REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIREMINISTRE
DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE SAAD DAHLEB BLIDA

Faculté des sciences Département : Informatique

MEMOIRE DE FIN D'ETUDES EN VUE DEL'OBTENTION DU DIPLÔME DE MASTER

EN INFORMATIQUE

Option : ingénierie des logiciels

CONCEPTION ET REALISATION D'UN CRYPTO SYSTEME BASEE SUR L'APPRENTISSAGE AUTOMATIQUE

Réalisé par :

- Larbi Younes
- Djahlat yasmina

Promotrice: Mme. Ghebghoub Yasmina
Présidente: M. Ferfera
Examinatrice: M. Midoun

Année Universitaire 2020/2021

REMERCIMENT

Avant tout, je remercie ALLEH le tout puissant de m’avoir guidé, aidéet donné la foi, laforce et le courage pour accomplir ce travail.
En préambule à ce mémoire, il m’est agréable de citer et adresser mesremerciements les plussincères aux personnes qui m’ont apporté leursaides et qui ont contribué à l’élaboration et aubon déroulement de ce travail :

A ma promotrice **M^{me} GHEBGHOUB**

« Mes remerciements les plus respectueux de m'avoir aidé dans ce travail, le regard critique, juste et avisé que vous aviez porté sur mes travaux ne peut que m'encouragerà être engagée dans mes recherches »

Aux membres de jury qui auront à juger et ce travail et d’avoir acceptéde l'examiner.
À la fin, il nous est agréable d'adresser nos vifs remerciements à tousceux qui m’ont aidé deprès ou de loin à élaborer ce mémoire.

Didicace

Je dédie cet humble acte en signe de reconnaissance et de respect auxpersonnes suivantes :
A ma mère et mon père qui sont venus pour moi et mes frères âme etâme
Les bougies de ma vie qui m'ont donné la vie et qui ont toujours été làpour moi.
Mes frères et mes sœurs .
« Lui qui m'a toujours entouré et m'a motivé sans cesse tout au long dece projet, et je lui souhaite un bel avenir plein de réussite »
Tout le monde avec le nom **LARBI** ET **DJAHLAT**
Aux personnes qui ont toujours été à mes côtés, mes bons amis, mescollègues.
À tous ces orateurs, j'exprime mon respect et ma gratitude.

Résume

Avec le temps, de nombreux domaines ont progressé, et le plus important de cesdomaines est le domaine de la technologie et d'Internet, et grâce à ce développement, le nombre d'utilisateurs d'Internet a considérablement augmenté et le nombre de développeurs dans ce domaine a augmenté, et donc les systèmes sur Internet se sont développés avec une grande différence en raison de la diversité de la mentalité et de la force des personnes dans ce domaine, et avec cela le nombrede pirates informatiques piratant les informations d'autres personnes a augmenté , etpour cela nous proposons dans ce travail un crypto système basé sur les algorithmesde l'apprentissage automatique afin de rendre la proposition plus intelligente d'un côté et d'autre cote afin d'aider les utilisateurs à choisir les méthodes de chiffrementles plus efficaces.

Mots clés : l'apprentissage automatique , algorithme , un crypto système , chiffrement ,

Abstract

Over time, many fields have progressed, and the most important of these fields is the field of technology and the Internet, and thanks to this development, the number of Internet users has increased significantly and the number of developers in this field has increased significantly, and therefore the systems on the Internet have developed with a big difference due to the diversity of mentality and strengthof people in this field, and with it the number of hackers hacking information fromother people has increased, and for that we explain in this brief the characteristics of information security and how to encrypt information in several ways through machine learning, which we will use to help ordinary Internet users to get the best option to encrypt information from hackers, and to provide many well-known encryption algorithms and access them fa target for the professional user familiar with these algorithms, and we will embody this idea in an easy-to-use application.

Keywords: machine learning, algorithm, crypto system, encryption

ملخص

بتقدم الزمن تقدمت تطورت الكثير من المجالات، و من اهم هذه المجالات مجال التكنولوجيا و الأنترنت ، و من خلال هذا التطور زاد كثيرا عدد المستخدمين لأنترنت ، و زاد كثيرا عدد المطورين في هذا المجال ، و بذلك تطورت كثيرا الأنظمة المتواجدة عبر الأنترنت و باختلاف كبير بسبب تنوع عقلية الناس و قوتهم في هذا المجال ، و بهذا دخل و زاد عدد المخترقين المتسللين الى معلومات الآخرين كثير من الناس ، و من اجل هذا نوضح في هذه المذكرة ميزات امان المعلومات و كيفية تشفير المعلومات بالعديد من الطرق من خلال التعلم اللي ، الذي سنستخدمه لمساعدة المستخدمين العاديين للأنترنت للحصول على افضل خيار لتشفير المعلومات ضد المتسللين ، و توفير العديد من خوارزميات التشفير المعروفة و الوصول اليها بسهولة بالنسبة للمستخدم المحترف ذو علم ب الخوارزميات ، و سنجسد هذه الفكرة في تطبيق سهل الاستخدام

الكلمات الرئيسية: التعلم الآلي ، الخوارزمية ، نظام التشفير ، التشفير

Table des matières :

Table des figures

Introduction général :

Le besoin de cacher des informations a préoccupé l'homme depuis le début de la civilisation. Le secret apparaît comme un facteur essentiel dans toute lutte pour le pouvoir, que ce soit dans un cadre militaire ou diplomatique. Aujourd'hui, de plus en plus d'applications dites civiles sont utilisées pour transmettre, stocker ou échanger des données existantes, entre deux ou plusieurs interlocuteurs, sur les réseaux de télécommunications. Détecter les actes de malveillance devient rapidement une nécessité. Les mesures préventives se sont avérées insuffisantes et ont conduit à la mise en place de systèmes de détection d'intrusions. L'intrusion est définie comme toute tentative qui pourrait compromettre l'exhaustivité, la confidentialité ou la disponibilité d'un réseau ainsi que toute tentative de contourner les dispositifs de sécurité sur le réseau ou l'appareil. Ces tentatives d'effraction peuvent être bénignes, extrêmement dangereuses et nocivespour l'entreprise.

Les laboratoires de recherche utilisent la cryptographie pour échanger des informations et des données confidentielles et confidentielles entre ces membres. Toutes ces informations et données doivent être cachées ou cachées afin qu'elles ne puissent pas être comprises par des tiers non autorisés, c'est pourquoi le cryptage est utilisé. En fait, chiffrement est un terme plutôt générique qui combine un groupe de techniques de chiffrement et de déchiffrement pour assurer la confidentialité et l'intégrité des données échangées dans les réseaux de communication existants.C'est le moyen le plus sûr de protéger ces données. Peu importe à quel point il est vulnérable aux attaques, vous pouvez le briser. Comme nous savons que les gens diffèrent dans leur niveau de connaissance, nous avons également pris en charge les utilisateurs réguliers en incluant l'apprentissage automatique dans le système de cryptage et ils peuvent obtenir le meilleur encodeurpour leurs fichiers.

Les techniques d'apprentissage automatique ont été largement appliquées ces dernières années pour la détection d'intrusions. Par exemple : arbres de décision, algorithmes et programmation génétique, Naif Bayes ,K plus proches voisins (KNN), réseaux de neurones ou encore SVM.

Généralement la cryptographie est utilisée pour rendre les documents plus confidentiel et elle est utilisée par des experts ou bien des professionnels de l'informatique et de la sécurité informatique. Dans le but de rendre la cryptographie plus utilisée même pour les personnes qui n'ont pas une expérience dans ce domaine. On propose un crypto système basé sur les algorithmesde l'apprentissage supervisé dont le but d'analyser les besoins de l'utilisateurs côté sécurité et de l'aider à prendre la bonne décision et choisir le meilleur algorithme de chiffrement

Nous allons proposer une approche intelligente basée sur la cryptographie et l'apprentissage automatique dans le but d'aider les utilisateurs à choisir les meilleurs algorithmes de chiffrements.Ce système divise les utilisateurs en deux catégories (Simple, Expert). Grace aux algorithmes d'apprentissage. Dans le cas d'un utilisateur simple, notre crypto récupère en entrée les critères desécurité proposées par le simple utilisateur et propose en sortie le meilleur algorithme de chiffrement pour crypter le document visé.

L'organisation du mémoire :

Cette thèse est organisée en quatre chapitres continus :

• **Premier Chapitre** : présente les concepts de base liés à la sécurité de l'information, lespropriétés de sécurité de l'information, les risques et les attaques potentielles contre les systèmes d'information et comment s'en protéger, les objectifs de sécurité de l'information, les concepts de cryptage et les types de cryptage ancien et nouveau, et nous présenterons les algorithmes de cryptage les plus importants.

• **Deuxième Chapitre:** passe en revue les concepts de base de l'apprentissage automatique, comment l'appliquer et ses domaines d'application, les types d'apprentissage automatique et les algorithmes les plus importants utilisés dans chaque type.

. Troisième Chapitre : Dans ce chapitre, nous avons présenté la solution proposée, une introduction sur KDD, ainsi que nous avons mis le point sur les critères de création de notre data set, et Dans ce chapitre, nous avons précisé le problème que nous voulons résoudre, et nous avons également expliqué la solution proposée en combinant codage et apprentissage automatique, et à la fin nous avons expliqué les outils utilisés dans ce travail.

CHAPITRE I

LA SECURIT INFORMATIQUE ET LA CRYPTOGRAPHIE

1 Introduction

Protéger l'information et maintenir sa sécurité nécessite de solides connaissances dans le domaine de la pénétration et du vol d'information, pour créer un système solide en termes de sécurité de l'information, nous avons donc discuté dans ce chapitre des termes liés à la sécurité de l'information et des méthodes de pénétration dans les systèmes d'information, et à l'autre extrémité de ce chapitre, nous avons parlé des anciennes et des nouvelles méthodes de cryptage, et nous avons expliqué comment sécuriser les informations grâce au cryptage symétrique et symétrique, et nous avons expliqué trois algorithmes de chiffrement, et nous avons également expliqué trois algorithmes de chiffrement symétriques, que nous avons utilisés dans notre proposition d'application .

2 Vocabulaire de base

La sécurité informatique utilise un vocabulaire bien défini que nous utilisons dans notre mémoire.
De manière à bien comprendre nos différents chapitres, il est nécessaire de définir certains termes:

2.1 La sécurité

La sécurité informatique est l'ensemble des moyens mis en œuvre pour minimiser la vulnérabilitéd'un système contre des menaces accidentelles ou intentionnelles [2].

Il faut distinguer la sécurité-innocuité (safety, en anglais), de la sécurité-confidentialité (Security,enanglais) [1] :

- La sécurité-innocuité vise à se protéger des défaillances catastrophiques, c'est-à-dire celles pour lesquelles des conséquences sont inacceptables vis-à-vis du risque encouru par les utilisateursdu système;

- La sécurité-confidentialité correspond à la prévention d'accès ou de manipulations non autorisées de l'information et concerne la lutte contre les fautes intentionnelles (virus, bombes logiques, chevaux de Troie, etc.). Elle vise également à garantir l'intégrité des informations fournies aux utilisateurs.

- Le concept de sécurité des systèmes d'information recouvre un ensemble de méthodes,

techniques et outils chargés de protéger les ressources d'un système d'information afin d'assurer ladisponibilité des services, la confidentialité et l'intégrité des informations [4].

▪ De plus, avec le développement de l'informatisation des échanges, la simple affirmation de la valeur d'une information n'est plus suffisante : il faut lui adjoindre des propriétés nouvelles, comme l'authentification, la traçabilité (ou la non-répudiation) ...

2.2 Les vulnérabilités

Ce sont les failles de sécurité dans un ou plusieurs systèmes. Ces vulnérabilités peuvent être organisationnelle (ex: pas de politique de sécurité), humaine (ex: pas de formation des personnels),logicielles ou matérielles (ex: utilisation de produits peu fiables ou non testés). Tout système vu dans sa globalité présente des vulnérabilités, qui peuvent être exploitables ou non. Les attaquesElles représentent les moyens d'exploiter une vulnérabilité. Il peut y avoir plusieurs attaques pourune même vulnérabilité mais toutes les vulnérabilités ne sont pas exploitables [5].

2.3 Les contre-mesures

Ce sont les procédures ou techniques permettant de résoudre une vulnérabilité ou de contrer une attaque spécifique (auquel cas il peut exister d'autres attaques sur la même vulnérabilité) [5].

2.4 Les menaces

Ce sont des adversaires déterminés capables de monter une attaque exploitant une vulnérabilité[5].

Pour d'autres définitions, consultez la norme ISO 7498-2 qui ne définit pas moins de 59 termes.

3 Objectifs de la sécurité

La sécurité d'un système d'information se traduit par une politique de sécurité regroupant des propriétés ne devant pas être mises en défaut par les utilisateurs du système. Une bonne politiquedesécurité doit préserver les aspects de [6] :

3.1 L'identification

L'utilisateur d'un système ou de ressources diverses possède une identité (une sorte de clé primaire d'une base de données) qui détermine ses lettres de crédits (credential) et ses autorisations d'usage. Cette dernière peut être déclinée de multiples manières, compte

utilisateur (login) d'un système d'exploitation ou techniques biométriques empreinte digitale, empreinte vocale …[6].

3.2 L'authentification

Cette opération consiste à faire la preuve de son identité. Par exemple on peut utiliser un mot de passe, ou une méthode de défi basée sur une fonction cryptographique et un secret partagé[6].

3.3 La confidentialité

C'est la garantie que les données échangées ne sont compréhensibles que pour les deux entités qui partagent un même secret souvent appelé association de sécurité (SA). Cette propriété implique la mise en œuvre d'algorithmes de chiffrements (par exemple RC4 ou DES) [6].

3.4 L'intégrité

L'intégrité des données (MAC, Message AuthentiCation). Le chiffrement évite les écoutes indiscrètes, mais il ne protège pas contre la modification des informations par un intervenant malintentionné. Des fonctions à sens unique (encore dénommées empreintes) telles que MD5 ou SHA1 réalisent ce service [6].

3.5 La non répudiation

Elle consiste à prouver l'origine des données. Généralement cette opération utilise une signature asymétrique en chiffrant l'empreinte du message avec la clé RSA privée de son auteur.

Afin de pouvoir sécuriser un système, il est nécessaire d'identifier les menaces potentielles, et donc de connaître et de prévoir la façon de procéder de l'ennemi. C'est pourquoi nous allons dans le point suivant analyser ce que nous appellerons "l'anatomie d'une attaque"[6] .

4 Anatomie d'une attaque

Dans cette section, nous allons présenter le squelette de toute attaque informatique. Cette squelette est constituée de cinq verbes anglophone : Probe, Penetrate, Persist, Propagate, Paralyze fréquemment appelés les "5 P". Détaillerons chacune de ces étapes [7] :

- **Probe** : consiste à collecter des informations sur un système cible par le biais d'outils. Par exemple, détermination de la version des logiciels utilisés par un scan des ports avec le programme Nmap ou encore les logiciels firewalk, hping ou SNMP Walk qui permettent de découvrir la nature d'un réseau.

- **Penetrate** : cette étape consiste à utiliser les informations récoltées pour s'introduire dansun réseau. Plusieurs techniques peuvent être utilisées telle que l'attaque par dictionnaire qui permet outrepasser les protections par mot de passe ou par l'exploitation des failles applicatives que nousverrons ci-après.

- **Persist** : création d'un compte avec des droits de super utilisateur pour pouvoir se ré infiltrer ultérieurement. Une autre technique consiste à installer une application de contrôle à distance capable de résister à un reboot (ex : un cheval de Troie).

- **Propagate** : cette étape consiste à observer ce qui est accessible et disponible sur le réseau local.

- **Paralyze** : cette étape peut consister en plusieurs actions. Le pirate peut utiliser le serveur pour mener une attaque sur une autre machine, détruire des données ou encore endommager le systèmed'exploitation dans le but de planter le serveur.

Afin de garantir une sécurité suffisante et atteindre les objectifs cités ci-dessus, il faut que toute attaque soit bloquée pendant l'une des 5 phases du squelette des "5P". Généralement, plus on se rapproche de l'étape suivante, plus le problème sera difficile et long à résoudre et que l'on se rapproche de la réussite de l'attaque. Nous pouvons dire donc que la protection des IDS intervientdès l'apparition ou début d'une attaque [7].

Afin de contrer ces attaques il est indispensable de connaître les principaux types d'attaques afindemettre en œuvre des dispositions préventives.

5 Les différents types d'attaques

Nous énumérons ci-dessous quelques grandes classes d'attaques des systèmes d'information quiillustrent la diversité des méthodes :

- Les attaques réseaux
- Les attaques applicatives
- Le déni de service
- Les attaques des données

En pratique, les attaques réelles sont souvent une combinaison des techniques ci-dessous et les scénarios peuvent être complexes [7] .

5.1 Les attaques réseaux

Les attaques réseaux s'appuient sur des vulnérabilités liées directement aux protocoles ou à leur implémentation. Il en existe un grand nombre. Néanmoins, la plupart d'entre elles ne sont que desvariantes des cinq attaques réseaux les plus connues aujourd'hui.
Les techniques de scan, Fragments attacks , IP Spoofing , ARP Spoofing , DNS Spoofing [9] [8]

5.2 Les attaques applicatives

Les attaques applicatives s'appuient principalement sur des vulnérabilités spécifiques aux applications utilisées. Ces failles peuvent être de natures diverses : problèmes de configuration, problèmes au niveau du code du logiciel, problèmes liés à de mauvaises interprétations de commandes ou de mauvaises exécutions de scripts.

5.3 Le Déni de service

Le déni de service est une attaque visant à rendre indisponible un service (application spécifique)ou la machine visée. Nous distinguerons deux types de déni de services, d'une part ceux dont l'origine est l'exploitation d'un bug d'une application et d'autre part ceux dus à une mauvaise implémentation d'un protocole ou à des faiblesses de celui-ci.

5.4 Les attaques des données

Les données transportées par le protocole applicatif (Contenu) peuvent constituer une menace pourl'intégrité du système qui les reçoit. Les principales attaques de ce type, nous trouvons: virus, ver,Applet Java, trojans…désignés par les codes malicieux ou Malwares [10].

6 Généralités sur la cryptographie

6.1 Terminologie

Une certaine confusion règne concernant les différents termes de la cryptographie, à cause en premier lieu de l'utilisation d'anglicismes (termes empruntés à l'anglais), ainsi nousallons définir la terminologie qui va être utilisée tout au long de l'étude afin d'éviter toute ambigüité [12] :

Clé : Une clé est un ensemble de paramètres utilisés en entrée d'une opération Cryptographique (chiffrement, déchiffrement).

Chiffrer ou chiffrement : transformation à l'aide d'une clé de chiffrement d'un message clairEn un message chiffré (cryptogramme), incompréhensible par des tiers n'ayant pas la Connaissance de la clé (en anglais Encryption). On utilise aussi le « crypter ».

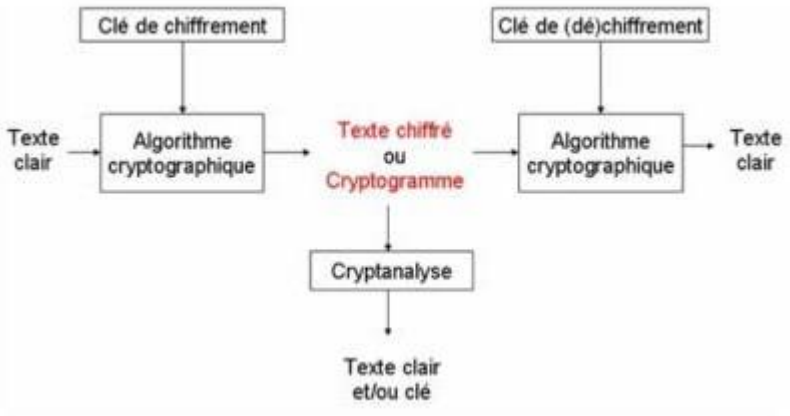


Figure 1 Protocole de chiffrement *[12]*

Déchiffrer ou déchiffrement : transformation qui consiste à retrouver les informations claires, à partir des informations chiffrées en utilisant la clé de déchiffrement.
Décrypter : retrouver le message clair correspondant à un message chiffré sans posséder la cléde déchiffrement.
Cryptosystème : Un Cryptosystème est constitué d'un algorithme cryptographique ainsi que toutes les clés possibles et tous les protocoles qui le font fonctionner.
Ceci dit, nous pouvons à présent donner une définition précise pour :

La cryptographie : Etymologiquement « écriture secrète », devenue par extension l'étude de cet art (donc aujourd'hui la science visant à créer des cryptogrammes, c'est-à-dire à chiffrer).
La cryptanalyse : science analysant les cryptogrammes en vue de les décrypter.
La cryptologie : C'est une science mathématique regroupant la cryptographie et la Cryptanalyse.

Plaintext : Terme anglais désignant le texte clair à chiffrer.

Ciphertext : Terme anglais désignant le texte chiffré. [11]

6.2 Techniques de chiffrement

Les premiers algorithmes utilisés pour le chiffrement d'une information ont été assez

rudimentaires dans l'ensemble, comme nous l'avons vu dans le paragraphe consacré à

L'histoire de la cryptographie. Ces anciennes méthodes de chiffrement ont constitué la premièrepierre

de l'édifice qui a mené vers les méthodes dites modernes. Comme nous allons

le voir dans la section qui suit ces méthodes dites modernes, au-delà de leur complexité se basent sur des principes artisanaux [13] .

6.3 Méthodes anciennes

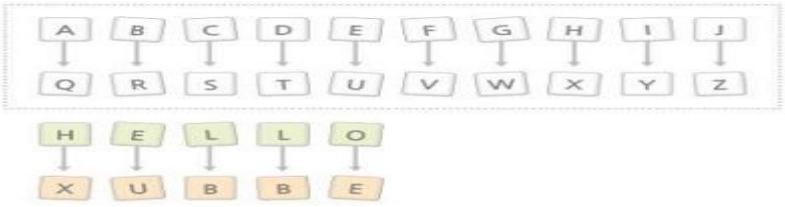
6.3.1 Chiffrement par substitution :

Le chiffrement par substitution consiste à remplacer systématiquement une lettre par Une autre. Citons l'exemple d'une fonction qui décale de n positions les lettres constituantUne phrase ce type de substitution est dit mono alphabétique.
Il existe un autre type de substitution, dit poly alphabétique qui introduit la notion de clés.

Exemples de chiffrement par substitution :

6.3.2 « Chiffrement de César » :

Cette méthode de chiffrement est considérée comme la plus ancienne méthode de



chiffrementPar

Figure 2 Chiffrement par substitution*[11]* .

substitution mono alphabétique. Son principe est assez élémentaire, il consistait Simplement à décaler les lettres d'un message de trois positions vers la droite dans l'alphabetlatin. La lettre A est ainsi transformée en Q, le B en R, etc [11] .

6.3.3 Chiffrement de Vigenère :

C'est l'un des premiers systèmes à introduire le concept de clé de chiffrement. Blaise de Vigenère (1523-1596) a repris le chiffre de César, ou le décalage utilisé change de lettre enlettre. Pour cela il a utilisé une table composée de 26 alphabets, écrits dans l'ordre, mais Décalés de ligne en ligne d'un caractère, mieux connue sous la table de Vigenère (voir figure 3).

Pour chiffrer un message, on choisit une clé qui sera un mot de longueur arbitraire. On Écrit ensuite cette clé sous le message à coder en la répétant aussi souvent que nécessaire pour Que sous chaque lettre du message à coder, on trouve une lettre de la clé. Pour coder, on Regarde dans le tableau l'intersection de la ligne de la lettre à coder avec la colonne de la lettreDe la clé [12].

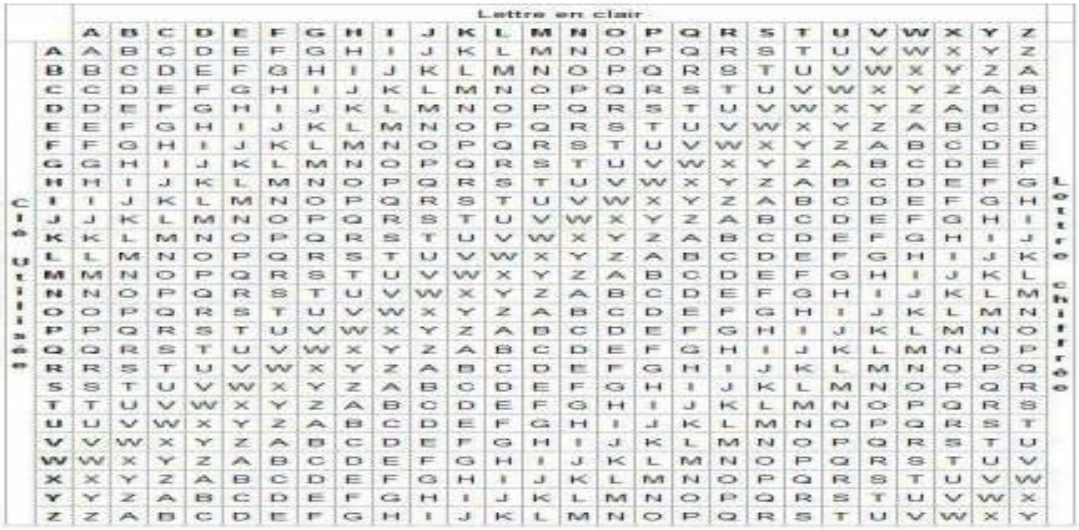


Figure 3 Table de Vigenère [12].

6.3.4 Permutation (transposition)

Chiffrement par permutation (Un chiffrement par transposition) est un chiffrement qui Consiste à changer l'ordre des lettres, le chiffrement par transposition demande de découper leTexte clair en blocs de taille identique. La même permutation est alors utilisée sur chacun desBlocs [14].

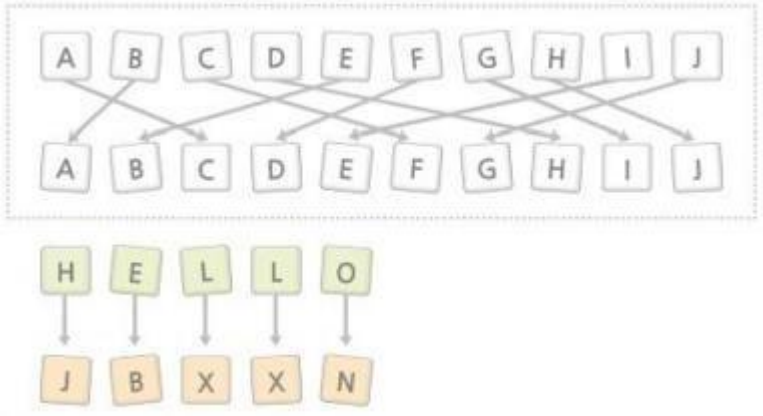


Figure 4 Chiffrement par transposition [14] .

6.3.5 Chiffrement par produit :

Le principe de ce mode de chiffrement consiste à utiliser des deux méthodes étudiées Précédemment, à savoir la substitution et la transposition. La combinaison de ces deux méthodeun chiffrement assez robuste qui fournit un bon niveau de sécurité, la plupart des

Algorithmes dechiffrement à clé symétrique comme nous allons le voir un peu plus loin Utilisent le chiffrement par produit [13] .

6.4 Méthodes modernes

6.4.1 Structure de chiffrement de Feistel 3 :

Cette structure est pratiquement la base de tous les algorithmes de chiffrement modernes.

La plupart des algorithmes de la fin du XX éme siècle, étaient des schémas de Feistel, (DES, Blowfish, RC5,…).

La structure de Feistel est assez simple et le chiffrement et déchiffrement sont similaires. Elle est basée sur des opérations de substitutions et de permutations avec une fonction Principale qui change de clé à chaque round [13] .

Chaque round applique :

1- une fonction F qui comporte des permutations via des P-BOX et des substitutions via des S-BOX.

2- Mixage linéaire via l'opération XOR.

3- Application de la clé de round qui est intégrée dans la fonction F via une opération XOR .

6.4.2 Fonctionnement de la cryptographie moderne

À côté de la fonction de chiffrement, qui permet de préserver le secret des données lors d'une transmission, et qui a été utilisée depuis très longtemps, la cryptographie moderne a développé denouveaux buts à atteindre et qu'on peut énumérer de manière non exhaustive: confidentialité, intégrité des données, authentification des divers acteurs, non-répudiation d'un contrat numérique, signature numérique, certification, contrôle d'accès, gestion des clés, preuve de connaissance.

La cryptologie moderne a pour l'objet l'étude des méthodes qui permettent d'assurer les services d'intégrité, d'authenticité et de confidentialité dans les systèmes d'information et de communication. Elle recouvre aujourd'hui également l'ensemble des procédés informatiques devant résister à des adversaires [11].

6.5 Les type de cryptage

6.5 .1 La cryptographie symétrique

Les chiffrements symétriques aussi connue sous le nom de cryptographie à clé secrète sont les héritiers des méthodes anciennes des cryptographies. On parle de cryptogra- phies symétriques lorsqu'un texte est crypté et décrypté avec la même clé, la clé secrète). L'expéditeur et le destinataire disposent chacun d'un algorithme et d'une clé qui doit être échangée entre eux pour respectivement chiffrer et déchiffrer des messages. Cette clé doit rester secrète sous peine qu'un tiers parvienne à déchiffrerles correspondances Le terme "symétrique" vient de cette particularité.L'échange des clés secrètes, qui doit se faire par un canal sécurisé, est souvent le pointfaible de ces méthodes de chiffrement. La figure ci-dessous illustre le principe de la cryptographie symétrique.

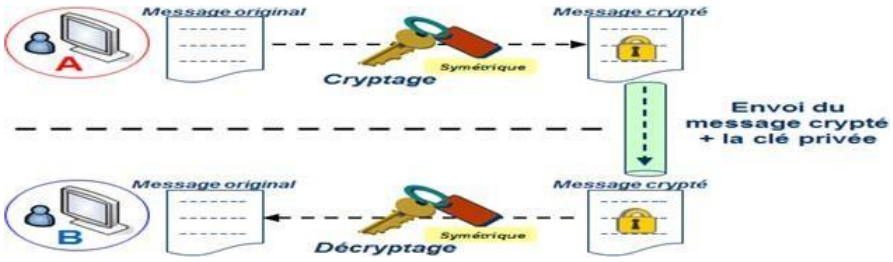


Figure 5 principe de chiffrement symétrique [16].

En général, les algorithmes de chiffrement symétrique sont très rapides notamment sur lesordinateurs. D'une part, ils ont une complexité moins élevée qu'un algorithme de chiffrement asymétrique et une simplicité d'implémentation ainsi une consommation de ressources et de bandepassante faible (rapidité de transmission). D'autre part, le processus de chiffrement est très rapideet sa réalisation peu complexe. Bien que très avantageuse, dans le chiffrement symétrique, la clé privée doit être connue de plus d'une personne, ce qui impose un risque de voir la clé récupérée parune personne tierce quand

elle est transmise d'un correspondant à l'autre. Toute personne interceptant la clé lors d'un transfert peut ensuite lire, modifier et falsifier toutes les informations cryptées ou authentifiéesavec cette clé.

Classiquement, on distingue deux types d'algorithmes de chiffrement symétrique : les algorithmes de chiffrement par blocs et les algorithmes de chiffrement à flot. Comme leur nom lesuggère, les algorithmes de chiffrement par blocs prennent en entrée des blocs de texte clair de taille fixe (typiquement 64 ou 128 bits) et retournent des chiffrés de même longueur. Cette catégorie d'algorithme de chiffrement symétrique est sans doute la plus utilisée en pratique avec, comme représentants, le DES et le standard actuel AES. Les algorithmes de chiffrement à flot, quant à eux,permettent de réaliser un chiffrement bit à bit d'un flot de données en entrée. La clé secrète permetd'initialisera génération d'une longue séquence de

bits, la suite chiffrant, à la manière d'un gêné-rater pseudo-aléatoire. Cette suite chiffrant est ensuite utilisée pour masquer les bits de texte clairsuivant le principe du chiffrement à masque jetable de VERNAM.

Les algorithmes de chiffrement à flot sont plus rapides que ceux chiffrant par blocs mais étaient généralement considérés comme beaucoup plus faibles de point de vue cryptographique.

Dans la suite, nous présenterons quelques algorithmes de chiffrement par bloc, à savoir, DES, IDEA, Blowfish et AES, et un algorithme de chiffrement à flot, le RC4 [16].

6.5.1.1 DES (Data Encryption Standard)

Au début des années 70, le développement des communications entre ordinateurs a né cécité la mise en place d'un standard de chiffrement de données pour limiter la prolifération d'algorithmes différents ne pouvant pas communiquer entre eux. Pour résoudre ce problème, L'Agence Nationale de Sécurité américaine (NSA) a lancé des appels d'offres. La société IBM a développé alors un algorithme nommé Lucifer, relativement complexe et sophistiqué. Aprèsquelques années de discussions et de modifications, cetalgorithme, devenu alors DES, fut adopté au niveau fédéral le 23 novembre 1976 .

Le cahier des charges était le suivant:

· L'algorithme repose sur une clé relativement petite, qui sert à la fois auchiffre-ment et au déchiffrement,

· L'algorithme doit être facile à implémenter, logicielle ment etmatériellement, et doit être très rapide,

· Le chiffrement doit avoir un haut niveau de sûreté, uniquement lié à la clé, et non la confidentialité de l'algorithme.

Les efforts conjoints d'IBM, qui propose Lucifer fin 1974, et de la NSA (National Se-curity Agency) conduisent à l'élaboration du DES (Data Encryption Standard). Le DESfut publié commestandard par le NBS le 15 janvier 1977 [17].

6.5.1.2 AES (Advanced Encryption Standard)

Advanced Encryption Standard aussi connu sous le nom de Rijndael est un algorithme de chiffrement symétrique établi comme standard par le NIST pour remplacer le DES

Principe

AES est un algorithme de chiffrement par blocs, les données sont traitéespar blocs de 128, 192

ou256 bits pour le texte clair et chiffré. La clé secrète a une longueur de 128, 192 ou 256 bits.

L'AES opère sur des blocs rectangulaires de 4 lignes et Nc colonnes, dont chaque termexi ; j (appelé octet ou byte) est composé de 8 bits (b = b7b6b5b4b3b2b1b0), et peut être repré- sentés algébriquementsous forme de polynômes de degrés ≤ 7 (b = b7X7+b6X6+b5X5+b4X4+b3X3+ b2X2 +b1X +b0),à coefficients dans {0, 1}. La clé peut être d'une longueurde 128, 192, ou 256 bits, de même pourle message clair et le message chiffré.

AES opère sur des matrices 4 * 4 (dans le cas ou la longueur du message = 128) dont les entrées sont des mots de 8 bits. On découpe le message clair en 16 blocs de 8 bits et

remplit en allant de haut en bas et de gauche à droite. Les quatre étapes d'une ronde sont (voirfigure 7) :

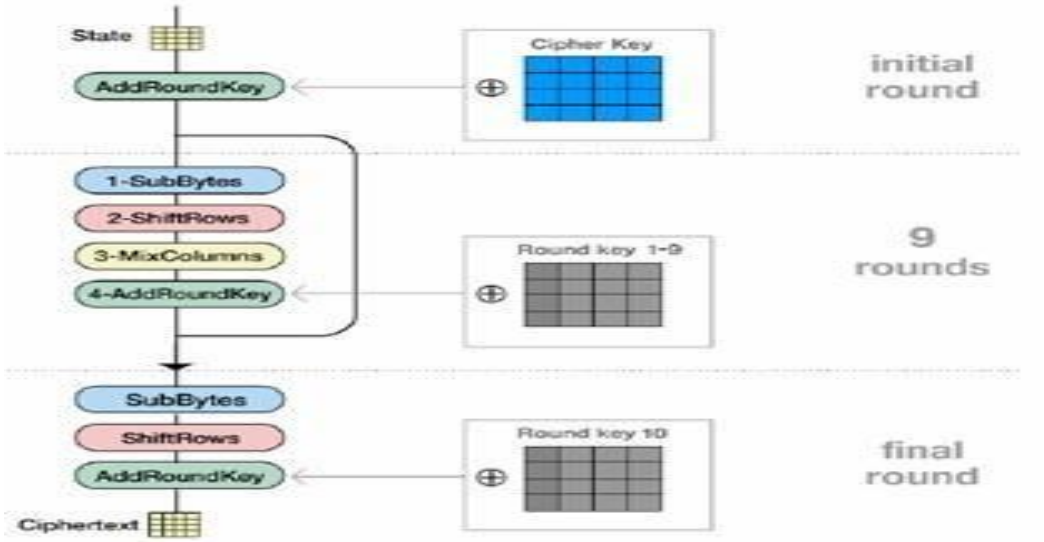


Figure 6 Chiffrement AES 128 bits .

- **SubBytes :** Chaque entrée est remplacée par un autre mot de 8 bits donné par un tableau decorrespondance ;

- **ShiftRows :** Les entrées sont décalées suivant un décalage circulaire à gauche d'un nombre decases dépendant de la ligne ;

- **MixColumns :** Chaque colonne est remplacée par une nouvelle colonne obtenueen transformantla colonne en un polynôme et en multipliant par un polynôme fixé ;

- **AddRoundKey :** Chaque entrée est remplacée par le XOR (ou exclusif) entre cette entrée etl'entrée correspondante dans une matrice 4 * 4 construite à partir de la clé [18].

Le nombre de tour dans AES (voir la figure 8) dépend de la taille du message clair(128, 192ou 256 bits) et la taille de la clé (128, 192 ou 256 bits).

		Tailles du Bloc : N_b x 32		
		128	192	256
Tailles de la Clé : N_k x 32	128	**10**	12	14
	192	**12**	12	14
	256	**14**	14	14
		AES N_b = 4	N_b = 6	N_b = 8

Figure 7 Nombre de tour avec AES .

6.5.1.3 XOR cipher

Origine :

Le cryptage XOR est un système de cryptage basique mais pas trop limité. Ainsi, il a beaucoup été utilisé dans les débuts de l'informatique et continue à l'être encore aujourd'hui car ilest facile à implémenter, dans toutes sortes de programmes [15].

Mécanisme :

Le XOR est un opérateur logique qui correspond à un "OU exclusif" : c'est le (A OU B) qu'onutilise en logique mais qui exclue le cas où A et B sont simultanément vrais. Voici sa table de vérité :

Table de vérité du XOR		
A	**B**	**(A XOR B)**
FAUX	FAUX	FAUX
FAUX	VRAI	VRAI
VRAI	FAUX	VRAI
VRAI	VRAI	FAUX

En informatique, chaque caractère du message à coder est représenté par un entier, le code ASCII. Ce nombre est lui-même représenté en mémoire comme un nombre binaire à 8 chiffres (les bits). On choisit une clé que l'on place en dessous du message à coder, en la répétant autantde fois que nécessaire, comme dans le cryptage de Vigenère. Le message et la clé étant convertien binaire, on effectue un XOR, bit par bit, le 1 représentant VRAI et le 0 FAUX. Le résultat enbinaire peut être reconverti en caractères ASCII et donne alors le message codé[15].

L'algorithme est complètement symétrique : la même opération est réappliquée au message final pour retrouver le message initial [15].

Remarque : Parfois, on applique une permutation circulaire aux bits du message final pour donner le message codé.

6.5 .2 La cryptographie Asymétrique

La cryptographie asymétrique, ou cryptographie à clé publique, est une méthode de chiffrement qui s'oppose à la cryptographie symétrique. Elle repose sur l'utilisation d'une clé publique (est mise à la disposition de quiconque) et d'une clé privée (gardée secrète), l'une permettant de coderle message et l'autre de le décoder. Ainsi, l'expéditeur peut utiliser la clé publique du destinatairepour coder un message que seul le destinataire (en possession de la clé privée) peut décoder, garantissant la confidentialité du contenu.

La figure suivante montre le principe de chiffrement et déchiffrement d'un message utilisentla cryptographie asymétrique.

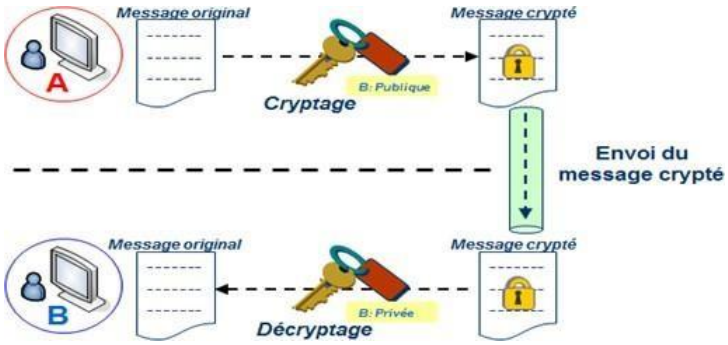


Figure 7 Principe du chiffrement asymétrique [16]

La cryptographie asymétrique offre de nombreux avantages, le premier avantage est d'améliorer la sécurité, elle permet d'échanger des messages de manière sécurisée sans aucun dispositif de sécurité. L'expéditeur et le destinataire n'ont plus besoin de partager des clés secrètes via une voiede transmission sécurisée. Les communications impliquent uniquement l'utilisation de clés publiques et plus aucune clé privée n'est transmise ou partagée[16].

Un autre avantage majeur des systèmes à clé publique est qu'ils permettent l'authentification des messages par signature électronique, ce qui rend impossible de décrypter le message dans le cas de son interception par une personne non autorisée[16].

L'inconvénient des systèmes à clé publique est leur vitesse qui est plus lent contraire- ment aux méthodes à clé secrète qui sont rapide. Ces méthodes sont peu recommandées pour les messages très longs [16].
Parmi les algorithmes de chiffrement symétrique, nous citons :

6.5.2.1 RSA (Rivest-Shamir-Adleman)

L'algorithme RSA a été inventé par Ron Rivest, Adi Shamiret Leonard Adleman en 1977. Le principe de base de RSA est de considérer un message comme un (grand) nombre entier et de fairedes calculs dessus pour le chiffrer. Il repose sur la factorisation en nombres premiers d'un

entier. Leprincipe de RSA est le suivant :

- Générer aléatoirement deux nombres premiers (p et q), puis les multiplier pour générer le nombre n,
- Déterminer ϕ(n) : ϕ(n) = (p - 1) * (q - 1),
- Choisir un entier naturel e premier avec ϕ(n) et strictement inférieur à ϕ(n),
- Calculer l'entier naturel d qui est strictement inférieur à ϕ(n) et e*d ≡ 1(modϕ(n)),
- Le couple (n,e) est la clé publique du chiffrement, alors que le couple (n,d) est saclé privée,
- Pour chiffrer un texte, nous calculons c avec : c = m^e mod n,
- Pour déchiffrer un texte chiffré, nous calculons m avec : m = c^d mod n.Où m est le message enclair et c le message crypté.

Avant d'être chiffré, le message original doit être décomposé en une série d'entiers Mde valeurscomprises entre 0 et n-1. Pour chaque entier M, il faut calculer c = m^e mod n. Le message chiffré est constitué de la succession des entiers c. Pour déchiffrer c, on utilise d,et on retrouve le message clair m par m = c^d mod n [18].

6.5.2.2 EL GAMAL

L'algorithme est décrit pour un groupe cyclique fini au sein duquel le problème de décision deDiffie-Hellman (DDH) est difficile. Des informations plus précises sont données dans la section Résistance aux attaques CPA.

On peut remarquer que *DDH* est une hypothèse de travail *plus forte* que celle du logarithme discret, puisqu'elle tient si jamais le problème du *logarithme discret* est difficile. Il existe parailleurs des groupes où le problème *DDH* est facile, mais où on n'a pas d'algorithme efficace pour résoudre le logarithme discret[4].

Comme il s'agit d'un schéma de chiffrement asymétrique, le crypto système est composé de troisalgorithmes (probabilistes) : **GenClefs**, **Chiffrer** et **Déchiffrer**.

Pour l'illustration, on va considérer que Bob veut envoyer un message à Alice. Mais ce message contient des informations sensibles, Bob ne veut donc pas qu'il soit compréhensible par une autrepersonne qu'Alice. Ainsi Bob va chiffrer son message.

Comme les schémas de chiffrement asymétrique sont en règle générale plus lents que leurs analogues symétriques, le chiffrement El Gamal est souvent utilisé en pratique dans le cadre d'un chiffrement hybride.

Une manière de voir ce schéma de chiffrement, est de faire un parallèle avec le protocole d'échange de clefs de Diffie-Hellman. L'algorithme de chiffrement consiste alors à envoyer un message chiffré C1 par masque jetable sous la clef partagée ht =gt-x, qui peut être calculé par Alicevu qu'elle dispose de C2=gt (voir illustration) [21].

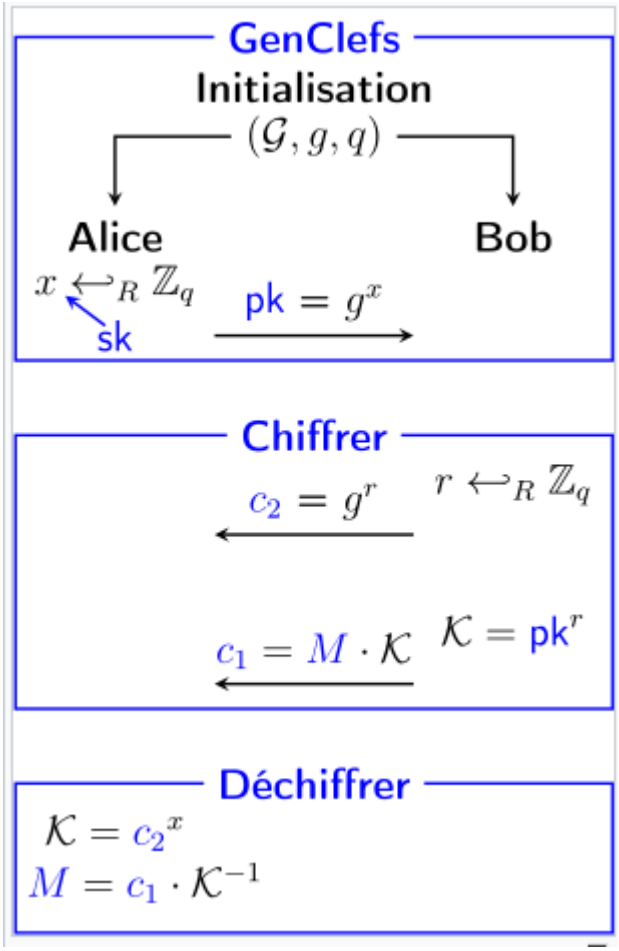


Figure 8 principe de chiffrement de el Gamal [21].

6.5.2.3 Algorithme ECC

La cryptographie basée sur les courbes elliptiques a permit de réduire la taille des Clés, elle nécessite des clés plus courtes en la comparant avec les techniques traditionnelles RSA pour avoir le même niveau de sécurité . Si nous comparons la taille des clés, pour atteindre le niveau de sécurité Equivalent à un AES de 256 bits, il est recommandé d'utiliser des clés de 512 bits pour ECC. Sachant que le crypto système RSA nécessite des clés avec la taille qui dépasse les 15000 bits. Dans vous trouvez plus de détails pour les recommandations proposées par l'Institut national des normes et de la technologie (NIST). Ces déférents avantages et recommandations proposés pour la cryptographie basée sur les courbes elliptiques ont amené à plusieurs implémentations e-cachés et rapides. Par conséquence ces cryptosystŁmes sont devenus plus demandés dans plusieurs domaines, tels que le militaire, les banques, les nouvelles technologies (smartphones) et d'autres.

ECC a gagné plus d'importance pour le futur des systèmes et des informations de sécurité américains classiques et non classiques ,il est devenu bien positionné
parmi les systèmes à clé publique, et il est prévu qu'il garde cette position dans le futur [20].

7 Conclusion

Le concepteur de crypto essaie de créer un chiffrement plus sécurisé, mais en même temps, lespirates tentent toujours de déchiffrer les messages.
Dans ce chapitre, nous avons identifié les concepts les plus importants liés à la sécurité de l'information et les risques qui la menacent. Nous avons expliqué comment protéger par le cryptage. Dans le deuxième chapitre, nous parlerons de l'apprentissage automatique et de ses types .

CHAPITRE II

L'APPRENTISSAGE AUTOMATIQUE

1 Introduction

Le grand nombre d'informations similaires et la similitude des nombreuses langues ont conduit à la création du domaine de l'apprentissage automatique, où dans ce chapitre nous définirons et aborderons son fonctionnement et présenterons ses types Apprentissage effectué etApprentissage non utilisé et expliquerons les différents algorithmes, dont l'algorithme k-plus proches voisins (k-PPV) qui nous intéresse beaucoup car notre application estbasée sur son fonctionnement.

2 Intelligence artificielle

L'intelligence artificielle est née dans les années 50, quand une poignée de pionniers du domainenaissant de l'informatique, ont commencé à se demander si les ordinateurs pouvaient être amenés à « penser », une question dont nous explorons encore aujourd'hui les ramifications. Une définition concise du champ serait la suivante : l'effort d'automatiser les tâches intellectuelles normalement effectuées par les humains. En tant que tel, l'IA est un domaine général qui englobel'apprentissage automatique et l'apprentissage en profondeur, mais qui comprend également beaucoup plus d'approches qui n'impliquent aucun apprentissage. Les programmes d'échecs initiaux, par exemple, ne concernaient que des règles codées en dur élaborées par des programmeurs et ne se qualifiaient pas comme apprentissage automatique. Pendant un temps assez long, de nombreux experts ont estimé que l'intelligence artificielle au niveau humain, pouvait être obtenue en faisant en sorte que les programmeurs fabriquent à la main un ensemble suffisamment large de règles explicites pour manipuler les connaissances. Cette approche est connue sous le nom d'IA symbolique et elle était le paradigme dominant de l'IA des années 50 età la fin des années 80. Elle a atteint son pic de popularité durant le boom des systèmes experts des années 80. Bien que l'IA symbolique se soit révélée appropriée pour résoudre des problèmes logiques bien définis, comme jouer aux échecs, il était difficile de trouver des règles explicites pour résoudre des problèmes flous plus complexes, tels que la classification des images, la reconnaissance de la parole et la traduction. Une nouvelle approche est apparue pour prendre la place de l'IA symbolique : l'apprentissage automatique. [22]

3 L’apprentissage automatique :

L'apprentissage automatique (Machine Learning) est un domaine de recherche en informatiquequi traite des méthodes d'identification et de mise en œuvre de systèmes et algorithmes par lesquels un ordinateur peut apprendre, ce domaine a souvent été associé à l'intelligence artificielle et plus spécifiquement l'intelligence computationnelle. L'intelligence computationnelle est une méthode d'analyse de données qui pointe vers la création automatiquede modèles analytiques. Autrement dit, permettant à un ordinateur d'élaborer des concepts, d’évaluer, prendre des décisions et prévoir les options futures. [23]

L'ensemble du processus d'apprentissage nécessite un ensemble de données comme suit :

- Ensemble de données pour l'entrainement : c'est la base de connaissance utilisée pour entrainer,notre l'algorithme d'apprentissage, pendant cette phase, les paramètres du modèle peuvent être réglés (ajustés) en fonction des performances obtenues.
- Ensemble de données pour le test : cela est utilisé juste pour évaluer les performances du modèlesur les données non vues.

Définition

Un programme d'ordinateur est capable d'apprendre à partir d'une expérience E et par rapport àun ensemble T de tâches et selon une mesure de performance P, si sa performance à effectuer une tâche de T, mesurée par P, s'améliore avec l'expérience E [24].

3.1 Modélisation

L’apprentissage automatique d’une machine toujours concerne un ensemble de tâches concrètes -
T. Pour déterminer la performance de la machine, on utilise une mesure de la performance P. Lamachine peut avoir à l’avance un ensemble d’expérience E ou elle va enrichir cet ensemble plustard.

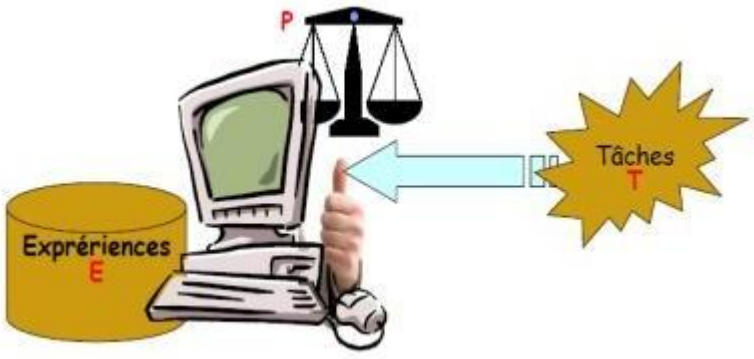


Figure 9 Schéma de modélisation d’une machine d’apprentissage [24

Donc, l’apprentissage automatique pour la machine est qu’avec l’ensemble de tâches T que la machine doit réaliser, elle utilise l’ensemble d’expériences E telle que sa performance sur T estaméliorée.

4 Domaines d’applications de l'apprentissage automatique :

L'apprentissage automatique s'applique à un grand nombre d'activités humaines et convient enparticulier au problème de la prise de décision automatisée. Il s'agira, par exemple:

- D'établir un diagnostic médical à partir de la description clinique d'un patient;
- De donner une réponse à la demande de prêt bancaire de la part d'un client sur la base de sa situation personnelle;
- De déclencher un processus d'alerte en fonction de signaux reçus par des capteurs ;
- De la reconnaissance des formes;
- De la reconnaissance de la parole et du texte écrit;
- De contrôler un processus et de diagnostiquer des pannes;

5 Types d'apprentissage

Les algorithmes d'apprentissage peuvent se catégoriser selon le mode d'apprentissage qu'ils emploient :

5 .1 Apprentissage supervisé

L'apprentissage supervisé est la tâche d'apprentissage automatique la plus simple et la plus connue. Il est basé sur un certain nombre d'exemples pré classifiés, dans lesquels est connu à priori la catégorie à laquelle appartient chacune des entrées utilisées comme exemples.

Dans ce cas, la question cruciale est le problème de généralisation, après l'analyse d'un échantillon d'exemples, le système devrait produire un modèle qui devrait fonctionner pour toutes les entrées possibles [22].

L'ensemble de données pour l'entrainement, est constitué de données étiquetées, c'est-à-dire d'objets et de leurs classes associées. Cet ensemble d'exemples étiquetés constitue donc l'ensemble d'apprentissage.

Afin de mieux comprendre ce concept, prenons un exemple : un utilisateur reçoit chaque jour ungrand nombre d'e-mails, certains sont des e-mails d'entreprises importants et d’autres sont des e- mails indésirables non sollicités ou des spam.

Un algorithme supervisé sera présenté avec un grand nombre d'e-mails qui ont déjà été étiquetéspar l'utilisateur comme spam ou non spam.

L'algorithme fonctionnera sur toutes les données étiquetées, faire des prédictions sur l'e-mail

etvoir si c’est un spam ou non.

Cela signifie que l'algorithme examinera chaque exemple et fera une prédiction pour chacun poursavoir si l'e-mail est un spam ou pas. La première fois, l'algorithme fonctionne sur toutes les données non étiquetées, la plupart des e-mails seront mal étiquetés car il peut fonctionner assez mal au début. Cependant, après chaque exécution, l'algorithme compare sa prédiction au résultat souhaité (l'étiquette). Au fur et à mesure, l'algorithme apprendra à améliorer ses performances et sa précision

Dans l'exemple que nous avons utilisé, nous avons décrit un processus dans lequel un algorithmeapprend à partir de données étiquetées (emails qui ont été catégorisés comme spam ou non- spam).

Dans certains cas, le résultat n'est pas nécessairement discret et il se peut que nous n'ayons pas un nombre fini de classes dans lesquelles classer nos données. Par exemple, nous essayons peut-être de prédire l'espérance de vie d'un groupe de personnes en fonction de paramètres de santé préétablis [22].

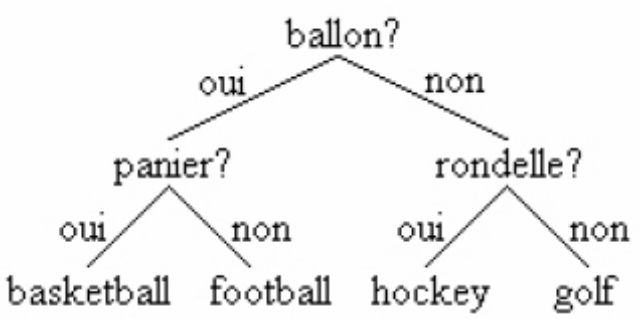
Dans ce cas, comme le résultat est une fonction continue (nous pouvons spécifier une espérance de vie comme un nombre réel exprimant le nombre d'années que la personne devrait vivre), nous ne parlons pas d'une tâche de classification mais plutôt d'un problème de régression.

Dans un problème de régression, l'ensemble d'apprentissage est une paire formée par un objet et une valeur numérique associée. Il existe plusieurs algorithmes d'apprentissage supervisé qui ont été développés pour la classification et la régression. Parmi tous, les arbres de décision, les règlesde décision, les réseaux de neurones et les réseaux bayé siens. [26] .

5.1.1 Les arbres de décision

L’arbre de décision, comme son nom l’indique, est un regroupement de fonctions locales structurées en forme d’arbre. En fait, les fonctions sont habituellement suffisamment localisées pour être de simples règles condition-action. Le nœud racine est associé à l’ensemble de données$D$, et chacun des autres nœuds est associé à un sous ensemble D_n. Un nœud est soit une feuille, soit un nœud interne utilisant une fonction **f** pour séparer son ensemble associé D_n entre au moins deux enfants. Chaque nœud de l’arbre représente donc

Figure 10 Exemple d’un arbre de décision [27]



un classifieur simplifié qui prend sa décision à partir d'un minimum de caractéristiques
L'idée de base de l'apprentissage des arbres de décision est de diviser récursivement et le plus efficacement possible les exemples de l'ensemble d'apprentissage à l'aide d'une fonction **f** qui minimise l'erreur empirique de cette classification locale jusqu'à ce que l'on obtienne des sous-ensembles d'exemples ne contenant (presque) que des exemples appartenant tous à une même classe. Il existe plusieurs méthodes pour trouver cette fonction, dont l'une est d'utiliser l'entropie [27].

5.1.2 Le Bayes naïf

Le Bayes naïf est un classifieur probabiliste [48] se basant sur le théorème de Bayes. Il classe les nouveaux exemples selon la probabilité de ces exemples à appartenir à chaque classe. En utilisant le théorème de Bayes, la probabilité qu'un exemple $\vec{d} = (d_1, d_2, \ldots, d_n)$ appartienne à la classe c_j est :

$$P(c_j|\vec{d}) = \frac{P(c_j)P(\vec{d}|c_j)}{P(\vec{d})}$$

Puisque P(d) est identique pour chaque classe, il est possible de l'enlever et cette équation devientalors:

$$P(c_j|\vec{d}) = P(c_j)P(\vec{d}|c_j)$$

Avec la supposition que les caractéristiques d'apprentissage sont indépendantes c'est à dire que laprobabilité qu'une caractéristique apparaisse dans un exemple est indépendante de la présence desautres caractéristiques. La probabilité de trouver l'exemple à l'intérieur d'une classe est égale au produit des probabilités de trouver ses caractéristiques :

$$P(c_j|\vec{d}) = P(c_j)\prod_{i=1}^{n} P(d_i|c_j)$$

Où n est le nombre de caractéristiques (attributs).

Bien que La supposition posée est fausse mais cela n'empêche pas un tel classificateur de présenter des résultats satisfaisants. Et surtout, elle réduit beaucoup les calculs nécessaires. Sans elle, il faudrait tenir compte de toutes les combinaisons possibles des caractéristiques des données,ce qui d'une part impliquerait un nombre important de calculs, mais aussi réduirait la qualité statistique de l'estimation, puisque la fréquence d'apparition de chacune des combinaisons serait très inférieure à la fréquence d'apparition des caractéristiques prises séparément.

Pour classifier une nouvelle donnée, la probabilité de chaque classe est calculée et la plus haute l'emporte. Avec ce classifieur, l'apprentissage se résume à établir la probabilité des caractéristiques à appartenir à chacune des classes. Selon l'initialisation des probabilités, il est possible de mettre à jour un tel classifieur régulièrement et sans trop de complications. Et malgréplusieurs détracteurs, le Bayes naïf est largement utilisé [28] [29].

5.1.3 Les réseaux de neurones

Les réseaux de neurones artificiels ont fait l'objet d'un intérêt soutenu de nombreux travaux depuisplus d'une vingtaine d'années grâce à de leur capacité à résoudre des problèmes non linéaires par apprentissage.

En 1943, Warren S. McCulloch et Walter Pitts [30] ont introduit la notion de Réseaux de NeuronesArtificiels (RNA). Leur but était de représenter l'activité électrique des cellules nerveuses du cerveau. Les réseaux qu'ils ont proposés, appelés réseaux neurologiques, étaient composés parl'interconnexion des petites unités élémentaires : les neurones formels (figure 12).

Un neurone formel : est un composant calculatoire faisant la somme pondérée des signaux reçusen entrée (calculant la quantité h) à laquelle nous appliquons une fonction de transfert, ici la fonction seuil (fonction de Heaviside), afin d'obtenir la réponse de la cellule (notée y).

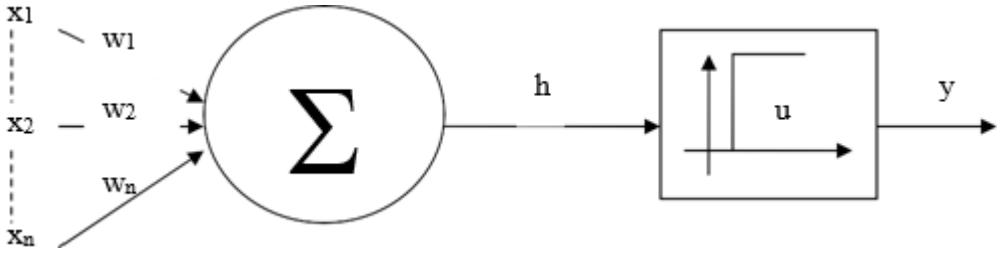


Figure 11 Le neurone Formel [30]

5.1.4 Les séparateurs à vaste marge

Les Séparateurs à Vaste Marge ou SVM est une technique inventée par Vladimir Vapnik . Le principe théorique des SVM comporte deux point fondamentaux : la transformation non linéairedes
données de l'espace d'entrée vers un espace dit de description de grande dimension et la détermination d'un hyperplan permettant une séparation linéaire optimale dans cet espace.

$$\underset{w,w_0}{\text{Arg max}}\ \min\left\{\|x - x_i\| \ :\ x \in \Re^d,\ (w^T x\ +\ w_O)\ =\ 0,\ i = 1,\ldots,m\right\}$$

L'hyperplan optimal est défini par le vecteur de poids w vérifiant l'équation :
Pour cet hyperplan, la marge vaut 1/||***w***||, et donc la recherche de l'hyperplan optimal revient àminimiser ||***w***||, soit à résoudre le problème suivant qui porte sur les paramètres ***w*** et *w*0 :

$$\begin{cases} \text{minimiser} & \frac{1}{2}\|w\|^2 \\ \text{sous les contraintes} & u_i(w^T x_i\ +\ w_O)\ \geq\ 1,\ \ i = 1,\ldots,m \end{cases}$$

Ce système peut être résolu avec des méthodes de programmation quadratique mais pour des valeurs de *d* assez petites. Pour des valeurs de *d* dépassant quelques centaines, il faut résoudre l'expression duale du problème qui est équivalente à la solution du problème original [31].

Après transformation, le problème devient celui de la recherche de paramètres α vérifiant lesystème d'équations suivant :

$$\begin{cases} \underset{\alpha}{\text{Max}} \left\{\sum_{i=1}^{m}\alpha_i\ -\ \frac{1}{2}\ \sum_{i,j=1}^{m}\alpha_i\alpha_j u_i u_j (x_i . x_j)\right\} \\ \alpha_i \geq 0,\ \ i = 1,\ldots,m \\ \sum_{i,j=1}^{m}\alpha_i u_i\ =\ 0 \end{cases}$$

L'hyperplan solution correspondant peut alors s'écrire :

$$h(x)\ =\ (w^* . x)\ +\ w_0^*\ =\ \sum_{i=1}^{m}\alpha_i^* u_i . (x_i . x_j)\ +\ w_0^*$$

Où *m* est le nombre d'exemple d'apprentissage, α est un vecteur de *m* variable et chaque composant

αi correspond à l'exemple d'apprentissage (xi, ui)

La méthode SVM a prouvé son efficacité dans le domaine d'apprentissage. Son succès est justifié par les solides bases théoriques qui la soutiennent et la plupart des techniques d'apprentissage possèdent un (trop) grand nombre de paramètres à fixer par l'utilisateur (structure d'un réseau de neurones, coefficient de mise à jour du gradient, . . .).

5.1.5 Le classifieur k-plus proches voisins (k-PPV)

k-PPV est un algorithme de la reconnaissance des formes qui a prouvé son efficacité faceau

Traitement de données textuelles.Cette méthode diffère des traditionnelles méthodes d'apprentissage car aucun modèle n'est induit à partir des exemples. Les données restent tellesquelles : elles sont simplement stockées en mémoire.

Pour prédire la classe d'un nouveau cas (où ranger un nouveau paquet ?), l'algorithme calcule sa similarité avec toutes les données déjà classées. Puis, il place les valeurs de similarité par ordre décroissant et ne garde que les k premières. La classe revenant le plus souvent parmi ces k paquetsest celle qui est attribuée au nouveau paquet.

Donc, comme son nom l'indique, la classification d'une nouvelle donnée dépend de ses k voisinsles plus proches. Cette notion de distance est plus facilement compréhensible avec l'exemple en deux dimensions de la figure 2. Dans cet exemple, la nouvelle donnée aura de fortes chances défaire partie de la classe noire puisqu'il côtoie de plus près des données de cette catégorie.

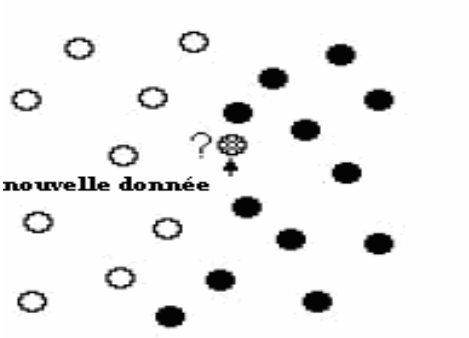


Figure 12 Exemple en deux dimensions du k-plus proches voisins [33]

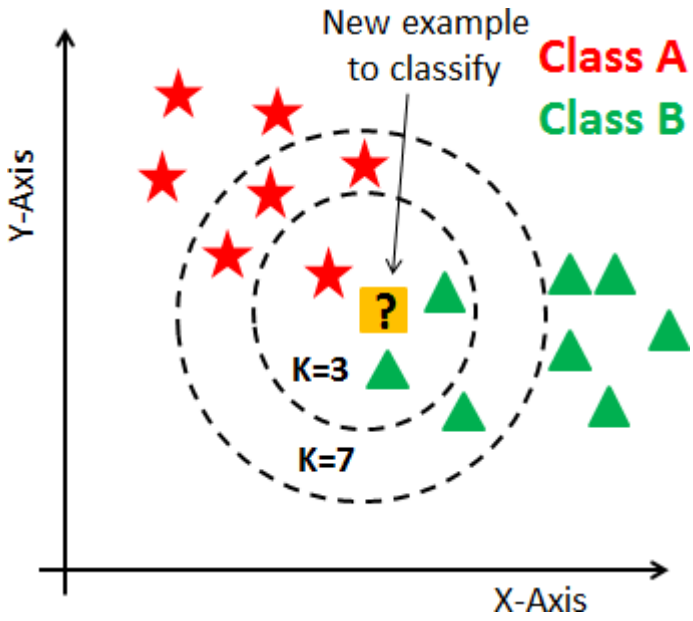


Figure 13 Exemple en deux dimensions du k-plus proches voisins [33]

La similarité entre deux données se calcule à l'aide d'une corrélation quelconque entre les deux vecteurs. Une telle corrélation peut être la distance euclidienne ou encore le cosinus de l'angle formé par les deux vecteurs.

Le choix de la constante k[14] est dépendant de la taille de l'échantillon et des classes, et influence les résultats de la classification. Lorsque k est petit, la classification est plus sensible à cause des données appartenant à une classe mais dont leur vecteur de représentation ressemble beaucoup plus à une autre. Par contre, lorsque k est trop grand, les classes ayant peu d'exemples peuvent être désavantagées par rapport à celles qui en ont plus.

Comme il a déjà été mentionné, le classifieur k-PPV ne construit aucune description explicite de la fonction à apprendre (dans notre cas, l'appartenance à une classe). La fonction n'est pas estimée qu'une seule fois pour tout l'espace, mais elle est estimée plutôt localement et différemment pourchaque nouvelle instance. L'absence d'apprentissage fait du k-PPV un parfait candidat pour les classifications qui doivent être constamment ajustées ou révisées. En fait, la mise à jour de la classification ne peut pas être plus simple. Il n'y a qu'à étiqueter et indexer les nouveaux exemples,et enlever les données nuisibles. En contrepartie, le temps de classement d'une nouvelle donnée peut être élevé (est proportionnel au nombre d'exemples et au nombre de caractéristiques utilisées), puisque c'est à ce moment que tout le calcul se fait. Cependant, une bonne indexation des exemples aide beaucoup à pallier ce problème. Malgré sa simplicité, le k-PPV est l'objet de plusieurs études [32] [33] [34].

5.2 Apprentissage non supervisé :

La deuxième classe d'algorithmes d'apprentissage automatique est appelée apprentissage non supervisé, dans ce cas, nous n'étiquetons pas les données au préalable, nous laissons plutôt l'algorithme arriver à sa conclusion. Ce type d'apprentissage est important car il est beaucoup plus commun dans le cerveau humain que l'apprentissage supervisé. Les algorithmes d'apprentissage non supervisé sont particulièrement utilisés dans les problèmes de clustering, dans lesquels, étant donné une collection d'objets, nous voulons être en mesure de comprendre etde montrer leurs relations. Une approche standard consiste à définir une mesure de similarité entre deux objets, puis à rechercher tout groupe d'objets plus similaires les uns aux autres, par rapport aux objets des autres clusters. Par exemple, dans le cas précédent des e-mails spam/ non spam, l'algorithme peut être capable de trouver des éléments communs à tous les spam (par exemple, la présence de mots mal orthographiés). Bien que cela puisse fournir une classificationmeilleure qu'aléatoire, il n'est pas clair que les spam/non spam puissent être facilement séparés. [32] [33] [34] .

En effet, le processus de « clustering » repose sur une mesure précise de la similarité des objetsque l'on veut regrouper. Cette mesure est appelée distance ou métrique. On distingue plusieurs algorithmes de « clustering », comme :

K-moyennes (KMeans). KMeans est un algorithme de partitionnement des données en K nombre de groupes ou clusters. Chaque objet sera associé à un seul cluster. Le nombre K est fixépar l'utilisateur.
Fuzzy KMeans**.** Il s'agit d'une variante du précédent algorithme proposant qu'un objet ne soit pas associé qu'à un seul groupe.
Espérance-Maximisation (EM**).** Cet algorithme utilise des probabilités pour décrire qu'un objet appartient à un groupe. Le centre du groupe est ensuite recalculé par rapport à la moyennedes probabilités de chaque objet du groupe.
Regroupement hiérarchique**.** Deux sous-algorithmes en découlent, à savoir d'une part le « bottom up » qui a pour fonction d'agglomérer des groupes similaires, donc en réduire le nombre (les rendre plus lisibles) et d'en proposer un ordre hiérarchique, et d'autre part, le « top down » qui fait le raisonnement inverse en divisant le premier groupe, récursivement, en sous-ensembles[35].

5.3 Apprentissage par renforcement :

L'apprentissage par renforcement est une approche de l'intelligence artificielle qui met

L'accent sur l'apprentissage du système à travers ses interactions avec l'environnement. Avec l'apprentissage par renforcement, le système adapte ses paramètres en fonction des réactions reçues de l'environnement, qui fournit ensuite un retour d'information sur les décisions prises.

Par exemple, un système qui modélise un joueur d'échecs qui utilise le résultat des étapes précédentes pour améliorer ses performances, est un système qui apprend avec le renforcement. La recherche actuelle sur l'apprentissage avec renforcement est hautement .

interdisciplinaire et comprend des chercheurs spécialisés dans les algorithmes génétiques, les réseaux de neurones, la psychologie et les techniques de contrôle. [33] .

La figure suivante résume les trois types d'apprentissage avec les problèmes connexes à résoudre
:

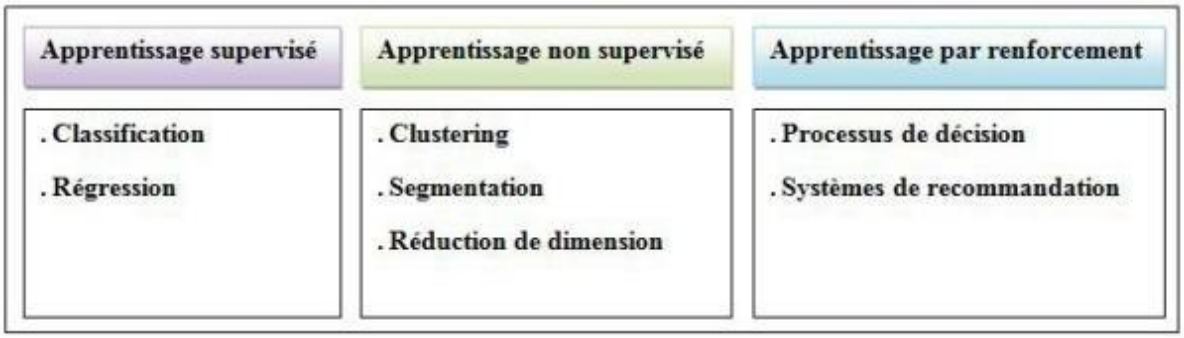


Figure 14 Types d'apprentissage [33]

6 CONCLUSION

L'apprentissage automatique a prouvé son efficacité dans divers domaines, y compris le domainedu codage. Nous avons clarifié les concepts les plus importants liés à l'apprentissage automatiqueet comment l'utiliser. Nous avons clarifié ses types et les algorithmes dont dépend chaque type.
Dans le troisième chapitre, nous parlerons de notre problème et de la solution que nous avons proposée, et nous expliquerons également comment faire le lien entre le cryptage et le machineLearning.

CHAPITRE III

IMPLÉMENTATION DE LA SOLUTION PROPOSÉE

1 INTRODUCTION

Dans ce chapitre, nous allons détailler notre approche qui consiste à présenter une solution qui vise à la fois de sécuriser les données et d'aider les utilisateurs prendre la meilleure décision. Généralement, la cryptographie est utilisée que par les experts malgré qu'elle offre une protection forte aux données surtout coté confidentialité, intégrité et disponibilité. Cela est à cause du manque d'expérience dans ce domaine. Dans ce chapitre, nous allons définir les étapes de notre proposition mais avant nous décrivons certains travaux qui s'intéressent par ce domaine.

2. Problématique de l'Approche proposée

la cryptographie est utilisée pour rendre les documents plus confidentiels. Généralement elle est utilisée par des experts ou bien des professionnels de l'informatique et de la sécurité informatique. Dans le but de rendre la cryptographie plus utilisée même par les personnes qui n'ont pas une expérience dans ce domaine. On propose un crypto système basé sur les algorithmes de l'apprentissage supervisé dont le but d'analyser les besoins de l'utilisateur côté sécurité et de l'aider à prendre la bonne décision et choisir le meilleur algorithme de chiffrement.

La figure suivant montre l'architecture de notre solution :

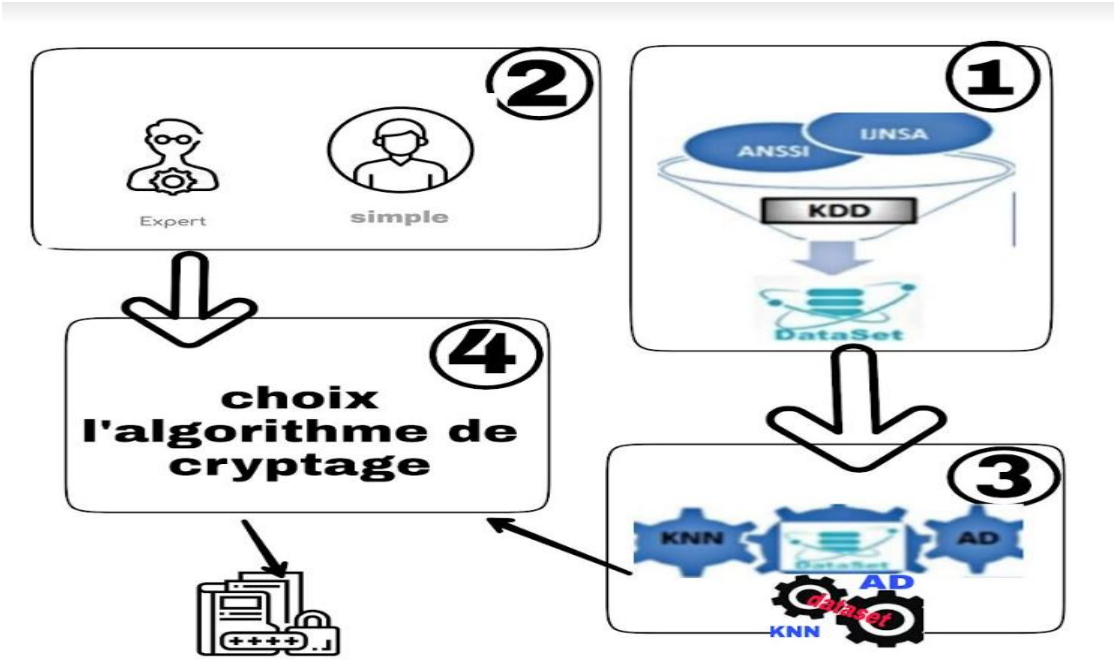


Figure 15 solution proposée

Ce système divise les utilisateurs en deux catégories (Simple, Expert).(les explications suivants sont conforme au numérotation dans la figure précédente :

Etape 1 : la création de notre dataset

Pour l'extraction des connaissances : on utilise la méthode KDD, l'extraction d'informations auparavant inconnues et intéressantes à partir de données brutes . Dans notre cas nous allons basé sur des descriptions proposés parl'IJNSA (era détail dans la section IV.8 et les recommandations de la sécurité dans le domaine de la cryptographie du ANSSI

Etape 2 : Dans cette étape, on peut avoir deux types de user :

• L'utilisateur expert : a des connaissances dans le domaine de la cryptographie, donc il a la capacité de choisir directement l'algorithme de cryptage a utiliser ;

• L'utilisateur simple, nous suivrons les étapes suivantes :

 L'utilisateur remplie le formulaire afin d'exprimer ces besoins de sécurité;

Etape 3 : Cette étape est la phase d'entraînement :

 Où on va utiliser la dataset crée dans l'étape (1) pour entraîner notre système avec l'un des algorithmes d'apprentissage (KNN ou bien AD) ;

 Le système va diriger l'utilisateur vers l'algorithme le plus adéquat selon leur besoin exprimés au cours de la première étape (2).

Etape 4: l'étape de chiffrement conformément aux besoins des utilisateurs.

3. L'extraction des connaissances avec l'approche KDD:

En raison de la croissance exponentielle des données, en particulier dans des domaines tels que les affaires, l'extraction manuelle de modèles étant devenue impossible au cours des dernières décennies. Par exemple, il est actuellement utilisé pour diverses applications telles que l'analyse de réseau social, la détection de fraude, la science, l'investissement, la fabrication, les télécommunications, le nettoyage de données, le sport, la recherche d'informations et en grande partie pour le marketing. KDD est généralement utilisé pour répondre à des questions telles que quels sont les principaux produits susceptibles d'aider à générer des profits élevés l'année prochaine dans Wal-Mart [42]

Ce processus comporte plusieurs étapes. Il commence par développer une compréhension du domaine d'application et de l'objectif, puis à créer un jeu de données cible. Viennent ensuite le nettoyage, le prétraitement, la réduction et la projection des données. La prochaine étape consiste à utiliser l'exploration de données (expliquée ci-dessous) pour identifier un motif. Enfin, les connaissances découvertes sont consolidées en visualisant et / ou en interprétant [42].

3.1 Définition :

Le KDD est un domaine de l'informatique qui traite de l'extraction d'informations auparavant inconnues et intéressantes à partir de données brutes. KDD est tout le processus consistant à essayer de donner un sens aux données en développant des méthodes ou techniques appropriées. Ce processus traite du mappage de données de bas niveau dans d'autres formes plus compactes, abstraites et utiles. Ceci est réalisé en créant de courts rapports, en modélisant le processus de génération de données et en développant des modèles prédictifs pouvant prédire des cas futurs [42].

3.2 L'exploration de données

Comme mentionné ci-dessus, l'exploration de données n'est qu'une étape du processus global de KDD. Il existe deux objectifs principaux d'exploration de données définis par l'objectif de l'application, à savoir la vérification ou la découverte. La vérification consiste à vérifier l'hypothèse de l'utilisateur concernant les données, tandis que la découverte détecte automatiquement des modèles intéressants. [42]
Il existe quatre tâches principales d'exploration de données: le regroupement (clustering), la classification, la régression et l'association (synthèse).

- Le clustering identifie des groupes similaires à partir de données non structurées.
- La classification est des règles d'apprentissage qui peuvent être appliquées à de nouvelles données.
- La régression consiste à rechercher des fonctions avec une erreur minimale pour modéliser les données.
- L'association recherche des relations entre les variables.

Ensuite, l'algorithme d'exploration de données spécifique doit être sélectionné. Selon l'objectif, différents algorithmes tels que la régression linéaire, la régression logistique, les arbres de décision et Naïve Bayes peuvent être sélectionnés. Ensuite, les modèles d'intérêt dans une ou plusieurs formes de représentation sont recherchés. Enfin, les modèles sont évalués en utilisant la précision prédictive ou la compréhensibilité [42].

3.3 Architecture typique d'une application basée sur la classification

La classification est une tâche très importante dans le Data Mining, mais le développement d'un outil de classification dans n'importe quel domaine doit être réalisé en appliquant d'autres phases d'extraction et de l'analyse d'informations qui sont montrées dans la figure suivante [43] :

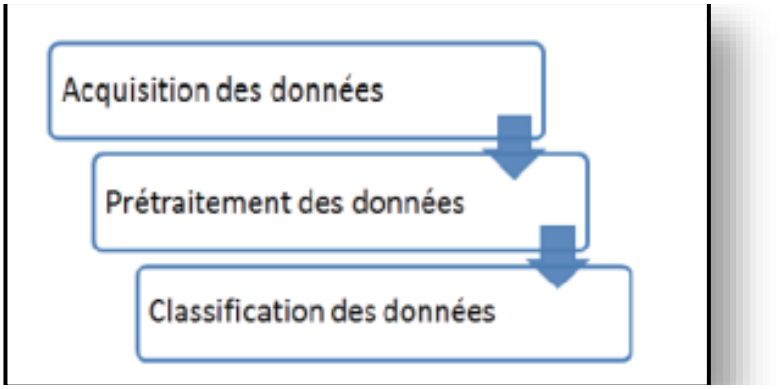


Figure 16 Processus du data mining [43]

- **Acquisition des données :**

D'une manière générale, il s'agit de mettre en place l'ensemble d'instrumentation (capteurs, matériel d'acquisition, etc.) de façon à reproduire le phénomène observé le plus fidèlement possible. [43]

- **Prétraitement de données :**

Cette phase correspond au filtrage des informations en ne conservant que ce qui est pertinent dans le contexte d'étude puisque ces données peuvent contenir plusieurs types d'anomalies (elles peuvent être omises à cause des erreurs de frappe ou à cause des erreurs dues au système lui même, elles peuvent être incohérentes donc on doit les écarter ou les normaliser...etc). Parfois on est obligé à faire des transformations sur les données pour unifier leur poids [43].

- **Classification des données :**

Dans cette étape, on doit choisir la bonne technique pour extraire les connaissances des données (les réseaux de neurones, les arbres de décision, les réseaux bayésiens...etc) [43].

4. Les sources de création de notre Dataset

Avant de crée notre dataset, il est important de signaler qu'on a base sur les deux supports suivant :

- **Le journal international de sécurité des réseaux et ses applications :**

L'**I**nternational **J**ournal of **N**etwork **S**ecurity & Its **A**pplications (IJNSA) est une revue bimensuelle en libre accès à comité de lecture qui publie des articles qui apportent de nouveaux résultats dans tous les domaines de la sécurité des réseaux informatiques et de ses applications. La revue se concentre sur tous les aspects techniques et pratiques de la sécurité et de ses applications pour les réseaux filaires et sans fil. L'objectif de cette revue est de rassembler des chercheurs et des praticiens du milieu universitaire et de l'industrie pour se

concentrer sur la compréhension des menaces de sécurité modernes et des contre-mesures, et sur l'établissement de nouvelles collaborations dans ces domaines[45].

- **Le guide d'agence national de sécurité des systèmes d'information**

Ce document rédigé par l'ANSSI présente un **« Guide de sélection d'algorithmes cryptographiques»**. Il est téléchargeable sur le site www.ssi.gouv.fr. Il constitue une production originale de l'ANSSI. Il est à ce titre placé sous le régime de la «Licence ouverte» publiée par la maison Etalab (www.etalab.gouv.fr). Il est par conséquent diffusable sans restriction.
L'objectif de ce document est d'apporter une aide à la sélection des mécanismes cryptographiques. Pour ce faire, il propose, sous forme des **recommandations**, des **notes d'implémentation** à destination des développeurs qui implémentent des mécanismes cryptographiques [46].

5. Critère de base pour créer notre dataset

Il existe différents types de méthodes cryptographiques qui peuvent être utilisées. Fondamentalement, la méthode cryptographique sélectionnée dépend des exigences de l'application telles que le **temps de réponse**, la **bande passante**, la **confidentialité** et **l'intégrité**. Cependant, chacun des algorithmes cryptographiques à ses propres points faibles et forts.
Dans cette section, nous présenterons le résultat de l'implémentation et de l'analyse appliquées sur plusieurs algorithmes cryptographiques tels que DES, 3DES, AES, RSA et blowfish. Aussi, nous montrerons les comparaisons entre les techniques cryptographiques précédentes en termes de performances, de faiblesses et de forces [47].
L'analyse a été effectuée sur la base des métriques suivantes :

- **Temps de cryptage**

Le temps nécessaire pour convertir un texte clair en texte chiffré est le temps de cryptage. Le temps de chiffrement dépend de la taille de la clé, de la taille du bloc de texte en clair et du mode. Dans notre expérience, nous avons mesuré le temps de cryptage en millisecondes. Le temps de cryptage affecte les performances du système [48]. Le temps de cryptage doit être moindre, ce qui rend le système rapide et réactif.
La figure suivante montre que l'algorithme blowfish enregistre le temps de cryptage le plus rapide et l'algorithme RSA enregistre le temps de cryptage le plus lent. Sur la base du temps de cryptage, nous sélectionnerons la technique Blowfish pour une évaluation plus approfondie [48].

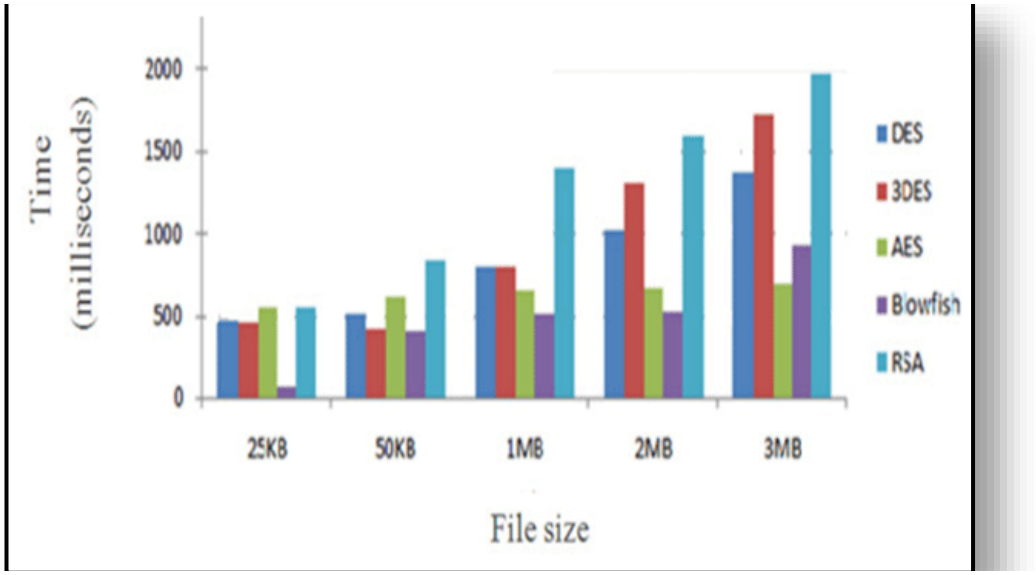


Figure 17 Temps de cryptage par rapport à la taille du fichier pour DES, 3DES, AES, Blowfish et RSA [48].

- **Temps de déchiffrement :**

Le temps de récupération du texte en clair à partir du texte chiffré est appelé temps de déchiffrement. Il est souhaité que le temps de décryptage soit moins similaire au temps de cryptage pour rendre le système réactif et rapide. Le temps de déchiffrement affecte les performances du système. Dans notre expérience, nous avons mesuré le temps de déchiffrement en millisecondes [48].
La figure suivante montre que le temps de décryptage pour tous les algorithmes est plus rapide que le temps de cryptage. En outre, l'algorithme Blowfish enregistre le temps de décryptage le plus rapide et l'algorithme RSA enregistre le temps de décryptage le plus lent. Sur la base de la fonction de temps de décryptage, nous sélectionnerons la technique du poisson-globe à prendre en compte au niveau d'évaluation suivant.

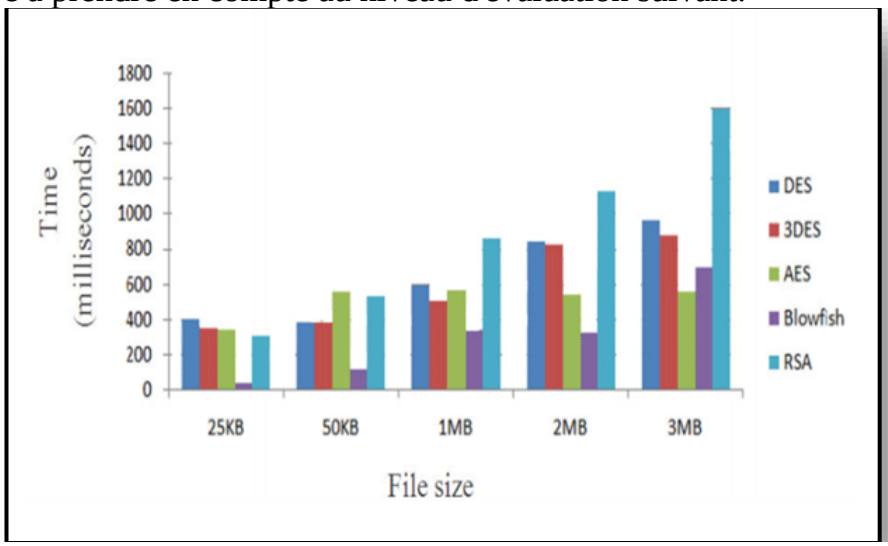


Figure 18 Temps de déchiffrement par rapport à la taille du fichier pour DES, 3DES, AES, Blowfish et RSA [48].

- **Mémoire utilisée :**

Différentes techniques de chiffrement nécessitent une taille de mémoire différente pour la mise en œuvre. Ce besoin en mémoire dépend du nombre d'opérations à effectuer par l'algorithme, de la taille de clé utilisée, des vecteurs d'initialisation utilisés et du type d'opérations. La mémoire utilisée a un impact sur le coût du système. Il est souhaitable que la mémoire requise soit aussi petite que possible [48].
La suite dans le tableau suivant présente cette mémoire utilisée pour les opérations unitaires pour toutes les techniques cryptographiques que nous avons étudiées. Blowfish a consommé moins de stockage de mémoire que les autres types, tandis que RSA utilise la mémoire la plus élevée.

Algorithme	Mémoire utilisée (Ko)
DES	18.2
3DES	20.7
AES	14.7
Poisson-globe	9.38
RSA	31,5

Figure 19 Comparaison de la mémoire utilisée [48].

- **Nombre de bits nécessaires pour encoder de manière optimale :**

le nombre de bits nécessaires pour encoder un caractère chiffré doit être inférieur. Depuis, le bit crypté sera transmis sur un réseau après codage ; cette métrique nous indique la bande passante requise pour la transmission. Si un bit chiffré est codé avec moins de bits, il consommera également moins de bande passante et moins de stockage. Par conséquent, cela a un impact sur le coût [48].
Le tableau suivant présente AES demande le plus grand nombre de bits à coder de manière optimale, tandis que DES demande le plus petit nombre de bits à coder de manière optimale [47].

Algorithme	Nombre moyen de bits requis pour encoder de manière optimale un octet de données chiffrées
DES	27
3DES	40
AES	256
Poisson-globe	128
RSA	44

Figure 20 longueur d'encodage optimale [48].

6. Les cas étudié par IJNSA:

Différents algorithmes de clés symétriques seront analysés pour diverses caractéristiques de fichiers, telles que différents types de données, la densité des données, la taille des données et la taille de la clé, et nous avons analysé la variation du temps de cryptage pour différents

algorithmes de chiffrement sélectionnés [47].
L'analyse succincte des différents algorithmes cryptographiques à clé symétrique pour divers paramètres est la suivante :

- **Etude de cas 1 : Fichiers avec différents types de données.**

Cette étude de cas a été réalisée pour vérifier si le cryptage dépend du type de données. Différents fichiers de type audio, image, texte et vidéo d'une taille de près de 50 Mo ont été choisis et le temps de cryptage de différents algorithmes de chiffrement est calculé pour ces types de données. Pour toutes les exécutions d'un algorithme de chiffrement spécifique, le paramètre variable est le type de données et les paramètres constants sont la taille de la clé et le mode de chiffrement par blocs. La taille de la clé et le mode de bloc sont maintenus à des paramètres minimaux. La taille de la clé de AES, DES, 3-DES, RC2, Blowfish, Skipjack, et RC4 est maintenue à des valeurs minimales de 128, 56, 112, 40, 32, 80 et 40 bits respectivement [47].

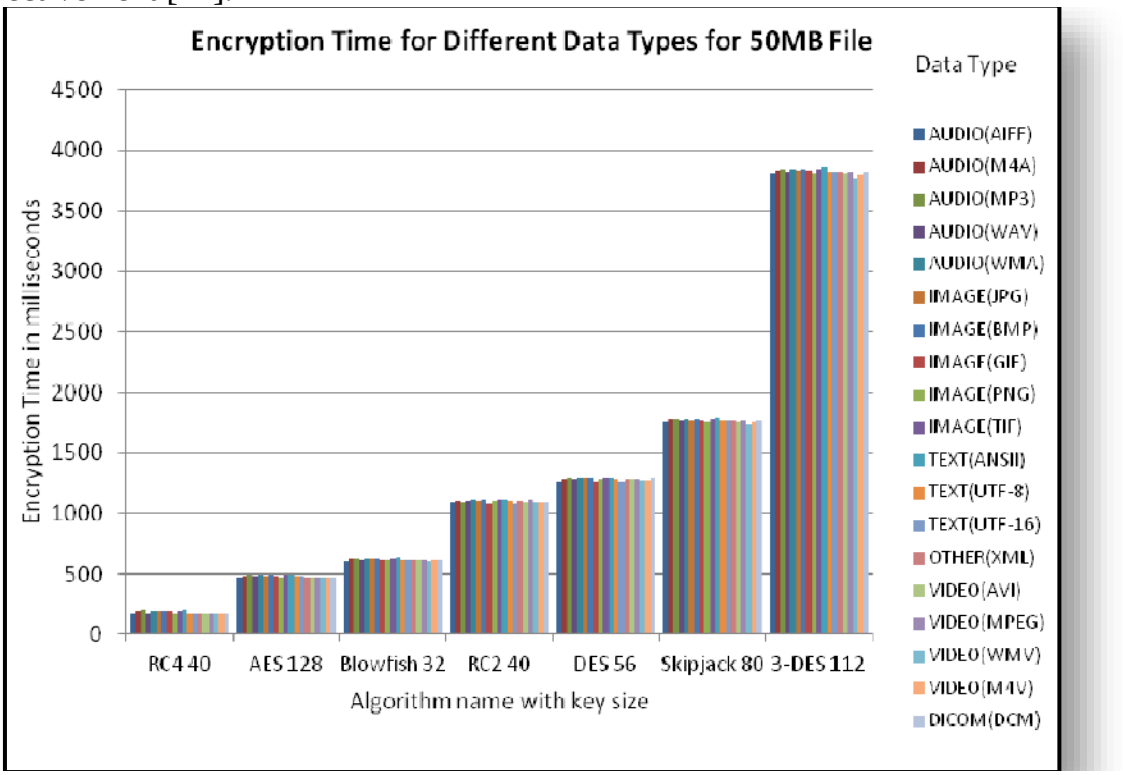


Figure 21 temps de cryptage Vs algorithme de cryptage pour les défirent type de données [47].

Observation : Dans la figure, on peut clairement voir que le temps de cryptage pour tous les types de données est presque le même. Le résultat montre que le temps de cryptage ne varie pas en fonction du type de données [47].

Alors : Le cryptage dépend uniquement du nombre d'octets du fichier et non du type des fichiers

- **Etude de cas 2 : Fichiers de données de même type mais de tailles différentes.**

Cette étude de cas est prise pour vérifier une fois de plus les observations obtenues dans

l'étude de cas 1. Le cas étude de cas 1 a révélé que le temps de cryptage dépend du nombre d'octets dans le fichier. Pour s'en assurer une autre étude est faite dans laquelle différents fichiers (BMP et FLV) de même type mais de tailles différentes sont donnés à crypter et leur temps de cryptage est calculé. Pour toutes les exécutions, la taille de la clé et le mode de cryptage par bloc sont fixés à des paramètres minimaux [47].
Le tableau suivant donne les détails des fichiers utilisés pour toutes les exécutions.

File Type	Varying Parameters (Data Size)	Constant Parameters
BMP	10.7MB, 50MB, 100MB	Data Type, Key size
FLV	50MB, 100MB, 482MB	

Figure 22 paramètres d'exécution pour des fichiers de taille défirent [47].

La figure suivante montre les résultats d'exécution pour les formats de fichiers BMP de différentes tailles respectivement

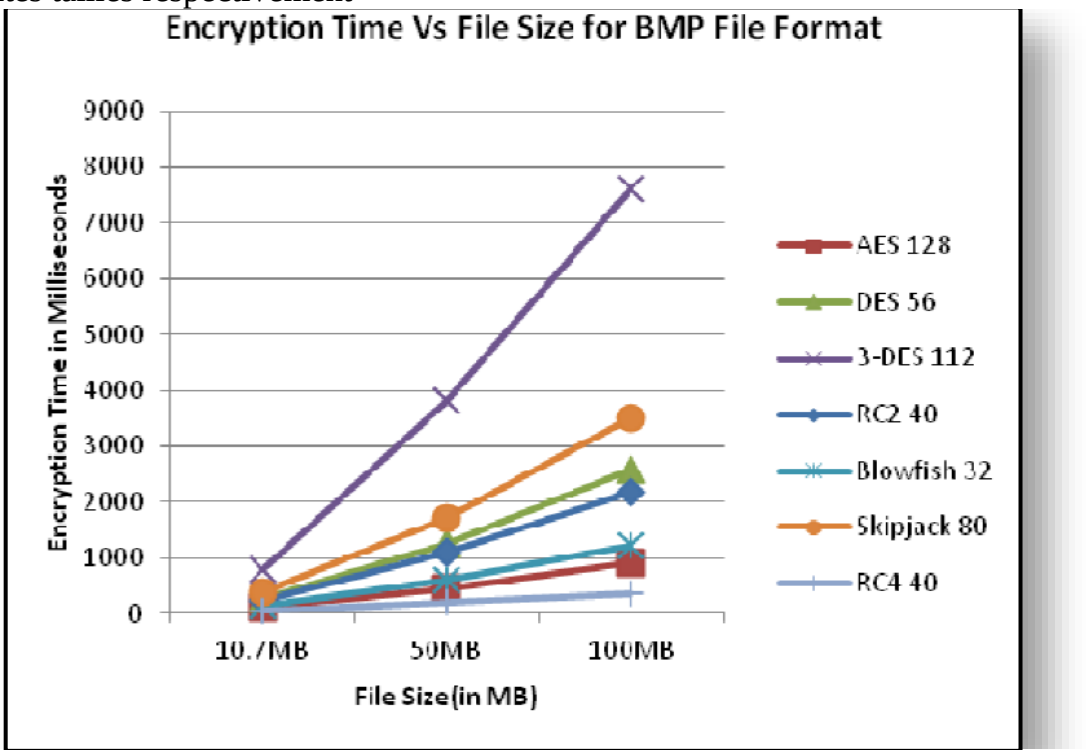


Figure 23 taille de fichier Vs temps de cryptage pour fichier BMP avec différent tailles [47].

La figure suivante montre les résultats d'exécution pour les formats de fichiers FLV de différentes tailles respectivement

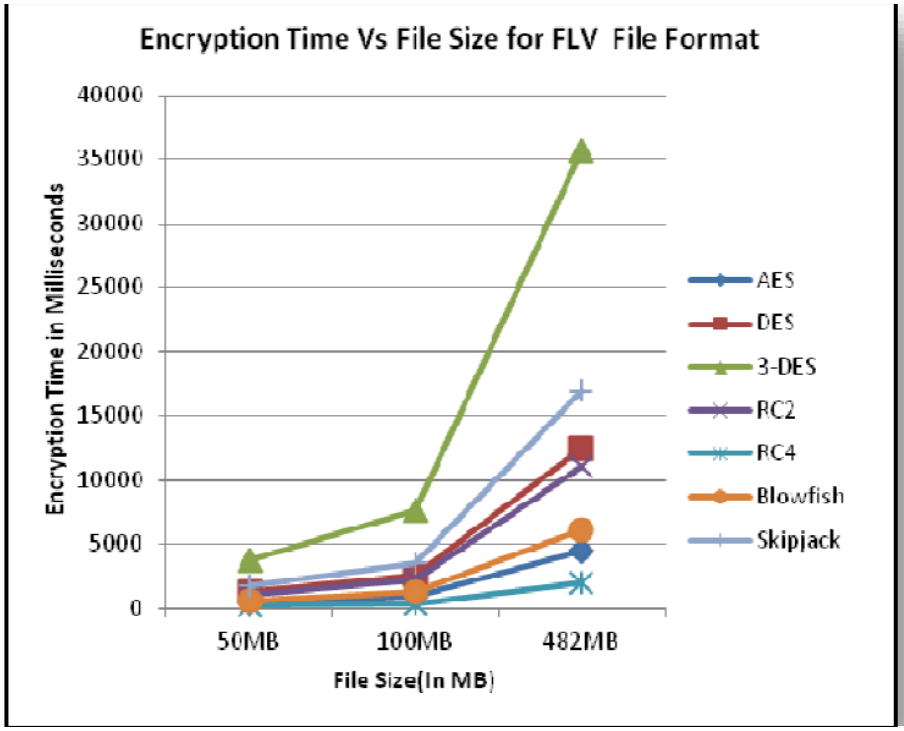


Figure 24 taille de fichier Vs temps de cryptage pour fichier FLV avec différent tailles [47].

Pour chaque algorithme de cryptage, les mêmes paramètres sont utilisés pour des fichiers de différentes tailles. Le tableau suivant montre le temps de cryptage de différentes tailles de fichiers de même type.

File Type	Size (In MB)	Encryption Time in Millisecond						
		AES	DES	3-DES	RC2	Blowfish	Skipjack	RC4
		128	56	112	40	32	80	40
BMP	10.7	101	272	788	238	133	381	40
	50	455	1253	3804	1095	614	1729	198
	100	909	2595	7628	2189	1223	3505	372
FLV	50	456	1268	3810	1112	629	1731	196
	100	918	2586	7631	2224	1267	3515	360
	482	4518	12529	35654	11038	6087	16941	1972

Figure 25 temps de cryptage de différentes tailles de fichiers de même type [47].

Observation : D'après les résultats du tableau et les figures, nous pouvons constater que le résultat pour différentes tailles de données varie proportionnellement à la taille du fichier de données.

Alors : Le temps de cryptage augmente lorsque la taille du fichier augmente en multiples de la taille des données.

➢ **Etude de cas 3 : Fichier avec différentes densités de données.**

Cette étude de cas a pour but de vérifier si le cryptage dépend de la densité des données ou non. Le taux de cryptage est évalué pour les deux fichiers de densité de données différente ; un fichier clairsemé de 69 Mo et un fichier dense de 58,5 Mo. Pour un algorithme de chiffrement, la taille de la clé et le mode de bloc sont maintenus au strict minimaux. Les résultats de l'exécution sont présentés dans le tableau suivant.

Algorithm Name	Dense (61392454 Bytes) AIFF file		Sparse (72000118 Bytes) AIFF file	
	Encrypt Time(ms)	Encryption Rate(MB/s)	Encrypt Time(ms)	Encryption Rate(MB/s)
AES 128	539	108.62	632	108.64
DES 56	1535	38.14	1800	38.14
3-DES 112	4363	13.42	5074	13.53
RC2 128	1283	45.63	1518	45.23
Blowfish 128	725	80.75	852	80.59
Skipjack 128	2040	28.70	2384	28.80
RC4 128	216	271.05	251	271.56

Figure 26 taux de cryptage pour les fichiers épars et denses [47].

7. Les critère (features) de notre dataset :

Dans la suite nous avons explique les critères (FEATURES) des notre dataset, la suivante montre une partie de notre dataset.

	connexion	frq	mode	vitessede	Tailledebloc	tailledecles	typecryp	nbcles	signaturedesmsgs	rappiditedecalcule	assu
0	non	elever	symitrique	tresrapide	128	128	bloc	1	defficilement	rapidedecalcul	
1	non	moy	symitrique	tresrapide	128	192	bloc	1	defficilement	rapidedecalcul	
2	non	faible	symitrique	tresrapide	128	256	bloc	1	defficilement	rapidedecalcul	
3	non	elever	symitrique	rapide	64	56	bloc	1	defficilement	rapidedecalcul	
4	non	moy	symitrique	rapide	64	56	bloc	1	defficilement	rapidedecalcul	
5	non	faible	symitrique	rapide	64	56	bloc	1	defficilement	rapidedecalcul	
6	non	elever	symitrique	pluslent	64	112	bloc	2	defficilement	rapidedecalcul	
7	non	moy	symitrique	pluslent	64	112	bloc	2	defficilement	rapidedecalcul	
8	non	faible	symitrique	pluslent	64	168	bloc	3	defficilement	rapidedecalcul	
9	non	elever	symitrique	rapide	64	128	bloc	1	defficilement	rapidedecalcul	
0	non	moy	symitrique	rapide	64	128	bloc	1	defficilement	rapidedecalcul	

Figure 27 dataset Algo_

Mode de connexion : connecter (sur un réseau) ou bien non connecter (en locale), le mode de connexion va influer au type de cryptage utilisé,

Fréquence : connaître la fréquence d'accès au fichier chiffré (élevé ou bien moyen) q, nous permet d'aider l'utilisateur à choisir l'algorithme approprié;
Mode: le mode de cryptage symétrique ou bien asymétrique;
Vitessde : vitesse de chiffrement (très rapide, rapide, lent…);
tailledebloc : les taille de bloc de chiffrement ;
typecry : le type de cryptage (par bloc ou bien par flux)
nbcles : nombre de clés à utiliser pour le chiffrement et déchiffrement (exp : pour RSA, nombre de clés = 3) ;
signaturedesmsg : est ce que l'algorithme de cryptage nous assure la signature des message ou non;
rapiditedecalcule : la vitesse de l'algorithme pour la génération des clés ;
assurelauthenticitedelidentite : est ce que l'algorithme de cryptage nous assurer l'authenticité de l'identité;
nbetiration : nombre des étirassions dans la phase d'exécution de chiffrement (exp pour le DES, le nombre détirassions = 16).

8. Environnement de développement :

Dans la suit nous parlerons du langage et outils de développement utilisés et nous présenterons également les résultats des tests d'évaluation de notre système

➢ Langage de programmation

Le langage de programmation Python a été créé en 1989 par **Guido van Rossum, aux Pays-Bas**. Le nom Python vient d'un hommage à la série télévisée **Monty Python's Flying Circus dont G. van Rossum** est fan. La première version publique de ce langage a été publiée en 1991[49].
La dernière version de Python est la version 3. Plus précisément, la version 3.7 a été publiée en juin 2018. La version 2 de Python est désormais obsolète et cessera d'être maintenue après le 1er janvier 2020. Dans la mesure du possible évitez de l'utiliser [49].
La **Python Software Foundation** est l'association qui organise le développement de Python et anime la communauté de développeurs et d'utilisateurs.

➢ caractéristiques du langage python

Ce langage de programmation présente de nombreuses caractéristiques intéressantes [49]:

- Il est multiplateforme. C'est-à-dire qu'il fonctionne sur de nombreux systèmes d'exploitation : Windows, Mac OS X, Linux, Android, iOS, depuis les mini-ordinateurs Raspberry Pi jusqu'aux supercalculateurs.
- Il est gratuit. Vous pouvez l'installer sur autant d'ordinateurs que vous voulez (même sur votre téléphone !).

- C’est un langage de haut niveau. Il demande relativement peu de connaissance sur le fonctionnement d’un ordinateur pour être utilisé.
- C’est un langage interprété. Un script Python n’a pas besoin d’être compilé pour être exécuté, contrairement à des langages comme le C ou le C++.
- Il est orienté objet. C’est-à-dire qu’il est possible de concevoir en Python des entités qui miment celles du monde réel (une cellule, une protéine, un atome, etc.) avec un certain nombre de règles de fonctionnement et d’interactions.
- Il est relativement simple à prendre en main.
- Enfin, il est très utilisé en bioinformatique et plus généralement en analyse de données.

Toutes ces caractéristiques font que Python est désormais enseigné dans de nombreuses formations, depuis l’enseignement secondaire jusqu’à l’enseignement supérieur.

➢ Anaconda Distribution

C’est une collection facile à installer des bibliothèques Python hautes performances, ainsi que Conda, notre outil de gestion des paquets et des environnements. Au-delà de la collection de paquets open source dans le programme d'installation d'Anaconda, vous pouvez utiliser Conda pour installer plus de 1,5k paquets (dont le langage R) du dépôt public d'Anaconda et plus de 20 000 paquets provenant de canaux paquets provenant des canaux communautaires, tels que Conda-forge et bioconda[50].
C’est un logiciel gratuit, disponible pour Windows, Mac OS X et Linux, et qui installera pour nous Python 3. Avec le gestionnaire de paquets conda, fourni avec Miniconda, nous pourront installer des modules supplémentaires qui sont très utiles en bioinformatique (NumPy, scipy, matplotlib, pandas, Biopython), mais également les notebooks Jupyter, Spyder[50].

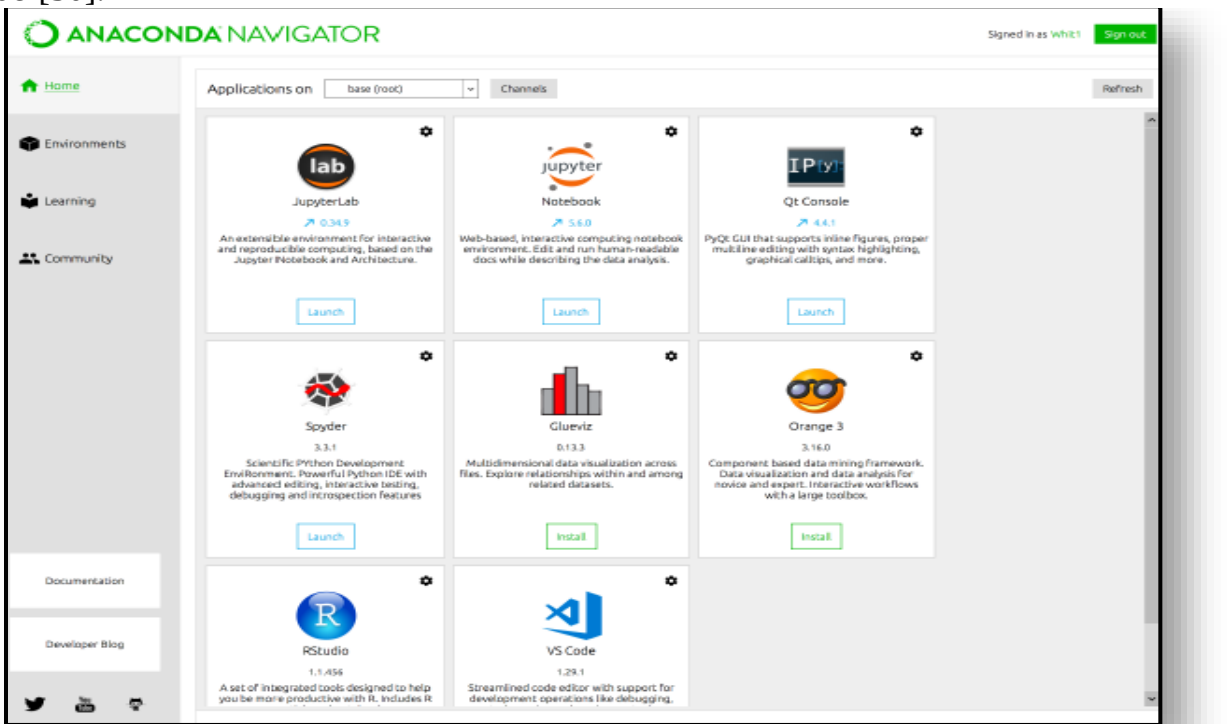


Figure 28 interface d'Anaconda

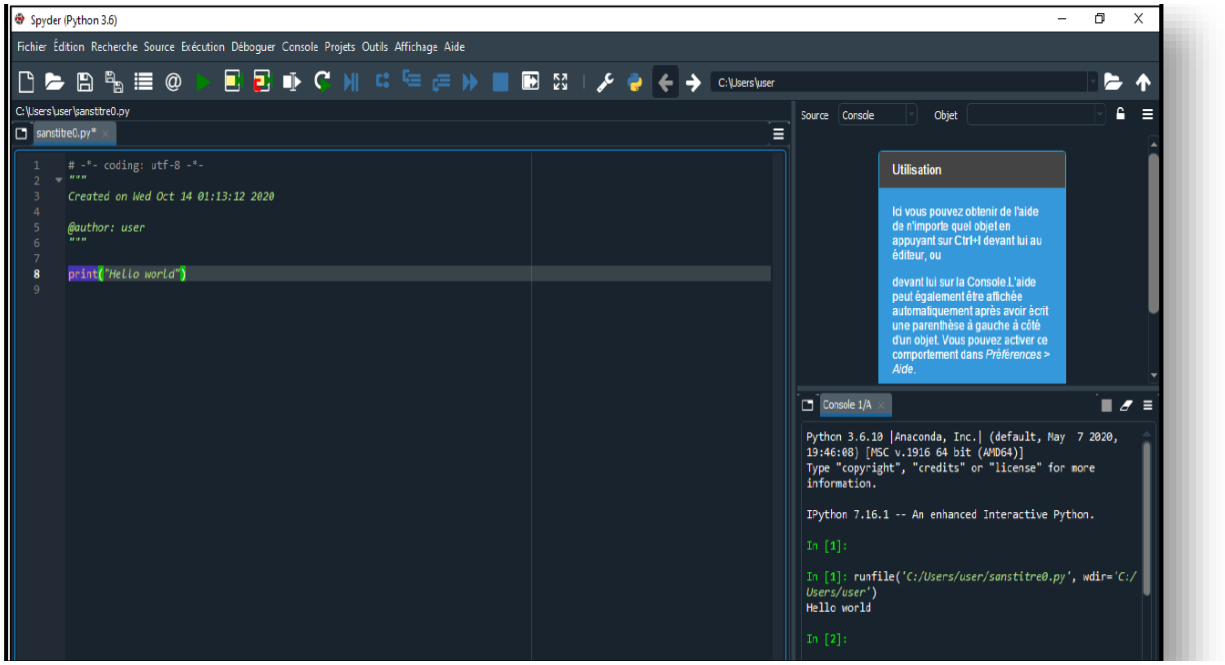


Figure 29 Environnement Spyder

9. Outils et librairies utilisés

Les bibliothèques que nous avons utilisées pour notre travail sont :

Pandas : elle permet la manipulation et l'analyse des données, nous l'avons utilisé pour lire le dataset.

Scikit-learn : c'est la principale bibliothèque d'outils dédiés au machine learning et à la science des données dans l'univers Python, nous avons utilisé plusieurs modules de cette dernière pour la création de notre modèle, le training et les tests.

Matplotlib : l'une des bibliothèques python les plus utilisées pour représenter des graphiques en 2D. Elle permet de produire une grande variété de graphiques et ils sont de grande qualité. Le module pyplot de matplotlib est l'un de ses principaux modules. Il regroupe un grand nombre de fonctions qui servent à créer des graphiques et les personnaliser (travailler sur les axes, le type de graphique, sa forme et même rajouter du texte).

Tkinter (de l'anglais Tool kit interface) : est le module graphique libre d'origine pour le

langage Python, permettant la création d'interfaces graphiques. L'un des avantages de Tkinter est sa portabilité sur les OS les plus utilisés par le grand public. Le module (classe) Tkinter contient des fonctions intégrées appelées méthodes permettant de réaliser des actions sur les fenêtres graphiques (objets).

NumPy

NumPy est une extension du langage de programmation Python, destinée à manipuler des matrices ou tableaux multidimensionnels ainsi que des fonctions mathématiques opérant sur ces tableaux.

Wordcloud : On les appelle WordCloud ou TagCloud, ce sont ces petits outils (site web et App) bien pratiques qui permettent de créer une représentation visuelle des mots (clés) d'une page web, d'un site internet, d'un livre, d'un texte, etc.

10. Implémentation

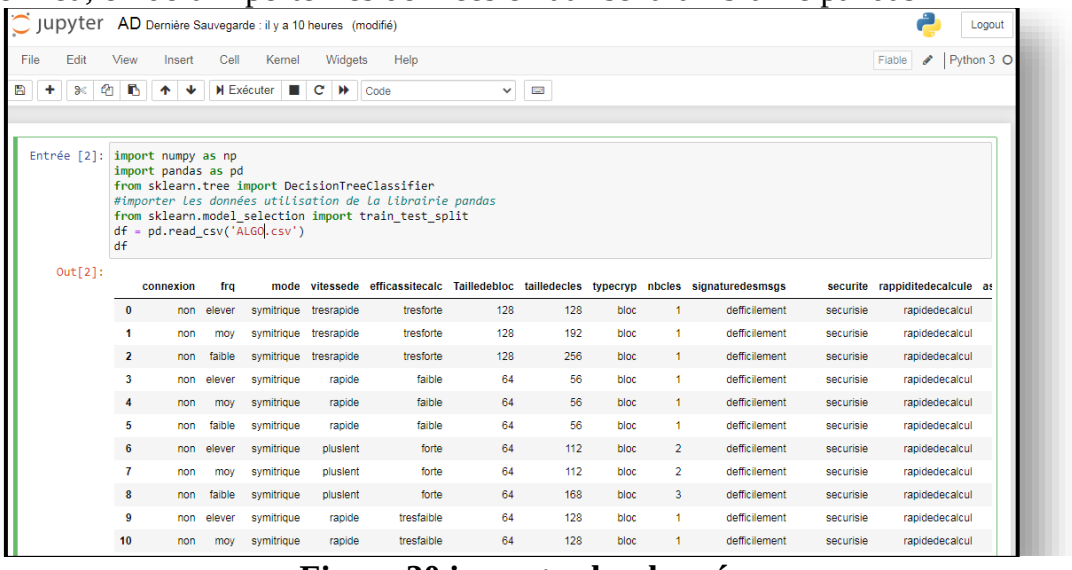
Dans notre application nous avons utilisés deux type d'algorithme d'apprentissage automatique (les arbres de décision et le KNN)

➢ Préparation de l'environnement de travail :

Sachant que les algorithmes d'apprentissage utilisées pour notre approche ne traites pas des données de type chaine de caractère, mais des données de type réel (Float), on doit préparer l'environnement d'apprentissage suivons les étapes qui suit :

a. Importer les données

En premier lieu, on doit importer les données en utilisent la librairie pandas



	connexion	frq	mode	vitessede	efficassitecalc	Tailledebloc	tailledecles	typecryp	nbcles	signaturedesmsgs	securite	rappiditedecalcule	as
0	non	elever	symitrique	tresrapide	tresforte	128	128	bloc	1	defficilement	securisie	rapidedecalcul	
1	non	moy	symitrique	tresrapide	tresforte	128	192	bloc	1	defficilement	securisie	rapidedecalcul	
2	non	faible	symitrique	tresrapide	tresforte	128	256	bloc	1	defficilement	securisie	rapidedecalcul	
3	non	elever	symitrique	rapide	faible	64	56	bloc	1	defficilement	securisie	rapidedecalcul	
4	non	moy	symitrique	rapide	faible	64	56	bloc	1	defficilement	securisie	rapidedecalcul	
5	non	faible	symitrique	rapide	faible	64	56	bloc	1	defficilement	securisie	rapidedecalcul	
6	non	elever	symitrique	pluslent	forte	64	112	bloc	2	defficilement	securisie	rapidedecalcul	
7	non	moy	symitrique	pluslent	forte	64	112	bloc	2	defficilement	securisie	rapidedecalcul	
8	non	faible	symitrique	pluslent	forte	64	168	bloc	3	defficilement	securisie	rapidedecalcul	
9	non	elever	symitrique	rapide	tresfaible	64	128	bloc	1	defficilement	securisie	rapidedecalcul	
10	non	moy	symitrique	rapide	tresfaible	64	128	bloc	1	defficilement	securisie	rapidedecalcul	

Figure 30 importer les données

Nous affichons les informations sur le type des variables.

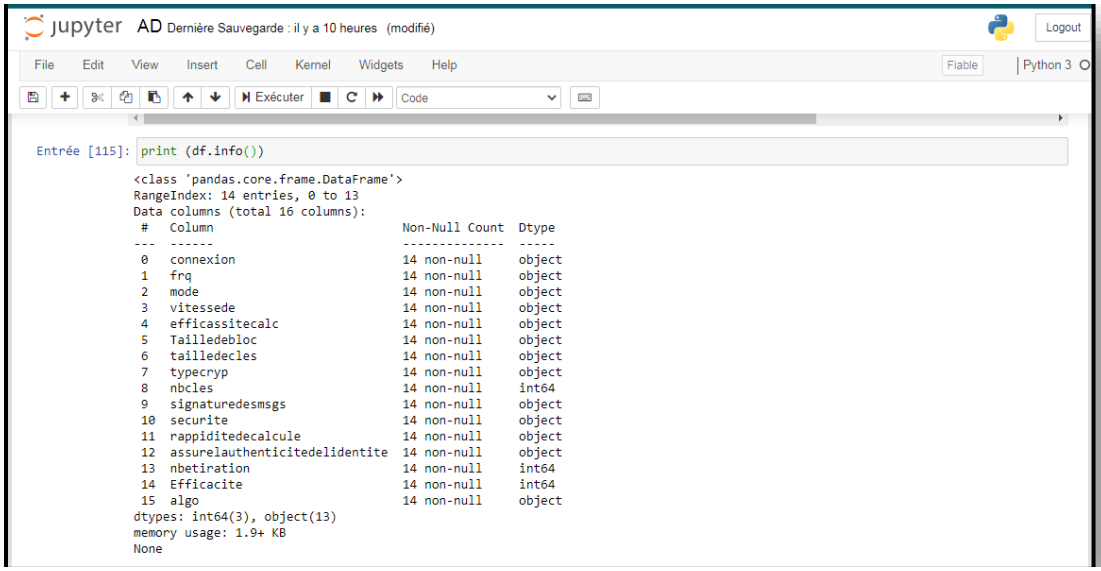


Figure 31 afficher des informations sur les données

b. Encodage et Normalisée de la variable cible

Le but de cette étape est de convertir les données de notre dataset depuis des données de type chaine de caractère a des données de type réel

Pour réaliser la dit étape, on utilise la librairie **sklearn.preprocessing** et la fonction **fit_transform().**

La figure suivante montre comme exemple la fonction de l'encodage pour les deux variables cible **connexion et frq.**

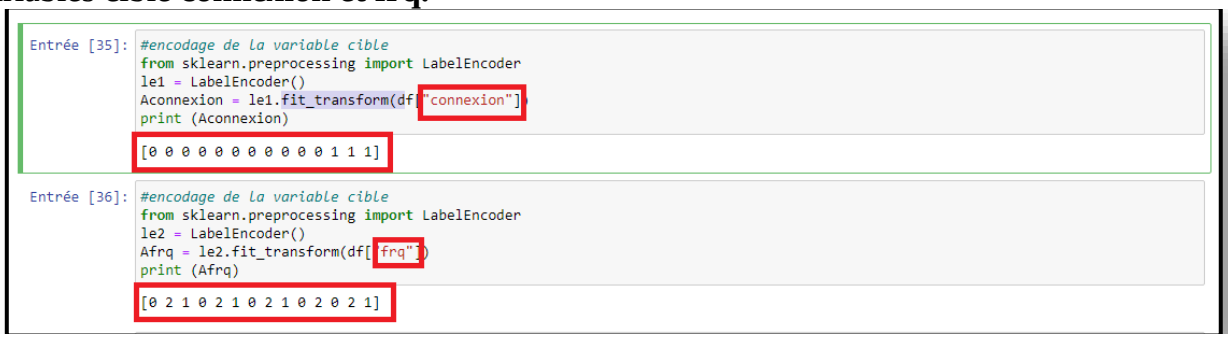
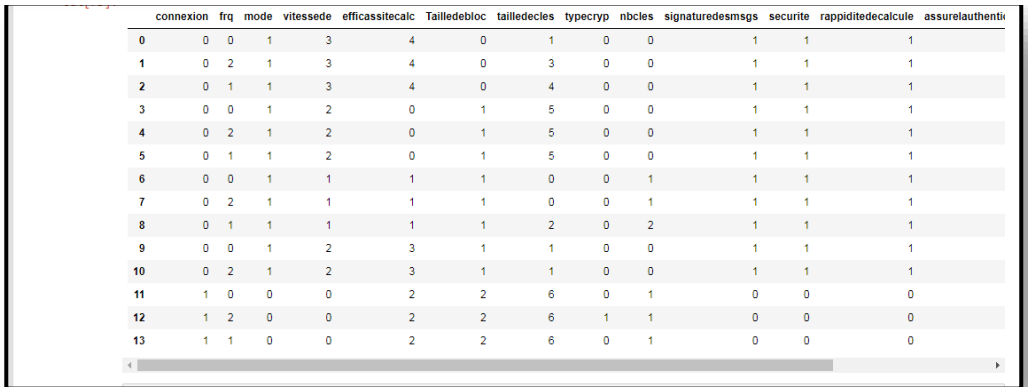


Figure 32 encodage des variables cible

Dans une boucle, on va répéter la fonction précédente sur tous les variables cible de notre dataset



	connexion	frq	mode	vitessede	efficassitecalc	Tailledebloc	tailledecles	typecryp	nbcles	signaturedesmsgs	securite	rappiditedecalcule	assurelauthenti
0	0	0	1	3	4	0	1	0	0	1	1	1	
1	0	2	1	3	4	0	3	0	0	1	1	1	
2	0	1	1	3	4	0	4	0	0	1	1	1	
3	0	0	1	2	0	1	5	0	0	1	1	1	
4	0	2	1	2	0	1	5	0	0	1	1	1	
5	0	1	1	2	0	1	5	0	0	1	1	1	
6	0	0	1	1	1	1	0	0	1	1	1	1	
7	0	2	1	1	1	1	0	0	1	1	1	1	
8	0	1	1	1	1	1	2	0	2	1	1	1	
9	0	0	1	2	3	1	1	0	0	1	1	1	
10	0	2	1	2	3	1	1	0	0	1	1	1	
11	1	0	0	0	2	2	6	0	1	0	0	0	
12	1	2	0	0	2	2	6	1	1	0	0	0	
13	1	1	0	0	2	2	6	0	1	0	0	0	

Figure 33 dataset encoder

c. Partition en échantillons d'apprentissage et de test

Nous cherchons à appliquer le schéma type de l'analyse prédictive : scinder les données en échantillons **d'apprentissage** et de **test**.

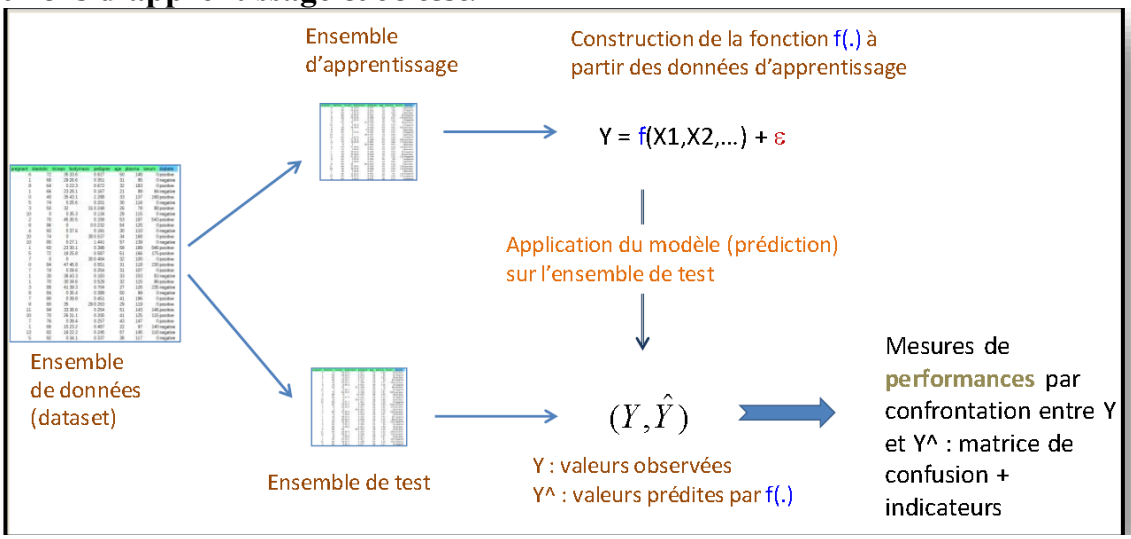


Figure 34 Schéma type de travail en analyse prédictive

Nous souhaitons réserver 80% observations pour l'apprentissage et 20% pour le test.

```
Entrée [76]: #su bdiviser les données en échantillons d'apprentissage et de test
             from sklearn.model_selection import train_test_split
             x_trainset, x_testset, y_trainset, y_testset = train_test_split(x,y, test_size=0.2, random_state=3)

Entrée [77]:
```

Figure 35 subdiviser les données en échantillons d'apprentissage et de test

11. Instanciation et modélisation

Afin de montrer les résultats obtenus pour les différents modèles, nous donnons dans ce qui suit les étapes à suivre pour chaque type d'algorithme d'apprentissage.

A. Modélisation avec les arbres de décision

Nous détaillerons dans ce qui suit, la création de notre arbre de Décision destinée

```
#Importation des modules nécessaires
from sklearn.tree import DecisionTreeClassifier, plot_tree
import matplotlib.pyplot as plt

#Déclaration de l'arbre de décision
Algo_Tree = DecisionTreeClassifier(max_depth=11)
#Entrainement de l'abre de décision
Algo_Tree.fit(x, y)
#Affichage de l'abre de décision obtenu après entraînement
plot_tree(Algo_Tree, feature_names= ['connexion','frq','mode','vitessede','efficassitecalc','Tailledebloc','tailledecles','type

plt.show()
```

Figure 36 programmation de l'AD

En suite, on doit afficher la précision de notre model

```
5]: #clf.predict(x)

1] Algo_Tree.score(x, y)

]: 1.0
```

Figure 37 pourcentage de la précision avec AD

On remarque que la précision de notre model est a 100%, alors on Affiche l'arbre à l'aide de sklearn.

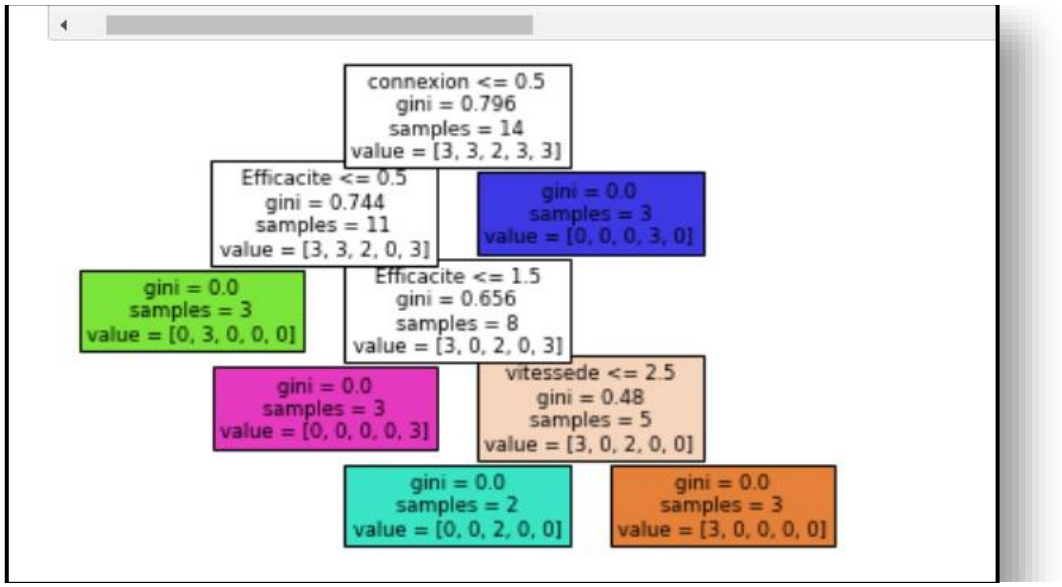


Figure 38 affichage l'AD de notre modèle

Tester le model

Supposant que les besoin de l'utilisateur simple sont comme suit :

- Mode de connexion : en local,
- La fréquence d'accédé au fichier est moyen ;
- Il ne se soucie pas d'assurer né la signature du document, né l'authentification ;
- Il veut d'utiliser une seul clé pour le chiffrement ;
- Il veut d'utiliser une clé de taille moyen ;
- Il ne se soucie pas à la rapidité de calcule,

Alors, avec ces paramètres, le résultat sera **DES (voir la figure suivante)**

```
[177]: prediction_algo_tree = Algo_Tree.predict([[0,1,1,1,1,1,2,0,2,1,1,1,0,6,1]])
       prediction_algo_tree

[177]: array([4], dtype=int64)
```

Figure 39 teste de l'AD

B. Modélisation avec l'algorithme de KNN

Nous détaillerons dans ce qui suit, la création de notre arbre de Décision destinée

```
[227]: from sklearn.neighbors import KNeighborsClassifier
       #instanciation et définition du k
       knn = KNeighborsClassifier(n_neighbors = 3)

[228]: knn.fit(x,y)

       C:\Users\shawlin\anaconda3\lib\site-packages\ipykernel_launcher.py:1: DataConversionWarning: A col
       a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
         """Entry point for launching an IPython kernel.

[228]: KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                            metric_params=None, n_jobs=None, n_neighbors=3, p=2,
                            weights='uniform')
```

Figure 40 programmation du KNN

En suite, on doit afficher la précision de notre model

```
[234]: knn.score(x,y)

[234]: 0.8571428571428571
```

Figure 41 pourcentage de la précision avec knn

On remarque que la précision de notre model est a 80%.

Tester le model

On utilise le même exemple de test pour les arbres de décision, Alors, le résultat sera aussi **DES (voir la figure suivante)**

```
[231]: prediction_algo_knn = knn.predict([[0,1,1,1,1,1,2,0,2,1,1,1,0,6,1]])
       prediction_algo_knn

[231]: array([4], dtype=int64)
```

Figure 42 tester le modèle KNN

Dans cette partie, nous allons décrire notre proposition ainsi les différentes techniques utilisée dans cette approche. On propose un crypto système basé sur l'apprentissage supervisé Généralement la cryptographie est utilisée pour rendre les documents plus confidentiel et elle est utilisé par des experts ou bien des professionnels de l'informatique et de la sécurité informatique et pour rendre la cryptographie plus utilisé même pour les personnes qui n'ont pas une expérience dans ce domaine. On propose un crypto système basé sur les algorithmes de l'apprentissage supervisé dont le but d'analyser les besoins de l'utilisateurs de notre système et de l'aider à prendre la décision et choisir le meilleur algorithme de chiffrement. Dans les éléments suivantes nous allons vous détailler notre proposition

12. Conclusion

Dans ce chapitre, nous avons présenté la solution proposé, une introduction sur KDD, ainsi que nous avons mis le point sur les critères de création de notre data set, et Dans ce chapitre, nous avons précisé le problème que nous voulons résoudre, et nous avons également expliqué la solution proposée en combinant codage et apprentissage automatique, et à la fin nous avons expliqué les outils utilisés dans ce travail.

Conclusion général :

L’objectif de ce travail est le développement un crypto système basé sur l’apprentissage automatique afin d’aider les utilisateurs simples à utiliser les algorithmes de cryptage. Dans une première partie, nous avons abordé la notion de la sécurité informatiques et la cryptographie. Cette partie a permis de mieux comprendre les notions de base liée à la sécurité informatique, le but principale de la cryptographie et comment ce dernier assurer les objectifs principaux de la sécurité informatique notamment l’intégrité, la confidentialité et l’authentification.

Ainsi , nous avons présenté une introduction au domaine de l’apprentissage automatique (Machine Learning), les défirent algorithme, l’utilisation de la méthode KDD pour crier note dataset, et l’approche de l’OCR afin d’améliorer le choix de notre algorithme d’apprentissage.

Enfin, ce mémoire se termine en abordant l'implémentation concrète de la sécurité des réseaux. Cette partie balaye les outils et les technologies disponibles pour assurer le développement des algorithmes d’AA, comme Python,

Ce travail de Master 2 nous a permis de :

- Approfondir nos connaissances théoriques et pratiques déjà acquises, maîtriser les nouvelles techniques et compléter notre initiale pour atteindre ainsi un niveau de perfection supérieur et de pouvoir apprendre d’autres nouveautés dans les différents domaines de la science en générale et de l’informatique en particulier.
- Découvrir le domaine de l’intelligence artificielle, plus précisément, l’apprentissage supervisé. Ce dernier utilise des outils statistiques pour découvrir des corrélations et établir des modèles dans des données.

Les références :

[1] Abou el kalam, A. Modèles et politiques de sécurité pour les domaines de la santé et des affairessociales. *Thèse de Doctorat, Laboratoire d'Analyse et d'Architecture des Systèmes du Centre National de la Recherche Scientifique*. Décembre 2003.

[2] Dubois, J. Classification automatique de courrier électronique. *Mémoire présenté pour l'obtention du grade de M. Sc. en informatique*, Université de Montréal. Juin 2002.

[3] Réseau CERTA. Système d'information: Qu'est-ce qu'un système d'information ?. Décembre 2005.

[4] Becquet, V. Le programme national : Sécurité des systèmes d'information. 2003.

[5] Scrap Aka, A.J. Le grand livre de SecuriteInfo.com. Février 2004.

[6] Urien, P. Introduction à la Sécurité des Réseaux. 2006.

[7] Burgermeister, D. et Krier, J. Les systèmes de détection d'intrusions. Juillet 2006.

[8] Detoisien, E. Les attaques réseaux". Juillet 2006.

[9] Viardin, A. Un petit guide pour la sécurité. Février 2006.

[10] Lubert, M. L'appliance de sécurité intégrée. Octobre 2005.

[11] R. Halit, M. Habachou « Conception et Réalisation d'un Cryptosystème Hybride ». Mémoire de fin d'études d'ingénieur en Electronique, Université MAMMERI Mouloud de Tizi Ouzou, 2008.

[12] S.B Hacini, M.T Inal « Implémentation d'Algorithmes de Cryptographie sousJava ». Mémoire de fin d'études de Licence, Université Abou Bakr Belkaid-Tlemcen, 2014.

[13] B. Martin, « Codage, Cryptologie et Applications », Presses Polytechniques et universitaires romandes, 2001.

[15] http://www.primenumbers.net/Renaud/fr/crypto/XOR.html .

[16] D. Doche G. Frey T. Lange K. Nguyen et F. Vercauteren R. Avanzi, H. Cohen. Handbook of elliptic and hyperelliptic curve cryptography. CRC Press, 2005.

[17] Neal Koblitz. Elliptic curve cryptosystems. Mathematics of Computation,[18] (177) :203209, January 1987.

[19] H. Lenstra. Factoring integers with elliptic curves. Annals of Mathematics,126 :649673, 1987.

[20] National Institute of Standards and Technology (NIST). Recommendation forkey management - part 1 : General (revised). NIST Special Publication Available online at : http ://csrc.nist.gov/publications/PubsSPs.html, pages 80057, 2007.

[21] Menezes, van Oorschot et Vanstone 1996] (en) A. J. Menezes, P. C. van Oorschot et S. A. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1996, 810 p. (ISBN 978-1-

4398-2191-6, lire en ligne [archive] [PDF]), « Chapitre 8.4 ElGamal public-key encryption ».

[22] CHOLLET Francois. Deep learning with python. 2017 .

[23] ZACCONE Giancarlo, MD REZAUL Karim, MENSHAWY Ahmed. Deep learning withtensorflow. 2017 .

[24] DJOKHRAB Ala eddine. Planification et optimisation de trajectoire d'un robot manipulateur à 6 ddl par des techniques neuro_floues. 2015 .

[25] P.Vincent, « Modèles à noyaux à structure locale », Thèse de Phd en informatique , Université de Montréal,2003.

[26] DJOKHRAB Ala eddine. Planification et optimisation de trajectoire d'un robot manipulateur à 6 .ddl par des techniques neuro_floues. 2015 .

[27] Luger, G.F. et Stubblefield, W.A. Artificial intelligence: structures and strategies for complexproblem solving. *Addison-Wesley*. pp. 824. 1998.

[28] Amor, N. B., Benferhat, S., and Elouedi, Z. Naive Bayes vs decision trees in intrusion detection systems. In *Proc. 19th ACM Symposium on Applied Computing*. pp. 420-424, 2004.

[29] Valdes, A. et Skinner, K. Adaptive model-based monitoring for cyber attack detection. *Proceedings of the Third International Workshop on Recent Advances in Intrusion Detection*. Vol. 1907. pp. 80-92. 2000.

[30] Mcculloch, W.C. et Pitts, W. A logical calculus of the ideas immanent in nervous activity.

Bulletin of mathematical biophysics. Vol. 5. pp. 115-133. 1943.

[31] Cornuéjols, A. Une nouvelle méthode d'apprentissage : Les SVM Séparateurs à Vaste Marge.
Bulletin de l'afia. No 51. Université de Paris-Sud, Orsay. Juin 2002.

[32] Geva, S. Boosting the Performance of Nearest Neighbour Methods with Feature Selection. *The5th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Hong Kong, Chine. pp. 210-221. 2001

[33] Han E.S., Karypis, G. et Kumar, V. Text categorization using weight adjusted k-nearest neighbor classification. *The 5th Pacific-Asia Conference on Knowledge Discovery and Data Mining*,Hong Kong, Chine, pp. 53-65. 2001.

[34] Liao, Y. et Vemur, V.R. Using text categorization techniques for intrusion detection. *The11thUSENIX Security Symposium*, San Francisco, California, Etats-Unis. pp.51-59. Août 2002.

[35] www.oracle.com/technetwork/java/index.html .

[36] Crosbie, M. et Spafford, E.H. Applying genetic programming to intrusion detection. *AAAISymposium on Genetic Programming*. pp. 1-8. 1995.

[37] Dasgupt, A.D. et Gonzalez, F.A. An intelligent decision support system for intrusion detectionand response. *Int'l Workshop on Mathematical Methods, Models and Architectrure*

For ComputerNetworks Security. Vol. 2052. pp. 1-14. 2001.
[38] Stein, G. et Bing, C. Decision Tree Classifier For Network Intrusion Detection With GA- based Feature Selection. *Proceedings of the 43rd annual Southeast regional conference*. Kennesaw, Georgia. Vol. 2. pp, 136 – 141. 2005.
[39] https://ichi.pro/fr/que-sont-les-algorithmes-d-arbre-de-decision-121064923742185
[40] https://research.aimultiple.com/author/cem.dilmegani/
[41] https://dataanalyticspost.com/lapprentissage-federe-exploite-les-dossiers-medicaux-sans-les-sortir-des-hopitaux/
[42] [EFFECTIVE USE OF THE KDD PROCESS AND DATA MINING FOR COMPUTER PERFORMANCE PROFESSIONALS Susan P. Imberman Ph.D. College of Staten Island,
[43] N. Labroche, Modelling of the chemical recognition system of ants for the unsupervised classification problem : application to web usage mining. Theses, Université François Rabelais Tours, Dec. 2003.
[44] www.authot.com/fr/2015/12/07/reconnaissancedetexte/
[45] https://airccse.org/journal/ijnsa.html
[46] www.ssi.gouv.fr Guide de sélection d'algorithmes cryptographiques l'ANSSI le 22/02/2021
[47] Une comparaison des algorithmes cryptographiques : DES, 3DES, AES, RSA et Blowfish pour deviner la prévention des attaques Mohammed Nazeh Abdul Wahid*, Abdulrahman Ali, Babak Esparham et Mohamed Marwan Université de création et de technologie de Limkokwing, Centre d'études supérieures, Cyberjaya, Malaisie
[48] Jeeva AL, Palanisamy V, Kanagaram K. Analyse comparative de l'efficacité des performances et des mesures de sécurité de certains algorithmes de chiffrement. Journal international de la recherche et des applications en ingénierie . 2012;2(3): 3033-30
[49] Abdesselem BEGHRICHE ; "De la Sécurité à la E-Confiance basée sur la Cryptographie à Seuil dans les Réseaux sans fil Ad hoc", mémoire de magistère; P : 1-18, 2008-2009
Dr. Nada Meskaoui, Nagi Wakim ; "Une approche techniques biométriques/agents pour la Sécurité des réseaux informatique"; Rapport ; 2006.