



MINISTÈRE DE L'INTÉRIEUR



DÉTECTION DES INFRACTIONS LIÉES AU NUMÉRIQUE À PARTIR D'UNE ANALYSE TEXTUELLE DES MANIÈRES D'OPÉRER

Service statistique ministériel de la sécurité intérieure - SSMSI

Maëlys Bernard - Zoé Gallos

1. PRÉSENTATION DU CONTEXTE ET DES DONNÉES

Les infractions liées au numérique : une définition peu opérationnelle

■ Définition Agence Nationale de la Sécurité des Systèmes d'Information : Actes contrevenant aux traités internationaux ou aux lois nationales, utilisant les réseaux ou les systèmes d'information comme moyens de réalisation d'un délit ou d'un crime, ou les ayant pour cible.

- Une criminalité protéiforme : le numérique un moyen plutôt qu'une fin
- Un phénomène délinquant difficile à quantifier

■ Une définition difficilement applicable opérationnellement au moment de la saisie des données

1. PRÉSENTATION DU CONTEXTE ET DES DONNÉES

Données issues du logiciel de rédaction des procédures de la gendarmerie nationale (LRPGN)

- Nature d'Infraction (NatInf) : La Natinf est la nomenclature des infractions créées par le ministère de la Justice.
Exemple libellé de Natinf : « Vol en bande organisée avec arme », « Destruction du bien d'autrui à raison de la religion », « Agression sexuelle »
 - ➔ Des listes de Natinf en lien avec les infractions liées au numérique
 - Liste de Natinf spécifiques : infractions explicitement liées au numérique
Exemple : Accès frauduleux dans un système de traitement automatisé des données, provocation directe à un acte de terrorisme commise au moyen d'un service de communication au public en ligne
 - Liste de Natinf génériques : infractions pouvant être commises ou non à l'aide des moyens numériques
Exemple : Falsification d'une attestation ou d'un certificat, provocation directe et public à commettre un acte de terrorisme
- Manière d'Opérer (ManOp) : Champ textuel libre qui permet aux gendarmes de résumer la procédure
Exemple de contenu d'une Manop : « Le ou les auteurs effectuent un achat sur le compte **** de la victime et modifient l'identifiant de son adresse mail. Le service contentieux de la société **** contacte la victime afin que celle-ci régularise la somme de 29859 euros suite à un achat. Elle leur explique ne pas être à l'origine de cet achat, son compte **** n'étant plus actif depuis environ 3/4 ans. »

1. PRÉSENTATION DU CONTEXTE ET DES DONNÉES

Écran représentant le logiciel de rédaction des procédures (LRPGN) de la gendarmerie nationale

« coche » cyber à cliquer (tout à fait à droite de l'écran) pour définir une infraction comme liée au numérique

The screenshot displays the LRPGN software interface. On the left, a yellow sidebar titled 'Cyber-espace' provides instructions: 'Il convient de cocher la case "CYBER" pour toutes les infractions pénales, tentées ou commises, au moyen ou à l'encontre d'un système d'information et de communication (principalement internet, les réseaux sociaux ou tout autre système numérique). Exemples : Escroqueries, abus de confiance, trafics, injures, menaces ou harcèlement, commis par voie électronique (via internet, courriel, réseau social...). Atteintes aux données (destruction, vol, recel, usurpation d'identité). Chantage à distance avec une vidéo ou photo compromettantes, corruption de mineur. Atteinte à l'intimité par fixation, enregistrement et transmission d'image. Prise de contrôle frauduleux des systèmes numériques et objets connectés... Lorsque vous cochez la case "CYBER", votre CRPJ sera automatiquement transmis au Centre de lutte Contre les Criminalités Numériques de Pontoise.'

The main interface shows the 'Analyse' section with the following fields and options:

- Compagnie ou escadron: []
- Ville compagnie ou escadron: []
- Nom unité principale: []
- Nom unité: []
- Cadre légal: []
- Enquête préliminaire: []
- Nmr Natinf: []
- Natinf: []
- Tentative:
- Récidive:
- Complicité:
- Libellé: []
- Qualification: []
- Prévu par: []
- Réprimé par: []
- Période: Entre le [01/2022 à 17:19] et le [01/2022 à 17:19]
- Nature de lieu: []
- URL: []
- Adresse: []
- Pays: France
- Commune - CP: []
- X: [] Y: []
- Ajouter aux Natinf favoris de la configuration:
- Liste associée: []

A red circle highlights the 'CYBER' checkbox, with a red arrow pointing to it and the text 'coche « cyber »'.

1. PRÉSENTATION DU CONTEXTE ET DES DONNÉES

Une détection des infractions liées au numérique parfois problématique

- Des Manop sans lien avec les infractions liées au numérique sont cochées « cyber »

Exemple de Manop cochée : « Le 19/01/2018 vers les 20H00 son agresseur, père de ses deux derniers enfants a défoncé la porte du domicile de Mme K il est entré à l'intérieur et a frappé plusieurs fois au visage la victime en lui donnant des gifles. Il voulait voir ses enfants qu'il n'a pas reconnu Mme K ne voulait pas. »

- Des Manop au contenu proche cochées ou non selon les cas

*Exemple de Manop cochée : « La victime va faire ses courses le 02/05 et s'aperçoit le 06/05 qu'il n'y a plus sa carte bancaire dans son portefeuille. Lorsqu'elle vérifie son compte bancaire elle constate 7 débits mentionnant **** dont elle n'est pas l'auteur. »*

Exemple de Manop non cochée : « La victime descend de son véhicule afin de se rendre à la banque et pense avoir fait tomber son portefeuille dans la rue. Un jour plus tard un paiement frauduleux est effectué avec la carte bancaire de la victime sur internet. »

- Amélioration de la détection grâce à un algorithme de machine learning supervisé et une labellisation des données de l'échantillon d'apprentissage

➔ *In fine* création d'une indicatrice de repérage des infractions liées au numérique plus robuste et cohérente

2. LA LABELLISATION : LES THÈMES

- Complexité de définir des thématiques pour la labellisation
- Comparaison avec d'autres nomenclatures
- Exploration statistique à partir de la coche « cyber »
 - Une nomenclature resserrée autour des thèmes principaux et qui n'est pas exhaustive

Labellisation thématique

Escroquerie, interaction avec l'outil numérique à motif crapuleux

Atteintes à la personne motif non crapuleux, (majeur/mineur)

Trouble à l'ordre public

Infraction au droit d'auteur

Non-respect des réglementations spécifiques au fonctionnement des outils numériques

2. LA LABELLISATION : LES DEGRÉS

- Pourquoi définir des degrés d'infractions liées au numérique ?
 - Des réalités d'expertise numérique dans la commission des infractions très différentes
 - Définition de 4 degrés :

Labellisation par degrés

Les supports numériques cibles de l'infraction

Infraction commise dans le cyberspace

Mise en relation par un moyen numérique

Pas de lien avec le numérique ou en dehors du champ des infractions liées au numérique

2. LA LABELLISATION : DEGRÉS D'ESCROQUERIE

Les supports numériques
cibles de l'infraction

Infraction commise dans le
cyberespace

Mise en relation par un
moyen numérique

Pas de lien avec le
numérique

« Le gérant d'une entreprise informatique est victime d'un piratage informatique sur son serveur. Il s'agit d'une attaque de type ransomware sous une variante de la version CrySIS ou bien Dharma. Il s'agit d'un virus qui crypte les dossiers les rendant inexploitable. le pirate demande alors une rançon de deux bitcoins, soit 18000 euros afin de les décrypter. »

« La victime souhaite acheter des visas pour partir [nom de pays]. Il se rend sur un site et se rend compte après avoir payé qu'il s'agit d'un site frauduleux. Le montant du préjudice est de 202 euros. »

« La victime contacte via [un site de vente en ligne] un vendeur pour l'achat de 04 billets pour [un parc d'attraction]. Il demande à un ami qui travaille sur [nom de ville] de payer le vendeur. La rencontre se fait [nom d'un lieu]. Le paiement se fait en liquide pour la somme de 150 euros. Les billets sont récupérés puis remis à la victime qui se rend au parc à l'issue. Elle passe les billets qui sont refusés à l'entrée au motif de faux billets. »

« Le ou les auteurs volent un chéquier et portefeuille range dans un sac à main. Celui-ci ouvert était accroché au crochet du caddie au niveau de la barre de manœuvre. »

2. LA LABELLISATION : DEGRÉS D'ATTEINTE AUX PERSONNES

Les supports numériques
cibles de l'infraction

Infraction commise dans le
cyberespace

Mise en relation par un
moyen numérique

Pas de lien avec le
numérique

« Le plaignant signale que le compte de sa fille âgée de 11 ans a été piraté au début du mois dernier. Il s'agit d'un compte [sur un réseau social] Depuis des photos aux caractères pornographiques sont mises en ligne via le compte piraté. Ces photos présente une jeune femme plus ou moins dénudée selon les clichés et dont on ne voit pas le visage. Il se peut qu'il s'agisse de la même personne sur chaque cliché. Sur la plupart des photos il y a la présence d'un commentaire explicite.»

« Depuis début février 2018 la victime est harcelée, injuriée et diffamée par l'une des employées de l'**** ou elle est directrice. Ces propos sont publiés publiquement sur [un réseau social] et envoyés par SMS à la victime par la mise en cause. La victime nous fournira un certificat médical mentionnant 02 jours d'ITT. »

« Le mis en cause prend attache avec la victime sur le site de rencontre ****. Un rendez-vous est fixé à [nom de ville]. Le mis en cause emmène la plaignante dans un appartement. Il la force à l'embrasser. La victime tente de repousser son agresseur mais s'épuise et subit l'agression. le mis en cause se trouvant ensuite dans la salle de bain la victime s'habille et quitte les lieux. »

« La victime éducatrice à l'institut thérapeutique éducatif pédagogique fait une remarque a un élève écoutant de la musique. Ce dernier refuse d'éteindre la musique puis s'énerve. Il attrape l'éducatrice et lui tire violemment les cheveux puis lui donne un coup de genou au niveau du visage »

NOMENCLATURE FINALE

Thèmes

Escroquerie, interaction avec l'outil numérique à motif crapuleux

Atteintes à la personne motif non crapuleux, (majeur/mineur)

Trouble à l'ordre public

Infraction au droit d'auteur

Non-respect des réglementations spécifiques au fonctionnement des outils numériques

Degrés

Les supports numériques cibles de l'infraction

Infraction commise dans le cyberespace

Mise en relation par un moyen numérique

Pas de lien avec le numérique

Les supports numériques cibles de l'infraction

Infraction commise dans le cyberespace

Mise en relation par un moyen numérique

Pas de lien avec le numérique

Toutes les infractions liées au numérique sans distinction de degré

Pas de lien avec le numérique

Toutes les infractions liées au numérique sans distinction de degré

Pas de lien avec le numérique

Toutes les infractions liées au numérique sans distinction de degré

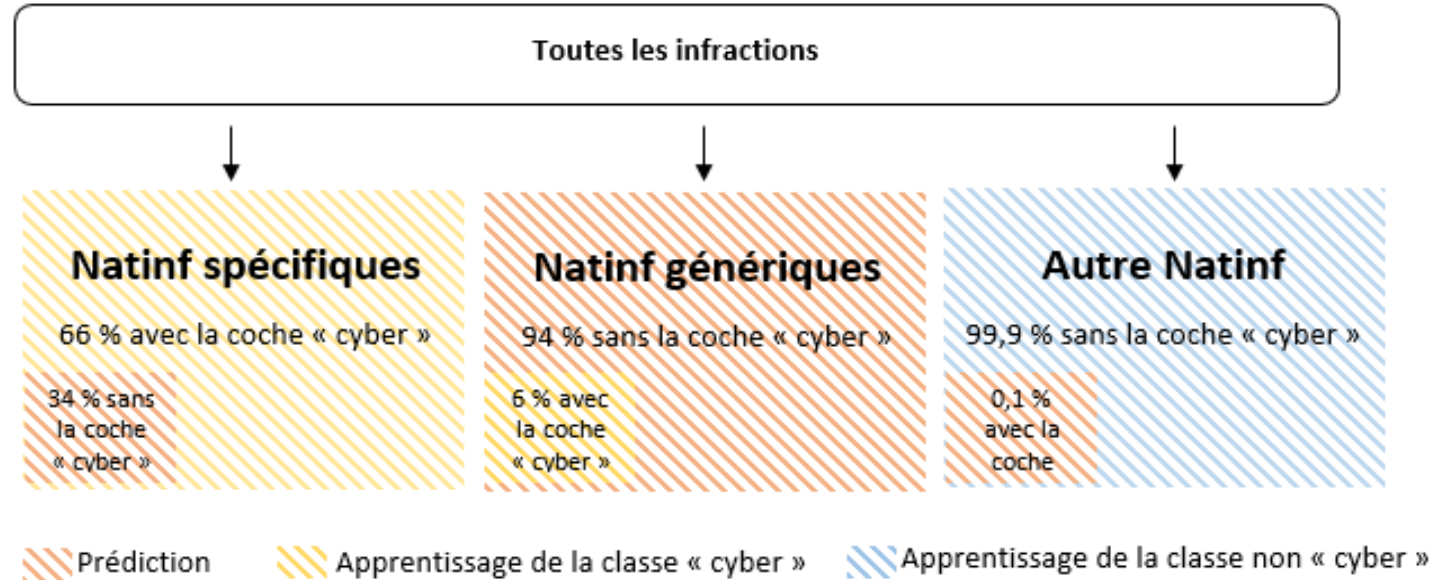
Pas de lien avec le numérique

3. CRÉATION D'UN PROTOCOLE DE LABELLISATION

- Écriture d'un document explicatif pour la labellisation :
- Tirage d'un premier échantillon de manop à labelliser
- Choix d'une interface de labellisation et écriture d'un tutoriel
- 1^{ère} phase de test en interne
 - Amendement labellisation
 - Amendement échantillonnage
 - Amendement du tutoriel
- Réécriture du document
- Formation des personnes qui labelliseront
 - Phase de labellisation
- En parallèle des travaux sur la labellisation, un premier travail de développement en machine learning a été effectué

4. APPRENTISSAGE POUR UNE CLASSIFICATION BINAIRE : CHOIX D'APPRENTISSAGE ET BIAIS

Choix d'apprentissage



Biais dû à l'apprentissage :

- Pour la classe « cyber » :
 - les Manop des infractions cochées ayant une Natinf générique peuvent ressembler à celles des infractions ayant une Natinf générique mais n'étant pas cochées
- Pour la classe « non cyber » :
 - les infractions avec une Natinf générique ne sont pas entraînées comme n'étant pas liées au numérique

➔ Risque de surestimation

5. PRÉTRAITEMENT

Prétraitement

- Nettoyage des manières d'opérer : suppression d'accents et caractères spéciaux, traitement de chaînes spéciales
- Réduction du nombre de mots : utilisation de listes de Stopwords enrichies des mots très récurrents et d'une partie du champ lexical de la gendarmerie, lemmatisation
- Recherche de bigrammes liés au numérique
 - Nombre de mots réduit de presque 70 %

Exemple de Manop avant le prétraitement : « *Faits commis le 31/12/2017 entre 14h et 18h30. Le ou les auteurs entrent par effraction par une fenêtre du salon en la forçant la grille de protection et soulevant le volet roulant. Puis la fenêtre à été ouverte avec un objet type pied de biche. Pas de vol constaté* » (49 mots)

Exemple de Manop après le prétraitement : « *fait commettre entrent effraction fenetre salon forcer grille protection soulever volet rouler fenetre objet pied biche vol constate* » (18 mots)

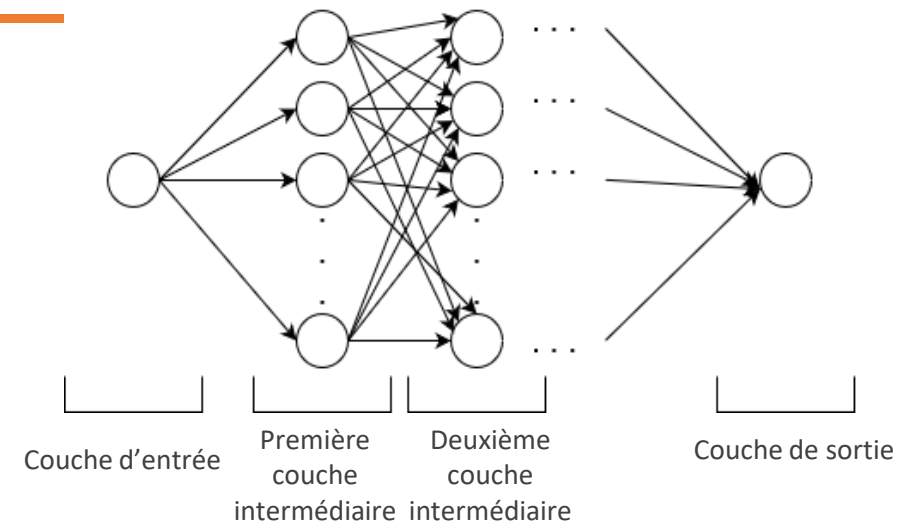
6. PRÉDICTION DES INFRACTIONS LIÉES AU NUMÉRIQUE

Passage du texte en valeur numérique

- Transformation de chaque mot du document en valeur numérique sous forme d'un vecteur
 - Méthode SkipGram de l'algorithme Word2vec : vecteur défini par les mots à proximité de ce mot : son contexte
 - Les mots qui ont un contexte proche auront des vecteurs proches
- Vecteur par Manop : moyenne des vecteurs de chaque mot de la Manop

Modèle pour prédire les infractions liées au numérique

- Multi layer perceptron
 - Réseau de neurones à une ou plusieurs couches cachées
 - **Couche d'entrée** : Vecteur numérique associé au document
 - **Couche de sortie** : Probabilité d'appartenir au numérique



7. RÉSULTAT DU PROCESSUS : MAUVAIS CLASSEMENTS

Mauvais classement dans la classe « cyber » : 3 % des infractions de cette classe sont mal classées

■ Manop ne faisant pas référence au numérique

Exemple : « Après une dispute familiale l'auteur bouscule sa femme contre un mur et tente de la retenir en la serrant par les bras lui occasionnant des contusions et des bleus. Certificat médical mentionnant 3 jours d'iTT »

■ Manop longue ne faisant référence que très rapidement à des faits liés au numérique

Exemple : « La victime est mise en relation via [réseau social] avec le vendeur d'une motocross. Un rendez-vous est fixé par la victime à proximité de son domicile. Le prix d'achat convenu est de 580 euros. La victime se rend seule à ce rendez-vous où elle est rejointe par un premier individu. Ce dernier l'attire sur un parking à l'abri des regards où 5 autres individus les attendent. La victime est alors agressée physiquement frappée à plusieurs reprises. Les mis en cause réussissent à arracher la sacoche de la victime à l'intérieur de laquelle était entreposé le numéraire, le téléphone portable et la carte nationale d'identité de la victime.[...] »

■ Manop faisant référence au numérique avec un champ lexical spécifique

Exemple : « Le ou les auteurs ouvrent deux trappes en bordure de chaussée espacées de 300 mètres, sectionnent un câble téléphonique 224 paires 8/10? de part et d'autres des trappes et le dérobe. Environ 224 clients ont été impactés suite à la coupure des communications électroniques. »

7. RÉSULTAT DU PROCESSUS : MAUVAIS CLASSEMENTS

Mauvais classement dans la classe « non cyber » : moins d'1 % des infractions de cette classe sont mal classées

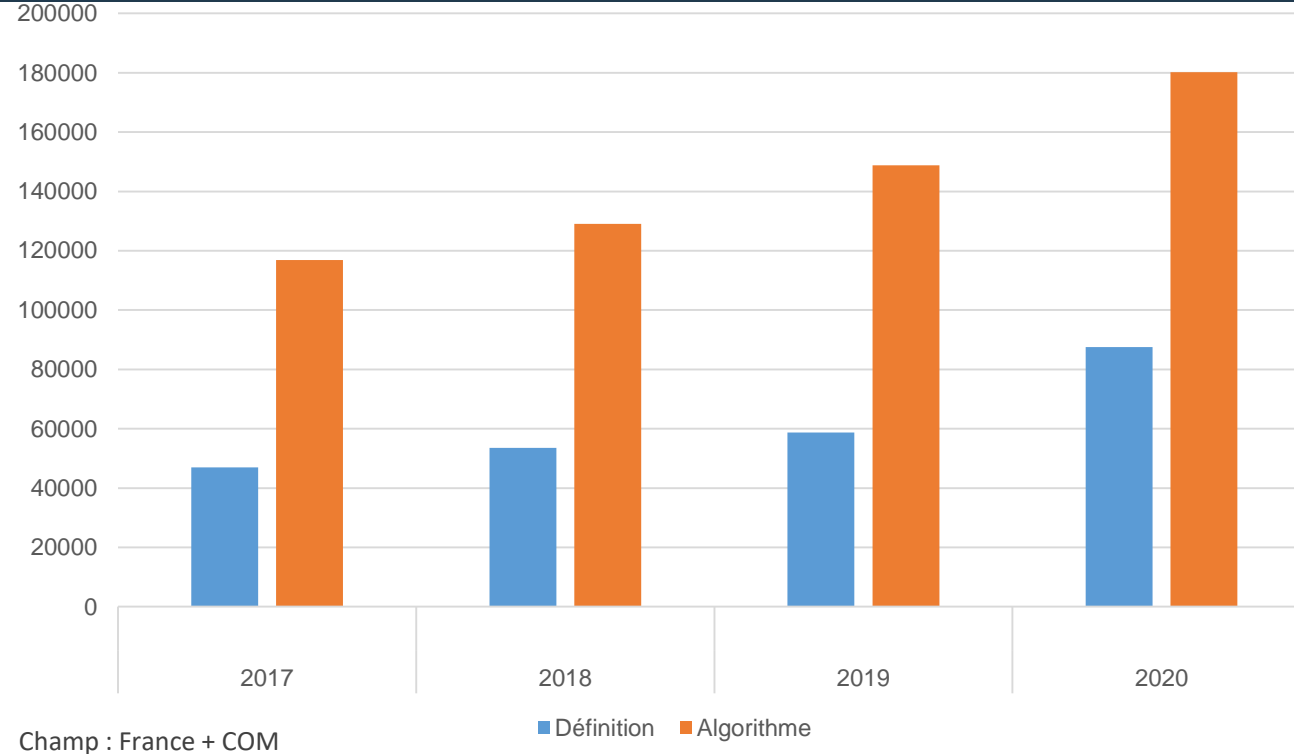
■ Manop faisant référence au numérique

Exemple : « La mise en cause fait une rencontre sur internet. Elle entretient une relation amoureuse pendant 1 an et demi sans jamais le rencontrer physiquement. Lui faisant croire qu'il doit recevoir de l'argent de la succession de ses parents il lui demande régulièrement pendant cette période de recevoir de l'argent par mandat cash [...]. »

■ Manop ne faisant pas référence au numérique avec des termes pouvant être souvent présents dans les Manop des infractions liées au numérique ou dans un contexte de mots liés à la cyber

Exemple : « La victime est présidente du conseil des parents d'élèves de la ****. A l'issue d'un conseil d'école houleux les parents d'élèves ont transmis un courrier diffamant la victime aux 170 parents d'élèves du groupe scolaire en l'espèce : « propos irrespectueux et diffamant », « menaces envers l'équipe éducative » et « excès de colère ». »

7. RÉSULTAT DU PROCESSUS : COMPARAISON ENTRE L'ALGORITHME ET LA DÉFINITION



Source : Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2017 et 2020, SSMSI.

- Entre 2 et 3 fois plus d'infractions liées au numérique estimées par l'algorithme que par la définition
→ attention au biais créé par l'apprentissage qui entraîne un risque de surestimation
 - Infractions dont le contenu de la Manop est cyber : oubli probable de la coche
 - Infractions dont la nature peut dépendre de l'interprétation du gendarme

8. APPLICATION À L'ENTRAÎNEMENT MULTI-CLASSES

- Après la labellisation, prédiction multi-classes pour les escroqueries et les atteintes aux personnes
 - Apprentissage ensembliste pour prédire chaque label l'un après l'autre par rapport au reste des données en classification binaire
 - Risque de mauvais classement avec des ressemblances dans le contenu de certaines Manop
 - Différents degrés d'escroqueries et d'atteintes aux personnes liés au numérique
 - Manop d'infractions pouvant contenir des mots liés au numérique alors que ces infractions ne sont pas considérées comme telles
Exemple : vols de colis suite à l'achat sur internet.
- attention particulière sur ces infractions considérées comme à risque de mauvais classement

CONCLUSION

■ Limites

- La détection des infractions liées au numérique est dépendante du contenu de la Manop
- Processus non adaptable à l'intégralité des infractions enregistrées par la police car les Manop sont très peu renseignées

■ Résultats sur les Manop de la gendarmerie permettront :

- Une estimation de la part des infractions liées au numérique dans le total des infractions enregistrées par la gendarmerie
- Enrichir la liste des Natinf spécifiques
- Redéfinir la liste des Natinf génériques
- Harmoniser la définition des infractions liées au numérique entre la police et la gendarmerie
 - amélioration du remplissage des variables liées au numérique
 - amélioration la détection des infractions liées au numérique

MERCI POUR VOTRE ATTENTION !

