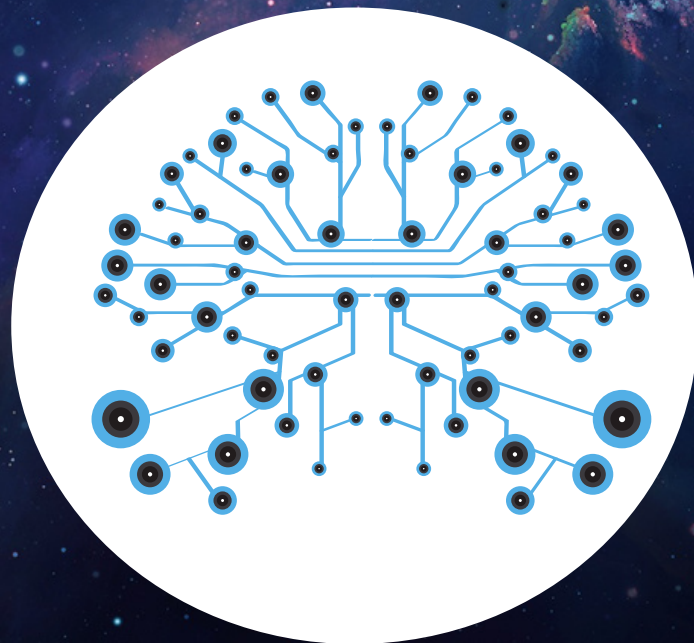




L'intelligence artificielle et la détection de fraude en assurance



SOMMAIRE

01.

Introduction

02.

Analyse et impact du phénomène de fraude en assurance

03.

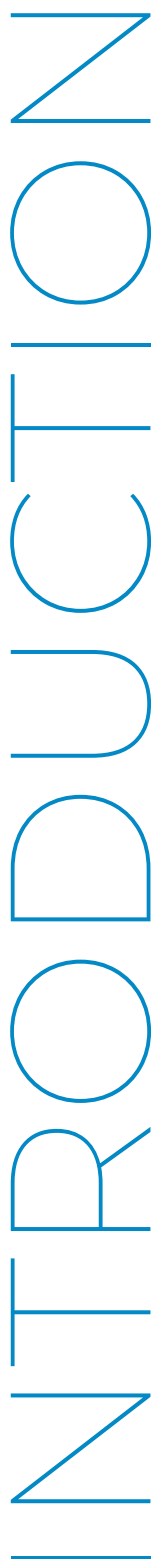
Méthodes de détection de fraude en assurance

04.

Usage du Machine Learning dans la détection de fraude en assurance

05.

Application du Machine Learning dans la détection de fraude en assurance



Les modèles de l'intelligence artificielle, peuvent-ils contribuer à la détection de fraude en assurance ?

Le phénomène de la fraude se définit par tout acte malhonnête intentionnel afin d'obtenir un avantage matériel ou moral. Cette pratique concerne tous les secteurs et les activités économiques avec des formes et types différents. Entre autres, la fraude comptable, cybercriminalité, la fraude de facture, la fraude fiscale, passant de la fraude bancaire jusqu'à la fraude dans le secteur de l'assurance.

En effet, l'assurance n'est pas à l'abri du fléau de la fraude qui est par définition complexe à contrôler. Les actes frauduleux représentent une entrave à la croissance des compagnies d'assurance étant donné qu'ils accroissent la charge des sinistres d'où une dégradation du ratio combiné ce qui provoque des pertes non négligeables en termes de rentabilité dans les différentes branches d'assurance notamment des biens. D'ailleurs, il y a eu environ **13 milliards** d'euros de réclamations frauduleuses en Europe en 2017 dont **2,5 milliards** d'euros de réclamations frauduleuses ont été détectés, en revanche 80% des fraudes sont non détectées selon *Insurance Europe*. Pour l'année 2020 on peut estimer approximativement le montant des réclamations frauduleuses en Europe à hauteur de **14,5 milliards** d'euros. Et c'est près d'un demi-milliard d'euros des fraudes identifiées en IARD en 2018 en France selon l'ALFA (l'Agence de Lutte contre la Fraude à l'assurance).

En vue de limiter l'impact financier négatif de la fraude, la détection des tentatives de fraude de manière permanente devient un sujet stratégique pour les compagnies d'assurance. Cependant, les techniques classiques utilisées montrent leurs insuffisances telles que l'analyse des dossiers par les gestionnaires des sinistres accompagnée par des risques opérationnels et la contrainte de temps. D'où, la nécessité d'élaborer des systèmes de détection de fraude plus sophistiqués et plus efficaces en faisant recours à des nouvelles technologies précisément l'**Intelligence Artificielle (IA)**. En effet, les assureurs peuvent se prémunir de la perte financière en exploitant le potentiel incontournable du **Machine Learning (ML)** à détecter des anomalies automatiquement avec un taux de précision important.

02. ANALYSE ET IMPACT DU PHÉNOMÈNE DE FRAUDE EN ASSURANCE

La diversité des catégories de fraude rend difficile de contrôler les actes frauduleux d'un point de vue légal ainsi que technique.

La fraude dans le secteur de l'assurance est tout acte exprimant la mauvaise foi de l'assuré qui cherche avec des moyens illégaux de s'enrichir sans cause en exécutant les garanties du contrat d'assurance. La fraude dans le secteur d'assurance concerne plusieurs produits particulièrement l'assurance non-vie et santé à différents niveaux :

La fraude à la souscription du contrat d'assurance :

L'assurée fournit des informations erronées ou incomplètes lors de la demande d'assurance dans le but d'avoir une prime d'assurance moins chère. En plus des fausses informations, il y a le cas de fraude d'identité en se présentant avec une identité falsifiée.

La fraude à la réclamation du sinistre :

Ce type de fraude concerne :

- L'exagération du montant du sinistre,
- La déclaration d'un sinistre inexistant dans la réalité,
- La provocation intentionnelle d'un sinistre,
- La déclaration d'un sinistre survenu avant la date d'effet du contrat d'assurance,
- La modification des circonstances d'un sinistre.

La fraude de multi-assurance :

Ce type de fraude consiste à couvrir un même risque auprès de plusieurs assureurs pour réaliser un profit additionnel lors d'un sinistre.



Article L113-8 Code des assurances France

"... Le contrat d'assurance est nul en cas de réticence ou de fausse déclaration intentionnelle de la part de l'assuré, quand cette réticence ou cette fausse déclaration change l'objet du risque ou en diminue l'opinion pour l'assureur, alors même que le risque omis ou dénaturé par l'assuré a été sans influence sur le sinistre."

Article L172-8 Code des assurances France

"... Les assurances cumulatives pour une somme totale supérieure à la valeur de la chose assurée sont nulles si elles ont été contractées dans une intention de fraude."

02. ANALYSE ET IMPACT DU PHÉNOMÈNE DE FRAUDE EN ASSURANCE

Tarifification



L'impact de la fraude se révèle au niveau de la tarification suite à l'ajustement des tarifs.

Afin d'équilibrer la charge sinistre, l'assureur ajuste le tarif à la hausse pour tenir en compte des sinistres frauduleusement déclarés.

De ce fait, la détection de la fraude permet d'établir des tarifs plus ajustés aux risques des assurés de bonne foi et par conséquent de proposer des tarifs plus compétitifs.

Provision



En pratique, l'indemnisation des sinistres avec des montants plus élevés et des sinistres non existants exige la constitution des dotations de provisionnement supérieures au provisionnement sans fraude. Dès lors, la détection de la fraude permet alors de réévaluer les provisions à la juste valeur.

Généralement, c'est les Provisions pour Sinistres à Payer ou PSAP selon le jargon assurance qui sont concernées.

Solvabilité II



Au niveau réglementaire de solvabilité, les assureurs sont invités à calculer les exigences en capital sous forme de **SCR** (Solvency Capital Requirement) et **MCR** (Minimum Capital Requirement). Puisque la détection des fraudes permet l'ajustement des tarifs et la hausse des provisions (BEL). Ceci permettra à tort d'impacter les mesures de risques, ainsi qu'apporter une fausse information sur la solvabilité de la compagnie d'assurance en question. En cas de fraude massive dans un portefeuille d'assurance, le Ratio de solvabilité pourrait être impacté à la baisse.

"... Lorsqu'un contrat d'assurance a été consenti pour une somme supérieure à la valeur de la chose assurée, s'il y a eu dol ou fraude de l'une des parties, l'autre partie peut en demander la nullité et réclamer, en outre, des dommages et intérêts."



Article L121-3
Code des assurances
France

03. MÉTHODES DE DÉTECTION DE FRAUDE EN ASSURANCE

Gestionnaires de sinistres



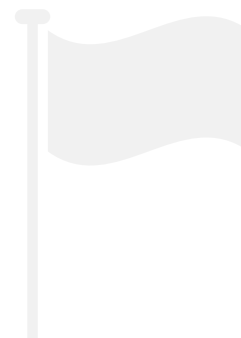
Pour faire face aux fraudes, les compagnies d'assurance font recours aux gestionnaires de sinistres. En réalité, les gestionnaires de sinistres peuvent identifier des cas douteux en analysant les éléments du sinistre et en appréciant le comportement de l'assuré lors de la déclaration du sinistre. Dès l'identification d'une situation anormale, le gestionnaire des sinistres signale ledit sinistre dans la perspective d'entamer des investigations.



Pourtant, cette méthode prend un temps important pour approfondir l'analyse ce qui impact la gestion dans sa globalité. Ainsi, le fléau de la fraude est plus complexe qu'une évaluation basée sur l'intuition humaine.

Drapeau rouge

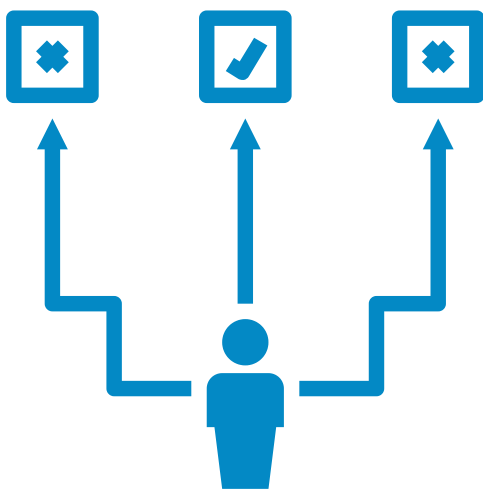
Les compagnies d'assurance utilisent ainsi un processus « drapeau rouge ». C'est un processus de détection de fraude à travers l'audit. Le principe de ce processus se base sur l'implémentation de certains indicateurs pour identifier si le sinistre est suspicieux ou non. Si la réclamation du sinistre est positive en réponse aux indicateurs choisis, une enquête approfondie s'ouvre à l'encontre de ce sinistre. En d'autres termes, si la réclamation reçoit un ou des drapeaux rouges la réclamation fait l'objet d'une enquête.



Cependant, cette méthode a l'inconvénient que les fraudeurs peuvent utiliser ces drapeaux pour contourner l'audit. En plus, certains cas de fraude peuvent s'échapper de la détection lors du processus d'audit étant donné que la méthode d'échantillonnage est fréquemment utilisée par l'audit.

03. MÉTHODES DE DÉTECTION DE FRAUDE EN ASSURANCE

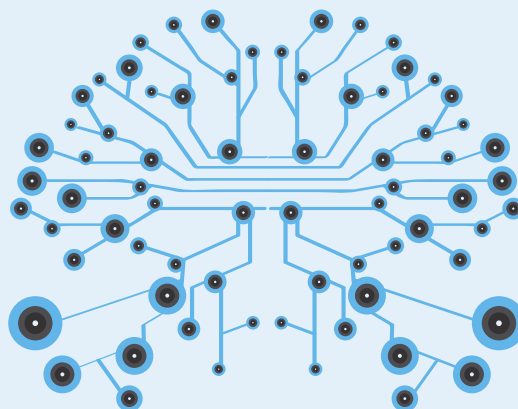
Règle décisionnelle



Il existe une approche distincte dite "règle décisionnelle" qui regroupe un ensemble de règles binaires. Ces règles sont définies pour chaque garantie de l'assurance. Si le sinistre vérifie une de ces règles, il est évalué comme un sinistre probablement frauduleux. À ce titre une investigation approfondie sera entamée concernant ledit sinistre.

Certainement, ce modèle est simple à mettre en place du fait qu'il ne nécessite pas une modélisation complexe. Mais il faut déterminer des centaines de règles pour les différentes garanties. Ainsi, il est dans l'incapacité de capter la complexité du fléau de fraude ce qui conduit à un modèle avec une faible spécificité d'où une observation approfondie est nécessaire.

La conception de méthodes basées sur l'intelligence artificielle est-elle une alternative pour la détection de fraude ?



Les méthodes classiques citées ci-dessus présentent des insuffisances, d'où une réflexion pour concevoir des méthodes plus sophistiquées et plus robustes est nécessaire. Le recours aux modèles automatiques basés sur l'Intelligence artificielle en l'occurrence le Machine Learning semble être d'une grande valeur ajoutée dans la détection de fraude au secteur de l'assurance.

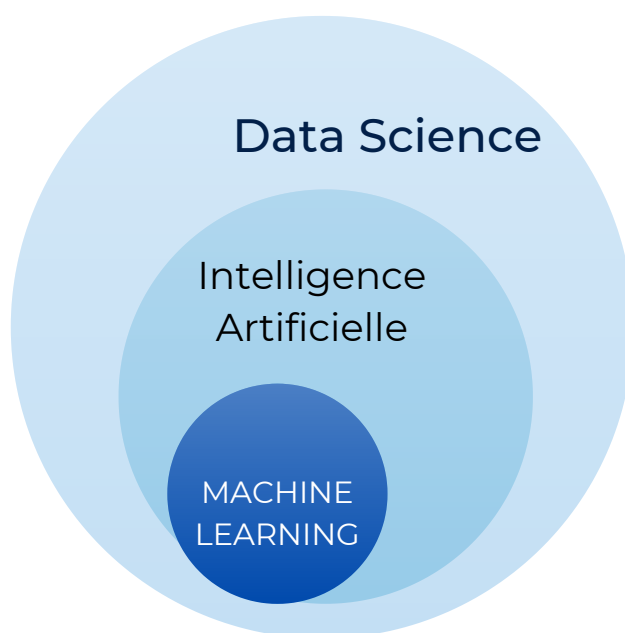
04. USAGE DU MACHINE LEARNING DANS LA DÉTECTION DE FRAUDE EN ASSURANCE

L'utilisation de Machine Learning devient de plus en plus présente dans tous les domaines y compris le secteur financier

Le machine learning (ML) représente l'ensemble des techniques permettant l'implémentation automatique des modèles afin de prédire des outputs à partir des inputs donnés. Cette technique est plébiscitée par l'industrie financière ces dernières années de par son efficacité à prédire.

La principale caractéristique du ML est l'apprentissage statistique. En effet, l'apprentissage automatique est la science permettant à la machine de prendre des décisions uniquement à partir des bases de données gigantesques sans lui spécifier ces règles de décisions au préalable.

L'usage des techniques de Machine Learning pour la détection de fraude repose sur la logique de détecter les points aberrants, atypiques et différents de l'ensemble des données traités.



04. USAGE DU MACHINE LEARNING DANS LA DÉTECTION DE FRAUDE EN ASSURANCE

Le Machine Learning englobe **4 techniques** qui diffèrent selon le type des données utilisées



L'apprentissage
supervisé

L'algorithme supervisé essaie de déduire une fonction ou des instructions permettant de prédire de nouvelles valeurs de sorties en corrigeant à chaque fois les valeurs prédites par rapport à des valeurs cibles. Ces algorithmes sont appliqués à des données libellées, dont nous connaissons les variables d'entrée et de sortie.



L'apprentissage
semi-supervisé

Ces algorithmes utilisent à la fois des données étiquetées et non-étiquetées pour ajuster le fonctionnement de classificateur avec des nouvelles données. Le transfert d'apprentissage est défini comme étant la capacité d'utiliser des connaissances et des compétences apprises dans des activités précédentes pour réaliser une nouvelle activité.



L'apprentissage par
renforcement

L'algorithme est en interaction permanente avec son environnement sans aucune information sur les actions à suivre. Le système essaie de trouver alors les actions à faire pour obtenir des récompenses. La récompense soit positive si l'action choisie est bonne, soit négative dans le cas contraire.



L'apprentissage
non supervisé

Les algorithmes sont appliqués à des données non libellées, il y a que des variables d'entrées sans réponses. L'algorithme non supervisé tente à trouver lui-même un modèle dans les données qui décrit mieux la structure des données sans aucune relation avec une variable cible. En effet, ce modèle permet de définir de classes des données homogènes non connues au préalable.

04. USAGE DU MACHINE LEARNING DANS LA DÉTECTION DE FRAUDE EN ASSURANCE

Dans le contexte de la détection de fraude en assurance, les assureurs principalement ne disposent pas des bases de données des sinistres étiquetées frauduleux. Ainsi, la nature du problème de détection de fraude est assimilée au problème de détection d'anomalie des points aberrants éloignant de la moyenne. De ce fait, un recours aux méthodes d'apprentissage non supervisé pour la détection de la fraude pourrait être pertinent.

Malgré la simplicité opérationnelle de l'apprentissage non supervisé et sa capacité à détecter des nouveaux types de fraudes, la validation de modèle reste compliquée en raison d'absence de la variable à expliquer. Pour pallier cet inconvénient, nous proposons d'effectuer du backtesting.

Dans le cas où les compagnies d'assurance disposent d'un historique des sinistres frauduleux, il est recommandé d'appliquer l'apprentissage supervisé. Cependant, il faut veiller à la taille de base des données, puisque la rareté des observations objectives (les fraudes) a un impact sur la performance des modèles. Pour remédier à ce problème des techniques tel que SMOTE peut rééquilibrer les observations des sinistres frauduleux et non frauduleux.



Parmi les méthodes de l'apprentissage non supervisé adaptées à la détection des valeurs aberrantes et par conséquent adaptées à la détection de fraude en assurance :

- Isolation Forest,
- Local Outlier Factor,
- Unsupervised Random Forest,
- Autoencoders
- k-Nearest Neighbors.
- SVM à une classe

04. USAGE DU MACHINE LEARNING DANS LA DÉTECTION DE FRAUDE EN ASSURANCE

L'utilisation de Machine Learning pour la détection de fraude présente un nombre important de bénéfices

1

Les modèles élaborés sont adaptables à toute nouvelle donnée, donc ils peuvent détecter des structures de fraude non identifiées par l'algorithme à partir des données structurées ou non,

2

Les algorithmes de ML peuvent exploiter des bases de données colossales ce qui permet son amélioration au fur et à mesure de l'élargissement de la base de données.

Ils sont aussi plus rapides dans l'évaluation automatique des sinistres en temps réel,

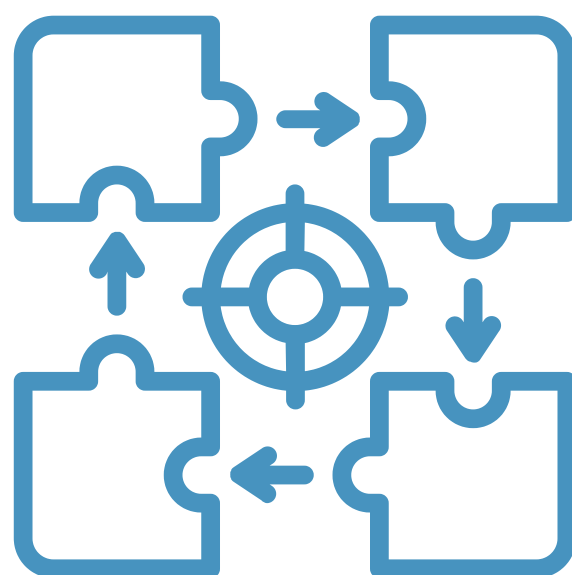
3

Dans le cas d'utilisation du ML, il n'y a qu'un seul coût à couvrir qui est celui des machines utilisées. Ainsi, cette solution semble moins chère par rapport à la détection des fraudes par des analystes.

04. USAGE DU MACHINE LEARNING DANS LA DÉTECTION DE FRAUDE EN ASSURANCE

Les méthodes supervisées

- Les machines à vecteurs de support,
- Les réseaux de neurones,
- AdaBoost,
- La régression logistique
- Naive Bayes,
- Les arbres de décision,
- Les forêts aléatoires,
- K plus proches voisins,...



Méthodes non supervisées

- Le partitionnement k-means,
- La classification dbscan,
- Les cartes auto-organisatrices,
- L'analyse en composantes principales (ACP).
- Les machines à vecteurs de support
- Partitionnement hiérarchique

04. USAGE DU MACHINE LEARNING DANS LA DÉTECTION DE FRAUDE EN ASSURANCE

Différents modèles du ML sont disponibles, mais quelques méthodes semblent plus efficaces

DECESION TREE

L'arbre de décision est une technique d'apprentissage supervisé qui peut être utilisée à la fois pour les problèmes de classification et de régression. Il s'agit d'un classificateur arborescent, où les nœuds internes représentent les caractéristiques d'un ensemble de données, les branches représentent les règles de décision et chaque feuille représente le résultat.

K NEAR NEIGHBORS

K-Nearest Neighbor est un algorithme supervisé le plus simple qui suppose la similitude entre le nouveau cas et les cas disponibles et place le nouveau cas dans la catégorie la plus similaire aux catégories disponibles.

ISOLATION FOREST

Cet algorithme non supervisé permet la détection des anomalies en isolant les données atypiques qui sont trop différentes de la plupart des autres données. Cet algorithme calcule pour chaque donnée un score d'anomalie qui reflète à quel point la donnée en question est atypique.

SVM ONE CLASS

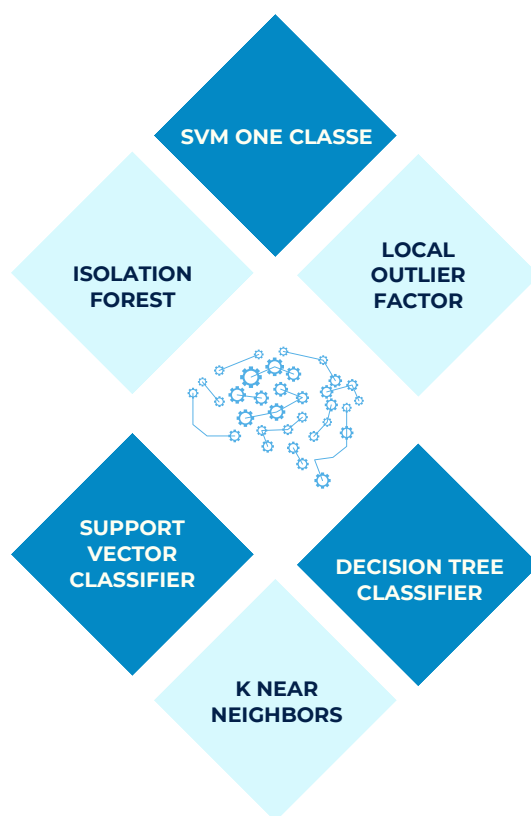
Une approche non supervisée qui est adaptée à la détection de nouveauté en classant les nouvelles données comme similaires ou différentes de l'ensemble d'entraînement en une seule catégorie.

05. APPLICATION DU MACHINE LEARNING DANS LA DÉTECTION DE FRAUDE EN ASSURANCE

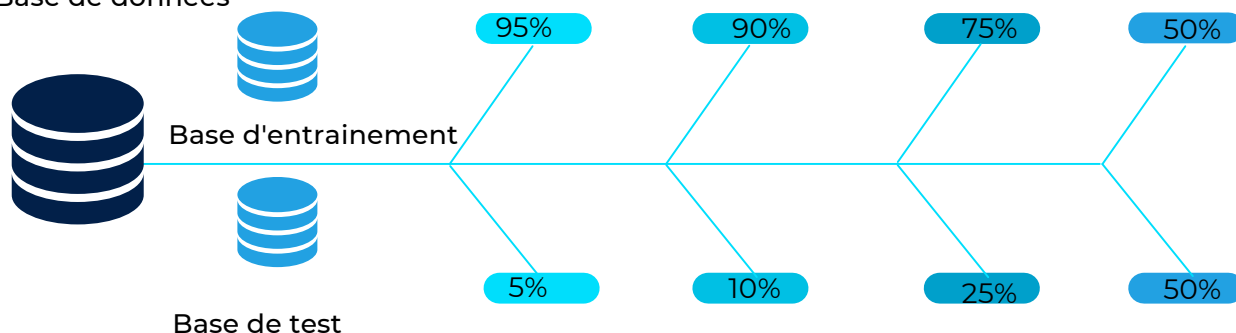
Afin de tester la performance d'utilisation de l'intelligence artificielle dans la détection de fraude en assurance, nous avons choisi six algorithmes de Machine Learning. D'une part, trois algorithmes appartiennent aux techniques d'apprentissage supervisé à savoir le classificateur d'arbre de décision (Decision Tree Classifier), classificateur de vecteur de support (Support Vector Classifier) et K les plus proches voisins (K Near Neighbors).

D'autre part, trois algorithmes appartiennent à la famille d'apprentissage non supervisé notamment Forêt d'Isolément (Isolation Forest), facteur de valeur aberrante locale (Local Outlier Factor) et Machine à Vecteur de Support une Classe (Support Vector Machine one Classe).

Nous avons appliqué ces méthodes sur une base de données des réclamations des sinistres en Open Data. Cette base se constitue de 1000 observations et 39 colonnes dont 38 variables explicatives et une variable à expliquer en l'occurrence « fraude reportée ».



Base de données

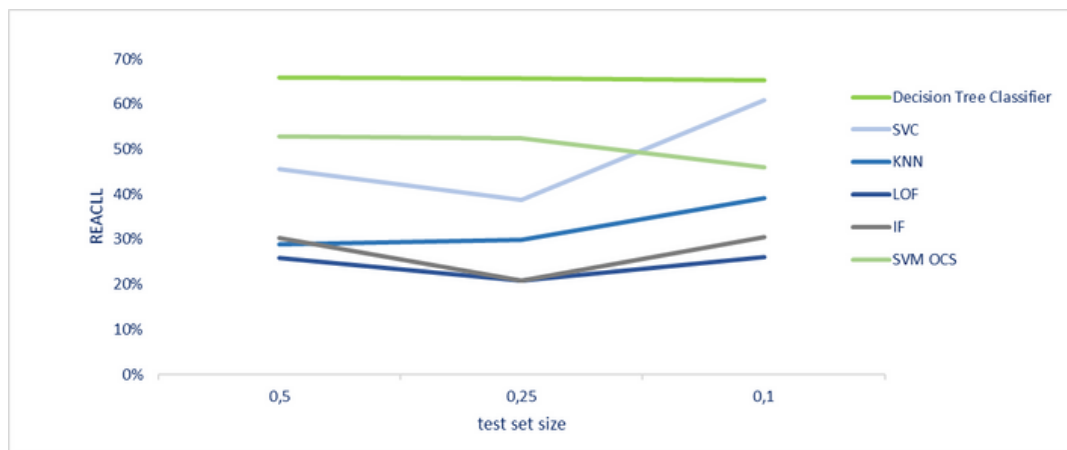


En plus de la comparaison des méthodes appliquées entre elles selon les mesures de performance, nous avons testé ces algorithmes sur des bases d'entraînement et de test de tailles différentes (5%, 10%, 25%, 50% de la base de données réservée au test).

05. APPLICATION DU MACHINE LEARNING DANS LA DÉTECTION DE FRAUDE EN ASSURANCE

Les premiers résultats de l'application des méthodes du ML sur un set de 1000 observations montrent que les algorithmes atteignent une meilleure performance lorsque 95% des données sont réservées à l'entraînement de l'algorithme.

Ainsi, le Classificateur d'arbre de décision et Classificateur de vecteur de support atteignent leurs meilleures performances à 95% des données entraînées et à 90% pour K les plus proches voisins.



L'étude montre aussi que l'algorithme SVM à une Classe a surmonté tous les algorithmes supervisés et non supervisés instaurés pour la détection de fraude selon la mesure de performance Sensibilité (Rappel ou Recall) qui indique le nombre de prévisions positives correctes sur le nombre total des observations positives réelles. Cet algorithme a atteint **70%** suivie par l'arbre de décision avec une Sensibilité à hauteur de **67%** et avec une Sensibilité de 66% pour le Support Vector Classifier.

En revanche, en queue de classement se trouve les K near Neighbors, LOF et IF, qui ont montré des mauvaises performances pour la détection des réclamations des sinistres frauduleuses avec une Sensibilité de **39%**, **26%** et **9%** respectivement. Parmi les raisons pour lesquelles ces dernières méthodes n'ont pas atteint leur meilleure performance, c'est la limitation relative à la base des données d'entraînement.

SVM One Class



Decision Tree



SVC



KNN



LOF



IF



05. APPLICATION DU MACHINE LEARNING DANS LA DÉTECTION DE FRAUDE EN ASSURANCE

Les mesures de performance

Mesures de performance	Recall	Accuracy	Precision	Spécificité	F1_score	Taux d'erreur	Test data size
SVM one Classe	0.7	0.53	0.81	0.46	0.52	0.47	0.5
Decision Tree Classifier	0.67	0.82	0.89	0.86	0.76	0.18	0.05
Support Vector Classifier	0.66	0.82	0.89	0.87	0.76	0.18	0.05
K near Neighbors	0.39	0.67	0.81	0.75	0.57	0.33	0.1
Local Outlier Factor	0.26	0.59	0.76	0.69	0.47	0.41	0.1
Isolation Forest	0.09	0.6	0.73	0.75	0.42	0.40	0.1

Dans l'ensemble des algorithmes de l'apprentissage non supervisé appliqués, SVM à une Classe arrive à prédire correctement 70% (Sensibilité égale à 0.7) des cas frauduleux en d'autres termes dans les 132 cas qui sont réellement frauduleux de la base de test, le modèle détecte 93 cas frauduleux. Ainsi, parmi les 210 cas qui sont détectés comme non frauduleux, le modèle a correctement prédit 81%. Ce modèle a un taux d'erreur à concurrence de 47% autrement dit, dans 500 observations de la base de test 236 sont incorrectement prédits.

Il a été observé dans cette application que parmi les méthodes de l'apprentissage supervisé, l'Arbre de Décision qui rapporte les meilleures performances. En effet, cet algorithme arrive à prédire 67% des cas frauduleux réels, cela veut dire que sur 12 cas réellement frauduleux il a détecté 8 cas correctement. Et parmi 37 cas prédits comme non frauduleux il a identifié correctement 33 cas non frauduleux avec une précision de 89%. De même, l'algorithme arbre de décision atteint le taux d'erreur le plus faible parmi tous les autres modèles concurrents. Ce taux se limite à 18%, alors parmi les 50 observations de la base test, il a repéré 9 cas incorrectement dont 5 cas comme frauduleux et 4 cas comme non frauduleux.



05. APPLICATION DU MACHINE LEARNING DANS LA DÉTECTION DE FRAUDE EN ASSURANCE

Dans l'ère de la technologie et de l'innovation les assurés de mauvaise foi cherchent toujours des nouvelles méthodes pour s'enrichir en activant les garanties de leurs contrats d'assurance. De ce fait les compagnies d'assurances sont dans l'obligation ainsi de suivre ce rythme d'innovation technologique afin de tirer parti des avantages d'utilisation des nouvelles techniques notamment l'intelligence artificielle.

La présente étude propose une nouvelle manière de procéder à l'identification des fraudes en ayant recours à des méthodes de Machine Learning. Notre application basée sur une comparaison de ces modèles supervisés et non supervisés a montré la possibilité d'identifier des réclamations des sinistres frauduleux par l'algorithme non supervisé SVM one Classe et l'algorithme supervisé Arbre de Décision à hauteur de 70% d'efficacité.

L'entrave à une évaluation parfaite des performances des modèles de détection de fraude en assurance issus de l'Intelligence Artificielle est la taille réduite de la base de données ce qui ne permet pas aux modèles mieux apprendre à partir de données passées. Ainsi, il est important la considération de la taille de base de données d'apprentissage.

Les compagnies d'assurance peuvent ainsi, utiliser un mixte des méthodes classiques de détection de fraude avec les méthodes basées sur l'IA.

GLOSSAIRE

Solvabilité II	Directive européenne qui a pour but l'harmonisation de la réglementation, la transparence de la communication financière des assureurs et la garantie de leur aptitude à honorer leurs engagements vis-à-vis de leurs assurées.
SCR	Solvency Capital Requirement correspond au capital dont la compagnie d'assurance a besoin pour faire face les engagements avec une probabilité de ruine égal à 0.5%
MCR	Minimum Capital Requirement est le niveau minimum de fonds propres en dessous duquel l'intervention de l'autorité de contrôle sera automatique
ALFA	Agence de Lutte contre la Fraude à l'Assurance.
IARD	Incendie, Accidents et Risques Divers.
ROC	Receiver Operating Characteristic.
IF	Isolation Forest .
SVC	Support Vector Classifier.
LOF	Local Outlier Factor.
KNN	K Near Neighbors.
SVM One Class	Support Vector Machine One Class.

A propos de iConcilio

iConcilio fait référence à un cabinet de conseil indépendant, société de droit français. Pour en savoir plus sur notre structure, consulter www.iconcilio.com/about.

iConcilio fournit des services professionnels dans les domaines du consulting et du financial advisory à ses clients des secteurs public et privé, quel que soit leur domaine d'activité. Notre équipe s'engage à fournir un service de grande qualité afin d'aider ses clients à répondre à leurs enjeux les plus complexes.

Nous mobilisons un ensemble de compétences diversifiées pour répondre aux enjeux de nos clients avec engagement et efficience. Nous intervenons dans les domaines d'assurance en actuariat et gestions des risques, management de projet, nouvelles technologies et veilles réglementaires.



+33 6 65 40 48 98
Contact@iconcilio.com



<https://www.linkedin.com/company/iconcilio-corporate>



<http://www.iconcilio.com/home>



EL AIDOUNI
AMINE
iCONCILIO Founder



KOUACH
YASSINE
Consultant Actuariat
et Responsable Revue