sinistres similaires sur des habitations éloignées (MRH) n'ayant aucun lien logique, mais présentant des caractéristiques techniques identiques.

4.3 Évaluation de la performance des modèles

4.3.1 Métriques de précision et de rappel

Mesurer la capacité du modèle à ne pas oublier de fraudeurs tout en évitant de suspecter des clients honnêtes.

4.3.2 Interprétabilité des résultats (XAI)

Expliquer pourquoi l'IA a flagué un dossier pour que l'enquêteur humain puisse prendre le relais.

Chapitre 5

Impacts Actuariels, Rentabilité et Perspectives

5.1 Analyse de la rentabilité économique (Business Case)

5.1.1 Le dilemme du coût de l'investigation

Est-il rentable d'engager des frais d'avocats et d'experts pour un sinistre à 1 200 €? Définition du ROI de l'IA.

5.1.2 Réduction des frais de gestion

Automatisation du tri pour concentrer les ressources humaines sur les dossiers à fort enjeu.

5.2 Conséquences sur la tarification et le provisionnement

5.2.1 Biais statistiques induits par la fraude

Analyse de la manière dont la fraude non détectée fausse la fréquence et le coût moyen des sinistres.

5.2.2 Impact sur la prime pure et l'équité

Modélisation de la baisse potentielle des primes pour les assurés honnêtes grâce à la réduction de la charge sinistre globale.

5.3 Conclusion et ouverture

5.3.1 Synthèse des travaux

5.3.2 Limites éthiques et réglementaires (RGPD)

5.3.3 Perspectives sur l'évolution de la fraude (Deepfakes)

Chapitre 6

Conclusion

6.1 Résumé des résultats

6.1.1 Synthèse des principaux résultats obtenus



6.2 ffhfhf



Annexes



Chapitre 3

Bibliographie