
Détection de la fraude en assurance par les méthodes d'intelligence artificielle :

Revue de la littérature

Fraud Detection in Insurance Using Artificial Intelligence Methods: A Literature Review

BELLAL Kamilia ¹

BENILLES Billel

Doctorante à l'Ecole Supérieure de Commerce

Professeur à l'Ecole Supérieure de Commerce

Laboratoire d'études pratiques en sciences commerciales

Laboratoire d'études pratiques en sciences commerciales

et sciences de gestion Koléa (Algérie)

et sciences de gestion Koléa (Algérie)

k_bellal@esc-alger.dz

b_benilles@esc-alger.dz

Date de soumission: 28/02/2026

Date d'acceptation: 06/05/2026

Date de publication: 05/31/2026

Résumé : L'utilisation de l'intelligence artificielle dans le secteur de l'assurance a considérablement évolué, transformant les processus de détection et de prévention des sinistres frauduleux. Ce travail propose une revue de la littérature portant sur l'application de l'IA pour identifier les différentes formes de fraude en assurance. Les résultats montrent que l'apprentissage supervisé, non supervisé et leurs algorithmes associés améliorent significativement la précision des contrôles antifraude. Cette performance repose sur l'exploitation des données contractuelles, des déclarations de sinistres et des historiques comportementaux. Les études indiquent que l'IA a radicalement transformé l'industrie de l'assurance en renforçant la détection des fraudes, mais elles soulignent également que ces outils restent confrontés à des limites, notamment la disponibilité et la qualité des données.

Mots clés : Fraude ; Assurance ; Intelligence Artificielle ; Apprentissage supervisé ; Algorithmes

Codes de classification Jel : G22, C52.

Abstract: The use of artificial intelligence in the insurance sector has evolved significantly, transforming the processes of detecting and preventing fraudulent claims. This work presents a literature review on the application of AI to identify different forms of insurance fraud. The results show that supervised and unsupervised learning, along with their associated algorithms, significantly improve the accuracy of fraud detection. This performance relies on the exploitation of contractual data, claim reports, and behavioral histories. Studies indicate that AI has radically transformed the insurance industry by enhancing fraud detection, but they also highlight that these tools still face limitations, notably the availability and quality of data.

Keywords: Fraud; Insurance; Artificial Intelligence; Supervised Learning; Algorithms

Jel Classification Codes : G22, C52.

*Corresponding author

1. Introduction :

L'industrie de l'assurance occupe une place centrale dans la gestion des risques auxquels sont exposés les particuliers et les entreprises. Néanmoins, elle est confrontée à une menace persistante : la fraude à l'assurance. Cette dernière constitue un enjeu complexe et coûteux, compromettant la stabilité des marchés d'assurance et impactant le montant des primes pour les assurés non fraudeurs. Dans ce contexte, la détection efficace de la fraude est devenue une priorité stratégique pour les compagnies d'assurance, les autorités de contrôle et la communauté scientifique. Traditionnellement, la détection des fraudes reposait sur un processus manuel, assuré par les gestionnaires de sinistres, qui analysaient les dossiers et prenaient des décisions basées principalement sur leur expérience et leur intuition. Ces méthodes présentent toutefois des limites importantes en termes de délais de traitement et de précision. Pour pallier ces contraintes, l'intelligence artificielle s'impose comme une technologie stratégique, offrant des solutions performantes pour l'identification précoce et fiable des comportements frauduleux.

L'exploitation des données massives (Big Data), combinée à la puissance de calcul et aux algorithmes d'apprentissage automatique (Machine Learning), a permis à l'IA de devenir un outil incontournable dans la lutte contre la fraude à l'assurance. Conscients de ce potentiel, les assureurs déploient progressivement ces technologies au sein de leurs systèmes de détection de fraude (Fraud Détection Systems). Parallèlement, de nombreuses recherches académiques et industrielles ont été menées pour évaluer l'efficacité de ces méthodes, identifier leurs avantages et limitations, et explorer les perspectives d'amélioration.

Cet article propose une synthèse des travaux de recherche portant sur l'usage de l'IA dans la détection de la fraude en assurance. Il se concentre sur les différentes techniques employées, incluant l'apprentissage supervisé, l'apprentissage non supervisé, ainsi que leurs divers algorithmes. L'étude examine également l'impact de ces approches sur l'efficacité opérationnelle des assureurs, la réduction des pertes financières et les limites de ces outils.

L'organisation de cet article est la suivante : la première section présente les concepts fondamentaux de la fraude à l'assurance ; la deuxième section offre un panorama général de l'intelligence artificielle, avec un focus sur le Machine Learning et ses principaux algorithmes ; enfin, la troisième section synthétise les recherches récentes sur la détection de la fraude à l'aide de l'IA.

2. La fraude en assurance : Concepts de base

La fraude est un phénomène particulièrement dangereux qui touche plusieurs domaines et affecte leur rentabilité, et l'un des secteurs les plus touchés est celui de l'assurance. Les compagnies d'assurance dans le monde sont actuellement confrontées à ce fléau, qui compromet à la fois leur rentabilité et leur solvabilité.

2.1 Définition de la fraude

Les définitions de la fraude sont nombreuses et variées. Pour mieux appréhender ce concept, il est utile d'en examiner quelques-unes. Selon le dictionnaire Petit Robert, la fraude correspond à « une action réalisée de mauvaise foi dans le but de tromper ». Elle désigne un acte illégal accompli pour tromper délibérément autrui, soutirer de l'argent contre sa volonté ou falsifier intentionnellement un document, portant ainsi atteinte aux droits ou aux intérêts d'autrui. Le Oxford English Dictionary définit la fraude comme « une tromperie illégale ou criminelle destinée à obtenir un gain financier ou personnel ».

Dans la littérature académique, la fraude est souvent décrite comme l'exploitation abusive des systèmes d'une organisation à but lucratif, sans que cela n'entraîne nécessairement de conséquences juridiques directes. Bien qu'aucune définition universellement acceptée de la fraude financière n'existe, certains auteurs la considèrent comme « un acte délibéré contraire à la loi, aux règles ou aux politiques, dans l'intention d'obtenir un avantage financier non autorisé » (Anuj & Panigrahi, 2012, pp. 37-38).

Le terme « fraude » est complexe à définir, car il englobe une grande variété de pratiques et d'idées reçues. On pense fréquemment aux fraudes fiscales, électorales, aux assurances ou encore à la fraude de consommateur, autant de situations largement médiatisées et souvent très coûteuses pour les entreprises.

Dans le cadre de cet article, l'accent sera mis sur la fraude en assurance (Gallet, 2014, p. 04).

2.2 La fraude en assurance et ces types :

La fraude en assurance constitue un problème majeur affectant la rentabilité et la solvabilité des compagnies.

Avant d'analyser ses conséquences, il est essentiel d'en définir clairement le concept et de distinguer ses différents types afin de mieux comprendre l'ampleur et la diversité de ce phénomène.

2.2.1 La fraude en assurance

La fraude en assurance recouvre la falsification de faits ou de déclarations ainsi que la dissimulation intentionnelle d'informations en vue d'obtenir un avantage indu. Elle touche l'ensemble des acteurs économiques : particuliers, entreprises privées, institutions publiques et, plus largement, l'État. Il s'agit d'une criminalité évolutive, qui s'est progressivement complexifiée avec le développement des technologies et la digitalisation des processus.

L'ampleur réelle de ses effets demeure difficile à quantifier, dans la mesure où une part significative des actes frauduleux échappe à la détection. À l'échelle internationale, la fraude s'est intensifiée au point de constituer une source majeure de pertes financières. Dans le secteur de l'assurance, elle se traduit par une augmentation des primes, une dégradation de la stabilité financière des entreprises et une réduction de l'accessibilité des produits d'assurance (Titilope Azeeza & Olowokudejob, 2025, p. 22).

2.2.2 Les types de fraude en assurance

La fraude en assurance se manifeste sous différentes formes et concerne l'ensemble des branches. On distingue principalement trois catégories : la fraude à la souscription, la fraude multi-assurances et la fraude aux sinistres (Jonathan, 2016).

a. La fraude à la souscription :

Elle vise principalement à réduire le montant de la prime. L'assuré communique volontairement des informations inexactes ou incomplètes lors de la déclaration du risque afin d'obtenir une tarification plus avantageuse. Il s'agit d'une fraude relativement simple à mettre en œuvre et souvent difficile à détecter, notamment en l'absence de dispositifs de vérification approfondis. Ce type de fraude se rencontre dans la plupart des branches, notamment en assurance automobile et en assurance vie.

b. La fraude multi-assurances :

Elle consiste pour un assuré à souscrire plusieurs contrats couvrant un même risque auprès de différentes compagnies, dans l'intention de percevoir plusieurs indemnités pour un même sinistre. Cette pratique, qui contrevient au principe indemnitaire, demeure complexe à identifier sans mécanismes de partage d'information ou bases de données communes entre assureurs.

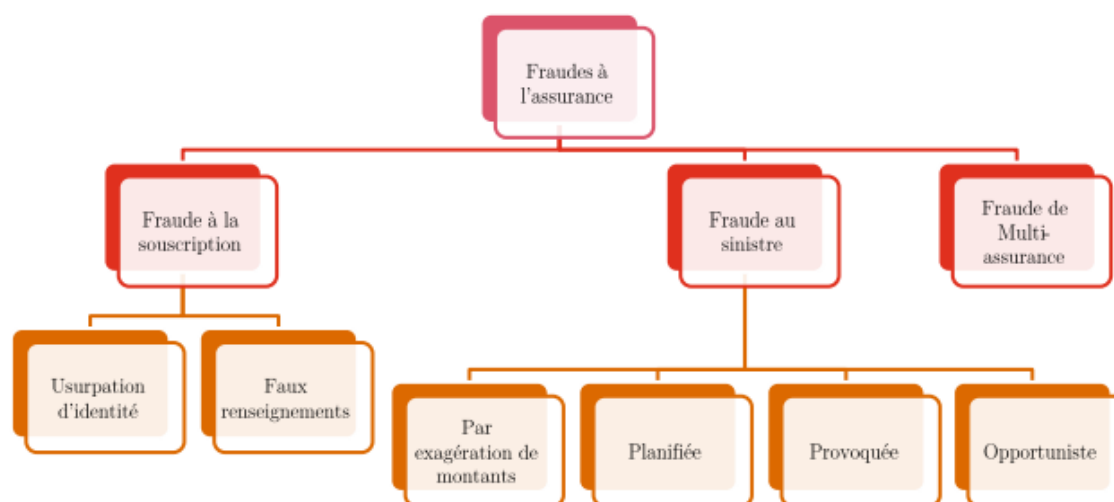
c. La fraude aux sinistres :

Elle intervient au moment de la déclaration ou du règlement du sinistre et peut prendre plusieurs formes :

- Par exagération du préjudice : majoration frauduleuse du montant des dommages par falsification de pièces justificatives (factures, devis) ou collusion avec un tiers (réparateur, prestataire de services) afin d'augmenter l'indemnité.
- Planifiée (sinistre fictif) : déclaration d'un événement dommageable qui ne s'est pas produit.
- Provoquée : survenance intentionnelle du sinistre par l'assuré dans le but d'obtenir une indemnisation.
- Opportuniste : intégration de dommages antérieurs non déclarés dans une déclaration de sinistre ultérieure afin d'en obtenir la prise en charge.

Ces différentes formes illustrent la diversité des schémas frauduleux et la nécessité de dispositifs de contrôle adaptés à chaque phase du cycle assurantiel. La figure suivante synthétise les différentes formes de la fraude en assurance.

Figure 01 : Les différents types de fraude en assurance



Source : Jonathan, Karsenty, La détection de fraudes à l'assurance, Mémoire présenté devant l'ENSAE Paris Tech pour l'obtention du diplôme de la filière actuariat et l'admission à l'Institut des Actuaire, 2016, p.8

2.3. L'impact de la fraude sur les acteurs de secteur assurantiel :

La fraude à l'assurance a des répercussions profondes et multidimensionnelles sur l'ensemble des acteurs du secteur. Pour les entreprises d'assurance, elle se traduit par une sinistralité artificiellement élevée entraînant un accroissement des charges techniques, notamment les frais d'expertise, d'investigation et de gestion des sinistres, ainsi que des investissements supplémentaires en dispositifs de détection et de conformité. Cette pression sur les provisions techniques et sur le résultat technique affecte la marge de solvabilité et limite la capacité d'allocation du capital. Du côté des assurés, l'alourdissement du coût des sinistres frauduleux se répercute dans la tarification par l'augmentation des primes pures et des chargements, favorisant parfois la sous-assurance ou le défaut d'assurance. Le renforcement des procédures de contrôle et d'instruction des dossiers peut rallonger les délais d'indemnisation des sinistres légitimes et altérer la qualité de service perçue. Pour les autorités de régulation et les organismes d'application de la loi, la lutte contre la fraude mobilise des ressources dédiées à l'audit, à l'enquête et aux poursuites, impliquant une réaffectation budgétaire et un coût d'opportunité, tout en nécessitant un renforcement des mécanismes de supervision prudentielle et de coopération interinstitutionnelle. Enfin, pour les

investisseurs et les marchés financiers, la hausse de la sinistralité imputable à la fraude pèse sur la rentabilité technique et financière des assureurs cotés, affecte les indicateurs de performance tels que le ROE et le résultat net, et peut entraîner une érosion de la valeur actionnariale, affaiblissant la confiance des investisseurs et augmentant la sensibilité du titre aux risques sectoriels. En définitive, la fraude à l'assurance perturbe l'équilibre actuariel des portefeuilles, compromet la performance technique des assureurs et génère des externalités négatives pour l'ensemble des parties prenantes du marché assurantiel (Baali, 2024, pp. 257-258).

3.L'Intelligence Artificielle IA :

3.1 L'émergence de l'IA :

L'expression « Intelligence Artificielle » a été officialisée en 1956, lorsque Marvin Minsky et John McCarthy (informaticien à Stanford) organisèrent le Dartmouth Summer Research Project on Artificial Intelligence (DSRPAI), un atelier d'environ huit semaines tenu au Dartmouth College dans le New Hampshire. Financé par la Rockefeller Foundation, cet événement, souvent considéré comme le point de départ du « printemps de l'IA », rassembla des personnalités appelées à devenir les pionniers du domaine, parmi lesquelles Nathaniel Rochester, futur concepteur de l'IBM 701 (premier ordinateur scientifique commercial), et Claude Shannon, père de la théorie de l'information. L'objectif du DSRPAI était de réunir des chercheurs de disciplines diverses afin de structurer un champ nouveau, destiné à créer des machines capables de simuler l'intelligence humaine. (Haenlein & Kaplan, 2019)

3.2 Définition de L'IA :

Il convient tout d'abord de préciser ce que recouvre la notion d'« intelligence ». Dans le langage courant, ce terme se réfère principalement à une capacité propre à l'être humain, perçue comme un trait distinctif qui le sépare de l'animal. L'intelligence est ainsi le plus souvent associée à la présence d'un esprit, ou plus exactement à certains processus mentaux permettant la compréhension, le raisonnement et la prise de décision.

Le terme « artificielle », quant à lui, pose moins de difficultés d'interprétation. Il désigne ce qui est créé, conçu ou inventé par l'homme, et plus largement tout ce qui ne relève pas d'un processus naturel, mais résulte d'une intervention technique ou technologique humaine.

L'intelligence artificielle peut être définie comme une discipline scientifique structurée autour d'un ensemble cohérent de connaissances, de méthodes et de techniques visant une finalité commune : la conception et la programmation de systèmes capables de modéliser certains processus cognitifs associés à l'intelligence humaine et de résoudre des problèmes complexes dans divers domaines du savoir. Elle s'appuie sur plusieurs champs

scientifiques, notamment la logique, les mathématiques, les statistiques, les probabilités et l'informatique, afin de reproduire sous forme computationnelle des capacités telles que le raisonnement, l'apprentissage, la prise de décision ou la reconnaissance de modèles. (Fouzi & Sadik , 2024, p. 11)

Dans cette perspective, est qualifié d'intelligent tout système capable de percevoir son environnement, de s'y adapter et de produire des réponses appropriées en fonction des situations rencontrées. L'IA vise précisément à doter les machines de cette aptitude adaptative et décisionnelle, en leur permettant d'accomplir des tâches traditionnellement réalisées par l'être humain grâce à ses facultés cognitives. (Mathivet, 2014)

Le développement contemporain de l'IA est indissociable de l'essor des big data, qui constituent le socle informationnel nécessaire à l'apprentissage et à l'amélioration des algorithmes. Les données massives représentent en effet la matière première sur laquelle reposent la conception, l'entraînement, la validation et l'optimisation des modèles d'intelligence artificielle. Au-delà de leur volume, leur qualité, leur fiabilité et leur accessibilité conditionnent la performance et la pertinence des systèmes développés.

Ainsi, la relation entre IA et big data est fondamentalement complémentaire : sans données massives, les solutions d'intelligence artificielle ne peuvent être entraînées ni déployées efficacement ; inversement, sans outils d'IA, les big data demeurent largement sous-exploitées. Correctement encadrées et mobilisées, ces deux dimensions offrent un potentiel considérable pour répondre à des problématiques complexes et relever des défis économiques, scientifiques et sociétaux majeurs. . (Mirjana , Aminata , & Garba, 2021)

3.3 L'importance de l'IA :

L'intelligence artificielle contribue à améliorer la qualité des décisions en réduisant certains biais humains grâce à des traitements algorithmiques fondés sur les données. Elle joue un rôle central en cybersécurité, automatise de nombreuses tâches répétitives et permet aux individus de se concentrer sur des activités à plus forte valeur ajoutée. En accroissant la rapidité d'exécution, la précision et la productivité, l'IA constitue un levier stratégique de performance pour les organisations. Elle renforce également la sécurité dans les environnements à haut risque et limite les erreurs humaines grâce à l'exécution rigoureuse de processus programmés.

Son développement repose largement sur l'exploitation des données massives et sur des infrastructures technologiques coûteuses. Malgré ses bénéfices, l'IA soulève plusieurs enjeux majeurs : impact potentiel sur l'emploi, risques d'usage malveillant, dépendance technologique croissante et défis éthiques liés à l'automatisation. En outre, si les machines excellent en efficacité opérationnelle, elles demeurent incapables de remplacer les dimensions relationnelles, émotionnelles et sociales propres à l'être humain, et restent limitées aux tâches pour lesquelles elles ont été conçues. (Sachin , Vinayak , & Zameer , 2020).

3.4. Les outils de la détection de la fraude en assurance: Le maching learning:

3.4.1. Maching Learning

Le machine learning constitue une branche de l'intelligence artificielle dédiée à la création d'algorithmes capables d'identifier automatiquement des motifs et des régularités à partir des données. Contrairement à la programmation traditionnelle, où les règles et relations fonctionnelles sont explicitement codées à partir de connaissances préalables, le machine learning s'appuie sur l'analyse de sous-ensembles de données pour estimer un modèle algorithmique. Ce modèle est capable de détecter des combinaisons complexes et inédites de variables et de pondérations, qui ne peuvent être déterminées analytiquement ni déduites à partir de principes théoriques a priori (Choi, Coyne, Kalpathy-Cramer, Chiang, & Campbell, 2020)

Le terme de machine learning a été introduit en 1959 par Arthur Samuel, considéré comme l'un des pionniers du domaine. Selon lui, il s'agit d'une approche permettant aux ordinateurs d'apprendre et de s'améliorer à partir de l'expérience, sans être explicitement programmés (Kouach, El attar, & El hachloufi, 2022). Aujourd'hui, le machine learning est largement utilisé dans la détection des fraudes, notamment dans le secteur des assurances.

Le processus d'apprentissage automatique se déroule généralement en quatre étapes principales: (Ayushi & Preeti, 2019):

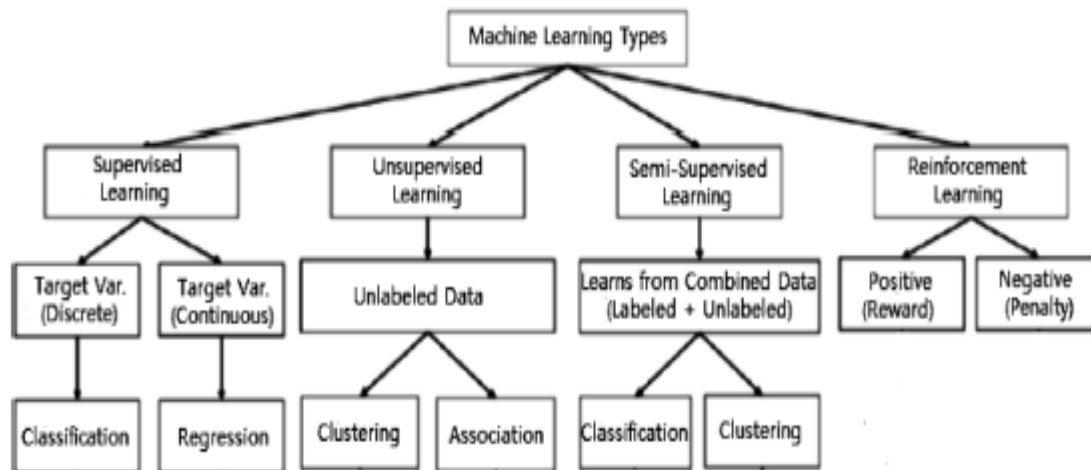
1. **Extraction des caractéristiques** des données pertinentes pour le problème étudié ;
2. **Sélection de l'algorithme** le plus adapté au type de données et à l'objectif ;
3. **Entraînement du modèle** et évaluation de sa performance à l'aide de métriques appropriées ;
4. **Utilisation du modèle entraîné** pour réaliser des prédictions ou des classifications sur de nouvelles données.

On distingue plusieurs types d'apprentissage automatique : (Sarker, 2021, pp. 4-5)

- **Apprentissage supervisé** : le modèle apprend une fonction ou une relation à partir de données d'entraînement étiquetées. Chaque exemple d'entrée X est associé à une sortie correspondante Y, contenant des étiquettes ou tags qui expliquent ou classifient l'exemple.
- **Apprentissage non supervisé** : aucune donnée n'est étiquetée, et l'algorithme cherche à identifier des structures cachées, des regroupements ou des relations au sein des données sans supervision humaine.
- **Apprentissage semi-supervisé** : les données disponibles combinent des éléments étiquetés et non étiquetés. Cette approche permet de construire des modèles de classification plus robustes et précis.

- **Apprentissage par renforcement** : l'agent apprend en interagissant avec son environnement, dans le but de maximiser les récompenses obtenues ou de minimiser les risques, en ajustant ses actions selon les résultats de ces interactions.

Figure 02: Les types de machine learning



Source: Sarker, Iqbal H, Machine learning: algorithms, real-world applications and research directions, SN Computer Science, 2(3), 2021, p.4

3.4.2 Les algorithmes de machine learning:

Le machine learning regroupe plusieurs algorithmes, parmi lesquels :

a- Random Forest : Cette méthode d'apprentissage statistique a été introduite par Leo Breiman (1984). L'apprentissage par arbres de décision constitue une approche classique de l'apprentissage statistique supervisé. Son objectif est de construire un modèle capable de prédire la valeur d'une variable cible à partir des valeurs de variables explicatives. Le principe de l'algorithme repose sur un partitionnement orthogonal de l'espace des variables explicatives, suivi de l'ajustement de modèles distincts sur chacune de ces sous-partitions. (Buari, 2019, p. 67)

Les forêts aléatoires se définissent comme un ensemble de modèles d'arbres de décision, où chaque arbre est construit à partir d'un vecteur aléatoire généré de façon indépendante mais selon une même distribution. L'erreur de généralisation de ce type de modèle tend presque sûrement vers une limite lorsque le nombre d'arbres devient très élevé. Cette performance globale dépend à la fois de la précision des arbres individuels et du degré de corrélation existant entre eux. (Breiman, 2001, p. 1)

b-Réseau de neurones artificiel (RNA): Un réseau de neurones artificiel (RNA), ou Artificiel Neural Network (ANN) en anglais, est un algorithme d'apprentissage automatique conçu sur le modèle des

réseaux neuronaux biologiques. Il est composé de nœuds (comparables aux corps cellulaires) reliés entre eux par des connexions (similaires aux axones et dendrites). À l'image des synapses biologiques qui se renforcent lorsque des neurones s'activent simultanément, les connexions entre nœuds d'un ANN sont pondérées en fonction de leur aptitude à fournir le résultat souhaité. (Maniati & al, 2022, p. 4)

c- Gradient Boosting Machine: Introduit par Freund et Schapire (1997), le boosting est une méthode générale visant à obtenir une prédiction hautement précise en combinant un grand nombre de prédictions initialement peu précises. Développé par Friedman (2001, 2002) ainsi que Natekin et Knoll (2013), le Gradient Boosting Machine (GBM) s'inscrit dans cette logique en construisant des modèles prédictifs à travers des ajustements itératifs (back-fitting) et des régressions non paramétriques. Plutôt que de se limiter à un seul modèle, le GBM débute par un modèle de base, puis améliore progressivement la précision en ajoutant de nouveaux modèles successifs, chacun étant optimisé par la minimisation d'une fonction de perte (Natekin et Knoll, 2013). (Zhiyuan, Danchen, & Thomas , 2019)

d- KNN(K Nearest Neighbors): Cet algorithme est une méthode d'apprentissage automatique relevant de l'apprentissage supervisé. Il se distingue par sa simplicité conceptuelle et par le fait qu'il appartient à la catégorie des algorithmes à apprentissage différé (lazy learning), contrairement aux approches dites « eager » qui construisent un modèle généralisé lors de la phase d'entraînement. Dans le cas du KNN, il n'y a pas véritablement de modélisation préalable complexe : l'algorithme ne procède pas à une estimation paramétrique ni à une abstraction des données. La phase d'apprentissage est donc réduite au strict minimum, ce qui la rend rapide. En revanche, l'ensemble (ou une large partie) des observations d'apprentissage doit être conservé, car elles sont mobilisées directement lors de la phase de prédiction. Le principe repose sur une mesure de proximité : pour prédire la valeur associée à une nouvelle observation (par exemple un assuré ou un sinistre), l'algorithme identifie les k observations les plus proches dans l'échantillon de référence.

En pratique assurantielle, cette méthode peut être exploitée de deux manières :

- **En classification**, lorsque l'objectif est d'affecter une observation à une catégorie (valeur discrète), telle que la détection d'un sinistre potentiellement frauduleux ou l'attribution d'un profil de risque. L'affectation repose généralement sur la règle de la majorité : la classe retenue est celle qui domine parmi les K Nearest Neighbors
- **En régression**, lorsque la variable cible est quantitative et continue, comme l'estimation du coût prévisionnel d'un sinistre ou le calcul d'une prime technique. La valeur prédite correspond alors en général à la moyenne ou, dans certains cas, à la médiane des valeurs observées chez les voisins sélectionnés.

Ainsi, le KNN constitue une approche non paramétrique adaptée aux problématiques de tarification, segmentation des risques et détection d'anomalies dans le domaine de l'assurance. (Ranjitha, Joshni, & Vaishnavi, 2023, p. 2)

E- La Machine à Vecteurs de Support (Support Vector Machine - SVM) : est une méthode d'apprentissage supervisé dédiée principalement aux tâches de classification. Elle repose sur la détermination d'un hyperplan ou de plusieurs hyperplans dans un espace à forte dimensionnalité, permettant d'assurer une séparation optimale entre des groupes d'observations distincts. Cette technique est particulièrement adaptée aux problématiques de classification binaire, notamment pour différencier deux catégories opposées telles que les cas de fraude et de non-fraude. (Sagar & Dhanalakshimi, 2025, p. 4)

F- Naïve Bayes : Le modèle Naïve Bayes (NB) est un algorithme d'apprentissage automatique rapide et fiable, particulièrement adapté aux grands jeux de données, qui permet d'estimer la précision prédictive lors des tests. La construction de ce modèle repose sur le théorème de Bayes, en supposant que les variables prédictives sont indépendantes les unes des autres. Dans le contexte de la détection de fraude, le modèle considère qu'un attribut présent dans une classe n'influence pas la présence d'un autre attribut. (Aslam, Imran unjra, Ftiti, Louhichi, & Tahira, 2022, p. 14)

4. Détection de la fraude en assurance par les méthodes d'IA : revue de littérature

Les travaux empiriques relatifs à la détection de la fraude en assurance demeurent relativement nombreux, bien qu'un nombre croissant d'études mobilisent des approches quantitatives et des techniques d'apprentissage automatique afin d'optimiser les dispositifs de maîtrise du risque de fraude, notamment en assurance automobile et en assurance de biens.

L'étude de (Aslam, Imran unjra, Ftiti, Louhichi, & Tahira, 2022) propose une approche pour la détection de fraude en assurance automobile utilisant des modèles prédictifs de machine learning (ML). Le jeu de données exploité, collecté en 2020 par le département d'analytique décisionnelle d'EXL Service, est public et open source, et comprend 33 variables, dont une variable cible binaire indiquant si un sinistre est frauduleux ou confirmé comme tel. Les variables incluent notamment : mois, semaine et jour de déclaration, marque du véhicule, âge de l'assuré, numéro du représentant ou courtier, franchise, nombre de sinistres antérieurs, changement d'adresse, nombre de véhicules assurés et numéro de police. Pour la construction des modèles, un échantillonnage aléatoire 80/20 a été appliqué, 80 % des données servant à l'apprentissage et 20 % à l'évaluation. La sélection des variables les plus pertinentes a été réalisée avec l'algorithme Boruta, qui a identifié neuf caractéristiques significatives : Age of Policy Holder, Age of Vehicle, Base Policy, Fault, Marital Status, Vehicle Category, Sex, Days Policy Claim et Policy Type.

Parmi celles-ci, Fault, Base Policy et Age of Policy Holder se révèlent être les plus déterminantes pour la détection de fraude. La fraude en assurance automobile étant un problème de classification binaire, trois modèles prédictifs régression logistique (LR), machine à vecteurs de support (SVM) et Naïve Bayes (NB) ont été appliqués et évalués. L'évaluation a été réalisée à l'aide de la matrice de confusion et de six métriques de performance : précision (accuracy), F1-score, précision (precision), rappel (recall), sensibilité (sensitivity) et spécificité (specificity). Un classement de l'importance des variables explicatives montre que Fault (responsabilité), Base Policy et Age of Policy Holder (âge de l'assuré) sont les facteurs les plus déterminants dans la détection de fraude. Les résultats ont montré que le SVM atteint la meilleure précision, tandis que la régression logistique obtient le F1-score le plus élevé, indiquant une cohérence supérieure dans la détection des fraudes. Chaque modèle présente des avantages spécifiques selon les métriques considérées, mais l'ensemble des trois modèles est jugé opérationnel et pertinent pour la détection de fraude en assurance automobile. Ainsi, (Kouach, El attar, & El hachloufi, 2022) proposent la mise en place d'un système de détection de la fraude en assurance automobile fondé sur le machine learning. Leur analyse repose sur un échantillon de 1 000 dossiers de sinistres comprenant 38 variables explicatives (type d'incident, nombre de véhicules impliqués, présence de témoins, etc.) et une variable cible binaire indiquant le caractère frauduleux ou non du sinistre avec l'utilisation de deux algorithmes : Isolation Forest et Local Outlier Factor. Les résultats mettent en évidence que les variables les plus discriminantes sont l'existence d'un témoin, l'ancienneté de la relation contractuelle entre l'assuré et la compagnie, le délai entre la souscription du contrat et la déclaration du sinistre, ainsi que le degré de gravité du sinistre. Les auteurs recommandent le recours à des méthodes d'apprentissage non supervisé, notamment le modèle Local Outlier Factor (LOF), afin d'identifier les comportements atypiques susceptibles de révéler une fraude. Dans la même optique, (Ben Kheder , 2021) adopte une approche d'apprentissage supervisé pour modéliser le risque de fraude à travers des algorithmes de classification tels que l'arbre de décision binaire et la forêt aléatoire. L'étude porte sur 1 183 assurés ayant souscrit une garantie « tous risques » sur la période 2015-2019. Les variables explicatives mobilisées sont d'ordre socio-démographique et contractuel (âge, profession, code postal, marque du véhicule, nature des dommages, existence d'un contrat de leasing). Les résultats indiquent qu'un assuré présentant un choc frontal, résidant dans la région du Grand Tunis, âgé de plus de 28 ans et disposant d'un véhicule non gagé présente une probabilité estimée de fraude de 87 %, ce qui illustre la capacité des modèles de classification à segmenter le portefeuille selon le niveau de risque de fraude. Par ailleurs (Hajraoui & Zahi, 2022), s'inscrivent dans une démarche hypothético-déductive visant à identifier les indicateurs les plus pertinents pour la conception d'un système automatisé de détection de fraude. À partir d'un échantillon de 1 000 observations incluant des dossiers frauduleux et non frauduleux, les auteurs appliquent successivement plusieurs modèles de régression logistique. Dans un premier temps, ils testent le modèle de Belhadji & Dionne (1994), intégrant notamment les variables relatives à une collision mineure

généralisant des coûts de réparation excessifs, à la survenance du sinistre peu après la souscription ou à l'approche de l'échéance contractuelle, à la présence de justificatifs suspects et au caractère mono-véhicule de l'accident. La modélisation retient principalement la variable relative aux coûts de réparation excessifs et celle liée au timing de survenance du sinistre par rapport à la vie du contrat. Dans un second temps, en se référant au modèle de Benedek & Laszlo (2019), ils examinent dix variables portant sur les caractéristiques du sinistre, du véhicule et de l'assuré (type d'incident, type de collision, localisation, âge et sexe de l'assuré, catégorie et marque du véhicule, nature de la garantie, etc.), dont six se révèlent statistiquement significatives. La combinaison des variables pertinentes issues des deux modèles montre que la collision mineure assortie de coûts excessifs et le type d'incident constituent les déterminants les plus explicatifs d'une suspicion de fraude.

(Kusuma, Brundhashree, Famida, Nagamani, & Nivedha, 2020) Proposent, quant à eux, une solution algorithmique alternative à la détection manuelle en mobilisant le classificateur Naïve Bayes. Leur modèle exploite principalement des variables temporelles et contractuelles telles que la date de survenance du sinistre, la date de déclaration, la date d'ouverture du dossier, la date d'effet du contrat, le montant de la prime et la fréquence des interactions entre l'assuré et l'assureur. À partir des données issues des rapports d'enquête, chaque dossier de sinistre est classé dans l'une des deux catégories : fraude ou absence de fraude, démontrant ainsi la pertinence des approches probabilistes dans le processus d'indemnisation. Les résultats montrent que l'algorithme de classification naïve bayes est une solution pour détecter les réclamations frauduleuses (Yücel, 2022) analyse pour sa part 5 589 dossiers de sinistres automobiles comprenant 25 variables relatives aux caractéristiques du véhicule (âge, valeur, couleur) et du conducteur (âge, genre, nombre de sinistres antérieurs). En mobilisant différentes techniques d'apprentissage automatique, l'étude conclut que le réseau de neurones artificiels (ANN) offre la meilleure précision dans la classification des demandes d'indemnisation frauduleuses. (Abhijeet, Rashmi, Amruta, & Nandinee, 2022) comparent plusieurs algorithmes d'apprentissage supervisé-régression logistique, gradient boosting, arbre de décision, forêt aléatoire et machines à vecteurs de support sur un échantillon de 1 000 observations comprenant 35 variables explicatives relatives au profil socio-professionnel et contractuel des assurés. Les résultats montrent que, dans le contexte étudié, la régression logistique présente la meilleure performance en matière de prédiction du caractère frauduleux des demandes d'indemnisation ainsi que de l'estimation du montant des sinistres frauduleux.

Dans le domaine de l'assurance habitation, (Kempa Severino & Yaohao, 2021) comparent neuf modèles d'apprentissage automatique à la régression logistique standard afin d'évaluer leur performance prédictive en matière de fraude. L'étude repose sur des données réelles collectées auprès d'une compagnie d'assurance brésilienne, comprenant 851 observations. Les variables explicatives incluent notamment l'écart en jours entre la

date d'effet ou d'échéance du contrat et la date de déclaration du sinistre, le capital assuré, le montant de la prime, l'âge, le sexe, la tranche de revenu ainsi que l'historique de sinistralité de l'assuré. Les résultats indiquent que la forêt aléatoire surpasse la régression logistique en termes de performance, tout en permettant d'établir un profil type des assurés fraudeurs.

Dans l'assurance santé l'étude de (Nabrawi & Alanazi, 2023) propose le développement d'un modèle prédictif capable d'identifier automatiquement la fraude à partir des réclamations d'assurance santé en Arabie Saoudite. La taille de l'échantillon a été constituée à partir de données collectées auprès de trois prestataires de soins de santé en Arabie Saoudite entre janvier 2022 et mai 2022. Le jeu de données comprenait des variables anonymisées pour toutes les instances et prestataires, permettant d'identifier les facteurs contributifs les plus significatifs avec une précision optimale. Le jeu de données, étiqueté et initialement déséquilibré, a été prétraité à l'aide de la technique SMOTE afin d'équilibrer les classes, et la sélection des variables a été effectuée à l'aide de l'algorithme Boruta, permettant d'exclure les caractéristiques non pertinentes. Trois modèles supervisés ont été appliqués : Random Forest (RF), Régression Logistique (LR) et Réseau de Neurones Artificiel (ANN). Les métriques de validation utilisées pour évaluer les performances comprenaient la précision globale (accuracy), la précision positive (precision), le rappel (recall), la sensibilité (sensitivity), la spécificité (specificity), le F1-score, ainsi que l'aire sous la courbe ROC (AUC). Les résultats montrent que le modèle Random Forest identifie comme principaux facteurs contributifs le type de police, l'âge du demandeur et le niveau d'éducation. Tous les trois modèles offrent des performances prédictives acceptables, mais l'étude recommande d'étendre l'analyse à un jeu de données plus large afin de renforcer la robustesse et la généralisation des résultats. Dans le même domaine l'étude de (Cherkaoui, Anoun, & Maizate, 2024) vise à proposer une évaluation plus précise des méthodes de détection de fraude en assurance santé en utilisant des données réelles et détaillées de demandes de remboursement. L'objectif principal est de comparer six algorithmes de classification supervisée couramment utilisés, y compris des réseaux de neurones, tout en analysant l'impact de deux méthodes de préparation des variables catégorielles. Cette étude a été réalisée sous la supervision d'experts en assurance, qui ont défini les règles d'inférence pour l'étiquetage des dossiers frauduleux et ont validé les résultats obtenus. L'article présente une étude de cas réelle au sein d'une compagnie d'assurance santé opérant dans plusieurs pays, désignée par la suite par « l'assurance ». L'objectif était de fournir aux auditeurs une solution capable d'identifier les dossiers présentant le plus fort risque de fraude ou d'abus afin de réduire le périmètre d'investigation et d'améliorer l'efficacité des contrôles. Les données utilisées correspondent à l'historique des demandes de remboursement et des décisions associées sur une période de deux ans (2018-2019) dans un pays où l'assurance traite en moyenne 17 000 demandes par mois. Chaque enregistrement correspond à une demande individuelle et comprend des variables de type catégoriel, numérique

et date telles que le code pathologie, le code d'intervention médicale, la spécialisation du prestataire de santé, la relation du patient avec l'assuré, le type de remboursement, la date de l'intervention, le montant réclamé et le montant remboursé. Le traitement des variables catégorielles dans cette étude a été limité à deux méthodes, le codage one-hot et l'entity embedding, qui n'ont toutefois pas permis d'exploiter pleinement toutes les informations pertinentes pour améliorer la précision et l'interprétabilité de la détection de fraude. Ces techniques doivent être complétées par d'autres approches afin d'enrichir les données d'entrée des algorithmes d'apprentissage automatique. Le benchmark comparatif a porté sur six méthodes de classification supervisée, à savoir la régression logistique, XGBoost, Random Forest, SVM, réseau de neurones avec codage one-hot et réseau de neurones avec entity embedding. L'implémentation a été réalisée à l'aide des bibliothèques scikit-learn, XGBoost et PyTorch. Pour comparer les performances des algorithmes sur l'ensemble de test, la métrique retenue est le score de précision moyenne basé sur la courbe précision-rappel, mieux adaptée aux tâches de détection d'anomalies.

Les résultats montrent que XGBoost fournit globalement les meilleures performances. Avec un taux positif de 20 % cet algorithme atteint un gain de 4,2, ce qui signifie que l'assurance améliore la détection des fraudes d'un facteur 4,2 et parvient à identifier 84 % des cas de fraude en se concentrant uniquement sur les 20 % des dossiers présentant les scores les plus élevés. Cette étude illustre ainsi l'efficacité des algorithmes supervisés dans la détection de fraude en assurance santé et souligne l'importance d'une préparation optimale des variables catégorielles ainsi que d'un benchmark comparatif pour sélectionner les modèles les plus performants.

Enfin L'étude menée par (Sagar & Dhanalakshimi , 2025) met en évidence que la détection des sinistres frauduleux constitue un défi stratégique pour les entreprises d'assurance, dans un contexte où les contrôles manuels et les systèmes fondés sur des règles métiers prédéfinies montrent leurs limites face à des schémas de fraude de plus en plus sophistiqués et évolutifs. Les auteurs proposent ainsi une approche fondée sur l'exploitation des données, en s'appuyant sur un jeu de données réel comprenant 1 000 dossiers de souscripteurs et 40 variables explicatives couvrant le profil sociodémographique des assurés, les caractéristiques des sinistres déclarés, les types d'incidents ainsi que les informations techniques relatives aux véhicules. La variable cible est binaire (fraud_reported : fraude déclarée / non déclarée), permettant une classification des dossiers en sinistres frauduleux ou légitimes.

Sur le plan méthodologique, l'étude mobilise deux algorithmes d'apprentissage supervisé, à savoir la Machine à Vecteurs de Support (SVM) et Extreme Gradient Boosting (XGBoost), afin de modéliser le risque de fraude. Un processus structuré de préparation des données est mis en œuvre, comprenant le traitement des valeurs manquantes, l'encodage des variables catégorielles, la normalisation des variables quantitatives et le rééquilibrage

des classes au moyen de la technique SMOTE pour corriger le déséquilibre entre sinistres frauduleux et non frauduleux. L'évaluation des performances repose sur des indicateurs standards en modélisation prédictive, notamment l'exactitude, la précision, le rappel, le score F1 et la matrice de confusion, assurant une analyse complète de la capacité discriminante des modèles.

Les résultats empiriques montrent que le modèle SVM présente une performance satisfaisante dans l'identification des sinistres non frauduleux, mais demeure peu performant dans la détection des cas frauduleux, en raison de sa sensibilité au déséquilibre des classes et de sa faible capacité à capter les signaux associés à la classe minoritaire. À l'inverse, le modèle XGBoost affiche des performances nettement supérieures, notamment en matière de rappel et de précision sur les sinistres frauduleux, traduisant une meilleure capacité de discrimination et une réduction plus efficace des faux négatifs. L'analyse de l'importance des variables révèle que des facteurs tels que le montant total réclamé, la gravité de l'incident, la disponibilité d'un rapport de police et la profession de l'assuré jouent un rôle déterminant dans l'évaluation du risque de fraude. En conclusion, XGBoost surpasse SVM sur la majorité des indicateurs de performance, en particulier dans l'identification des dossiers frauduleux relevant de la classe minoritaire, ce qui en fait un outil plus adapté à un dispositif opérationnel de gestion du risque de fraude en assurance. Toutefois, le modèle présente certaines limites : il repose exclusivement sur des données historiques structurées, nécessite un réentraînement périodique afin de s'adapter à l'évolution des techniques frauduleuses et demeure fortement dépendant de la qualité et de la fiabilité des données exploitées. Malgré ces contraintes, l'étude souligne l'intérêt de l'intégration de solutions d'apprentissage automatique dans les systèmes d'aide à la décision des assureurs afin de renforcer l'efficacité et la performance des dispositifs de lutte contre la fraude.

Tableau N°01 : Tableau récapitulatif des travaux empiriques sur la détection de la fraude en assurance

Auteurs	Titre de l'article	Année	Modèle utilisé	Résultats principaux
Aslam, Imran Unjra, Ftiti, Louhichi, & Tahira	Insurance Fraud Detection: Evidence from Artificial Intelligence and Machine Learning	2022	Régression logistique (LR), SVM, Naïve Bayes (NB)	SVM meilleure précision, LR meilleur F1-score ; Fault, Base Policy et âge de l'assuré sont determinants
Kouach, El Attar, & El Hachloufi	Auto insurance fraud detection using unsupervised learning	2022	Non supervisé : Isolation Forest et Local Outlier Factor	Variables discriminantes : témoin, ancienneté contrat, délai souscription- déclaration, gravité du sinistre Local Outlier Factor meilleur algorithme

Auteurs	Titre de l'article	Année	Modèle utilise	Résultats principaux
Ibtihel Ben Kheder	Modélisation de la fraude sur les contrats d'assurance automobile à l'aide d'une approche d'apprentissage automatique	2021	Arbre de décision binaire, Forêt aléatoire	Assuré avec choc frontal, âge >28 ans, véhicule non gagé : probabilité de fraude estimée 87 %
Hajraoui & Zahi	Indicateurs pertinents de détection automatique de fraude : Cas des compagnies d'assurance automobile	2022	Démarche hypothético-déductive Régressions logistiques (Belhadji & Dionne, Benedek & Laszlo)	Collision mineure avec coûts excessifs et type d'incident : principaux déterminants de suspicion de fraude
Kusuma et al.	Detecting Fraud Vehicle Insurance Claims using Machine Learning Technique	2020	Naïve Bayes	Naïve bayes est une solution pour détecter les réclamations frauduleuses Variables temporelles et contractuelles pertinentes ; approche probabiliste efficace pour classification fraude / non fraude
Ahmet Yücel	A novel data processing approach to detect fraudulent insurance claims for physical damage to cars	2022	Réseau de neurones artificiel (ANN)	ANN offre la meilleure précision pour la classification des demandes frauduleuses
Abhijeet et al.	Fraud detection and analysis for insurance claim using machine learning	2022	Régression logistique, Gradient Boosting, Arbre de décision, Forêt aléatoire, SVM	Régression logistique meilleure performance pour prédiction de fraude et estimation des montants

Revue de la littérature

Matheus Kempa Severino & Peng	Machine learning algorithms for fraud prediction in property insurance: Empirical evidence using real-world microdata	2021	Régression logistique, 9 modèles ML	Forêt aléatoire meilleure performance ; profil type des assurés fraudeurs identifié
Nabrawi & Alanazi	Fraud detection in healthcare insurance claims using machine learning	2023	Random Forest, Régression logistique, ANN	Random Forest identifie type de police, âge, niveau d'éducation ; tous les modèles offrent performances acceptables
Cherkaoui, Anoun, & Maizate	A benchmark of health insurance fraud detection using machine learning techniques	2024	Régression logistique, XGBoost, Random Forest, SVM, ANN (One-hot et Entity Embedding)	XGBoost meilleures performances ; identification de 20 % des dossiers les plus risqués couvre 84 % des fraudes
Sagar & Dhanalakshmi	Insurance Fraud Detection Using Machine	2025	SVM, XGBoost	XGBoost surpasse SVM, meilleure détection de dossiers frauduleux, importance du montant réclamé, gravité, rapport de police et profession
Sagar & Dhanalakshmi	Fraud Detection in Insurance Claims using Data-driven Analytics	2025	SVM, XGBoost	XGBoost surpasse SVM, meilleure détection de dossiers frauduleux, importance du montant réclamé, gravité, rapport de police et profession

Source: Établi par nos soins

5. Discussions et analyses:

Les résultats des différentes études montrent une certaine convergence sur les variables les plus discriminantes pour la détection de fraude. Par exemple, dans l'assurance automobile, des facteurs tels que l'âge de l'assuré, la responsabilité (Fault) et la nature du contrat (Base Policy) apparaissent systématiquement comme des indicateurs

significatifs. De même, dans l'assurance santé, le type de police, le montant réclamé et les caractéristiques du prestataire ou du patient jouent un rôle déterminant. Ces éléments confirment l'importance de l'intégration de variables à la fois socio-démographiques, contractuelles et transactionnelles pour améliorer la précision des modèles.

Par ailleurs, l'efficacité relative des modèles varie selon les métriques utilisées. Alors que le SVM se montre performant en précision globale, des modèles comme Random Forest ou XGBoost obtiennent de meilleurs résultats sur le rappel et la détection de la classe minoritaire, ce qui est crucial dans les contextes de fraude où les dossiers frauduleux représentent une faible proportion du total. Cela souligne l'importance d'un benchmark comparatif et de l'utilisation de métriques adaptées pour évaluer les modèles, notamment dans les situations de déséquilibre des classes.

Enfin, certaines études soulignent les limites des techniques actuelles, notamment la dépendance à la qualité des données, le besoin de prétraitements sophistiqués pour les variables catégorielles, et la nécessité d'un réentraînement périodique pour suivre l'évolution des schémas de fraude. Ces observations suggèrent que, bien que les modèles d'apprentissage automatique offrent des performances prometteuses, leur mise en œuvre opérationnelle requiert une intégration dans un cadre décisionnel plus large, combinant expertise humaine, règles métier et surveillance continue.

En conclusion, la littérature montre que l'adoption de modèles prédictifs basés sur l'apprentissage automatique permet de renforcer la détection de fraude en assurance, mais leur efficacité dépend fortement de la qualité des données, de la sélection des variables et du choix du modèle en fonction des métriques pertinentes. Les recherches futures pourraient explorer l'intégration de méthodes hybrides, combinant apprentissage supervisé et non supervisé, ainsi que l'exploitation de données externes pour améliorer la robustesse et la généralisabilité des modèles.

6.Conclusion :

La détection des fraudes en assurance est devenue une nécessité, compte tenu de l'augmentation des pratiques frauduleuses dans le monde et de leur impact sur la solvabilité des compagnies d'assurance. Les méthodes traditionnelles sont devenues lourdes et chronophages, ce qui a conduit de nombreux chercheurs à se tourner vers des approches plus sophistiquées.

Dans cet article, nous présentons un aperçu général de la fraude en assurance, ainsi que de l'intelligence artificielle et de ses outils, en mettant l'accent sur le machine learning et ses principaux algorithmes. Nous proposons

également une revue de littérature sur la détection de fraude dans les compagnies d'assurance à l'aide de l'IA. Cette dernière montre que l'utilisation de ces technologies a profondément transformé la manière dont les compagnies d'assurances et les institutions financières luttent contre ce fléau, en offrant des outils capables de détecter rapidement et efficacement les activités frauduleuses, réduisant ainsi les pertes financières et protégeant la réputation des organisations.

L'une des principales forces de l'IA réside dans sa capacité à analyser de grandes quantités de données en temps réel, ce qui permet d'identifier des schémas et des anomalies qui échapperaient à une surveillance humaine. Les algorithmes d'apprentissage automatique s'adaptent et évoluent pour repérer les nouvelles techniques de fraude à mesure qu'elles apparaissent. De plus, l'IA améliore la précision de la détection en réduisant le nombre de faux positifs, permettant aux équipes de sécurité de se concentrer sur les cas les plus critiques. Elle offre également la possibilité de détecter des fraudes complexes et sophistiquées.

Cependant, l'IA n'est pas une solution miracle. Elle nécessite une mise en œuvre et une maintenance rigoureuse, ainsi qu'une formation continue pour rester efficace face à l'évolution des schémas de fraude. La protection de la vie privée et la sécurité des données des clients constituent également des enjeux majeurs. Ainsi, la combinaison de l'expertise humaine avec les capacités de l'IA constitue la meilleure approche pour établir une défense solide contre la fraude, permettant aux entreprises et aux institutions financières de protéger plus efficacement leurs actifs et leurs clients.

L'ensemble des études examinées souligne l'importance de l'IA dans la lutte contre la fraude au sein des compagnies d'assurance. Néanmoins, elles mettent en évidence certains défis pour obtenir des résultats fiables et satisfaisants, tels que le manque de données adaptées, la taille limitée des échantillons, les ressources insuffisantes, le déséquilibre des classes et les contraintes liées à la confidentialité.

7. Références

1. Abhijeet, U., Rashmi, B., Amruta, K., & Nandinee, M. (2022). Fraud detection and analysis for insurance claim using machine learning. *IEEE*.
2. Anuj, S., & Panigrahi, P. (2012). A Review of Financial Accounting Fraud Detection based on Data Mining Techniques. *International Journal of Computer Applications*, pp. 37-38.
3. Aslam, F., Imran unjra, A., Ftiti, Z., Louhichi, W., & Tahira, S. (2022). Insurance Fraud Detection: Evidence from Artificial Intelligence and Machine Learning. *Research in international business and finance*, 62, p. 14.

4. Ayushi , C., & Preeti , G. (2019). Machine Learning and Deep Learning. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(12), p. 1.
5. Baali, A. (2024). La fraude à l'assurance : État des lieux - Cas de l'Algérie. *Revue Dirassat Iqtissadiya*, 11(1), pp. 257-258.
6. Ben Kheder , I. (2021). Modélisation de la fraude sur les contrats d'assurance automobile à l'aide d'une approche d'apprentissage automatique. *Cahier d'IFID N°03*.
7. Breiman, L. (2001). Random Forests. *Statistics Department, University of California, Berkeley*, 45(1), p. 1.
8. Buari, A.-M. (2019). *Provisionnement individuel en assurance non vie dans le cas de garanties à développements courts, modélisation non paramétrique par des méthodes de machine Learning*. Mémoire en vue de l'obtention de diplôme d'un actuaire ,univeersité de bretagne occidental.
9. Cherkaoui, O., Anoun, H., & Maizate, A. (2024). A benchmark of health insurance fraud detection using machine learning techniques. *IAES International journal of artificiale intelligence*, 13(2).
10. Choi, R., Coyne, A., Kalpathy-Cramer, J., Chiang, M., & Campbell, J. (2020). Introduction to Machine Learning, Neural Networks and Deep Learning. *Translational vision science & technology*, 9(2).
11. Fouzi, M., & Sadik , A. (2024). Avantages et risques de l'intelligence artificielle sur la recherche scientifique en sciences humaines et sociales. *Revue echos jeunes chercheurs*, 1(2), p. 11.
12. Gallet , O. (2014). Halte au fraudes. *Guide pour Mangers et auditeurs*, 04.
13. Haenlein, M., & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the past, presnt and future of Artificial Intelligence. *California Management Review*, 61(4).
14. Hajraoui, G., & Zahi, J. (2022). Indicateurs pertinents de détection automatique de fraude : Cas des compagnies d'assurance automobile. *Alternatives managériales et économiques*, 4.
15. Jonathan, K. (2016). La détection de fraudes à l'assurance. *Mémoire présenté devant l'ENSAE Paris Tech pour l'obtention du diplôme de la filière Actuariat et l'admission à l'Institut des Actuaire*, 7-8.
16. Kempa Severino, M., & Yaohao , P. (2021). Machine learning algorithms for fraud prediction in property insurance: Empirical evidence using real-world microdata. *Machine Learning with Applications*, 5.
17. Kouach, Y., El attar, A., & El hachloufi, M. (2022). Auto insurance fraud detection using unsupervised learning. *Alternatives Managériales Economiques*, 4.
18. Kusuma, Brundhashree, U., Famida, Nagamani, K., & Nivedha, S. (2020). Detecting Fraud Vehicle Insurance Claims using Machine Learning Technique. *IJESC*, 10(6).
19. Maniati , & al. (2022). A Neural Network approach for integrating banks' decision in shipping finance. *Cogent Economics & Finance*, 10, p. 4.

20. Mathivet, V. (2014). *L'Intelligence Artificielle pour les développeur : concepts et implémentations en C#*. ENI Editions.
21. Mirjana , S., Aminata , A., & Garba, N. (2021). Emerging technology trends: Artificial intelligence and big data for development 4.0. *International Telecommunication Union*.
22. Nabrawi , E., & Alanazi, A. (2023). Fraud detection in healthcare insurance claims using maching learning. *Risks*, 11(9).
23. Ranjitha, H., Joshni , P., & Vaishnavi , C. (2023). Vehicle Insurance Fraud Detection Using Machine Learning. *International Journal of Advanced Scientific Innovation*, 5, p. 2.
24. Sachin , B., Vinayak , P., & Zameer , M. (2020). Advantages and disadvantages of Artificial Intellegence . *National seminar on "Trends in geography, commerce, IT and sustainable development"*. I.C.S. College Khed .
25. Sagar , A., & Dhanalakshimi , M. (2025). Insurance Fraud Detection Using Machine. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 13, p. 4.
26. Sarker, I. (2021). Machine learning: algorithms, real-world applications and research directions. *SN Computer Science*, 2(3), p. 4.
27. Titilope Azeeza, F., & Olowokudejob, F. (2025). Examining Motor Insurance Claims Fraud Detection and Prevention: Evidence of the Nigerian Insurance Companies. *AYBU Business*, 5(2), p. 22.
28. Yücel, A. (2022). A novel data processing approach to detect fraudulent insurance claims for physical damage to cars. *Journal of new results in science*, 11(2).
29. Zhiyuan, H., Danchen, L., & Thomas , L. (2019). Gradient Boosting Machine: A Survey. *Eprint ArXiv:1908.0695*, p. 1.