Journées de méthodologie statistique de l'Insee 2022

DETECTION DES INFRACTIONS LIEES AU NUMERIQUE A PARTIR D'UNE ANALYSE TEXTUELLE DES MANIERES D'OPERER

Maëlys BERNARD, Zoé GALLOS

Ministère de l'Intérieur, Service statistique ministériel de la sécurité intérieure

maelys.bernard@interieur.gouv.fr
zoe.gallos@interieur.gouv.fr

Résumé

La cybercriminalité est une forme de criminalité en plein essor, et la lutte contre celle-ci est un enjeu important pour les services de police et de gendarmerie. Elle s'avère toutefois particulièrement difficile à quantifier. Elle a été définie par un groupe de travail piloté par le Service statistique ministériel de la sécurité intérieure (SSMSI) en 2014 comme regroupant toutes les infractions pénales tentées ou commises à l'encontre ou principalement au moyen d'un système d'information et de communication (SIC). La cybercriminalité n'est donc pas un champ infractionnel dont les contours seraient définis uniquement par des natures d'infractions. C'est surtout le mode opératoire qui fera d'une infraction qu'elle est qualifiée de « cyber » ou non. Dès lors, le phénomène est approché en utilisant des variables annexes associées aux infractions et caractérisant celles-ci. Néanmoins, ces variables ne sont pas toujours bien renseignées, et reposent sur l'interprétation du gendarme ou du policier. Ainsi, avec ces seules informations, le SSMSI n'est aujourd'hui pas en mesure de diffuser un indicateur fiable de la cybercriminalité.

Afin de pallier ce problème et de mesurer plus finement la cybercriminalité, un travail d'analyse textuelle appuyé par des techniques de *machine learning* a été implémenté en collaboration avec le SSP Lab. Cette analyse repose sur le texte contenant la description de la manière d'opérer, qui comporte quelques dizaines de mots. Ces manières d'opérer sont des résumés des procédures et sont remplies par les gendarmes et les policiers dans les logiciels de rédaction des procédures. La saisie de cette zone textuelle étant obligatoire en gendarmerie, on se limite à ce champ dans un premier temps.

La labellisation d'un échantillon d'infractions sera menée par des experts-métier, allant au-delà de la simple distinction entre « cyber » et « non-cyber » et cherchant à distinguer une typologie propre à la cyberdélinquance, identifiant par exemple ce qui relève de moyens numériques pour des infractions de droit commun ou ce qui relève au contraire d'attaques purement cybercriminelles (rançongiciels, contenus illicites sur le web, etc.). En sus certaines thématiques, les escroqueries et les atteintes aux

personnes, seront affinées par degré d'utilisation des moyens numériques dans la commission de l'infraction.

Les travaux préparatoires ont été menés en utilisant un échantillon d'apprentissage défini à partir d'une indicatrice croisant la coche « cyber » et des informations sur les natures d'infraction. Ils permettent de dérouler l'ensemble des étapes du processus de traitement. Il est nécessaire dans un premier temps de rendre le texte utilisable pour l'analyse : le texte a donc été découpé en *token* par des phases de prétraitements.

Le passage du texte en valeur numérique est réalisé par *word embedding* : l'algorithme *Word2Vec* est ici utilisé. La projection du document est ensuite réalisée par la moyenne des vecteurs projetés associés aux mots composant chaque document. L'apprentissage est ensuite réalisé sur le jeu de données test.

Enfin, une méthode de machine learning est appliquée pour prédire la cybercriminalité. Pour cela, plusieurs méthodes ont été testées telles que la régression logistique, les forêts aléatoires, les réseaux de neurones ou encore la méthode du Gradient Tree Boosting.

Toutes ces méthodes ont donné des résultats similaires quant à l'estimation de la cybercriminalité au sein des procédures enregistrées par la gendarmerie.

La prochaine étape sera de tester cet algorithme en s'appuyant sur un échantillon d'apprentissage labellisé. En outre, il s'agira de produire, non plus une classification binaire (cyber/non cyber), mais une prédiction multi-classes. Il sera alors nécessaire d'adapter l'algorithme en utilisant, par exemple, un apprentissage ensembliste, en veillant toutefois à certains points. Avec la classification binaire actuelle les contenus textuels des manières d'opérer associée à certaines infractions « non cyber » pourraient être proches de ceux associés à des infractions liées au numérique, rendant le processus de détection délicat. Ce problème sera également présent lors des prédictions multi-classes, notamment lors de la prédiction par degré d'utilisation du numérique sur le thème des escroqueries et des atteintes aux personnes.

Abstract

Cybercrime is not an offence field whose scope would be only defined by the nature of the offence. It's the manner of commission which will qualify the offence as cyber or not. Nevertheless the variables available at SSMSI to pinpoint cybercrime is not all the time well filled because of the security forces insight of the subject. This project, started thanks to a partnership with SSP LAB, INSEE and SSMSI, intends to enhance the detection of the phenomena so-called « cybercrime » or digital offences within the data from security forces' complaint drafting software. This article aims to explain the key stages of the project: implementing an algorithm in Natural Language Processing and build a data labelling grid enabling to refine the variable beyond a binary distinction cyber versus non-cyber by important themes inside digital offences.

Avant-propos

La cybercriminalité est une forme de criminalité en plein essor, et la lutte contre celle-ci est un enjeu important pour les services de police et de gendarmerie. Elle s'avère toutefois particulièrement difficile à quantifier. Les infractions constatées sont enregistrées par les forces de sécurité à l'aide de leurs logiciels de rédaction de procédures. En l'état actuel, les données ainsi collectées ne permettent pas une comptabilisation fiable de la cyberdélinquance car la définition de ce phénomène complexe demeure en partie floue et que le processus de saisie permettant d'identifier une infraction dite

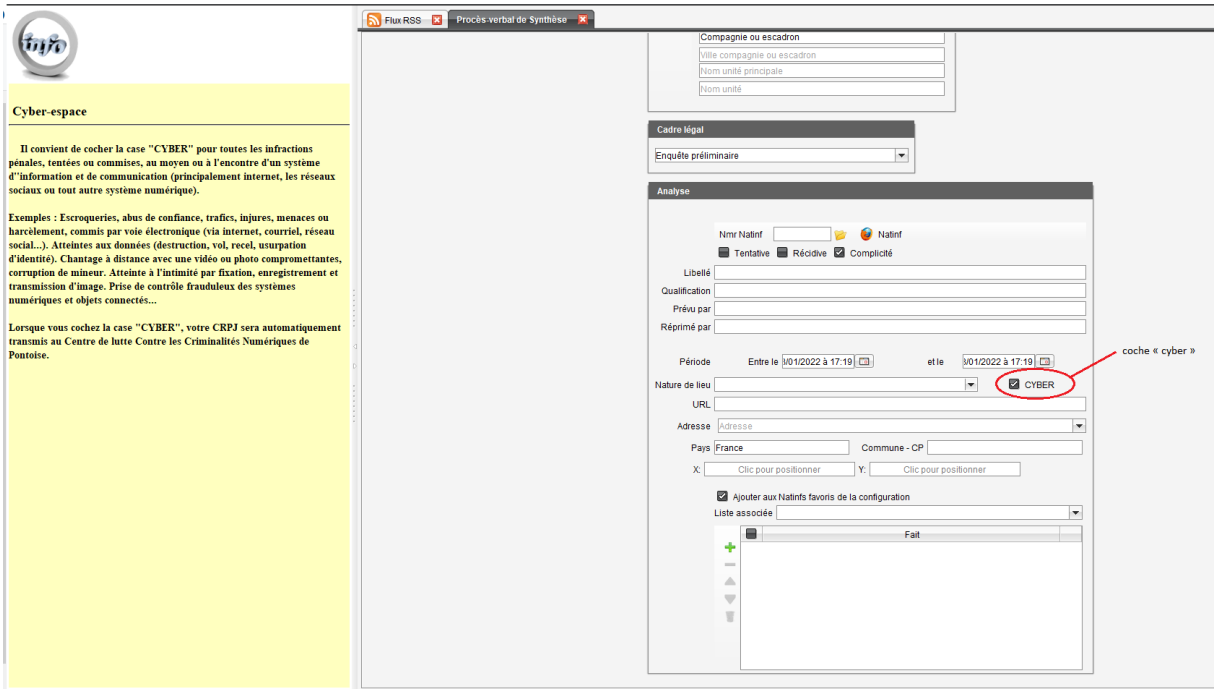
cybercriminelle laisse une place à des différences de pratiques des forces de sécurité pour cette caractérisation. Pour répondre aux enjeux croissants de pilotage de la réponse des forces de sécurité mais aussi d'information au public sur les phénomènes de cyberdélinquance, le SSMSI conduit un travail visant à mieux caractériser et mieux repérer les infractions liées au numérique[1], en vue d'un suivi statistique fiable et pertinent. L'amélioration du repérage de ce type d'infractions s'appuie sur la mise en œuvre d'un algorithme *de Natural Language Processing*. L'algorithme sera entraîné à détecter différentes catégories d'infractions liées au numérique sur un échantillon issu d'un protocole de labellisation des données.

1. Un projet pour l'amélioration de la détection des infractions liées au numérique

1.1. Une détection des infractions liées au numérique parfois problématique

Chaque année, les policiers et les gendarmes enregistrent entre 3 et 4 millions d'infractions de crimes et délits. Lors de l'enregistrement de celles-ci, ils saisissent diverses informations dans leurs logiciels de rédaction des procédures. Les infractions, en particulier, sont qualifiées en sélectionnant une « nature d'infraction [2]» (Natinf) dans la nomenclature des infractions. Les Natinf peuvent avoir des libellés aussi variés que « Vol en bande organisée avec arme », «Destruction du bien d'autrui à raison de la religion », « Agression sexuelle ». Grâce à la Natinf, le SSMSI peut répartir les infractions dans la classification internationale des infractions à des fins statistiques (ICCS)[3].

Figure 1 : Représentation de la coche « cyber » dans le logiciel de rédaction des procédures



[1] Suite à des travaux exploratoires visant à mieux appréhender le périmètre de la « cyberdélinquance », le SSMSI préfère utiliser l'expression « infractions liées au numérique ».

2 La Natinf est la nomenclature des infractions créées par le ministère de la Justice en 1978 pour les besoins de l'information du casier judiciaire et des juridictions pénales. Elle recense la plupart des infractions pénales en vigueur ou abrogées, et évolue au gré des modifications législatives et réglementaires.

3 La classification internationale des infractions à des fins statistiques est une nomenclature créée par l'ONUDC (Office des Nations unies contre les drogues et le crime) afin de regrouper les infractions selon des catégories homogènes.

À la Natinf les forces de sécurité peuvent adjoindre d'autres variables comme l'appartenance ou non de l'infraction au champ du numérique. Pour la gendarmerie, il existe une coche « Cyber » qui doit être sélectionnée dans le logiciel de rédaction (*figure 1*) lorsque l'infraction relève des infractions liées au numérique. Les gendarmes disposent d'un champ de texte libre pour rédiger un court résumé de la procédure. Ce champ est appelé manière d'opérer ou Manop[4]. Ci-dessous deux exemples de 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. »*

« Le ou les auteurs fracturent la porte d'entrée de la maison fouillent partiellement et dérobent 200 euros en espèces. »

D'après les données de la gendarmerie, auxquelles le SSMSI a accès[5], la qualification « cyber » des infractions ne semble pas toujours renseignée de façon homogène. En effet, certaines infractions qualifiées de « cyber » ne laissent aucun doute à la lecture de la Manop sur l'absence de lien avec le numérique de l'infraction.

Par exemple dans la Manop suivante : « *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. »*

La qualification de « cyber » de ce type d'infraction laisse à penser à une erreur au moment de la saisie. Au-delà de ce premier constat, les différences d'appréciation de la qualification « cyber » d'une infraction peuvent entraîner un peu d'hétérogénéité dans les données disponibles au sein des infractions liées au numérique. En effet, le SSMSI a identifié des cas dits « tangents » au périmètre des infractions liées au numérique. Par exemple, certaines Manop concernent des vols de carte bleue suivi d'un achat sur internet.

La Manop suivante a été cochée « cyber » par la gendarmerie : « *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*. »

Tandis que cette Manop ne l'a pas été : *« 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. »*

L'infraction constituée d'un vol de carte bleue puis d'un achat sur internet pose question au regard du périmètre des infractions liées au numérique. De façon plus générale cela interroge sur la définition de la « cyberdélinquance ».

Le SSMSI a piloté en 2014 un groupe de travail qui a proposé une définition de la cyberdélinquance comme regroupant toutes les infractions pénales tentées ou commises à l'encontre ou principalement

[4] Cette étude ne porte que sur les données de la gendarmerie, les données textuelles de la police étant trop parcellaires.

[5] Le SSMSI n'a pas accès à tous les éléments constitutifs d'un dossier d'enquête. Certaines données comme les procès-verbaux ne sont pas disponibles au sein du SSMSI. Le constat sur la fiabilité des données est réalisé à partir des données disponibles.

au moyen d'un système d'information et de communication (SIC)[1]. A noter que cette définition est similaire à celle formulée par l'Agence Nationale de Sécurité des Systèmes d'Information (A.N.S.S.I.)[6]. Cette définition laisse une part d'interprétation aux agents qui rédigent les plaintes. Ainsi sont qualifiés de « cyber » des faits aussi différents que des attaques par rançongiciel[7], du phishing, du harcèlement, des arnaques à la romance, de la pédopornographie, des prélèvements bancaires suspects ou encore des prises de contact par internet entraînant une infraction en dehors du cyberespace. La qualification « cyber » est donc mobilisée pour un ensemble d'infractions très vaste qui recoupe presque l'ensemble du champ infractionnel tant la pénétration des outils numériques dans la délinquance est importante.

Fort de ce constat, le présent projet propose d'améliorer la détection de la délinquance liée au numérique à partir du champ textuel appelé Manop. Le but premier est de contrecarrer les éventuelles erreurs de saisies et également de pouvoir repérer les infractions liées au numérique y compris dans les cas où les gendarmes n'ont pas mentionné le caractère cyber ni à travers une Natinf spécifique ni par la coche « cyber », mais l'ont bien fait apparaître dans la Manop. En deuxième intention, il s'agit d'aller plus avant vers une variable plus fine qu'une simple détection binaire des infractions liées au numérique. *In fine*, ce projet permettrait d'aboutir à la création d'une variable de caractérisation des infractions liées au numérique utilisant mieux l'information contenue dans les procédures (y compris les zones textuelles), et comportant plusieurs modalités, afin de traduire la diversité des situations de cyberdélinquance. Une telle variable serait plus cohérente avec les besoins de suivi statistique et d'étude des phénomènes de criminalité numérique.

Le projet se développe selon deux processus. Le travail d'utilisation des algorithmes en Natural Language Processing a été fait en parallèle d'un travail de labellisation des données d'entraînement. Il s'agit d'effectuer les différentes étapes du prétraitement, puis l'entraînement d'un algorithme de Machine Learning sur les textes issus des Manop. Ce travail a été initié en collaboration avec le SSP Lab de l'Insee (Julie Djiriguian).

Les algorithmes utilisés étant supervisés, une labellisation des données de l'échantillon d'apprentissage est souhaitable pour assurer la qualité de la prédiction. Cette étape de labellisation sera détaillée dans la troisième partie de l'article.

1.2. Présentation du contenu des Manop

Les Manop sont donc des petits textes de quelques dizaines à un peu plus d'une centaine de mots. Ce champ est presque systématiquement rempli par la gendarmerie, a contrario de la police[8], pour laquelle ce champ n'est pas obligatoire dans le logiciel de rédaction des procédures.

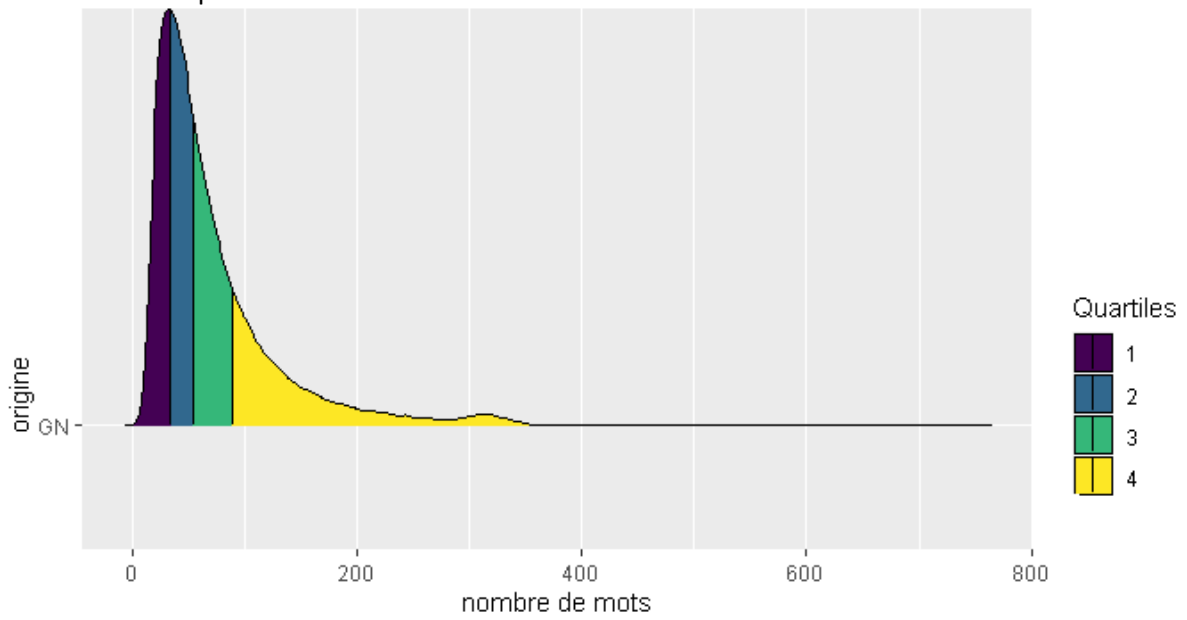
Parmi les Manop renseignées par la gendarmerie en 2018, moins de 0,1 % contiennent moins de 5 mots. De plus, la moitié des procédures enregistrées avec une Manop non vide, ont plus de 50 mots (*figure 2*). L'étendue interquartile de la distribution, c'est-à-dire l'étendue après avoir retiré les 25 % des valeurs les plus faibles ainsi que les 25 % des valeurs les plus élevées pour neutraliser les valeurs extrêmes, a une valeur de 52 mots.

[6] L'agence nationale de la sécurité des systèmes d'information, créée en 2009, a pour mission d'apporter son expertise et son assistance technique aux administrations et aux entreprises. [2]

[7] Le terme « rançongiciel » (ou ransomware en anglais) est une contraction des mots « rançon » et « logiciel ». Il s'agit donc par définition d'un logiciel malveillant bloquant l'accès à l'ordinateur ou à ses fichiers dont le but est d'obtenir de la victime le paiement d'une rançon. [3]

[8] Les manières d'opérer rédigées par les policiers ne sont que très peu renseignées dans les procédures enregistrées, pour un tiers seulement entre 2018 et 2020. Parmi celles-ci 17 % contiennent moins de 5 mots. Le champ d'étude se limitera donc aux Manop enregistrées par la gendarmerie.

Figure 2 : Distribution des procédures enregistrées par les gendarmes en fonction du nombre de mots des manières d'opérer.



Lecture : 25 % (Q1) des Manop enregistrées par la gendarmerie ont moins de 33 mots.
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie en 2018.

1.3. La qualification « cyber » d'une Manop étendue pour toutes les infractions d'une procédure

Le champ relatif à la manière d'opérer est présent au niveau de la procédure alors que la nature d'infractions et la coche « cyber » renseignée par la gendarmerie sont au niveau de l'infraction. Une procédure peut en réalité contenir plusieurs infractions. Toutefois, en 2018, presque 90 % des procédures enregistrées par la gendarmerie contiennent 1 seule infraction. Par ailleurs, parmi les procédures avec plus d'une infraction, seules 1 % ont au moins une infraction avec la coche « cyber » et une autre sans la coche. Ainsi, au vu de ce faible effectif de procédures, il a été choisi de dupliquer la manière d'opérer pour toutes les infractions d'une même procédure.

2. Méthode

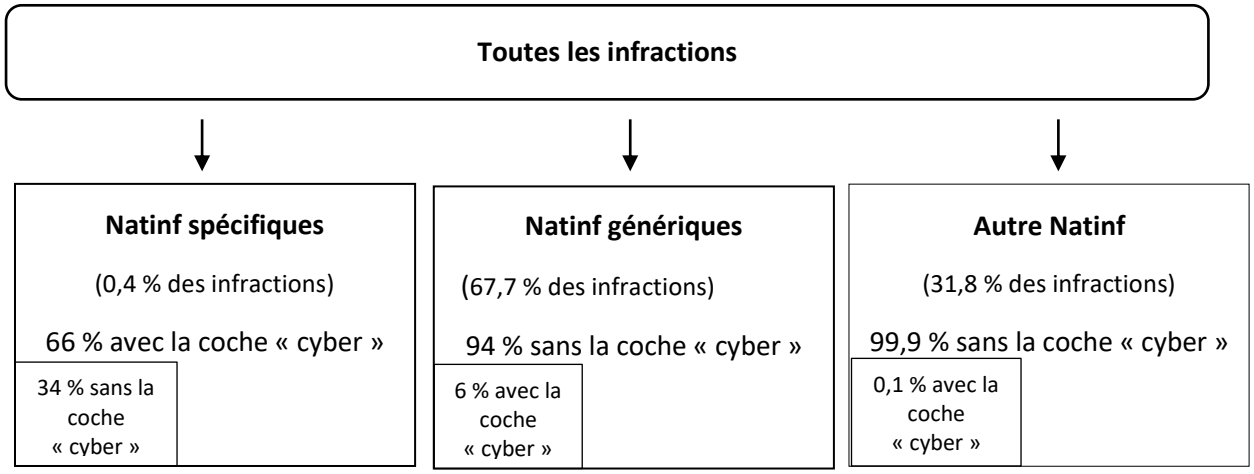
2.1. Choix de l'apprentissage

Le but de l'analyse est de détecter les infractions liées au numérique à partir d'informations contenues dans les Manop. Pour ce faire, l'apprentissage supervisé est utilisé. Il s'agit d'entraîner le modèle sur des données étiquetées pour que, appliqué sur l'ensemble des données, l'algorithme puisse prédire les infractions liées au numérique. Dans un premier temps, avant la labellisation, la coche « cyber » saisie par les gendarmes ainsi que les infractions qualifiées par leur nature d'infractions seront utilisées comme permettant de caractériser les infractions liées au numérique.

Le groupe de travail piloté par le SSMSI en 2014 a permis de définir les infractions liées au numérique par deux listes : une liste de Natinf dites spécifiques et une liste de Natinf dites génériques. La liste des

Natinf spécifiques comprend les infractions qui relèvent explicitement du numérique[9]. La liste de Natinf génériques comprend quant à elle les infractions qui ne sont pas clairement des infractions liées au numérique mais qui peuvent le devenir si elles ont eu lieu sur internet ou dans le cadre d'un SIC. Cette liste est mise à jour annuellement, en ajoutant les Natinf contenant au moins 20 infractions cybercriminelles dans l'année, repérées par les variables disponibles telles que la coche « cyber » pour la gendarmerie, le mode opératoire, le contexte de la procédure, ainsi que la nature de lieu de l'infraction pour la police. Ainsi pour être considérées comme des infractions liées au numérique, au sens de la définition proposée par le groupe de travail, les Natinf génériques doivent être croisées avec la coche « cyber ».

Figure 3 : Répartition des infractions en fonction du type de Natinf



Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie en 2018.

D'une part, un des différents croisements possibles entre la coche « cyber » et les natures d'infractions constituera les données d'apprentissage pour la classe « Cyber ». D'autre part, les infractions non cochées et n'appartenant pas aux listes de Natinf spécifiques et génériques constitueront les données d'apprentissage pour la classe « non cyber ».

Ainsi, plusieurs variantes pour l'apprentissage des données en vue de la détection des infractions liées au numériques ont été testées (*figure 4*).

Le premier croisement risque d'être trop restrictif. Ce croisement ne couvre que le champ des atteintes au système de traitement automatisé des données ainsi que quelques autres Natinf comme le harcèlement en ligne ou les infractions liées à la pédopornographie. Les infractions concernant les escroqueries en ligne, par exemple, apparaissant dans la liste des Natinf génériques, ne seront jamais considérées comme liées au numérique lors de l'entraînement. Ainsi, l'algorithme risque de ne pas être correctement entraîné pour ce type d'infractions, et de ne pas bien classer ces infractions. Le nombre d'infractions prédites comme liées au numérique risque donc d'être sous-estimé.

[9] En 2018, seules 66 % des infractions ayant une Natinf spécifique ont été cochées « cyber ». On observe donc un désalignement entre cette coche qualifiant l'infraction de « cyber » et le choix d'une nature d'infraction liée à la délinquance numérique.

Figure 4 : Différents croisement testés pour les données d'apprentissage

Variante	Différents croisements pour l'apprentissage de la classe cyber	Apprentissage de la classe « non cyber »	Problème sur la prédiction
1	Natinf spécifiques * coche « cyber » *3 500 infractions en 2018*	Natinf ni spécifiques ni génériques et n'étant pas cochées « cyber » *397 400 infractions en 2018*	Risque de sous-estimation de la prédiction
2	Natinf spécifiques + Natinf génériques * coche « cyber » *53 000 infractions en 2018*		Risque de surestimation de la prédiction
3	Natinf spécifiques * coche « cyber » + Natinf génériques * coche « cyber » *51 100 infractions en 2018*		Risque de surestimation de la prédiction moins importante que pour le deuxième croisement

Lecture : En 2018, 3 500 infractions ont une Natinf spécifique et sont cochées « cyber » par le gendarme. Ces infractions, qui constituent le premier apprentissage testé pour la classe « cyber », risquent de sous-estimé la prédiction.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie en 2018.

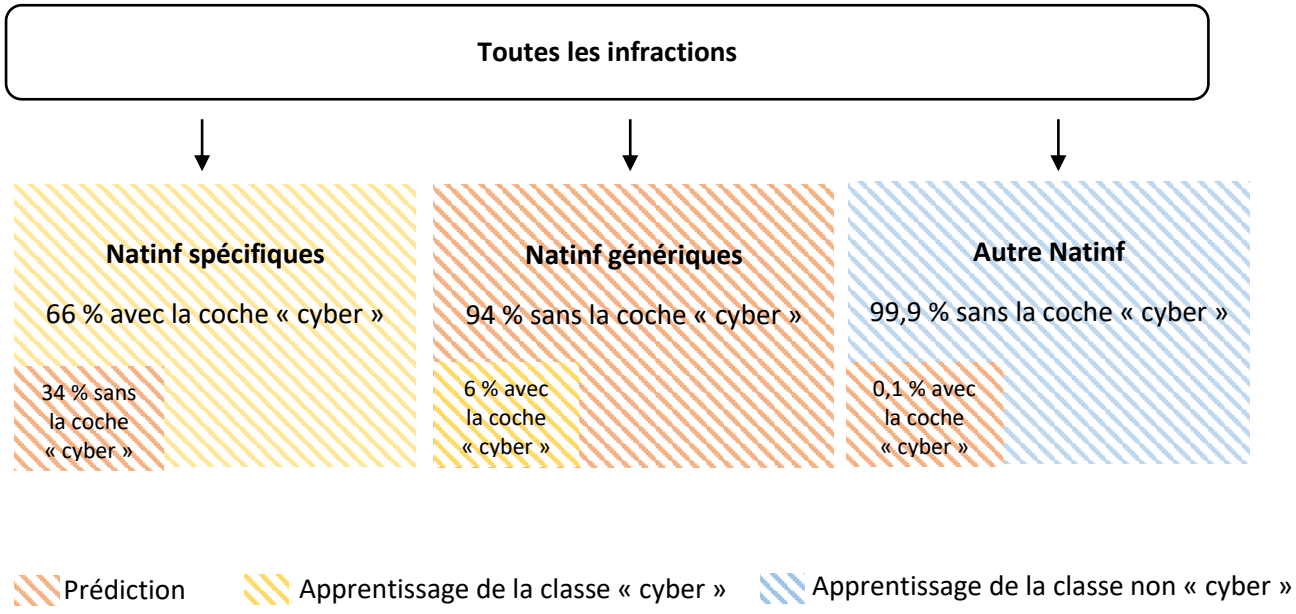
Le deuxième croisement peut être quant à lui trop souple, car il est difficile de savoir, si les infractions ayant une Natinf spécifique mais n'étant pas cochées « cyber » relèvent d'un oubli de coche ou d'une mauvaise classification des Natinf. Il n'est donc pas certain qu'il s'agisse d'infractions liées au numérique. Par exemple, l'infraction de la procédure associée à la Manop « *Le ou les auteurs s'introduisent dans le local technique d'alimentation du parc éolien de ***. Ils coupent des câbles électrique paralysant l'alimentation du système éolien Aucune effraction constatée* », est associée à une Natinf spécifique et n'est pas cochée « cyber ». Or, dans la manière d'opérer rien ne fait référence au numérique. De plus, les manières d'opérer des Natinf génériques cochées « cyber » peuvent ressembler aux manières d'opérer des Natinf génériques qui ne sont pas cochées « cyber ». Les données pourront donc être entraînées avec des termes non cyber, auquel cas la prédiction des infractions liées au numérique sera surestimée. Après une expertise réalisée sur un ensemble de Natinf génériques cochées « cyber » d'un côté et non cochées « cyber » de l'autre, il s'est avéré les Manop des Natinf génériques cochées « cyber » contenaient bien des mots très spécifiques au numérique, les distinguant ainsi des Manop des infractions non cochées « cyber » avec une Natinf générique. Dans certains cas, la Manop est cependant imprécise, et aucun élément lié au numérique n'apparaît. Dans ce cas, les Manop des infractions génériques même si elles sont cochées « cyber » seront proches de celles des infractions génériques non cochées « cyber ».

Enfin, la 3ème proposition d'apprentissage semble être celle qui correspond le mieux à la réalité, malgré un risque de surestimer la prédiction des infractions liées au numérique. En effet, le problème évoqué précédemment lors de l'entrainement des infractions ayant une Natinf générique et étant cochées « cyber » par le gendarme, qui peuvent ressembler aux infractions ayant une Natinf générique mais n'étant pas cochées « cyber », est également présent pour cet apprentissage. Le biais créé par cet apprentissage devrait néanmoins être moins important que pour la 2ème proposition d'apprentissage car la surestimation potentielle créée par les Natinf spécifiques non cochées « cyber » est retirée.

Par ailleurs, au sein de l'apprentissage « non cyber », qui ne comporte que les autres Natinf non cochées « cyber » par le gendarme, un biais peut être présent. En effet, les Natinf génériques non cochées « cyber » ne sont pas incluses dans l'apprentissage « non cyber ». Ainsi, si le champ lexical de certaines infractions est uniquement présent dans des infractions ayant une Natinf générique, ce type d'infraction ne sera jamais entraîné comme n'étant pas lié au numérique. Par exemple, si toutes les infractions de viols avaient une Natinf générique, ces infractions, lorsqu'elles sont cochées « cyber » seraient bien entraînées avec l'apprentissage « cyber ». Néanmoins, les infractions de viols non cochées « cyber » ne seraient pas entraînées. Ainsi le champ lexical relatif au viol ne sera jamais entraîné comme n'étant pas lié au numérique, et l'algorithme risquera donc de prédire, presque systématiquement, les viols comme étant un fait lié au numérique. Il y a donc un risque de surestimation des faits liés au numérique pour ce type d'infractions, même si celui-ci risque d'être assez faible car les autres Natinf (ni spécifiques, ni génériques) englobent un champ très large de la délinquance.

Ainsi, aucun de ces croisements ne permet d'obtenir un apprentissage sans biais. C'est pourquoi une labellisation de l'échantillon est nécessaire pour ne pas créer de biais par la définition des données d'apprentissage. Cette labellisation sera détaillée dans la dernière partie de ce document.

Figure 5 : Echantillon d'apprentissage et échantillon de prédiction en fonction du type de Natinf



Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie en 2018.

Le processus de détection des infractions liées au numérique a néanmoins été réalisé avec la troisième variante d'échantillon d'apprentissage sur les données de 2018, pour produire de premiers résultats avant cette labellisation.

Après l'entraînement, l'objectif est de prédire les infractions liées au numérique sur les infractions génériques et spécifiques non cochées « cyber » par le gendarme ainsi que les autres Natinf cochées « cyber » par le gendarme (*figure 5*).

Les données d'apprentissage sont ici déséquilibrées. En effet, la classe « cyber » contient beaucoup moins d'infractions que la classe « non cyber ». En 2018, par exemple, la classe « cyber » contenait 51 000 infractions, alors que la classe « non cyber » en contenait 397 000.

Les données déséquilibrées créent un biais en faveur de la classe majoritaire (la classe « non cyber »). En effet, les algorithmes d'apprentissage font l'hypothèse que les classes sont équilibrées et cherchent à minimiser le taux d'erreur, c'est-à-dire le taux de mauvais classement. Ainsi les classificateurs seront bons pour la classe majoritaire mais peu précis dans la classe minoritaire, car manquer un négatif n'aura pas la même conséquence que de manquer un positif. En effet, ne pas classer une infraction non liée au numérique comme telle aura moins d'impact que d'oublier de classer une infraction liée au numérique comme telle. Par exemple, si 1 500 infractions sont mal classées à la fois dans la classe « cyber » et « non cyber », le taux d'erreur total sera de 0,7 %. Le taux d'erreur de la classe « non cyber » est bon, de 0,4 %. Néanmoins, le taux d'erreur pour la classe « cyber » est nettement moins bon (2,9 %).

Dans cette étude, ce qui importe le plus est de prédire correctement les infractions liées au numérique, c'est-à-dire de classer correctement les infractions dans la classe minoritaire, à la fois ne pas classer à tort comme liées au numérique des infractions qui ne le sont pas, et ne pas exclure de la classe cyber des infractions qui le sont. Il est donc nécessaire de prendre en compte ce déséquilibre de classe dans le processus.

Afin de réduire ce déséquilibre, la classe « cyber » a été enrichie des données de 2019 et 2020. Ainsi le rapport entre la classe minoritaire (« cyber ») et majoritaire (« non cyber ») qui était de 13 % passe alors à 49 %. La classe « cyber » qui contenait 8 fois moins d'infractions que la classe « non cyber », n'en contient dorénavant plus que 2 fois moins.

2.2. Le prétraitement

Les manières d'opérer sont rédigées manuellement par les gendarmes. Afin d'être utilisable pour l'analyse et compréhensible par l'algorithme, le texte va être découpé en mots, *tokens*, puis transformé en valeur numérique.

Il s'agit, néanmoins, de saisies manuelles qui peuvent contenir des erreurs. De plus, l'information recherchée est noyée au milieu d'un texte.

Tout d'abord, un nettoyage des manières d'opérer a été effectué avec notamment la suppression des accents et caractères spéciaux. De plus, certaines chaînes spéciales (numéro de téléphone, adresse URL, adresse mail) ont été remplacées par des chaînes de caractères communes. Par exemple, les adresses mails ont été remplacées par « AdresseMail ».

Le nombre de mots présents dans l'ensemble des manières d'opérer peut être très important, il est donc nécessaire de réduire la dimension. Pour cela une liste de mots-outils (articles, prépositions, pronoms,etc.) ainsi qu'une liste de mots non indispensables ont été utilisées pour supprimer les mots correspondants dans les Manop. Il s'agit de la liste des *stopwords* présent dans les packages *spacy* et *NLTK* de Python, enrichie par une liste de *stopwords* du package R du même nom. Par ailleurs, la liste des mots non indispensables contient, entre autres, les mots les plus récurrents, ainsi qu'une partie du champ lexical de la gendarmerie tels que les mots « auteurs », « victimes », « gendarmes », « plaintes », non discriminants pour l'analyse car ils ne permettront pas de distinguer une infraction liée au numérique ou non.

Par ailleurs, pour réduire davantage le nombre de mots, une lemmatisation des mots, qui consiste à ramener un mot sous sa forme canonique, a également été effectuée permettant ainsi de considérer que différentes formes d'un même mot (pluriels, conjugaison) sont équivalentes.

Enfin, une recherche de bigrammes ayant un lien fort avec le numérique a été effectuée afin d'améliorer le repérage de ce type de délinquance. En effet, certains mots peuvent ne pas avoir de

sens « cyber » individuellement, mais lorsqu'ils sont associés, peuvent être en lien avec le numérique. Par exemple, les mots « réseau » et « social », n'ont a priori pas de lien individuel avec le numérique. Néanmoins, le bigramme « Réseau social » est plus fréquemment associé au numérique. Cette liste a été formée en analysant les manières d'opérer des infractions liées au numérique. Les mots ont été joints 2 à 2 et l'occurrence de chaque bigramme a été calculée. Les bigrammes les plus récurrents et ayant un sens cyber ont été conservés.

Le prétraitement a permis de réduire de presque 70 % le nombre de mots pour l'ensemble des manières d'opérer. Ainsi, avant le prétraitement les manières d'opérer contenaient en moyenne 69 mots dont 51 mots différents, après elles en contiennent en moyenne 24, dont 21 différents.

Après le prétraitement, une analyse textuelle simple (nuage de mots) a été effectuée afin de repérer les thèmes présents dans chaque catégorie de Natinf.

Figure 6 : Nuage de mots des Manop des infractions ayant une Natinf spécifique enregistrées par la gendarmerie en 2018

Lecture : Le nuage de mots contient les 100 mots les plus fréquents dans les Manop des infractions ayant une Natinf spécifique. La taille des mots est proportionnelle à leur occurrence.
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie en 2018.

Les mots des manières d'opérer des infractions ayant une Natinf spécifique sont fortement liés au numérique. En effet, les mots les plus fréquents sont « contact », « compte », « OrgInternet[10] », « pirat[11] » qui appartiennent au champ lexical du numérique (*figure 6*). Par ailleurs, les mots « donnees », « ordinateur » appartenant aussi fortement au numérique sont visibles, et apparaissent régulièrement dans les manières d'opérer.

S'agissant des mots présents dans les Manop des infractions ayant une Natinf générique, ils renvoient au vocabulaire de la délinquance en général, avec des mots les plus fréquents comme « derober », « vol », « domicile », « constate », « euro » (*figure 7*). Les mots « telephone », « compte », « bancaire » apparaissent néanmoins assez fréquemment, et peuvent faire référence au numérique.

[10] « OrgInternet » fait référence aux chaînes de caractères liées à internet telles que les différents réseaux sociaux, sites internet, etc. Par exemple, Facebook et Facebook.com ont été remplacés par « Orginternet ».
[11] « Pirat » fait référence à la classe du mot piratage.

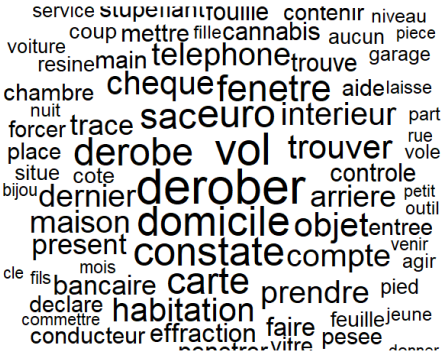
Figure 7 : Nuage de mots des Manop des infractions ayant une Natinf générique enregistrées par la gendarmerie en 2018

Lecture : Le nuage de mots contient les 100 mots les plus fréquents dans les Manop des infractions ayant une Natinf générique. La taille des mots est proportionnelle à leur occurrence.
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie en 2018.

Figure 8 : Nuage de mots des Manop des infractions, ayant une Natinf générique et étant cochées « cyber » par le gendarme, enregistrées par la gendarmerie en 2018

Lecture : Le nuage de mots contient les 100 mots les plus fréquents dans les Manop des infractions ayant une Natinf générique et étant cochées « cyber » par le gendarme. La taille des mots est proportionnelle à leur occurrence.
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrées par la gendarmerie en 2018.

Figure 9 : Nuage de mots des Manop des infractions, ayant une Natinf générique et n'étant pas cochées « cyber » par le gendarme, enregistrées par la gendarmerie en 2018

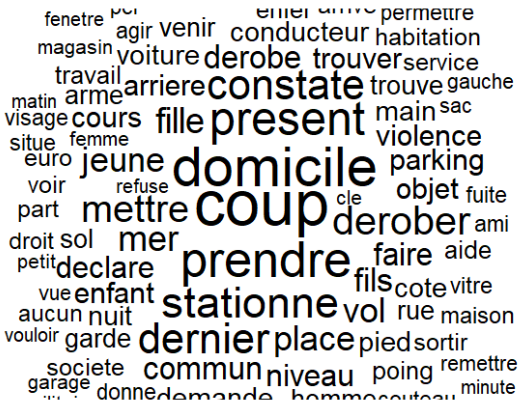


Lecture : Le nuage de mots contient les 100 mots les plus fréquents dans les Manop des infractions ayant une Natinf générique et n'étant pas cochées « cyber » par le gendarme. La taille des mots est proportionnelle à leur occurrence.
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie en 2018.

En se concentrant sur les manières d'opérer des infractions ayant une Natinf générique et étant cochées « cyber » par le gendarme, le champ lexical des Manop renvoie à un champ lexical du numérique (*figure 8*). En effet, des mots très fortement liés au numérique sont les plus fréquent tels que « compte », « euro », « orginternet », « mail », etc. Les mots des Manop pour les infractions ayant des Natinf génériques non cochées « cyber » renvoient, quant à eux, à des thèmes très généraux de

la délinquance (*figure 9*) comme c'est également le cas pour les mots des manières d'opérer des procédures dites « non cyber » (c'est-à-dire des procédures pour lesquelles les Natinf des infractions ne sont ni génériques, ni spécifiques) (*figure 10*).

Figure 10 : Nuage de mots des Manop des infractions, ayant une autre Natinf (ni spécifique, ni générique) enregistrées par la gendarmerie en 2018



Lecture : Le nuage de mots contient les 100 mots les plus fréquents dans les Manop des infractions ayant une autre Natinf. La taille des mots est proportionnelle à leur occurrence.
Champ: France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie en 2018.

2.3. Apprentissage des données

2.3.1. Echantillon apprentissage/test

Afin d'entraîner, d'optimiser et de tester l'algorithme les données ont été découpées en 3 échantillons : apprentissage, validation et test.

Tout d'abord, l'échantillon d'apprentissage permet d'entraîner le modèle afin de minimiser son erreur. Ensuite, l'échantillon de validation permet de choisir les valeurs optimales des hyper paramètres des modèles, notamment pour le modèle permettant le passage du texte en valeur numérique. Enfin, l'échantillon test est utilisé pour évaluer la performance de l'algorithme. En effet, comme l'algorithme est entrainé avec les données d'apprentissage pour minimiser les erreurs, il est important de ne pas tester la performance de l'algorithme sur ces mêmes données.

2.3.2. Passage du texte en valeur numérique

Après le nettoyage du texte et la réduction du nombre de mots, le texte est ensuite passé en valeur numérique sous forme d'un vecteur par prolongement lexical, en utilisant l'algorithme *Word2Vec[4]*. Il s'agit d'un réseau de neurones à une couche cachée.

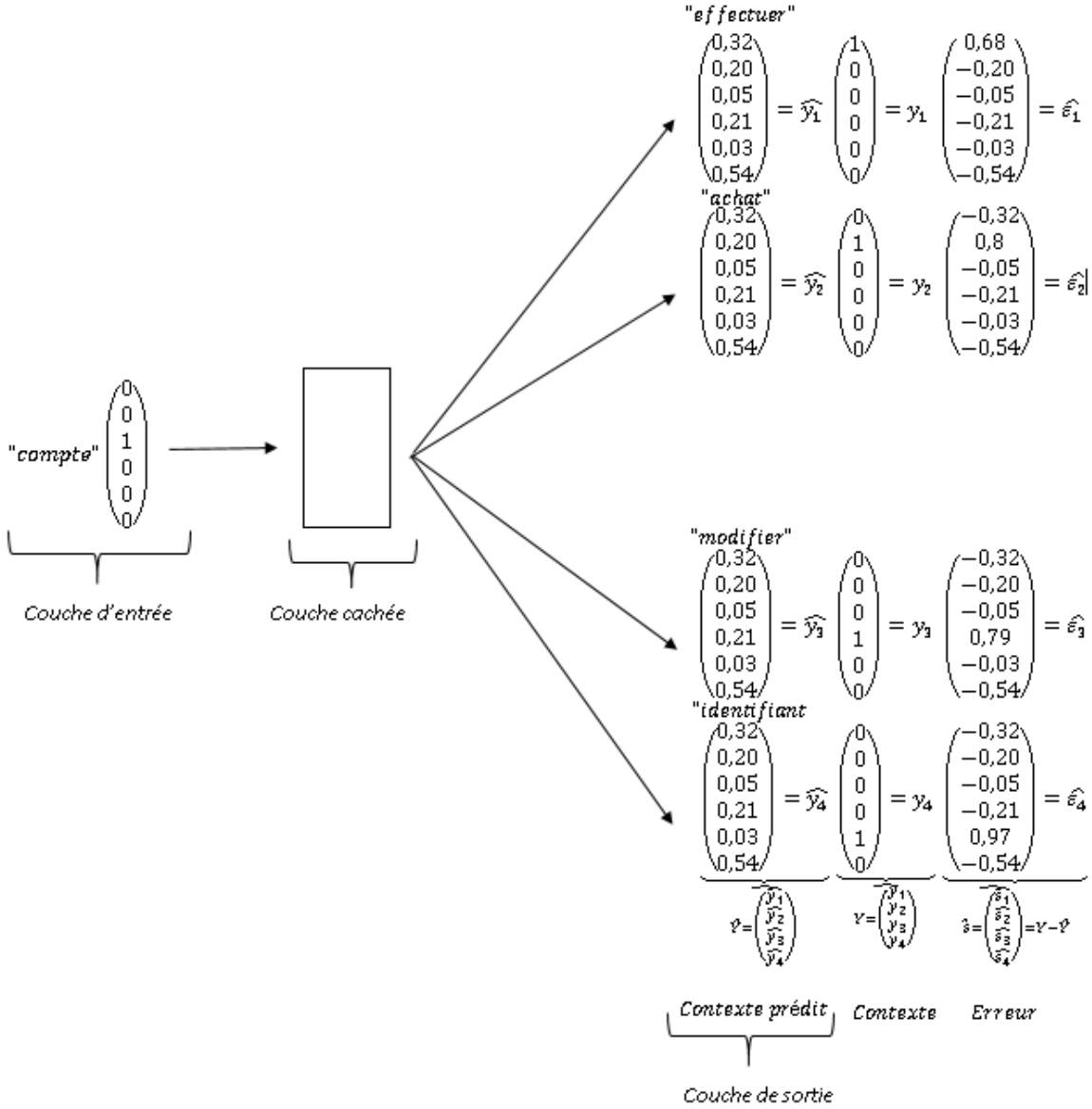
Le réseau va apprendre des représentations vectorielles pour que les mots qui ont un contexte[12] proche aient un vecteur proche. La méthode utilisée est le *Skip-Gram* qui consiste, à partir d'un mot, à prédire son contexte. Le vecteur du mot sera ainsi défini par les mots à proximité de ce mot dans les Manop.

[12] Le contexte d'un mot se rapporte au p mots qui entourent ce mot. p est généralement compris entre 4 et 15.

Plus explicitement, chaque mot du corpus[13] est stocké dans un dictionnaire. Pour chaque mot du dictionnaire, les poids synaptiques du réseau sont ajustés pour que la valeur prédite soit la plus proche possible de la valeur à prédire, c'est-à-dire pour que le contexte prédit soit le plus proche possible du contexte du mot. Pour ce faire, les poids synaptiques sont corrigés par retro propagation du gradient [5]. Il s'agit d'une méthode qui consiste à corriger les poids, de la dernière jusqu'à la première couche, proportionnellement à l'impact qu'il a sur l'erreur. Cet algorithme est effectué pour l'ensemble des mots des données d'entrainement.

La projection du document est ensuite réalisée par la moyenne des vecteurs projetés associés aux mots composant chaque document.

Figure 11 : Exemple de la méthode Skip-Gram de l'algorithme Word2Vec pour le passage du texte en valeur numérique



[13] Le corpus est composé de l'ensemble des Manop de nos données d'apprentissage.

Par exemple, en considérant pour simplifier le corpus suivant :

Avant le prétraitement	*« Le ou les auteurs effectuent un achat sur le compte de la victime et modifient l'identifiant de son adresse mail »*
Après le prétraitement	*« effectuer achat compte modifier identifiant adressemail»*

Le dictionnaire, noté D, est ainsi constitué de tous les mots uniques du corpus après le prétraitement. D={effectuer, achat, compte, modifier, identifiant, adressemail}

En prenant, par exemple, un contexte de taille p=2 (2 mots avant et 2 mots après) et en considérant le mot « compte », celui-ci aura pour contexte : C=({effectuer, achat, modifier, identifiant}).

La couche d'entrée du réseau de neurones sera donc ce mot avec un encodage *one-hot* [14] (ici $\begin{pmatrix}0\\0\\1\\0\\0\\0\end{pmatrix}$).

La couche de sortie est constituée d'un vecteur par mot du contexte. Le contexte prédit $\hat{Y}$ (figure 11) sera ensuite comparé aux valeurs du contexte Y afin que l'erreur $\hat{\varepsilon}$ soit la plus faible possible. On effectue ceci pour l'ensemble du dictionnaire.

2.3.3. Choix du modèle pour prédire les infractions liées au numérique

Après avoir nettoyé le texte, et l'avoir rendu interprétable pour l'algorithme, différentes méthodes de prédiction peuvent être utilisées. Nous avons testé plusieurs modèles tels que le *Support Vector Machines* (SVM), le *Multi Layer Perceptron* (MLP), l'*Extreme gradient boosting* (XGBoost), la régression logistique, etc. Pour la suite de l'étude, le modèle MLP a été retenu, il s'agit du modèle le plus performant.

2.3.3.1. Multi Layer Perceptron

La méthode MLP est un réseau de neurones qui est composé de plusieurs couches. Contrairement au réseau de neurones précédent qui comporte une seule couche cachée, MLP peut contenir plusieurs couches cachées. Dans ce réseau de neurones, la couche d'entrée est constituée des vecteurs numériques associés à chaque document, une ou plusieurs couches intermédiaires cachées et enfin une couche de sortie qui correspondra à la probabilité d'appartenir ou non au champ du numérique. La phase d'apprentissage pour ce modèle va permettre d'apprendre et d'ajuster les poids synaptiques pour que la valeur prédite soit la plus proche possible de la valeur à prédire, c'est-à-dire que la probabilité que l'infraction relève du numérique pour la classe « cyber » soit élevée, ou que la probabilité que l'infraction relève du numérique soit faible pour la classe « non cyber ».

Pour ce faire, les poids synaptiques sont dans un premier temps choisis aléatoirement. La première étape consiste à calculer, pour chaque document, la valeur en sortie, puis dans un second temps la valeur de sortie sera comparée à la valeur à prédire. Les poids seront ensuite corrigés par la retro propagation du gradient. En effectuant ceci pour l'ensemble des données d'apprentissage, les poids des réseaux de neurones seront équilibrés et l'erreur sera ainsi minimisée.

[14] L'encodage one-hot consiste à transformer une variable en un vecteur composé de 0 et de 1. Le vecteur de taille n, sera composé d'un seul 1, correspondant, ici, à la position du mot dans la Manop.

2.3.3.2. Optimisation des hyper paramètres

Les hyper paramètres vont contrôler la précision du modèle. Ainsi, afin d'améliorer la précision et l'efficacité du modèle, il est nécessaire de choisir les valeurs optimales des hyper paramètres. Pour ce faire, la méthode de recherche par grille (*Grid Search*) a été utilisée. Il s'agit d'une méthode d'optimisation exhaustive qui testera toutes les combinaisons possibles. Pour chaque hyper paramètre, on choisit une liste de valeurs, et la méthode va permettre de déterminer la combinaison d'hyper paramètres optimale.

2.3.3.3. Performance du modèle

Les données d'apprentissage étant légèrement déséquilibrées, les métriques utilisées pour choisir le modèle et calculer la performance de celui-ci seront des métriques qui ne sont pas sensibles au déséquilibre de classe, comme le F1-score ou le rappel.

Commençons par représenter la matrice de confusion :

		Classes réelles	
		Négative	Positive
Classes prédites	Négative	Vrais négatifs	Faux négatifs
	Positive	Faux positifs	Vrais positifs

Dans notre étude, il est aussi « grave » de classer une infraction comme liée au numérique alors qu'elle ne l'est pas, c'est-à-dire d'avoir un faux positif, que d'oublier de classer une infraction liée au numérique, c'est-à-dire avoir un faux négatif. Le F1-Score permet donc de prendre en compte ces deux aspects, en calculant la moyenne harmonique de la précision et du rappel.

$$F1 - score = 2 * \frac{precision * rappel}{precision + rappel}$$

D'une part, le rappel est la part de vrais positifs parmi le nombre total de positifs, c'est-à-dire la proportion d'infractions bien classées comme liées au numérique par notre modèle parmi le nombre total d'infractions liées au numérique. Il s'agit donc de la proportion d'infractions liées au numérique correctement prédites.

$$rappel = \frac{Vrais\ positifs}{Vrais\ positifs + Faux\ négatifs}$$

D'autre part, la précision est la part de vrais positifs parmi le nombre total de positifs prédits par le modèle.

$$precision = \frac{Vrais\ positifs}{Vrais\ positifs + Faux\ positifs}$$

Figure 12 : Performance du modèle MLP

	Sans optimisation des hyper paramètres	Avec optimisation des hyper paramètres
F1-Score	0,96	0,97
Rappel	0,96	0,97

Lecture : Avec l'optimisation des hyper paramètres, 97 % des infractions liées au numérique sont correctement prédites avec le modèle MLP (rappel).
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie en 2018.

L'optimisation des paramètres a permis d'améliorer très légèrement les résultats. Ainsi, le F1-score et le rappel du modèle MLP optimisé valent 0,97 ; c'est-à-dire que 97 % des infractions liées au numérique sont correctement prédites. Une infraction est prédite « cyber » lorsque la probabilité de prédiction du modèle MLP est supérieure à 0,5.

2.4. Résultat du processus

2.4.1. Analyse des infractions mal classées par l'algorithme pendant l'apprentissage

2.4.1.1. La classe « cyber »

Figure 13 : Proportion d'infractions mal classées au sein de la classe cyber en distinguant les infractions ayant une Natinf spécifique et celles ayant une Natinf générique

	Données d'entraînement 2018-2020				
	Nombre d'infractions total de la classe « cyber »	**Proportion**	**Mauvais classement**	**Part dans le total des mauvais classements**	**Part de mauvais classement au sein du type de Natinf**
Spécifiques	17 664	9 %	932	14 %	5 %
Génériques	178 067	91 %	5 595	86 %	3 %
Total	195 731	100 %	6 527	100 %	3 %

Lecture : 5 % des infractions avec une Natinf spécifique sont mal classées.
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2017 et 2020.

Au sein de la classe « Cyber », la grande majorité des infractions dont la classification est erronée concerne les infractions ayant une Natinf générique. En effet, dans les données d'entraînement 86 % des infractions mal classées ont une Natinf générique (*figure 13*). Néanmoins la part d'infractions ayant une Natinf générique représente 91 % des données d'entraînement de la classe « Cyber ». En

rapportant le nombre d'infractions mal classées au nombre total d'infractions de son type de Natinf dans les données d'entraînement, les infractions ayant une Natinf spécifique sont celles qui sont le moins bien classées : 5 % des infractions étant cochées « cyber » avec une Natinf spécifique sont mal classées ; cela concerne 3 % des infractions cochées « cyber » avec une Natinf générique.

Au sein des infractions génériques mal classées, certaines infractions sont très fréquentes : on peut retrouver, par exemple, les infractions d'appropriations frauduleuses et chantages concernant des vols (vol aggravé par deux circonstances, vol à la roulotte, vol par effraction) mais également les violences suivies d'incapacité n'excédant pas 8 jours par personne étant ou ayant été conjoint, concubin ou partenaire lié à la victime par un pacte civil de solidarité : la totalité de ces infractions sont mal classées (*annexe 1*). Néanmoins, ces infractions sont très peu présentes dans les données et représentent moins de 0,1 % des observations.
Les Manop des procédures concernant les violences suivies d'incapacité n'excédant pas 8 jours par personne étant ou ayant été conjoint, concubin ou partenaire lié à la victime par un pacte civil de solidarité ne font pas référence à des faits liés au numérique. Par exemple, dans la Manop « *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* », aucun mot ne permet de déterminer qu'il s'agit d'un fait lié au numérique. Pour les vols aggravés par 2 circonstances, dans la grande majorité des cas, les Manop ne font également pas référence à des faits liés au numérique. Pour la seule infraction dont la Manop fait référence au numérique, il s'agit d'une prise de contact sur un réseau social. Par 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 ou 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. […]* »

S'agissant des infractions spécifiques mal classées, les infractions concernant l'interruption volontaire des communications électroniques, les viols commis par une personne mise en contact avec la victime par réseau de télécommunications subissent, quasi-systématiquement, une mauvaise classification dans les données d'entraînements (*annexe 2*). Ce sont, là aussi, des infractions peu présentes dans les données. Les Manop faisant référence aux infractions d'interruption volontaire des communications électroniques font toutes référence à des mots liés au numérique. Par exemple l'infraction de la procédure associée à la Manop « *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. »* est bien une infraction liée au numérique. Néanmoins, comme précédemment, il s'agit d'un champ lexical spécifique, qui ne doit, probablement, pas intervenir dans les Manop d'autres infractions. Comme l'effectif est faible, l'algorithme n'a pas suffisamment été entraîné pour bien classer ces infractions.
Pour les viols commis par une personne mise en contact avec la victime par réseau de télécommunications, les Manop évoquent effectivement la mise en contact par téléphone, mais se concentrent davantage sur l'infraction de viol. Ainsi, l'information pouvant permettre de classer l'infraction comme étant liée au numérique est un peu « noyée » par les informations concernant les viols qui ont plus de poids.

2.4.1.2. La classe « non cyber »

En s'intéressant maintenant à la classe « non cyber », seules 0,4 % des infractions sont mal classées. Comme précédemment, certaines infractions sont assez fréquemment mal classées par l'algorithme, néanmoins les proportions sont plus faibles. Par exemple, les infractions de recel de bien provenant

d'abus de confiance par personne recouvrant des fonds ou des valeurs pour le compte tiers, sont presque toutes mal classées (*annexe 3*). Néanmoins, il s'agit d'infractions issues d'une seule procédure, avec la même Manop. Cette Manop fait, en revanche, référence à des faits liés au numérique : « *La mise en cause fait une rencontre sur internet. Elle entretient une relation < amoureuse > avec cet homme pendant un 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 et de le réexpédier par même voie au BENIN. Elle fait ainsi transiter environ 60 000€ qui proviennent d'escroqueries diverses commises par le biais d'internet* ». En effet, elle contient des termes liés à la thématique « cyber » comme « internet » ou « mandat cash ». De plus, presque deux tiers des infractions concernant l'expédition de correspondance à découvert contenant une diffamation est mal classé. Les Manop de ces infractions, comme par exemple «*La victime est président du conseil des parents d'élèves de la ****. A l'issu 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 ».* » ne fait pas nécessairement référence au numérique, mais certains termes comme « courrier » peuvent être souvent présents dans les Manop des infractions liées au numérique et dans un contexte de mots liés au numérique.

2.4.2. Résultats de la prédiction sur les autres catégories

3 catégories d'infractions vont être étudiées :
- les infractions génériques non cochées « cyber » par le gendarme
- les infractions spécifiques non cochées « cyber » par le gendarme
- les autres infractions (ni spécifiques ni génériques) non cochées « cyber » par le gendarme

2.4.2.1. Infractions non cochées « cyber » avec une Natinf spécifique

S'agissant des infractions avec une Natinf spécifique et n'étant pas cochées « cyber » par le gendarme, l'analyse se limitera à la période 2017-2018 puisqu'au-delà, les infractions associées à des Natinf spécifiques sont automatiquement cochées « cyber » dans le logiciel de rédaction des procédures de la gendarmerie. Entre 2017 et 2018, seules 13 % des infractions ayant une Natinf spécifique et n'étant pas cochées « cyber » sont prédites « non cyber » par l'algorithme.

Une analyse des champs lexicaux (sous forme de nuages de mots) des manières d'opérer a été réalisée pour :

1- Les infractions avec une Natinf spécifique, non cochées « cyber » par le gendarme et prédites « non cyber » par l'algorithme (*figure 14*)
2- Les infractions avec une Natinf spécifique, non cochées « cyber » par le gendarme et prédites « cyber » par l'algorithme (*figure 15*)
3- Les infractions avec une Natinf spécifique, cochées « cyber » par le gendarme et prédites « cyber » par l'algorithme (*figure 16*)

Les champs lexicaux des deux derniers types d'infractions sont très proches et très fortement liés au numérique avec des termes les plus fréquents tels que « compte », « orgInternet », « pirat », etc. Le champ lexical des infractions non cochées « cyber » et prédites « non cyber » par l'algorithme est assez différent avec des termes les plus fréquents comme « jeune », « fille », « sexuel » et les mots pouvant faire référence à des faits liés au numérique sont moins présents.

Figure 14 : Nuage de mots des Manop des infractions enregistrées par la gendarmerie, ayant une Natinf spécifique non cochées « cyber » par le gendarme et prédites « non cyber » par l'algorithme

Lecture : Le nuage de mots contient les 100 mots les plus fréquents dans les Manop des infractions ayant une Natinf spécifique, non cochées « cyber » et prédites « non cyber ». La taille des mots est proportionnelle à leur occurrence.
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2017 et 2020.

Figure 15 : Nuage de mots des Manop des infractions enregistrées par la gendarmerie, ayant une Natinf spécifique non cochées « cyber » par le gendarme et prédites « cyber » par l'algorithme

Lecture : Le nuage de mots contient les 100 mots les plus fréquents dans les Manop des infractions ayant une Natinf spécifique, non cochées « cyber » et prédites « cyber ». La taille des mots est proportionnelle à leur occurrence.
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2017 et 2020.

Figure 16 : Nuage de mots des Manop des infractions enregistrées par la gendarmerie, ayant une Natinf spécifique cochées « cyber » par le gendarme et prédites « cyber » par l'algorithme

Lecture : Le nuage de mots contient les 100 mots les plus fréquents dans les Manop des infractions ayant une Natinf spécifique, cochées « cyber » et prédites « cyber ». La taille des mots est proportionnelle à leur occurrence.
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2017 et 2020.

En étudiant les Manop des infractions non cochées « cyber » et prédites « non cyber » par l'algorithme pour les 3 Natinf les plus fréquentes, à savoir « modification de données résultant d'un accès frauduleux à un système de traitement automatisé», « accès frauduleux dans un système de traitement automatisé de données » et « diffusion de l'image d'un mineur à caractère pornographique via un réseau de télécommunications », il s'agit dans la majorité des cas d'infractions liées au numérique qui peuvent concerner des accès frauduleux aux ordinateurs, boîte mail, comptes (réseaux sociaux, sites internet,…) comme c'est le cas dans la Manop : *« La plaignante constate sur son ordinateur portable qu'un tiers a accédé à plusieurs fichiers à des dates et heures auxquelles elle se trouvait au travail. Elle soupçonne fortement son ex-mari informaticien. Cependant aucune effraction n'a été commise au domicile. »*

Il peut également s'agir de diffusions de photos ou vidéos à caractère sexuel qui ont suivi un acte consenti ou non. Dans la plupart de ces infractions, la Manop est principalement concentrée sur l'infraction ou le contexte non cyber ce qui peut expliquer la prédiction « non cyber » comme dans la Manop : «*Courant novembre 2017 un mercredi après-midi la victime joue à la console chez un copain. Deux autres copains arrivent. Ils jouent à un jeu de qui arrivera à enlever de force un vêtement à l'autre. Le jeu se retourne contre la victime qui dit avoir été déshabillée de force. Le dernier garçon prend une photo de la victime et la diffuse au collège via [réseau social]* »

Il y a néanmoins des infractions qui ne semblent pas être liées au numérique au vu du contenu de la Manop : « *Le directeur d'agence s'aperçoit que de la colle est présente tout autour du dispositif d'insertion de carte bancaire sur le DAB extérieur. Apres plusieurs jours il n'a pas eu de porte à sa connaissance de clients victimes d'utilisations ou retraits frauduleux.*».

Figure 17 : Probabilité de prédiction du total des infractions enregistrées par la gendarmerie (en %)

	Q1	Médiane	Q3
Total des infractions prédites « non cyber » par l'algorithme	0,04	0,20	0,85
Total des infractions prédites « cyber » par l'algorithme	90,43	99,30	99,93

Lecture : 50 % (médiane) des infractions prédites « non cyber » par l'algorithme ont une probabilité de prédiction inférieure à 0,20 %
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2017 et 2020.

Figure 18 : Probabilité de prédiction des infractions avec une Natinf spécifique, non cochées « cyber » et prédites « non cyber » par l'algorithme (en %)

	Q1	Médiane	Q3
Infractions avec une Natinf spécifique non cochées prédites « non cyber » par l'algorithme	1,25	9,52	25,80

Lecture : 50 % (médiane) des infractions avec une Natinf spécifique, non cochées « cyber » prédites « non cyber » par l'algorithme ont une probabilité de prédiction inférieure à 9,52 %
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2017 et 2020.

Par ailleurs, la probabilité de prédiction « cyber », issue du MLP, des infractions ayant une Natinf spécifique et qui ont été prédites « non cyber » est assez élevée par rapport à la probabilité de prédiction du total des infractions prédites. En effet, alors que 75 % des infractions prédites « non cyber » par l'algorithme ont une probabilité de prédiction inférieure à 0,85 % (*figure 17*), 75 % des infractions spécifiques ont une probabilité de prédiction supérieure à 1,25 % et 25 % ont une probabilité de prédiction supérieure à 25,8 % (*figure 18*).

En conclusion, les infractions ayant des Natinf spécifiques sont dans une très grande majorité des cas des infractions liées au numérique. Les infractions non cochées « cyber » sont donc très probablement des oublis de coche. De plus, même si ces infractions sont prédites « non cyber » par l'algorithme, leurs probabilités de prédictions restent assez élevées.

2.4.2.2. Infractions non cochées « cyber » avec une Natinf générique

Entre 2017 et 2020, 10 % des infractions ayant une Natinf générique et non cochées « cyber » par le gendarme ont été prédites « cyber » par l'algorithme.

Une analyse des infractions pour les 3 Natinf génériques les plus présentes dans la base (qui représentent presque 50 % des infractions ayant une Natinf générique) a été réalisée afin d'étudier plus finement ces infractions. Il s'agit :

- des escroqueries
- des vols simples
- des vols par effraction dans un local d'habitation ou un lieu d'entrepôt

Tout d'abord, la part d'infractions prédites « cyber » au sein de chaque Natinf est très variable. Alors que les infractions d'escroqueries non cochées « cyber » sont très souvent prédites « cyber » par l'algorithme (64,5 %) (*figure 19*), celles concernant les vols simples et les vols par effraction dans un local d'habitation ou un lieu d'entrepôt sont plus rarement prédites « cyber » avec respectivement 4,5 % et 0,02 % des infractions non cochées « cyber » prédites « cyber » par l'algorithme.

Figure 19 : Part des infractions prédites « cyber » par l'algorithme parmi les infractions non cochées « cyber » par le gendarme (en %)

	Part des infractions prédites « cyber » par l'algorithme parmi les infractions non cochées « cyber » par le gendarme (en %)
Escroqueries	64,5
Vols simples	4,5
Vols par effraction dans un local d'habitation ou un lieu d'entrepôt	0,02

Lecture : 64,5 % des infractions d'escroqueries non cochées « cyber » sont prédites « cyber » par l'algorithme.
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2017 et 2020.

Figure 20 : Probabilité de prédiction des infractions non cochées « cyber », prédites « cyber » par l'algorithme pour les 3 Natinf génériques les plus fréquentes (en %)

	Q1	Médiane	Q3
Escroqueries non cochées « cyber » prédites « cyber » par l'algorithme	94,60	99,56	99,94
Vols simples non cochés « cyber » et prédits « cyber » par l'algorithme	68,84	86,44	97,07
Vols par effraction non cochés « cyber » et prédits « cyber » par l'algorithme	59,37	76,84	92,02

Lecture : 25 % (Q1) des infractions d'escroqueries non cochées « cyber » et prédites « cyber » par l'algorithme ont une probabilité de prédiction inférieure à 94,60 %
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2017 et 2020.

Les escroqueries

Figure 21 : Nuage de mots des Manop des infractions d'escroqueries enregistrées par la gendarmerie, prédites « cyber » par l'algorithme

Lecture : Le nuage de mots contient les 100 mots les plus fréquents dans les Manop des infractions d'escroqueries, prédites « cyber ». La taille des mots est proportionnelle à leur occurrence.
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2017 et 2020.

Figure 22 : Nuage de mots des Manop des infractions d'escroqueries enregistrées par la gendarmerie, non cochées « cyber » par le gendarme et prédites « non cyber » par l'algorithme

Lecture : Le nuage de mots contient les 100 mots les plus fréquents dans les Manop des infractions d'escroqueries, non cochées « cyber » et prédites « non cyber ». La taille des mots est proportionnelle à leur occurrence.
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2017 et 2020.

Dans un premier temps, les champs lexicaux des escroqueries prédites « cyber » par l'algorithme (cochées et non cochées « cyber » par le gendarme) font bien référence au numérique (« virement », « orginternet », « compte ») (*figure 21*). Celui des infractions non cochées « cyber » et prédites « non cyber » par l'algorithme ne fait, en revanche, pas référence au numérique (*figure 22*). Après avoir analysé quelques Manop d'escroqueries non cochées « cyber » par le gendarme et prédites « cyber » par l'algorithme, elles sont, dans de nombreux cas, des infractions d'escroqueries en ligne ou par téléphone comme on le voit par exemple avec la Manop « *Victime reçoit un appel l'avisant qu'elle a gagné la somme de 1500 euros mais elle doit rappeler un numéro en 08. Elle s'exécute et au bout d'une demi-heure elle s'aperçoit que c'est une arnaque.* ». Néanmoins certaines infractions non cochées « cyber » et prédites « cyber » concernent des escroqueries bancaires (vols de cartes bancaires, utilisation de chèques volés) qui ne sont pas des infractions liées au numérique mais dont le champ lexical des Manop est proche de celui des Manop des infractions « cyber ».

Les Manop des infractions non cochées « cyber » et prédites « non cyber » par l'algorithme ne font, effectivement, pas référence à des faits liés au numérique. Il s'agit principalement d'escroqueries liées aux cartes bancaires et aux chèques.

Les probabilités de prédictions des escroqueries non cochées « cyber » et prédites « cyber » sont aussi élevées que les probabilités de prédictions du total des infractions prédites « cyber » par l'algorithme, alors que certaines infractions ne relèvent pas du numérique (*figure 17 et figure 20*). Cela peut s'expliquer par des mots dans les manières d'opérer souvent présents dans les Manop des infractions « cyber ».

Les vols simples

S'agissant des vols simples non cochés « cyber » et prédits « cyber » par l'algorithme, il s'agit, dans de nombreux cas, de vols de colis suite à l'achat sur internet ou de vols de carte bancaire suivi de son utilisation. Il ne s'agit donc pas nécessairement d'infractions liées au numérique. Néanmoins, comme dans les Manop de ces infractions des mots pouvant faire référence au numérique sont souvent présents (les sites internet par exemple), l'algorithme associe ces Manop à des infractions liées au numérique (exemple : *« La victime consulte un site de vente de vêtements en ligne [site internet] sur lequel elle a effectué une commande. Elle constate qu'un colis est répertorié "livré" depuis le 02 juin or elle n'a rien reçu dans sa boite aux lettres. »)*

Même si ces infractions sont prédites « cyber » leur probabilité de prédiction reste néanmoins plus faible que pour le total des infractions prédites « cyber », avec 25 % des infractions qui ont une probabilité de prédiction inférieure à 69 % (*figure 20*), contre 90 % pour le total (*figure 17*).

Les vols par effraction dans un local d'habitation ou un lieu d'entrepôt

Enfin, le champ lexical des vols par effraction non cochés « cyber » mais prédits « cyber » par l'algorithme est également associé au compte et à la carte bancaire mais des mots tels que « orginternet », « ordinateur », « ligne » (*figure 23*) peuvent laisser penser que ces infractions ont eu lieu dans le numérique. En analysant les quelques Manop, il s'agit dans un tiers des cas d'infractions liées au numérique tels que des vols avec une mise en contact sur internet (« *A la suite d'une annonce sur ***** la victime reçoit un appel l'acheteur ne se présente jamais mais a réussi à récupérer l'adresse de la victime au préalable. Quatre jours après les deux moto cross ont été dérobés*. ») par exemple. Néanmoins, dans le reste des cas, il ne s'agit pas d'infractions liées au numérique. Il est principalement question de vols par effraction pour lesquels la victime a été avertie par mail (« *Les victimes reçoivent un appel téléphonique suivi d'un mail de leur système d'alarme leur informant l'intrusion de deux individus dans leur domicile. Durant cette période les victimes sont à la commune*

de [nom de commune] dans une zone ou le réseau téléphonique ne passe pas. Lorsqu'ils consultent leur messagerie il est 17 heures 30 minutes. Ils se rendent directement à la brigade de gendarmerie de [nom de la ville] pour nous aviser de ce cambriolage. »), de vols d'ordinateurs et ou de vols d'objets avec des objets retrouvés sur des sites internet par exemple. Les Manop de ces infractions contiennent donc des mots liés au numérique alors qu'elles ne sont pas considérées comme telles.

Figure 23 : Nuage de mots des Manop des infractions de vols par effraction enregistrées par la gendarmerie, non cochées et prédites « cyber » par l'algorithme

Lecture : Le nuage de mots contient les 100 mots les plus fréquents dans les Manop des infractions de vols par effraction, non cochées « cyber » et prédites « cyber ». La taille des mots est proportionnelle à leur occurrence.
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2017 et 2020.

De plus, de la même façon que les vols simples, même si ces infractions sont prédites « cyber » leur probabilité de prédiction reste néanmoins plus faible que pour le total des infractions prédites « cyber », avec 25 % des infractions qui ont une probabilité de prédiction inférieure à 59 % (*figure 20*).

En conclusion, dans certains cas, les infractions ayant une Natinf générique non cochée « cyber » et prédite « cyber » par l'algorithme sont effectivement des infractions liées au numérique, notamment avec les escroqueries, où la majorité des infractions prédites « cyber » semblent effectivement être des infractions liées au numérique. Néanmoins, la classification binaire « cyber », « non cyber » semble moins bien adaptée pour certaines infractions comme celles liées aux cartes bancaires, puisque les Manop des infractions bancaires liées et non liées au numérique semblent avoir un champ lexical assez proche. Ce problème est aussi présent entre les arnaques sur les sites internet et les vols de colis suite à un achat sur un site internet. Ce problème sera plus facilement corrigé avec la labellisation qui est décrite dans la suite de ce document.

Le seuil relatif à la probabilité de prédiction, qui a ici été choisi à 0,5 pour classer une infraction comme « cyber » pourra être adapté en fonction du label.

2.4.2.3. Infractions cochées cyber avec une autre Natinf (ni spécifique, ni générique)

Concernant, les autres Natinf (ni spécifique ni générique), entre 2017 et 2020, 48 % de ces infractions cochées « cyber » par le gendarme ont été prédites « cyber » par l'algorithme. Le champ lexical de ces infractions fait référence à des mots liés au numérique tels que « orginternet », « compte », « site », « réseau social » (*figure 24*).

Figure 24 : Nuage de mots des Manop des autres infractions enregistrées par la gendarmerie, cochées « cyber » et prédites « cyber » par l'algorithme

Lecture : Le nuage de mots contient les 100 mots les plus fréquents dans les Manop des autres infractions, cochées « cyber » et prédites « cyber ». La taille des mots est proportionnelle à leur occurrence.
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2017 et 2020.

Figure 25 : Nuage de mots des Manop des autres infractions enregistrées par la gendarmerie, cochées « cyber » et prédites « non cyber » par l'algorithme

Lecture : Le nuage de mots contient les 100 mots les plus fréquents dans les Manop des autres infractions, cochées « cyber » et prédites « non cyber ». La taille des mots est proportionnelle à leur occurrence.
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2017 et 2020.

Le champ lexical des autres infractions cochées « cyber » mais prédites « non cyber » est assez différent (*figure 25*) mais peut quand même être lié au numérique comme le mot « vente » qui peut faire référence à de la vente frauduleuse sur internet. En analysant un échantillon de Manop, celles contenant le mot « vente » concernent, effectivement dans la très grande majorité des cas, des ventes frauduleuses sur internet ; ces infractions ne sont peut-être pas assez entraînées par l'algorithme. Néanmoins les autres Manop (qui ne font pas référence à la vente) ne semblent pas être liés au numérique.

Par ailleurs, les probabilités de prédiction de ces autres infractions liées à la vente sont plus élevées que pour les autres infractions non liées à la vente cochées « cyber » et non prédites « cyber » avec 25 % des infractions liées à la vente qui ont une probabilité de prédictions supérieure à 23 % contre 8 % pour les autres infractions (*figure 26*).

Après avoir analysé les résultats de l'algorithme sur l'échantillon d'apprentissage ainsi que sur l'échantillon de prédiction, une comparaison entre la définition proposée par le groupe de travail piloté par le SSMSI en 2014 et les prédictions de l'algorithme a été réalisée.
Pour rappel, la définition proposée par le groupe de travail considère qu'une infraction est liée au numérique si sa Natinf appartient à la liste des Natinf spécifiques ou qu'elle appartient à la liste des Natinf génériques et qu'elle est cochée « cyber » par le gendarme, ou qu'elle a eu lieu sur internet.
L'ensemble des infractions enregistrées chaque année avec une Manop non vide par la gendarmerie a été prédite par l'algorithme.

Figure 26 : Probabilité de prédiction des autres infractions cochées « cyber » et prédites « non cyber » par l'algorithme (en %)

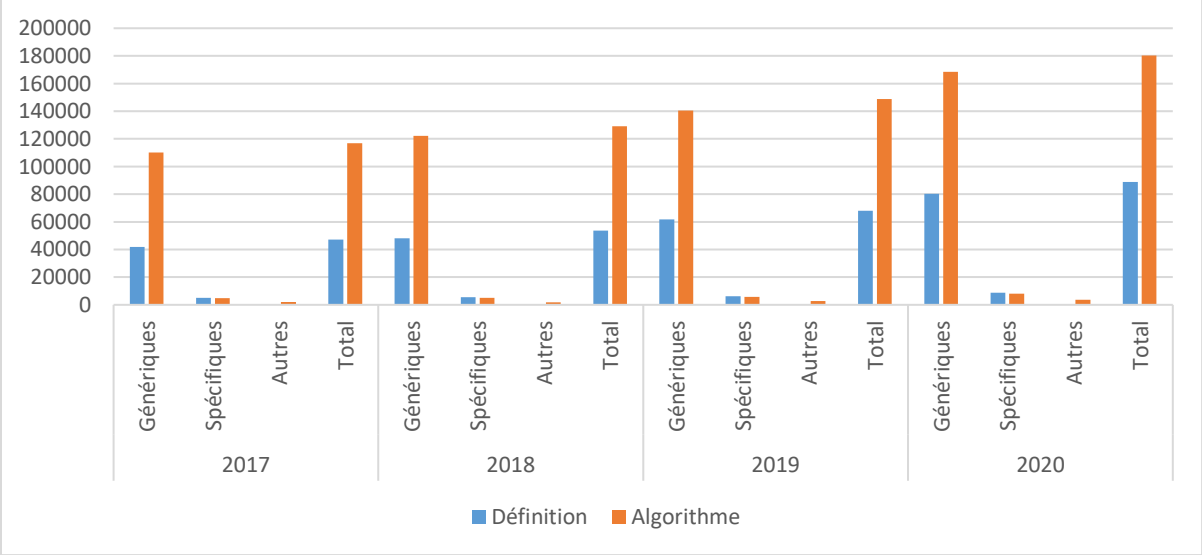
	Q1	Médiane	Q3
Autres infractions cochées « cyber », prédites « non cyber » par l'algorithme	0,08	1,36	10,88
Autres infractions cochées « cyber » relatives à la vente et prédites « non cyber » par l'algorithme	2,78	7,79	22,90
Autres infractions cochées « cyber » non relatives à la vente et prédites « non cyber » par l'algorithme	0,05	0,61	8,23

Lecture : 75 % (Q3) des autres infractions cochées « cyber » et prédites « non cyber » par l'algorithme ont une probabilité de prédiction inférieure à 10,88 %
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2017 et 2020.

Avertissement : Les résultats suivants sont des résultats expérimentaux sur les données enregistrées par la gendarmerie dans le cadre d'un travail méthodologique.

2.4.3. Comparaison entre le nombre d'infractions liées au numérique selon la définition par le groupe de travail et selon l'algorithme

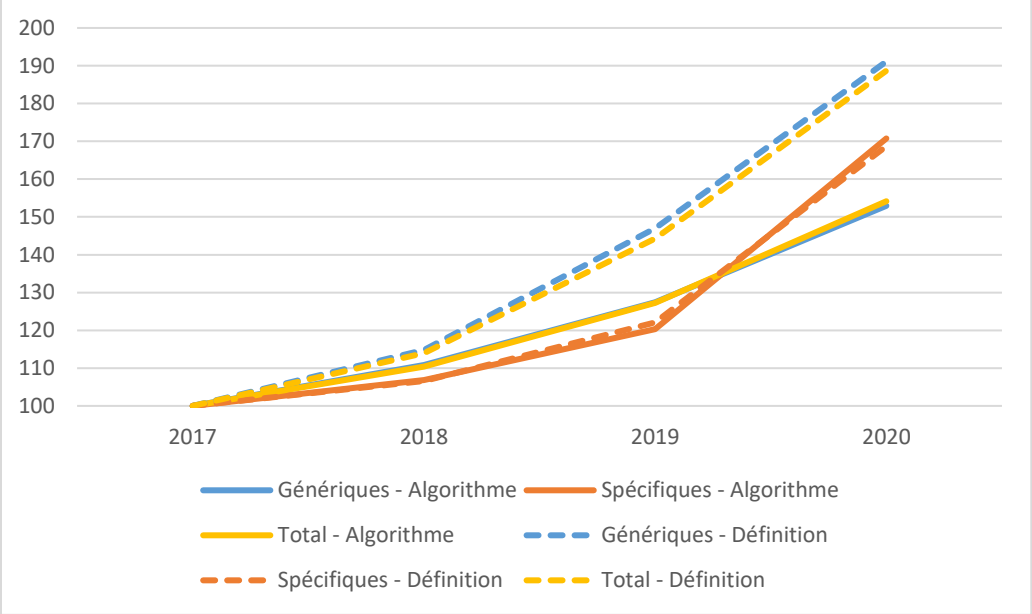
Figure 27 : Comparaison du nombre d'infractions liées au numérique avec l'algorithme et selon la définition par type de Natinf



Lecture : En 2017, 47 000 infractions sont liées au numérique au sens de la définition du groupe de travail et 117 000 selon l'algorithme.
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2017 et 2020.

En comparant le nombre d'infractions liées au numérique au sens de la définition préconisée par le groupe de travail au nombre prédit par l'algorithme, des écarts très marqués sont observés. Hormis pour les infractions ayant une Natinf spécifique, où le nombre obtenu avec la définition est très proche du nombre obtenu par l'algorithme, le nombre d'infractions liées au numérique estimé par l'algorithme est très largement supérieur au nombre obtenu avec la définition du groupe de travail (*figure 27*). Sur la période 2017-2020, pour les infractions avec une Natinf générique, le nombre d'infractions liées au numérique selon l'algorithme est environ deux fois plus élevé que celui obtenu avec la définition. Néanmoins, comme évoqué et vu précédemment, la prédiction « cyber » par l'algorithme des infractions ayant une Natinf générique risque d'être surestimée au vu de l'apprentissage choisi. C'est pourquoi une labellisation et une validation des résultats seront nécessaires afin de ne pas trop surestimer le nombre d'infractions prédites « cyber ». Cette validation sera effectuée par l'analyse d'un échantillon de procédures par des personnels métiers (policiers et gendarmes). Par ailleurs, certaines infractions n'ayant ni une Natinf spécifique ni une Natinf générique ont été prédites « cyber » par l'algorithme (en moyenne 2 500 par an), alors qu'elles ne sont pas considérées comme telles avec la définition. Ainsi, le nombre total d'infractions liées au numérique a plus que doublé passant de 54 000 avec la définition à 129 000 avec l'algorithme en 2018.

Figure 28 : Comparaison de l'évolution du nombre d'infractions liées au numérique au sens de la définition et prédites par l'algorithme entre 2017 et 2020 (données en base 100 en 2017)



Lecture : En 2017 et 2019, le nombre d'infractions liées au numérique au sens de la définition du groupe de travail a augmenté de de 44 % et de 27 % avec l'algorithme.
Champ : France + COM.
Source : SSMSI, Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2017 et 2020.

Alors que le nombre d'infractions, sur l'ensemble du périmètre de la délinquance enregistrée, a globalement diminué en 2020, notamment en lien avec la crise sanitaire et les mesures gouvernementales mises en place, le nombre d'infractions liées au numérique enregistrées par la gendarmerie est en nette hausse depuis 2017, en particulier en 2020.
La hausse est davantage marquée pour l'estimation des infractions liées au numérique au sens de la définition du groupe de travail (*figure 28*) que par l'algorithme. En effet, le nombre d'infractions liées au numérique au sens de la définition a augmenté de 89 % entre 2017 et 2020 tandis que le nombre d'infractions liées au numérique prédit par l'algorithme n'a augmenté que de 54 % dans le même temps. Cette hausse d'infractions au sens de la définition, peut s'expliquer par une hausse du phénomène mais peut également refléter des changements de pratique de la part des gendarmes dans les logiciels de rédaction avec une meilleure connaissance du phénomène par exemple. Il y a donc une amélioration

probable du remplissage des variables permettant de détecter ce phénomène. L'évolution du nombre total d'infractions liées au numérique reflète essentiellement celle du nombre d'infractions génériques liées au numérique : la hausse du nombre d'infractions génériques relevant du numérique est davantage marquée pour l'estimation faite avec la définition entre 2017 et 2020.
Les nombres d'infractions spécifiques obtenus avec les deux estimations suivent, en revanche, la même évolution entre 2017 et 2020.

3. La labellisation

Afin d'améliorer la détection des infractions liées au numérique il est possible de fiabiliser l'échantillon d'entraînement en le définissant par un plan de labellisation. Ce plan de labellisation permettra d'affiner la détection en établissant des catégories thématiques à l'intérieur du champ des infractions liées au numérique.

3.1. Une labellisation multi-classes des infractions liées au numérique afin de mieux appréhender une réalité complexe

La détection des infractions liées au numérique peut être améliorée, non seulement par le reclassement de certaines infractions en « cyber » ou non « cyber », mais aussi, par une meilleure catégorisation des infractions : il s'agit de passer d'une variable indicatrice binaire (« cyber » versus non « cyber ») à une variable catégorielle correspondant aux principaux champs des infractions liées au numérique. Une lecture des typologies existantes a été mise en perspective avec la classification internationale des infractions à des fins statistiques (ICCS) afin de produire une première grille de labellisation multi-classes.

3.1.1. Les typologies existantes

Le rapport produit par le ministère de l'Intérieur sur l'état de la menace liée au numérique propose deux répartitions des infractions liées au numérique les plus fréquentes. La première est issue des données du Centre de lutte contre les criminalités numériques (C3N) de la gendarmerie[15]. Il s'agit d'une répartition des dix natures d'infractions les plus utilisées dans les comptes rendus de police judiciaire en lien avec le numérique. Ces comptes rendus sont marqués « cyber » grâce à la même coche que celle qui nous permet d'identifier les Manop comme « cyber ».

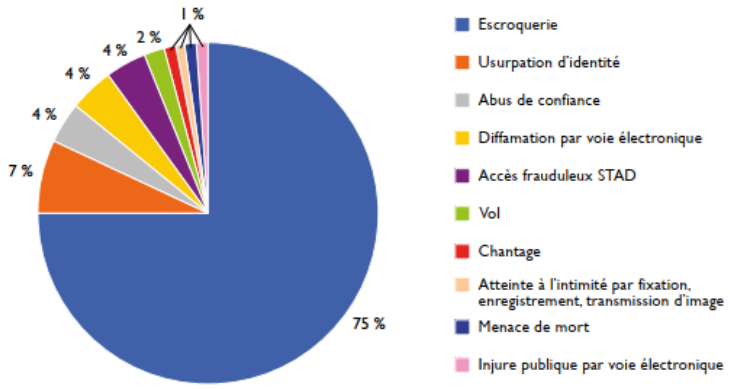
Cette répartition (*figure 29*) donne une première idée des infractions liées au numérique les plus courantes. Avec 75 % des infractions relevant de l'escroquerie, ce phénomène délinquant est de très loin majoritaire. Ensuite viennent les usurpations d'identité et les abus de confiance. Les accès frauduleux dans les Systèmes de Traitement Automatisé de Données (STAD) sont les infractions qui nécessitent l'expertise la plus importante de la part des enquêteurs. Elles ne représentent que 4 % du champ des infractions liées au numérique ce qui pose question sur le classement d'infractions comme les rançongiciels. Ce sont des atteintes aux STAD mais commises dans le but de réaliser une escroquerie ou extorsion. Il n'est pas précisé comment ce type d'infraction est classé ici.

Issue du même rapport, une autre répartition des infractions est proposée (*figure 30).* Les données sont issues de la plateforme PHAROS exploitée par l'Office central de lutte contre la criminalité liée aux technologies de l'information et de la communication (OCLCTIC). Cette plateforme regroupe les signalements faits sur le site www.internet.signalement.gouv.fr par des internautes et des

[15] Les données du C3N sont proches des données du SSMSI car elles proviennent également des logiciels de rédaction des procédures

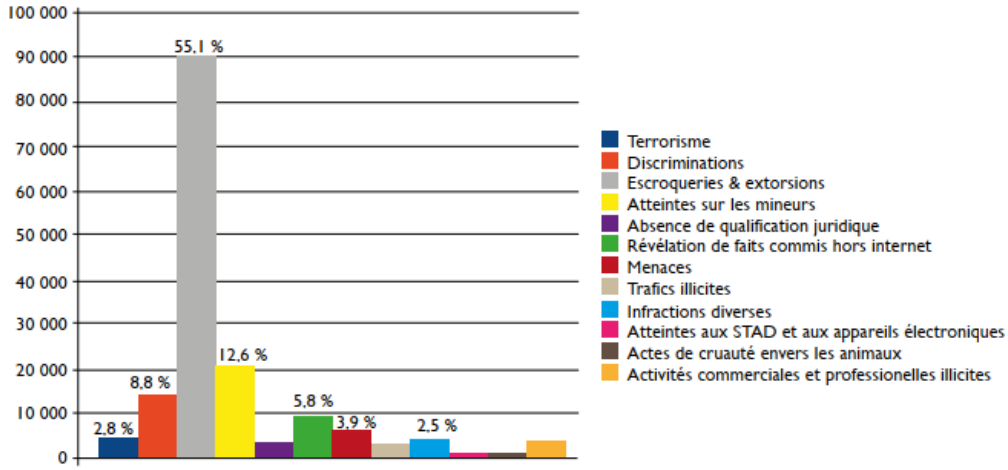
professionnels du numérique. La figure ci-dessous propose une répartition des signalements par catégorie.

Figure 29 : Répartition des comptes rendus de police judiciaire (CRPJ) Cyber 2018[16]



Lecture : En 2018, 4 % des infractions liées au numérique sont des abus de confiance
Source GN – C3N [6]

Figure 30 : Répartition des signalements PHAROS par catégorie 2018



Lecture : En 2018 12,6 % des signalements faits sur la plateforme PHAROS sont des atteintes sur les mineurs
Source : plateforme PHAROS [6]

Là encore les escroqueries sont de très loin le plus grand motif de signalement d'infractions liées au numérique. Toutefois on peut observer des divergences avec la répartition du C3N car le deuxième signalement le plus fréquent ce sont les atteintes sur les mineurs, puis les discriminations. La part des infractions qui relèvent uniquement des atteintes aux personnes est plus importante dans cette répartition avec les atteintes sur les mineurs 12,6 %, les discriminations 8,8 % et les menaces 3,9 %. Le pourcentage des infractions qualifiées d'atteintes aux STAD est beaucoup plus bas ici par rapport à la répartition du C3N (<1 % vs 4 %). Les infractions qui agissent ou altèrent les supports logiciels des STAD sont plus souvent un moyen d'atteindre un but infractionnel (motif crapuleux, trouble à l'ordre public, etc.), qu'un but de commission d'une infraction. Ainsi, comme pour le graphique produit par le C3N,

[16] L'acronyme STAD utilisé dans les différentes figures signifient système de traitement automatisé de données

le classement des infractions qui relèvent à la fois des atteintes aux STAD et d'une autre nature d'infraction interroge.

3.1.2. La nomenclature ICCS : avantages et limites

Dans l'existant, l'autre voie de réflexion explorée est de définir des catégories de Manop selon la classification internationale des infractions à des fins statistiques (ICCS : International Classification of Crime for Statistical Purposes). La nomenclature statistique française des infractions (NFI) a été réalisée à partir de l'ICCS construite par l'Office des Nations Unies contre la drogue et le crime. Cette classification a adapté l'ICCS au contexte français par un groupe de travail interministériel (Justice et Intérieur). Labelliser en suivant les catégories de cette nomenclature (ICCS) présente l'avantage de s'inscrire dans la continuité d'un cadre existant, d'établir des comparaisons et de balayer l'ensemble du champ infractionnel.

Figure 31 : Tableau des catégories de l'ICCS

	CATÉGORIES DE NIVEAU 1
1	Actes entraînant ou visant à entraîner la mort
2	Actes portant atteinte ou visant à porter atteinte à la personne
3	Actes préjudiciables à caractère sexuel
4	Actes visant des biens avec violence ou menaces contre une personne
5	Actes visant uniquement des biens
6	Actes faisant intervenir des drogues contrôlées ou d'autres substances psychoactives
7	Actes relevant de la fraude, de la tromperie ou de la corruption
8	Atteintes à l'ordre public, à l'autorité et aux dispositions juridiques de l'État
9	Atteintes à la sécurité publique et à la sûreté de l'État
10	Atteintes au milieu naturel
11	Autres actes illégaux

Source : Classification internationale des infractions à des fins statistiques, ONUDC, mars 2015 [6]

L'ICCS est une nomenclature généraliste dont le champ porte sur l'ensemble des infractions connues à ce jour. Elle présente l'avantage d'un classement cohérent des infractions en différentes catégories. Tout en maintenant les réserves que l'on peut émettre sur la coche « cyber » quelques explorations statistiques ont été conduites afin de mieux comprendre ces catégories ICCS au regard des infractions liées au numérique. Par exemple, dans la catégorie 1 des *Actes entraînant ou visant à entraîner la mort*, 10,5 % des infractions qui sont des incitations au suicide ont été cochées « cyber » par la gendarmerie tandis que seulement 1,9 % des Autres actes entraînant ou visant à entraîner la mort ont été cochés « cyber » par les gendarmes. Autre exemple, dans la catégorie 10 *Atteintes au milieu naturel* seul le trafic de faune et de flore, coché « cyber » à hauteur de 21,4 % de ses infractions, semble pouvoir faire partie des infractions liées au numérique.

Cependant les libellés de ces catégories ICCS masquent la réalité des infractions les plus souvent commises en lien avec le cyberespace. À titre d'exemple, la catégorie 3 regroupe des actes préjudiciables à caractère sexuel aussi différents que la pédopornographie ou le harcèlement sexuel sur majeur. Par ailleurs tous les autres types de harcèlement sont eux compris dans la catégorie 2. Cette division thématique au regard des infractions liées au numérique amèneraient à créer des catégories trop fines. Le contenu de la Manop pourrait indiquer un classement de l'infraction dans deux catégories à la fois comme par exemple dans des menaces (catégorie 2 selon l'ICCS) et du harcèlement sexuel sur majeur (catégorie 3). Autre exemple, le vol de données, considéré comme une

atteinte à la personne, est classé dans la catégorie 2. Or le vol de données est le plus souvent une atteinte aux STAD et ces infractions sont, elles, catégorisées dans la section 9. Proposer une typologie calquée sur l'ICCS ne paraît donc pas être la meilleure solution pour appréhender plus finement les problématiques liées à la délinquance numérique. Cependant, il paraît nécessaire de s'en inspirer pour conserver une cohérence avec l'existant sur la criminalité.

3.1.3. Proposition de labellisation multi-classes du SSMSI

À ce stade une connaissance plus qualitative des Manop permettrait de définir spécifiquement les infractions liées au numérique et d'établir les situations tangentes à ce champ. L'analyse qualitative des Manop peut permettre de déterminer plus spécifiquement la frontière entre les infractions liées au numérique et celles qui n'en sont pas. Un certain nombre d'entre elles pourront être plus difficiles à catégoriser et ainsi être considérées comme tangentes au champ des infractions liées au numérique. Afin d'avoir un aperçu du contenu des Manop deux processus de lecture ont été opérés.

Un premier travail sans définition préalable de la « cyber délinquance » afin de produire une labellisation binaire des Manop a été réalisé. Les Manop de l'échantillon test qui ont été classées en « cyber / non cyber » par l'algorithme ont été relues par une personne membre des forces de sécurité et labellisées de façon binaire afin de donner un premier aperçu des différences d'appréciations entre les prédictions de l'algorithme et l'interprétation des infractions liées au numériques par un expert.

À partir de ce premier travail une deuxième lecture analytique de ce même échantillon de Manop par un chargé d'étude spécialiste a été faite en comparant la prédiction de l'algorithme et la labellisation humaine. De cette deuxième lecture analytique, il ressort plusieurs constats :

- Certaines infractions se produisent dans le cyberespace et de façon concomitante dans la vie réelle. Par exemple, plusieurs Manop concernant des agressions physiques étaient précédées ou suivies de harcèlement ou d'injure sur internet.
- Il existe un halo d'infractions qui « gravitent » autour des outils numériques. C'est principalement là que s'observent les dissensions entre les deux lectures (première labellisation, la relecture analytique et les prédictions de l'algorithme).
- Parmi le halo des infractions qui impliquent des outils numériques et où les dissensions apparaissent des thématiques ressortent, parmi lesquelles :
    - Les atteintes matérielles[17]
    - Les infractions liées à l'environnement bancaire[18]
    - La captation d'image sans diffusion[19]

Par la suite, des Manop sélectionnées selon leur classement dans l'ICCS ont été lues. Parmi les Manop cochées « cyber » une dizaine ont été tirées au sort par catégorie ICCS. Cette étape a eu pour but de comprendre le lien entre le contenu de la Manop et le classement dans la Natinf et l'ICCS.

[17] La plupart des infractions très spécifiques au numérique comme les piratages, les rançongiciels, les attaques par déni de service distribué, sont en fait des atteintes logicielles et non des atteintes au support physique des technologies de l'information et de la communication. Le fait de voler un téléphone portable est-il une infraction liée au numérique ? Ou encore de détruire une antenne 5G ?

[18] La plupart des actes bancaires ou mouvement d'argent s'observent sur internet. Par exemple, certains mouvements bancaires qui semblent frauduleux, sont constatés en ligne mais leur origine est inconnue, d'autres fois, il s'agit d'un vol de carte bancaire en réel suivi d'un achat en ligne. Ainsi le simple fait de constater un mouvement d'argent suspect sur un compte en ligne plutôt que sur un relevé de compte papier constitue-t-il une infraction liée au numérique ?

[19] Le fait de filmer ou photographier quelqu'un sans son consentement sans en faire la diffusion par quelque moyen que ce soit constitue-t-il une infraction liée au numérique ?

À partir de ces différents éléments d'analyse, le SSMSI propose la labellisation thématique suivante :

Figure 32 : Grille de labellisation thématique des infractions liées au numérique du SSMSI

Thème	Détail
Escroquerie, interaction avec l'outil numérique à motif crapuleux	Du rançongiciel au phishing en passant par l'arnaque à la romance et les arnaques par prise de contact et paiement par moyen frauduleux
Atteintes à la personne (majeur/ mineur), motif non crapuleux	Harcèlement, injure, menace, atteinte à l'intimité de la vie privée, pédopornographie
Trouble à l'ordre public	Incitation à la haine, apologie du terrorisme
Infraction au droit d'auteur	Téléchargement illégal, infraction à la loi Hadopi, publication de musique sans déclaration à la SACEM.
Non-respect des réglementations spécifiques au fonctionnement des outils numériques	Non publication des mentions légales, non-respect des obligations de déclarations (Données CNIL)

3.2. Définir des degrés de l'utilisation des outils numériques pour certaines classes de la labellisation

Comme évoqué précédemment un certain nombre d'infractions gravitent dans un halo « cyberdélinquant » où les moyens numériques sont peu utilisés mais rendent tout de même possible la commission d'une infraction. Le SSMSI propose d'affiner la nomenclature thématique avec une deuxième entrée par degrés d'utilisation des outils numériques dans la commission des infractions. Les degrés proposés sont les suivants :

Figure 33 : Grille de labellisation des infractions par degré d'utilisation des moyens numériques

Degrés	Détail
Les supports numériques cibles de l'infraction	L'outil numérique est un moyen de commission de l'infraction et ce moyen nécessite d'agir sur le fonctionnement du support numérique au niveau logiciel, même de façon légère. Les atteintes aux STAD, les piratages, et toutes les atteintes logicielles sont concernées. Par ailleurs, les intrusions frauduleuses dans un support numérique sans altération de son fonctionnement sont aussi considérées dans cette catégorie. *Par exemple : rançongiciel, intrusion dans un compte par vol de mots de passe*
Infraction commise dans le cyberespace	L'infraction a été réalisée dans le cyberespace sans porter atteinte au support logiciel des outils numériques et n'a pas nécessité d'intrusion frauduleuse. *Par exemple : arnaque à la romance, phishing, harcèlement*
Mise en relation par un moyen numérique	Toutes les utilisations des outils numériques pour *in fine* concrétiser une infraction en dehors du cyberespace. *Par exemple : toutes les situations où deux individus prennent contact au préalable par un moyen numérique. Puis une infraction est commise (escroquerie, agression).*
Pas de lien avec le numérique ou en dehors du champ des infractions liées au numérique	Toutes les infractions sans lien avec le numérique et les outils numériques NB : les atteintes matérielles (vol de matériel informatique, etc.), les vols de colis qui ont été commandés sur internet sont considérés comme sans lien avec le numérique

Affiner par degré d'importance de l'outil numérique dans la commission de l'infraction permet de suivre des phénomènes délinquants fondamentalement différents les uns des autres mais dont le point commun est de rendre possible des infractions qui n'auraient pas été commises autrement. Les catégories affinées par degré sont les escroqueries et les atteintes à la personne car la proportion[20] des infractions liées au numérique dans ces catégories est importante. La lecture des Manop a par ailleurs montré des réalités différentes dans l'utilisation des moyens numériques pour commettre ces infractions.

Figure 34 : Grille de labellisation des escroqueries par degré d'utilisation des moyens numériques

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 ou en dehors du champ des infractions liées au numérique
Définition	Toutes les escroqueries qui agissent sur les supports de communication ou les STAD	Toutes les escroqueries qui sont commises dans le cyberespace	Les escroqueries qui nécessitent qu'une mise en contact par moyen numérique sans pour autant être réalisées de façon dématérialisée.	Toutes les escroqueries sans lien avec le numérique et les outils numériques
Exemple	Rançongiciel	Arnaque faux site internet	Mise en contact par internet pour commettre une escroquerie, exemple vente d'une voiture de particulier à particulier avec paiement en faux chèques	Infraction sans lien avec le numérique
Manop	*« 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 inexploitables. 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 accroche au crochet du caddie au niveau de la barre de manœuvre. »*

[20] Les résultats sont issus des statistiques exploratoires présentées dans le paragraphe 3.1.2, réalisées grâce à la coche cyber.

Figure 35 : Grille de labellisation des atteintes à la personne par degré d'utilisation des moyens numérique

Atteintes à la personne (majeur/ mineur), motif non crapuleux	**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 ou en dehors du champ des infractions liées au numérique**
Définition	Tous les comportements malveillants envers des personnes physiques qui agissent sur les supports numériques	Tous les comportements malveillants envers des personnes physiques commis dans le cyberespace	Toutes les mises en relation / contacts qui ont abouti à une atteinte à la personne non dématérialisée	Toutes les atteintes à la personne sans lien avec le numérique et les outils numériques
Exemple	Piratage de compte (réseaux sociaux par exemple)	Harcèlement, pédopornographie, diffusion d'image ou vidéo sans consentement	Prise de contact en vue de / suivie d'une agression	Infraction sans lien avec le numérique
Manop	*« 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 cliche. Sur la plupart des photos il y a la présence d'un commentaire explicite. Sur une photo en particulier y figure un sexe féminin. Du fait que les photographies sont diffusées sur le compte d'une jeune fille celle-ci sont visibles pour ces proches eux aussi mineurs. »*	*« 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. »*

Les trois autres thèmes ne seront pas déclinés en degrés car les effectifs pour chaque degré seraient trop petits.

Figure 36 : Grille de labellisation thématique sans degré d'utilisation des moyens numériques

Thème	Détail	Manop
Trouble à l'ordre public	Incitation à la haine, apologie du terrorisme	*« Entre le 13 février 2018 et le 27 mars 2018 un individu âgé d'environ 20 ans poste sur [un réseau social] des propos et photographies faisant l'apologie d'actes de terrorisme et menaçant de destruction dangereuse pour les personnes. Le 28 mars 2018 un homme résidant à Toulouse ayant connu l'individu et qui a pu voir ces propos et photographies sur [un réseau social] faisant l'apologie de DAESH est entendu. L'ensemble des écrits et images faisant l'apologie du terrorisme visibles sur [un réseau social] de l'individu et accessibles par la personne nous signalant les faits nous sont transmises. »*
Infraction au droit d'auteur	Téléchargement illégal, infraction à la loi Hadopi, publication de musique sans déclaration à la SACEM.	*« De janvier 2016 à mars 2017 la commission de protection des droits de l'HADOPI constate plusieurs téléchargements illégaux de fichiers vidéos et audios. Pour ce faire plusieurs logiciels de téléchargement* frauduleux de type pe*er to* pe*er ont été utilisés. Les adresses IP des téléchargements illégaux correspondent à un client identifie par l'opérateur internet [Nom opérateur internet] sur la commune de [Nom de la commune et adresse]. Au total 9 fichiers ont été téléchargés illégalement par le mis en cause durant la période des faits. »*
Non-respect des réglementations spécifiques au fonctionnement des outils numériques	Non publication des mentions légales, non-respect des obligations de déclarations (Données CNIL)	*« Depuis février 2015 le mis en cause par l'intermédiaire de sa société et [d'un site d'avis en ligne] propose ses services de gestion de données et de stockage à des restaurateurs afin de vérifier la véracité des avis laissés sur [un site de tourisme]. Plusieurs plaintes sont déposées par des clients de restaurant et une par un restaurateur lui-même qui condamnent les pratiques douteuses du mis en cause constitutifs d'une infraction délictuelle. L'enquête révèle également un manquement à la CNIL. »*

3.3. Le processus de labellisation

L'un des problèmes majeur du renseignement de la coche « cyber » par les gendarmes, concerne l'approche du phénomène et de sa définition par ces mêmes agents. Le périmètre des infractions liées au numérique peut différer selon les personnels métiers qui saisissent les données administratives des plaintes[21]. Le processus de labellisation impliquera qu'un certain nombre de personnes internes au SSMSI lisent des Manop et les labellisent selon la nomenclature établie. Il s'agit donc de définir un

[21] Il faut se référer à l'analyse produite en 1.1 où certaines Manop comportent des infractions très similaires et ne sont pas toujours cochées « cyber ».

protocole rigoureux afin de ne pas reproduire l'écueil d'avoir plusieurs interprétations possibles selon les différents personnels métiers ou les chargés d'études qui labelliseront les Manop.

Etapes du processus de labellisation

- Ecriture d'un protocole Alpha de labellisation à partir de la nomenclature thématique et de la nomenclature par degrés. Un exemple de structure de ce protocole est disponible dans l'annexe 4 de l'article. Ce document propose une organisation des définitions, exemples et arbitrages nécessaires qui permettent de comprendre les différentes catégories pour les personnels métiers qui labelliseront.
- Premier test de labellisation sur un échantillon de Manop par une équipe resserrée en interne
- Effectuer un bilan de la labellisation
    - Lecture et analyse de la labellisation effectuée
    - Echange sur les difficultés, les cas tangents et les erreurs de labellisation
- Amendement du protocole de labellisation pour écrire un protocole beta
- Formation des personnels métiers au protocole beta et à l'interface de labellisation

Tirage de l'échantillon à labelliser

L'échantillon à labelliser doit couvrir toutes les catégories définies dans notre typologie[22]. À chaque Manop est attribuée une Natinf, elle-même classée dans la nomenclature ICCS. Il est donc possible d'apparier les deux nomenclatures (Natinf et ICCS) afin de tirer un échantillon de Manop selon les thématiques de l'ICCS. Le tirage de l'échantillon sera fait à partir des catégories ICCS[23] et sous catégories ICCS[24]. De plus la typologie proposée par le SSMSI étant dérivée de la nomenclature ICCS chaque catégorie de la labellisation peut ainsi trouver une correspondance thématique dans l'ICCS.

Le premier échantillon doit être réduit pour être lisible rapidement par tous les membres de la première équipe de labellisation et faciliter les échanges. Malgré un nombre restreint d'exemples cet échantillon doit tout de même avoir suffisamment de Manop pour rencontrer le maximum de cas différents.

Afin de définir le ratio entre les Manop liées au numérique et celles qui ne le sont pas, on ne se fiera qu'à la coche « cyber » pour ce tirage. Cette coche étant insuffisante à elle seule pour caractériser les infractions liées au numérique, la labellisation devrait permettre d'améliorer le classement des infractions entre ces deux catégories. Etant donné qu'il y a plusieurs groupes à labelliser à l'intérieur des infractions liées au numérique il est plus intéressant de surreprésenter les infractions de la coche « cyber ». Ce ratio entre « cyber / non cyber » ne sera pas celui de l'échantillonnage final pour l'entraînement de l'algorithme, compte tenu des reclassements effectués.

On peut considérer pour ce premier échantillon d'environ 600 Manop à labelliser une répartition entre 200 Manop non cochée « cyber » et environ 400 Manop cochées « cyber ». Chacune des 5 catégories[25] à l'intérieur des infractions liées au numérique doit avoir un nombre minimum de Manop à labelliser.

La sélection aléatoire des 400 Manop cochées « cyber » s'effectuera selon un plan de sondage stratifié par catégories ICCS avec des surreprésentations appliquées à certaines d'entre elles : les allocations

[22] Escroqueries, atteintes aux personnes, troubles à l'ordre public, infraction au droit d'auteur, non-respect des réglementations spécifiques au fonctionnement des outils numériques
[23] Comme vu précédemment en section 3.1.2, il existe 11 catégories ICCS
[24] Il existe un peu plus d'une centaine de sous-catégorie ICCS
[25] Escroqueries, atteintes aux personnes, troubles à l'ordre public, infraction au droit d'auteur, non-respect des réglementations spécifiques au fonctionnement des outils numériques

ne seront pas exactement proportionnelles aux effectifs. Selon les répartitions des infractions liées au numérique du C3N et de l'OCLCTIC, les escroqueries sont le phénomène délinquant le plus important. Il est possible qu'une grande variété d'escroqueries se produise par les moyens numériques. Afin de rencontrer suffisamment d'exemples à labelliser dans les catégories importantes, il est nécessaire de les représenter selon leur poids. Un autre point d'attention se porte sur certaines catégories qui auront très peu de Manop disponibles. Il est important d'avoir un minimum de ces Manop à lire ce qui amènera à surreprésenter un peu les plus petites catégories. L'échantillon de 400 Manop liées aux infractions numériques est indicatif. Ce nombre permet de définir un premier ordre de grandeur suffisant mais pas trop important non plus de Manop à lire pour la première phase de labellisation. Ainsi si l'on surreprésente les plus petites catégories il faudrait logiquement sous représenter les plus importantes. Il n'est pas nécessaire pour ce premier tirage d'être trop strict sur le respect de cet ordre de grandeur.

Compte tenu de ces points d'attention la meilleure méthode pour échantillonner les Manop cochées « cyber » serait un plan de sondage stratifié avec des surreprésentations pour les catégories de Manop à effectif faible.

L'échantillonnage de la labellisation finale s'appuiera sur le processus qui a permis de tirer l'échantillon de la labellisation test. La répartition entre infractions liées au numérique ou non pourra être rééquilibrée selon les besoins de l'algorithme afin d'arriver à un ratio plus proche de 50/50. La coche « cyber » sera de nouveau utilisée pour cet aspect du tirage. Dans l'hypothèse où après la labellisation test un nombre significatif de Manop sont passées d'une classe à l'autre il sera possible d'appliquer un correctif au moment du tirage de la labellisation finale. Selon les conclusions tirées de la labellisation test, la méthode du plan stratifié avec surreprésentation des catégories les plus petites pourra être amendée ou changée.

La phase de labellisation

La phase de labellisation par des personnels métiers du SSMSI est une étape qui nécessitera de limiter au maximum les biais d'interprétation des différentes personnes impliquées dans le processus. Il existe certaines pratiques permettent d'améliorer cette étape pour en réduire les biais au minimum. Toutefois celles-ci sont assez peu documentées[26] tant dans le secteur de la recherche en Machine Learning que par les entreprises qui y ont recours pour le fonctionnement de leurs algorithmes[27].

Sur les Territoires de la République française une expérience a été conduite par Etalab afin de construire le premier jeu de données ouvert de questions/réponses pour la francophonie. Le projet PIAF « Pour des intelligences artificielles francophones » vise à améliorer les performances des utilisateurs de l'IA en traitement automatique du langage en mettant à disposition des données d'entraînement de qualité en français. La réalisation de ce projet a nécessité des séances « d'annotation » de morceaux de texte en français. La méthode était contributive et sur la base du

[26] À ce jour peu d'articles ont été publiés sur les protocoles de labellisation des données et la formation des personnels qui labelliseront. On peut toutefois citer en exemple un article disponible sur OpenAI [7] qui explique brièvement le processus de recrutement de travailleurs temporaires sélectionnés selon certaines de leurs compétences afin d'annoter des textes selon leur sentiment. Peu d'informations sont ajoutées à la formation de ces annotateurs et la vérification des données labellisées.

[27] Sur le site du laboratoire de Google sur l'AI [8] presqu'aucun article ne porte sur comment améliorer la labellisation des données. L'article le plus proche du sujet porte sur une évaluation comparative de deux types de jeux de données (beaucoup de données mal labellisées versus peu de données très qualitatives) [9]. Un autre exemple peut être cité d'après le service d'Amazon Mechanical Turk [10] dont le nom fait directement référence à une marionnette. Dans la présentation de ce service, l'intérêt est plutôt porté sur la facilité d'utilisation et la réduction des coûts que sur la manière de réduire les biais de labellisation des annotateurs.

volontariat. Certains enseignements dans le déroulé du processus d'annotation peuvent inspirer le protocole qui doit être développé au sein du SSMSI. Parmi ces enseignements, on peut citer l'utilisation d'une interface de labellisation simple d'utilisation, une formation pédagogique du procédé de labellisation, une explication claire des enjeux et aboutissements du projet.

Le processus de labellisation de ce projet devrait éviter certains des écueils du projet PIAF contributif car les personnels qui procéderont à la labellisation des données travaillent au sein du SSMSI. Ces agents sont déjà sensibilisés au travail des données ce qui facilitera le travail d'explication des enjeux et de formation. Pour pallier le manque de documentation sur les processus de formation des personnels qui labellisent, il est possible de s'inspirer dans une certaine mesure des méthodes pédagogiques utilisées pour former les enquêteurs. En effet, cet autre domaine de la statistique requiert l'utilisation de personnels à former pour produire le travail le plus objectif et similaire possible. Les enquêteurs sont sensibilisés aux tenants et aboutissants de l'étude et à éviter les interprétations personnelles lors de la passation des questionnaires. Ainsi dans un processus d'annotation des données, bien saisir le but de la labellisation et limiter les biais d'interprétation personnelle pour les personnels sont également des prérequis nécessaires. Sur le site de l'INED [11], une fiche explicative est disponible sur la formation des enquêteurs. Si les différences entre les deux procédés sont nombreuses, cela permet de partir d'une base de travail pour construire la formation des personnels qui labelliseront.

Actuellement, en matière de labellisation peu d'autres solutions sont disponibles. L'alternative principale aux équipes métiers de labellisation ou d'un projet contributif volontaire aurait été de faire appel à une plateforme de micro-travail. Compte tenu du caractère sensible des données du ministère de l'Intérieur et l'absence de suivi méthodologique[28] des personnels qui labellisent, cette solution n'était pas envisageable.

3.4. Les points d'attention de l'entraînement multi-classes

Certains points d'attention sont à observer dans le développement de ce processus de labellisation.

En l'état d'avancement de l'implémentation l'algorithme est testé pour produire une classification binaire. Après la labellisation, il s'agira de produire une prédiction multi-classes pour deux des catégories de la typologie (Escroqueries, atteintes à la personne). Plusieurs méthodes de Machine Learning sont à même de répondre au problème. Il est envisageable de prédire l'un après l'autre, chaque label par rapport au reste des données en classification binaire. Plusieurs algorithmes sont plus performants pour une prédiction binaire que pour une prédiction multi-classes. La combinaison d'un ensemble de k méthodes $M_1, M_2, M_3, \dots, M_k$ est appelé apprentissage ensembliste[29].

Le deuxième point d'attention concerne les infractions dont le contenu des Manop « non cyber » pourrait être très proche de celui des infractions liées au numérique. Grâce à Word2Vec, les mots sont projetés en vecteur numérique. Le vecteur est composé des probabilités que les mots adjacents à un mot soient effectivement adjacents à ce mot[30]. Ainsi les mots liés au numérique s'ils ont des contextes proches auront des signatures proches. Lors d'une classification binaire entre une situation « cyber »

[28] Antonio A. Casilli a mené une enquête sur les « travailleurs du clic ». Le processus de labellisation des données, entre autres tâches nécessaires au fonctionnement des algorithmes de Machine Learning, est souvent opéré par des travailleurs précaires dans des pays où la pauvreté est importante (Inde, Pakistan, Philippines, Madagascar).

[29] L'apprentissage ensembliste est connu sous le nom d'*Ensemble Learning* dans la littérature anglophone. La méthode ensembliste s'applique dans plusieurs contextes comme avec les méthodes d'échantillonnage dites bootstrap ou bagging ou encore lorsque l'algorithme XGBoost est utilisé car il s'agit d'un ensemble d'arbres de décision.

[30] Voir section 2.3.2 sur la méthode word2vec

versus « non-cyber » une même infraction aura des champs lexicaux proches et la distinction se fera pour beaucoup sur les mots liés au numérique. Par exemple, dans un contexte de harcèlement commis dans la rue ou sur internet les mots liés au harcèlement seront proches tandis que la variation apportée par les mots du lieu de commission sera plus ténue. Ce type de configuration rend le processus de détection plus délicat.

Si ce raisonnement est vrai pour la classification binaire actuelle, il l'est aussi pour des prédictions multi-classes par degré où par définition le contexte lexical sera proche. L'algorithme devra prédire sur les thèmes de l'escroquerie et des atteintes aux personnes en distinguant à l'intérieur de ces catégories l'importance des moyens numériques dans la commission de l'infraction. À l'intérieur de l'une ou l'autre catégorie les mots en lien avec le thème seront proches. Par exemple, les mots en lien avec l'escroquerie comme l'argent ou les moyens de paiement seront présents dans les trois degrés d'escroqueries liées au numérique. Cependant selon le degré du moyen numérique utilisé les infractions varient. En effet, si l'on reprend le détail des escroqueries du tableau de la figure 24 dans la section 3.2, les contextes explicatifs autour de l'infraction diffèrent. Dans le cas d'un rançongiciel les mots du champ lexical de l'informatique sont plus présents, tandis que dans le cas d'une mise en contact des mots liés à une rencontre physique sont eux présents.

Autre exemple concernant la proximité des champs lexicaux, certaines infractions comme le vol de colis suite à un achat sur internet ne sont pas considérées comme des infractions liées au numérique selon notre nomenclature pour autant les mots « internet », ou les sites de e-commerce seront présents dans la Manop. Le vocabulaire sera ainsi proche d'autres types d'escroquerie qui seront considérées comme appartenant au champ des infractions liées au numériques. Il serait judicieux de suivre ces infractions considérées comme à risque de mauvais classement en les marquant d'une variable afin de vérifier par la suite le taux d'erreur véritable.

Conclusion

Le projet initié en collaboration avec le SSP Lab de l'INSEE, a pour objectif d'améliorer la détection des phénomènes dits « cyberdélinquants » ou infractions liées au numérique au sein des données issues du logiciel de rédaction des procédures de la gendarmerie. Pour ce faire deux étapes ont été réalisées en parallèle : une implémentation d'un algorithme en traitement de texte automatisé ainsi qu'une labellisation des données.

Tout d'abord, malgré les potentiels biais créés par l'échantillon d'apprentissage, l'algorithme de traitement de texte automatisé a donné des résultats encourageants sur l'estimation des infractions liées au numérique. Il a permis de détecter de nombreuses infractions liées au numérique non définies comme telles par les variables annexes retenues par le groupe de travail.

Pour pallier le biais créé lors de l'apprentissage, une grille de labellisation des données a été construite. Cette labellisation permettra de faire fonctionner l'algorithme avec un échantillon d'apprentissage labellisé et également d'affiner la caractérisation au-delà d'une distinction « cyber » versus « non cyber », selon des thématiques importantes dans les infractions liées au numérique. De plus, certaines thématiques, à savoir les escroqueries et les atteintes aux personnes, seront affinées par degré d'utilisation des moyens numériques dans la commission de l'infraction. Avec cette labellisation, il s'agira d'effectuer une prédiction multi-classes ; l'algorithme de prédiction binaire implémenté dans ce projet pourra être facilement adaptable pour cette prédiction multi-classes.

Néanmoins, même si les résultats sont encourageants pour la suite, ce processus de détection des infractions liées au numérique comporte des limites. D'une part la détection de ces infractions est dépendante du contenu de la Manop. En effet, le contenu des Manop de certaines infractions

considérées comme liées au numérique n'y font pas référence (le gendarme n'en a pas fait mention dans la Manop). Le processus ne permettra donc pas de les détecter. D'autre part, le processus, qui a été testé sur les infractions enregistrées par la gendarmerie, n'est pas applicable sur les données de la police, car très peu de Manop sont renseignées du fait de leur caractère non obligatoire.

Les résultats obtenus grâce à la labellisation permettront, tout de même, d'avoir une estimation de la part des infractions liées au numérique dans le total des infractions enregistrées par la gendarmerie. Par ailleurs, la grille de labellisation des données permettra de définir plus précisément et clairement le phénomène, ainsi que d'harmoniser la définition entre la police et la gendarmerie. Ceci pourra donc permettre un meilleur remplissage des variables permettant de détecter les infractions liées au numérique et ainsi, une meilleure quantification du phénomène.

Référence

[1] Razafindranovona T., Moreau A. Les défis de la mesure statistique de la cybercriminalité, *Revue de la Gendarmerie Nationale*, n°266, 4e trimestre 2019, janvier 2020

[2] https://www.ssi.gouv.fr/

[3] https://www.ssi.gouv.fr/entreprise/glossaire/r/

[4] Mikolov T., Chen K., Corrado G., Dean J., Efficient Estimation of Word Representations in Vector Space, *arXiv:1301.3781v3 [cs.CL]*, 7 Sep 2013

[5] Rumelhart D.E., Hinton G.E. et Williams R.J., Learning Internal Representations by Error Progration, *Parallel Distributed Processing*, Edition MIT Press, pp.318-362

[6] Ministère de l'Intérieur, Délégation ministérielle aux industries de sécurité et à la lutte contre les cybermenaces, Etat de la menace liée au numérique en 2019, Mai 2019, rapport n°3, p88 & p90

[7] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C.L. Wainwright, P. Mishkin, P. Christiano, J. Leike, R Lowe, et al. (Février 2021) *Training language models to follow instructions with human feedback. OpenAI,* https://cdn.openai.com/papers/Training_language_models_to_follow_instructions_with_human_feedback.pdf

[8] https://deepmind.com/research/

[9] Nanda, N., Gowal, S., Uesato, J. (Juillet 2021). *An Empirical Investigation of Learning from Biased Toxicity Labels*. DeepMind. https://deepmind.com/research/publications/2021/An-Empirical-Investigation-of-Learning-from-Biased-Toxicity-Labels.

[10] Amazon Mechanical Turk. https://www.mturk.com/

[11] Institut national d'études démographiques. *Formation des enquêteur.trice.s et gestion de la collecte.* Ined. https://www.ined.fr/fr/ressources-methodes/methodologie-enquete/les-choix-methodologiques/formation-des-enqueteur-e-s/

Bibliographie :

Bird S., Klein E., Loper E., Natural Language Processing with Python, *O'Reilly*, Section 3.7 Regulars Expression for Tokenizing Text, pp.109-112 , Juin 2009

Buda Mateusz, Maki Atsuto, Mazurowski Maciej A. A systematic study of the class imbalance problem in convolutional neural networks, Neural Networks, Chapter 1 pp.1-2 & Chapter 2 pp.3-13, Octobre 2017

Chollet F., Deep Learning with Python, *Manning Edition*, Part 1 Chapter 3 pp56-60 & Part 2 Chapter 6 pp.178-195, 2017

Dangeti, P., Statistics for Machine Learning, Packt Publishing, Juillet 2017

Parizeau, M., Réseaux de neurones. GIF-21140 et GIF-64326, Université de Laval, Chapitre 5 pp.39-68, 2006

Sun, Y.; Wing, A.K.C.; Kamel, M.S. Classification of Imbalanced Data: A Review, International Journal of Pattern and Artificial Intelligence, Novembre 2011, Vol. 23, pp.687–719, 2009

Annexe :

Annexe 1 : Proportion d'infractions cochées « cyber » mal classées avec une Natinf générique pour les 5 Natinf les moins bien classées

Libellé Natinf	Nombre d'infractions mal classées	Proportion dans le total des infractions mal classées ayant une Natinf générique (en %)	Nombre total d'infractions	Proportion de la Natinf dans le total des infractions de la classe « cyber » ayant une Natinf générique (en %)	Proportion d'infractions mal classées au sein de la Natinf (en %)	Exemple de Manop
Violence suivie d'incapacité n'excédant pas 8 jours par une personne étant ou ayant été conjoint, concubin ou partenaire lié à la victime par un pacte civil de solidarité	11	0,17	11	0,01	100	« 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. »
Viol	8	0,12	8	0,00	100	Le mis en cause prend attache avec la victime sur le site de rencontre [site de rencontre]. Un rendez-vous est fixé à [ville]. Le mis en cause emmène la plaignante dans un appartement. Il la force à l'embrasser, [explication du viol]. 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.
Vol par effraction dans un local d'habitation ou un lieu d'entrepôt	74	1,13	74	0,04	100	« Les auteurs des faits forcent à coups de pieds la porte d'entrée du local peinture de la mairie de [ville]. Aucun objet n'a été dérobe. Présence de 3 traces de pas sur la partie basse de la porte »
Vol à la roulotte	39	0,60	39	0,02	100	« La victime se rend dans sa caravane parquée dans son jardin clos le samedi 22 décembre 2018 à 09h00. C'est alors qu'elle constate qu'une vitre de toit a totalement été arrachée les placards sont ouverts dont un se trouve au sol. Cependant aucun objet n'a été volé. »
Vol aggravé par deux circonstances	11	0,17	11	0,01	100	« Les mis en cause commettent de nombreux vols avec ou sans effraction dans poulaillers jardins potagers au cours desquels ils dérobent poules coqs oies plants pour le potager ainsi que de l'outillage. L'un des auteurs est en récidive de conduire d'un véhicule sans permis. »

Note : Seules les Natinf ayant plus de 5 infractions dans la classe « cyber » ont été conservées.
Lecture : 100 % des infractions relatives au viol sont mal classées par l'algorithme.
Champ : France + COM.
Source : Base des infractions de crimes et délits enregistrés par la gendarmerie entre 2018 et 2020

Annexe 2 : Proportion d'infractions cochées « cyber » mal classées avec une Natinf spécifique en fonction des 5 Natinf les moins bien classées

Libellé Natinf	Nombre d'infractions mal classées	Proportion dans le total des infractions mal classées ayant une Natinf spécifique (en %)	Nombre total d'infractions	Proportion de la Natinf dans le total des infractions de la classe « cyber » ayant une Natinf spécifique (en %)	Proportion d'infractions mal classées au sein de la Natinf (en %)	Exemple de Manop
Interruption volontaire des communications électroniques	9	0,14	9	0,05	100	« Le ou les auteurs ouvrent une trappe puis une autre située à quelques centaines de mètre plus loin. Ces trappes se trouvent sur le côté de la chaussée. Ils les coupent et dérobent 600 mètres de câbles de l'opérateur [nom opérateur]»
Viol commis par une personne mise en contact avec la victime par réseau de télécommunications	26	0,40	27	0,15	96	« En février 2019 Mme. (P-1) entre en contact avec le mis en cause via une application de rencontre. Apres quelques échanges ils décident de se rencontrer un soir au domicile de Mme. (P-1). Apres quelques verres sans alcool et une discussion dans le salon les 2 protagonistes se sont embrassés à plusieurs reprises. Le mis en cause mène alors la victime dans sa chambre ou ils ont une relation sexuelle que Mme. (P-1) décrit comme violente et non consentie." »
Proxénétisme aggravé auteur mis en contact avec la victime par réseau de communications électroniques	14	0,21	19	0,11	74	« La victime se prostitue au sein de l'hôtel [nom d'hôtel] depuis novembre 2018. Les surveillances démontrent qu'elle est protégée par l'auteur. Celui-ci se charge également d'alimenter le site internet [site internet] en photographies et réserve les chambres d'hôtels. La zone d'action se concentre sur la moitié de la France et cela depuis novembre 2017. La victime travaille 10 jours par mois et le nombre de prestations journalières est de 5 à 10 clients pour un minimum de 80 ? par clients. L'auteur bénéficie en partie des subsides de cette prostitution. »
Exposition non autorisée d'un dispositif technique ayant pour objet la captation de données informatiques	6	0,09	9	0,05	67	« Le fils des victimes signale que le réseau WIFI de ses parents chez qui il réside un appareil ne leur appartenant pas se connecte. Il remarque également que la lumière du garage peut s'allumer sans raison. La piscine installée sur le terrain clôture de la propriété a été vidée en partie sans anomalie repérée. »
Détention sans motif légitime d'équipement d'instrument de programme ou donnée conçu ou adapté pour une atteinte au fonctionnement d'un système de traitement automatisé de données	5	0,08	9	0,05	56	« Un phénomène de JACKPOTTING est détecté sur le grand ouest de la FRANCE. il s'agit de vols de billets présents dans les DAB. les auteurs fracturent le plafonnier à l'aide d'un petit pied de biche puis sortent le câble reliant l'écran à un ordinateur présent dans le DAB. ils branchent ensuite ce câble à leur PC portable puis à l'aide de partage de connexion prennent contact avec un hacker installe à l'étranger qui prend la main sur la suite de l'opération. Il fait sauter les sécurités puis la personne présente récupère les billets qui sortent du DAB. »

Note : Seules les Natinf ayant plus de 5 infractions dans la classe « cyber » ont été conservées.
Lecture : 100 % des infractions relatives à l'interruption volontaire des communications électroniques sont mal classées par l'algorithme.
Champ : France + COM.
Source : Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2018 et 2020.

Annexe 3 : Proportion d'infractions appartenant à la classe « non cyber » mal classées pour les 5 Natinf les moins bien classées

Libellé Natinf	Nombre d'infractions mal classées	Proportion dans le total des infractions mal classées ayant une autre Natinf (en %)	Nombre d'infraction total	Proportion de la Natinf dans le total des infractions de la classe « non cyber » (en %)	Proportion d'infractions mal classées au sein de la Natinf (en %)	Exemple de Manop
Recel de bien provenant d'abus de confiance par personne recouvrant des fonds ou des valeurs pour le compte de tiers	10	0,60	12	0,00	83	« La mise en cause fait une rencontre sur internet. Elle entretient une relation < amoureuse > avec cet homme pendant un 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 [site de transfert d'argent] et de le réexpédier par même voie au BENIN. Elle fait ainsi transiter environ 60 000? qui proviennent d'escroqueries diverses commises par le biais d'Internet »
Expédition de correspondance à découvert contenant une diffamation – P&T	6	0,36	10	0,00	60	« La victime est président du conseil des parents d'élèves de [association]. A l'issu 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 « accès de colère » »
Blanchiment concours à une opération de placement dissimulation ou conversion du produit d'une escroquerie commise en bande organisée	5	0,30	12	0,00	42	« Entre septembre et décembre 2017 les stations-services [nom de la station] font l'objet d'un piratage de type skimming. Au total 970 cartes bancaires sont compromises pour un préjudice de 171.81817 euros tentes dont 73.63176 euros réalisés. Les retraits sont effectués en [nom du pays]. »
Recel de chèque contrefaisant ou falsifié	18	1,08	46	0,01	39	« La MEC se voit proposer l'encaissement d'un chèque de 361570€ sur son compte pour en restituer une somme par mandat cash [site de transfert d'argent]. Ne pouvant le faire car sous curatelle elle donne les coordonnées de sa fille qui encaisse le chèque et lui remet la somme de 3500€. La MEC ne se souviens pas du montant renvoyé par mandat »
Menace de mort matérialisée par écrit, image ou autre objet, commise en raison de la race, l'ethnie, la nation ou la religion	9	0,54	23	0,01	39	« Suite à une dispute familiale la victime reçoit un sms de la part d'un de ses petit frère contenant des menaces de mort à son encontre ainsi que envers ces enfants sur fond de religion musulmane. »

Note : Seules les Natinf ayant plus de 10 infractions dans la classe « non cyber » ont été conservées.
Lecture : 60 % des infractions relatives à l'expédition de correspondance à découvert contenant une diffamation sont mal classées par l'algorithme.
Champ : France + COM.
Source : Bases des infractions de crimes et délits enregistrés par la gendarmerie entre 2018 et 2020.

Annexe 4 : Déroulement de la labellisation en phase 1

Un exemple de structure du document de labellisation

Comme décrit dans le corps du document dans la partie *3.1.3. Proposition de labellisation multi-classes du SSMSI,* la labellisation se déroulera en deux phases. La première phase permettra de tester le protocole de labellisation. Ci-dessous un exemple de structure du protocole de labellisation.

Dans la section 1 est décrit le déroulement la labellisation avec quelques consignes. La labellisation multi-classes proposée par le SSMS sera déclinée dans la section 2. Le document présente une structure possible du document qui reste à enrichir. Ensuite des consignes de priorisation entre les différentes catégories sont présentées dans la section 3 et un récapitulatif sera décrit dans la section 4.

SOMMAIRE

Section 1 - Consigne de labellisation pour la première phase d'annotation :

La labellisation se déroulera en deux temps. Le but de la première phase est de tester le processus de labellisation des infractions liées au numérique par thématique et degré d'utilisation des outils numériques. Plus précisément il s'agira de tester les deux grilles (deux ou trois niveaux de degrés d'utilisation des outils cyber).

La première phase de la labellisation se déroulera à petite échelle, avec entre 100 et 200 MANOP à labelliser par personne. Quatre personnes sont à mobiliser pour cette première phase. Chacune labellise en autonomie avec pour ressource les documents *Protocole de labellisation des infractions liées au numérique – niveau 3* ainsi qu'avec le *Tutoriel d'utilisation de l'interface de labellisation*.

Deux séries de MANOP seront à labelliser. Ainsi deux personnes labellisent la même série de MANOP. Le but n'est pas d'échanger pendant la phase d'annotation mais bien de labelliser en autonomie. Ainsi aucun arbitrage ne sera fait durant cette phase. Chaque cas qui pose question et s'avère difficile à labelliser doit être signalé. Le processus de report des cas tangents sera détaillé ultérieurement. Une réunion de mise en commun et d'échanges sera organisée à l'issue de cette phase afin de prendre en compte les retours, et discuter des éventuels changements et des deux protocoles de labellisation. C'est lors de cette réunion que des arbitrages seront fait sur les cas tangents et que les personnes qui auront labellisé les mêmes MANOP pourront échanger sur ces cas.

Section 2 – Thématiques à labelliser (protocole en cours d'écriture) :

Thème 1 - Escroquerie :

E1 Escroqueries commises par des moyens numériques complexes

Résumé	
Label : Escroqueries commises par des moyens numériques complexes	***Degrés :*** Moyen numérique les plus complexes
Code à annoter : E1	***Motif :*** Escroquerie, crapuleux
Cible : Personne physique, personne morale, Etat	

Définition

Liste des infractions inclues

Plus en détail

Exemple de MANOP à inclure

Infractions exclues et situations tangentes

Exemple de MANOP tangente

E2 Escroqueries commises uniquement dans le cyberespace mais n'ayant pas nécessité d'intrusion ou piratage

Résumé	
Label : Escroqueries commises dans le cyberespace	***Degrés :*** Lieu du cyberespace uniquement, sans action sur les supports numériques
Code à annoter : E2	***Motif :*** Escroquerie, crapuleux
Cible : Personne physique, personne morale, Etat	

Définition

Liste des infractions inclues

Plus en détail

Exemple de MANOP à inclure

Infractions exclues et situations tangentes

Exemple de MANOP tangente

E3 Escroqueries facilitées ou rendues possible grâce aux outils numériques

Résumé	
Label : Escroqueries commises en réel mais facilitées par des moyens numériques	***Degrés :*** Interaction numérique et contact réel
Code à annoter : E3	***Motif :*** Escroquerie, crapuleux
Cible : Personne physique, personne morale, Etat	

Définition

Liste des infractions inclues

Plus en détail

Exemple de MANOP à inclure

Infractions exclues et situations tangentes

Exemple de MANOP tangente

Thème 2 – Atteinte à la personne :

A1 Atteintes aux personnes commises à l'aide de moyens numériques complexes

Résumé	
Label : Atteintes aux personnes commises par des moyens numériques complexes	***Degrés :*** Moyen numérique les plus complexes
Code à annoter : A1	***Motif :*** Atteinte à la personne (pas de motif crapuleux)
Cible : Personne physique	

Définition

Liste des infractions inclues

Plus en détail

Exemple de MANOP à inclure

Infractions exclues et situations tangentes

Exemple de MANOP tangente

A2 Atteintes aux personnes commises dans le cyberespace

Résumé	
Label : Atteintes aux personnes commises dans le cyberespace	***Degrés :*** Lieu du cyberespace uniquement, sans action sur les supports numériques
Code à annoter : A2	***Motif :*** Atteinte à la personne (pas de motif crapuleux)
Cible : Personne physique	

Définition

Liste des infractions inclues

Plus en détail

Exemple de MANOP à inclure

Infractions exclues et situations tangentes

Exemple de MANOP tangente

A3 Atteintes aux personnes commises à la fois dans le réel et le cyberespace ou facilitées par les outils numériques

Résumé	
Label : Atteintes aux personnes commises en réel mais facilitées par des moyens numériques	***Degrés :*** Interaction numérique et contact réel
Code à annoter : A3	***Motif :*** Atteinte à la personne (pas de motif crapuleux)
Cible : Personne physique	

Définition

Liste des infractions inclues

Plus en détail

Exemple de MANOP à inclure

Infractions exclues et situations tangentes

Exemple de MANOP tangente

Thème 3 – Trouble à l'ordre public :

TOP Infractions qui sont des troubles à l'ordre public et à la sécurité de l'état

Résumé	
Label : Trouble à l'ordre public	***Degrés :*** Pas de déclinaison par degré
Code à annoter : TOP	***Motif :*** Motif non crapuleux
Cible : Etat, société civile, personnes morales	

Définition

Liste des infractions inclues

Plus en détail

Exemple de MANOP à inclure

Infractions exclues et situations tangentes

Exemple de MANOP tangente

Thème 4 – Infraction au droit d'auteur :

AER Non-respect des droits d'auteur et des réglementations spécifiques liés aux outils numériques

Résumé	
Label : Infraction au droit d'auteur et aux réglementations des outils numériques	***Degrés :*** Pas de déclinaison par degré
Code à annoter : AER	***Motif :*** Fraude, appropriation, sans motif, usage sans motif crapuleux
Cible : Personnes morales, personnes physiques, collectivité publique	

Définition

Liste des infractions inclues

Plus en détail

Exemple de MANOP à inclure

Infractions exclues et situations tangentes

Exemple de MANOP tangente

Thème 6 – Atteintes sur les supports numériques sans motif apparent :

ASM Atteinte sur les supports numériques ou commise dans le cyberespace sans motif apparent

Résumé	
Label : atteinte sur les supports numériques sans motivation ou à motivation différente des précédentes catégories	***Degrés :*** Pas de déclinaison par degré
Code à annoter : ASM	***Motif :*** pas de motif apparent, ou motif différent de crapuleux, atteinte à la personne, trouble à l'ordre public.
Cible : Personnes morales, personnes physiques	

Définition

Liste des infractions inclues

Plus en détail

Exemple de MANOP à inclure

Infractions exclues et situations tangentes

Exemple de MANOP tangente

Thème 7 – Pas de lien avec les infractions liées au numérique

NC Absence de lien avec les infractions liées au numérique

Résumé	
Label : Sans lien avec les infractions liées aux outils numériques	***Degrés :*** Pas de déclinaison par degré
Code à annoter : NC	***Motif :*** Tout motif accepté
Cible : Personnes morales, personnes physiques, collectivité, Etat…	

Définition

Liste des infractions inclues

Plus en détail

Exemple de MANOP à inclure

Infractions exclues et situations tangentes

Exemple de MANOP tangente

Section 3 - Règles de priorisation :

Si une infraction liée au numérique est à la fois une escroquerie ou une atteinte à la personne elle doit être classée comme une escroquerie.

Plus généralement si une infraction est à la fois une escroquerie vs tout autre thématique, il faut la labelliser comme une escroquerie.

Si une infraction est à la fois une atteinte à la personne vs tout autre motif (escroquerie non comprise), il faut la labelliser comme une atteinte à la personne.

Les infractions commises en réel mais avec une conséquence possible sur internet ne sont pas dans le champ des infractions liées au numérique.

Les infractions au support matériel des outils numériques (vol de téléphone, ordinateur ou tablette, détérioration de câble, casse de serveur, etc…) ne rentrent pas dans le périmètre des infractions liées au numérique.

Section 4 – Grille récapitulative :

Label	Code à annoter	Exemple
Escroqueries commises par des moyens numériques complexes	**E1**	*Rançongiciel, virus*
Escroqueries commises dans le cyberespace	**E2**	*Arnaque à la romance, aux bitcoins, aux faux sites internet*
Escroqueries commises en réel mais facilitées par des moyens numériques	**E3**	*Arnaque à la vente en ligne et au paiement par chèque ou faux billets*
Atteintes aux personnes commises par des moyens numériques complexes	**A1**	*Intrusion dans boîte ou compte réseaux sociaux*
Atteintes aux personnes commises dans le cyberespace	**A2**	*Cyber-harcèlement*
Atteintes aux personnes commises en réel mais facilitées par des moyens numériques	**A3**	*Agression en ligne et en réel*
Trouble à l'ordre public	**TOP**	*Menace contre l'Etat, terrorisme*
Infraction au droit d'auteur et aux réglementations des outils numériques	**AER**	*Non figuration des mentions légales sur site internet, téléchargement illégal*
Atteinte sur les supports numériques sans motivation ou à motivation différente des précédentes catégories	**ASM**	*Infraction dans le cyberespace et intrusion dans STAD sans motif ou autre motif que ceux déjà développés*
Sans lien avec les infractions liées aux outils numériques	**NC**	*Infractions qui sont sans lien avec le numérique où dont le caractère « cyber » n'est pas suffisamment constitué*