République algérienne démocratique et populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Kasdi Merbah de Ouargla

Mémoire présenté pour l'obtention Du diplôme de Master

préparée par :

Mili Mebarka

Oussif Faiza Maroua

Filière : Informatique

Option : fondamentale

Thème :

Machine learning for spam comment Filtering in Social Networks

Président : Dr.Ben Kadour M.Kamal

Examinatrice : Dr.ben Khelifa Randa

Superviseur : Dr.Bachir Said

Année Universitaire 2021/2

Dédicaces

On dédie ce mémoire à ma chère mère, qui m'a donné la vie «lemima Djamila», à mon cher père «Masoud»

à toute la famille, à mes frères Bilal, Ayoub et Samir,

à mes filles, Laila , Rabia et Salima, à la petits poussins, Noor et Adam.

A mes amis et collègues, surtout amis d'enfance,

Amal, Habiba et Siham, à tous mes professeurs,

depuis l'école primaire, mon professeur «Juwaida », jusqu'à l'université, merci à tous les honorables

professeurs qui ont contribué à mon éducation.

À toutes personnes malades à mon pays et partout dans le monde. Que dieu les bénisse. Je dédie ce modeste travail.

Oussif Faiza Maroua,

Dédicaces

Oh Dieu, louange à toi avant que tu sois satisfait, et louange à toi si tu es satisfait, et louange à toi après que tu sois satisfait.

Au confort de mes yeux, à celle qui a mis le paradis sous ses pieds... à celle qui m'a donné la vie ....ma chère mère, que Dieu la protège et prenne soin d'elle

À celui qui augmente mon affiliation avec lui et son souvenir avec fierté, et à celui qui veille la nuit pour mon éducation et mon éducation, mon cher père, que Dieu prolonge sa vie.

A tous mes frères bien-aimés et amis les plus chers, à tous ceux que le destin me réunira dans les jardins de l'étude et fera d'eux des frères.

Nous demandons à Dieu de faire de ce travail un phare pour chaque étudiant de la connaissance.

Mili Mebarka,

Remerciement

Au nom d'Allah, le Très Miséricordieux, le Très Miséricordieux, et la Prière et la Paix soient sur le Seigneur des Messagers, notre prophète Muhammad, la paix soit sur lui.

Je remercie Dieu pour toutes les opportunités, les épreuves et la force qui m'ont été offertes pour terminer la rédaction de la thèse. J'ai tellement vécu au cours de ce processus, non seulement du point de vue académique mais aussi du point de vue personnel.

Tout d'abord, je tiens à remercier sincèrement mon directeur de thèse, le ***Dr.Bashir Saaid****, pour ses conseils, sa compréhension, sa patience et ses encouragements à terminer cette thèse. Ce fut un grand plaisir de l'avoir comme encadreur.*

Je tiens à remercier mes parents, mes sœurs, ma famille, mes amis ainsi que mes collègues, ainsi que tous ceux qui m'ont aidé à mener à bien ce projet.

Merci à tou

Table des matières

Table des figures

Liste des tableaux

Résumé

Ce projet présente une méthodologie de détection des commentaires de spam sur les vidéos YouTube. Le but de cette recherche est de comparer les résultats de plusieurs algorithmes de deep learning déjà programmés pour détecter les commentaires de spam, les commentaires d'utilisateurs commentant leurs intentions promotionnelles, ou les commentaires de ceux les utilisateurs qui ne sont pas pertinents pour le sujet de la vidéo..

La politique de monétisation introduite par YouTube pour la chaîne de son utilisateur et la publicité de différentes publicités sur les vidéos YouTube a attiré un grand nombre d'utilisateurs. Cette augmentation d'un grand nombre d'utilisateurs a également entraîné une augmentation des utilisateurs malveillants dont le travail consiste à créer des robots automatisés pour commenter et s'abonner à différentes chaînes YouTube. Les commentaires de ces utilisateurs malveillants nuisent à la publicité de la chaîne et affectent également l'expérience normale de l'utilisateur.

YouTube travaille également sur ce problème en utilisant différentes méthodes pour limiter ces types de commentaires malveillants de bots automatisés en bloquant ces commentaires. Ces types de méthodes sont inefficaces dans la mesure où les spammeurs ont découvert différentes méthodes pour contourner ces approches heuristiques.

Dans ce travail, différentes techniques utilisées pour la classification des commentaires indésirables sont analysées et les résultats obtenus dans chaque approche sont comparés pour s'attaquer à ce problème majeur.

Mots-clés : spam de commentaire, YouTube, Classification, Réseaux sociaux.

Abstract

This project presents a methodology for detecting spam comments on YouTube videos.The purpose of this research is to compare the results of several deep learning algorithms that are already programmed to detect spam comments, comments from users commenting on their promotional intentions, or comments from those users that are not relevant to the topic of the video.

The monetization policy introduced by YouTube for its user's channel and advertisement of different ads on YouTube videos has attracted a large number of users. This increase in a large number of users has also lead to an increase in malicious users whose job is to create automated bots for commenting and subscription to different YouTube channels. These malicious users' comments hurt the channel publicity and also affect the normal user's experience.

YouTube is also working on this issue by using different methods to limit these kinds of automated bots malicious comments by blocking those comments.These kinds of methods are ineffective so far as spammers have found out different methods to bypass those heuristic approaches.

In this work, different techniques used for classification of spam comments are analyzed and the results obtained in each approach are compared to tackle this major issue.

Key-words : comment spam, YouTube, Classification, Social networks.

الملخص

، والغرض من هذا البحث هو مقارنة نتائج YouTube يقدم هذا المشروع منهجية لاكتشاف التعليقات غير المرغوب فيها على مقاطع فيديو العديد من خوارزميات التعلم العميق التي تمت برمجتها بالفعل لاكتشاف التعليقات غير المرغوب فيها ، أو التعليقات من المستخدمين الذين .يعلقون على نواياهم الترويجية ، أو التعليقات من هؤلاء. المستخدمين غير ذوي الصلة بموضوع الفيديو
عددًا كبيرًا YouTube لقناة مستخدميه وإعلان إعلانات مختلفة على مقاطع فيديو YouTube جذبت سياسة تحقيق الدخل التي قدمها موقع من المستخدمين. أدت هذه الزيادة في عدد كبير من المستخدمين أيضًا إلى زيادة عدد المستخدمين الخبثاء الذين تتمثل مهمتهم في إنشاء المختلفة. تضر تعليقات المستخدمين الخبيثة بالدعاية على القناة وتؤثر أيضًا على YouTube روبوتات آلية للتعليق والاشتراك في قنوات .تجربة المستخدم العادية
أيضًا على حل هذه المشكلة من خلال استخدام طرق مختلفة للحد من هذه الأنواع من التعليقات الخبيثة من الروبوتات عن YouTube يعمل طريق حظر تلك التعليقات ، هذه الأنواع من الأساليب غير فعالة حتى الآن حيث اكتشف مرسلو الرسائل غير المرغوب فيها طرقًا مختلفة .لتجاوز تلك الأساليب الاستكشافية
في هذا العمل ، يتم تحليل التقنيات المختلفة المستخدمة في تصنيف التعليقات المزعجة ومقارنة النتائج التي تم الحصول عليها في كل نهج .لمعالجة هذه القضية الرئيسية

الكلمات الرئيسية: التعليقات غير المرغوب فيها ، يوتيوب ، التصنيف ، الشبكات الاجتماعية

Introduction générale

Aujourd'hui, les technologies couvrent tous les domaines de la vie humaine et élargissent les plates-formes de communication avec une zone appropriée et à faible coût. Les organisations publicitaires et à but lucratif utilisent cette large zone d'audience et la plate-forme à faible coût pour envoyer des informations et des cibles souhaitées au spam.En plus de créer des problèmes pour les utilisateurs, cela constitue également une menace pour la productivité,la fiabilité et la sécurité du réseau.Nous décrivons les commentaires de spam comme ceux qui contiennent une intention convaincante ou qui sont jugés inappropriés dans le contexte d'une vidéo particulière.

La perspective de la monétisation par la publicité sur les canaux de médias sociaux populaires au fil des ans a attiré un nombre de plus en plus important d'utilisateurs. Cela a à son tour conduit à la croissance d'utilisateurs malveillants qui ont commencé à créer des bots automatisés, qui sont efficaces dans la distribution organisée à grande échelle de mes-sages de spam sur plusieurs canaux à la fois. YouTube lui-même s'est attaqué à ce problème avec quelques méthodes limitées de blocage des commentaires indésirables constitués de liens.Ces méthodes se sont avérées exceptionnellement improductives car les spammeurs ont trouvédes stratégies pour contourner ces méthodes expérimentales.

Les algorithmes de classificationd'apprentissage en profondeur standard fonctionnent, mais il est toujours possible d'obtenir une meilleure précision en utilisant de nouvelles méthodes. Dans ce travail, nous visons à iden-tifier ces retours en mettant en œuvre des algorithmes d'apprentissage automatique tels que svm, naive baise , random forest ,n_gram qui se sont avérés très efficaces pour détecter et par lasuite combattre les commentaires de spam. Ce travail est organisé en trois principaux chapitrescomme suit :

Chapitre 1 :ce chapitre présente un apercu de l'analyse des réseaux sociaux et des ses application en apprentissage automatique ainsi qu'en traitement du langage naturel (TALN).

Chapitre 2 : Dans ce chapitre nous Donnons un aperçu du système de filtrage du spam, nous avons préparée la liste des travaux connexes et nous avons identifié le problème que nous abordons dans ce mémoire.

Chapitre 3 : Dans ce chapitre nous définissons l'ensemble de données, les algorithmes, les outils,les bibliothèques et les logiciels que nous avons dû utiliser et nous discutons les résultats du modèle proposé.

Chapitre 1

Analyse des réseaux sociaux

1.1 Introduction

Les réseaux sociaux (ARS)font partie de la vie quotidienne des gens. Des millions d'utilisateurs sont connectés via les réseaux sociaux en ligne. L'analyse des réseaux sociaux est une tâche difficile en raison du grand nombre d'utilisateurs et de l'énorme quantité de données. L'avènement des technologies de techniques d'apprentissage a permis d'effectuer une analyse rigoureuse des réseaux ARS . Dans le domaine de l'analyse des réseaux sociaux, de nombreuses études ontété menées en utilisant des techniques d'apprentissage profond sous différents angles. Dans ce chapitre, nous définissons le concept d'analyse des réseaux sociaux et le cas de recherche pourles applications Diverses analyses de réseaux sociaux utilisant des techniques d'apprentissage

1.2 Définition de réseau social

Un réseau social est un ensemble de nœuds socialement liés par une ou plusieurs relations. Les membres du nœud, ou du réseau, sont des entités connectées Grâce à notre étudede la relation de ses modèles. ces unités sont principalement des individus ou des organisations,mais en principe toutes les connexions à d'autres entités peuvent être vérifiées en tant que nœuds [5].

1.3 Définition de l'analyse des réseaux sociaux

Le RS est une technique d'exploration de données qui révèle la structure et le contenu d'un ensemble d'informations en le représentant comme un ensemble d'objets ou d'entités interconnectés et liés. La structure RS se compose de nœuds connectés à d'autres nœuds associés par des liens. Les nœuds du réseau sont les personnes et les groupes, tandis que les liens montrent des relations ou des flux entre les nœuds. Lorsque deux nœuds se connectent, cela crée une arête. Un chemin d'accès fait référence à une collection de nœuds connectés par un lien.

La figure 1.1 montre un réseau entièrement connecté où tous les nœuds du réseau social sont connectés [1].

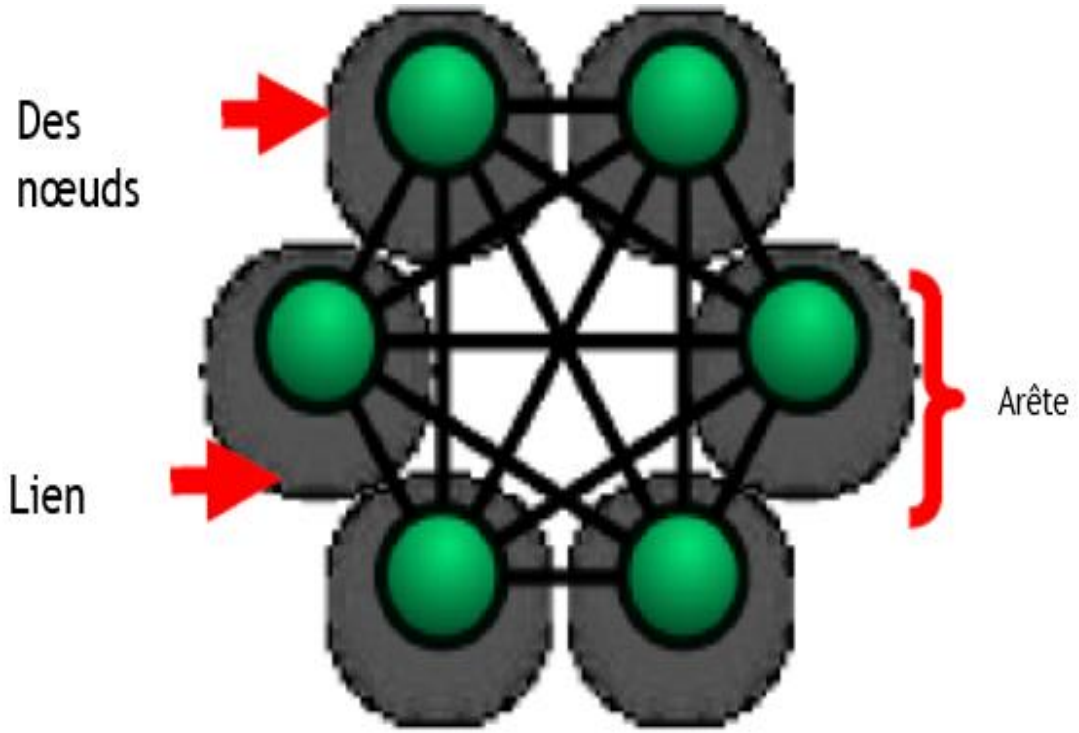

Figure 1 la structure d'ARS

1.4 Traitement automatique du langage naturel (TALN)

La quantité de contenu généré par les utilisateurs dans les médias sociaux augmente de façon exponentielle. Les données textuelles ne peuvent pas être traitées efficacement par une machinecomme avec d'autres formats de données. Une machine doit comprendre langage naturel des humain etle langage pour analyser le contenu du textuel. Le traitement automatique du langage naturel (TALN) aide les machines à comprendre langage naturel des humain et le langage dans le contenu textuel généré par les réseaux sociaux.

Le flux de contenu des médias sociaux vers un système de stockage de données volumineuses et l'analyse par ML

et TALN sont illustrés dans la figure 1.3 Ces derniers temps, le d'approntissage automatique etl'intelligence artificielle jouent un rôle essentiel dans l'engagement de millions d'utilisateurs demédias sociaux. Des études récentes montrent que les clients sont plus fidèles aux entreprises qui leur répondent rapidement. Les robots ou les programmes d'approntissage automatique comprennent automatiquement les requêtes des clients à l'aide de la TALN et y répondent sur-

le-champ. Cette avancée aide les entreprises à fidéliser leurs clients et à établir des relations plus solides avec eux [4].

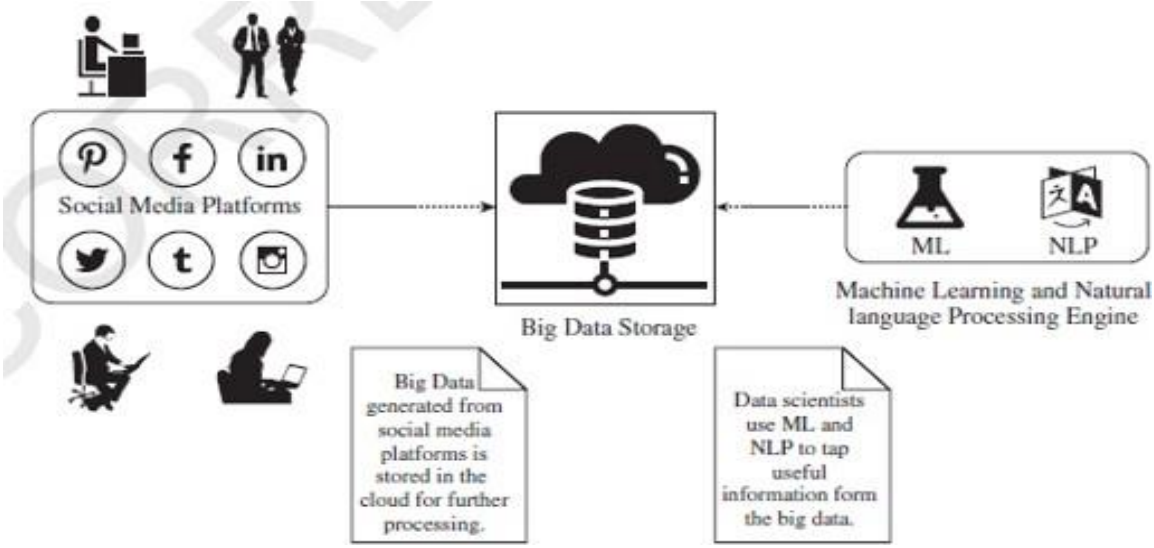

Figure 2 Flux de travail du Big Data, de d'apprentissage automatique et des réseaux sociaux

1.1 Application de l'analyse des réseaux sociaux

ARS fournit de nombreux services à ses utilisateurs tels que des appels, des messages, des alertes et des notifications de publication de contenu et d'offres d'emploi. Les utilisateurs connectés via ARS peuvent également connaître les sentiments et les opinions des utilisateurs sur un sujet ou un événement. La figure 1.2 ci-dessous montre d'autres applications ARS.

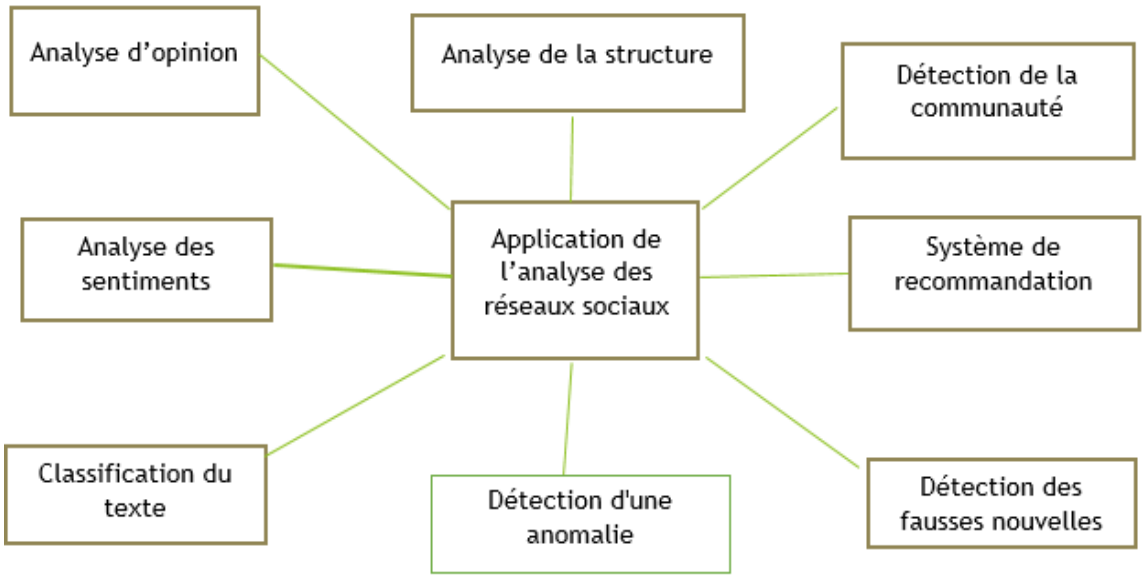

Figure 3 application de l'analyse des réseaux sociaux

1.1.1 Analyse d'opinion

Une opinion est un point de vue ou une déclaration qui peut ne pas être étayée par des preuves. Souvent, les gens discutent d'un sujet ou partagent du contenu sur les réseaux sociaux en ligne et leurs abonnés ou amis peuvent améliorer leurs opinions sur le sujet. Un phénomène affectant les opinions des gens est appelé influence sociale et son analyse est appelée analyse de l'influence sociale.

Ses applications incluent la recommandation de contenus, d'utilisateurs, de produits et de publicités. Les méthodes classiques d'analyse de l'influence sociale incluent des règles générées manuellement pour extraire des caractéristiques. Ces méthodes nécessitent une connaissance du domaine et sont souvent limitées par les connaissances des experts du domaine. L'utilisation de techniques d'apprentissage en profondeur surmonte les limites des méthodes classiques car elle ne nécessite pas l'extraction de fonctionnalités via des scripts écrits manuellement [2].

1.1.2 Analyse des sentiments

L'analyse des sentiments classe les émotions des utilisateurs à partir du texte qu'ils partagent sur les réseaux sociaux et les sites de microblogging. Un exemple de ceci pourrait être la classification des clients satisfaits, neutres et mécontents à partir des données de rétroaction. Cette méthode vise à comprendre le contenu généré par l'utilisateur et à décider de son émotion avec des méthodes de calcul ou statistiques. La technologie Web est l'avancée technologique la plus importante de la dernière décennie. Cela a changé la façon dont les gens pensent et la façon dont ils achètent des articles. Lo et Potdar (2009).

1.1.3 Classification du texte

Les données non structurées sous forme de texte sont omniprésentes : e-mails, chats, pages Web et blogs en ligne.

Ces dernières années, il y a eu une croissance exponentielle du nombre de documents numériques et de textes complexes qui nécessitent une compréhension plus approfondie des méthodes d'apprentissage automatique pour classer avec précision les textes dans de nombreuses applications. Le problème de la classification a été largement étudié dans les domaines de l'exploration de données, de l'apprentissage automatique, des bases de données et de la recherche d'informations avec des applications dans divers domaines, tels que l'organisation de documents et le diagnostic médical, et le filtrage des groupes de discussion. La classification de texte consiste à attribuer des catégories prédéfinies à des documents texte (Sebastiani, 2002). C'est l'une des tâches fondamentales dans le domaine du traitement automatique du langage naturel (TALN) avec de larges applications telles que l'analyse des sentiments (Tan et al., 2011 ; Wang et al., 2018c), l'étiquetage des sujets (Joachims, 1998 ; Hingmire et al. ., 2013 ; Dieng et al., 2016), la détection de spam (Kolcz, 2005 ; Renuka et Visalakshi, 2014) et la catégorisation des intentions (Sappelli et al., 2016 ; Lampert et al., 2008) [3].

1.1.4 Détection de fausses nouvelles

L'utilisation croissante des réseaux sociaux en ligne les a rendus vulnérables aux rumeurs et aux fausses nouvelles. Les fausses nouvelles peuvent avoir des conséquences néfastes sur la vie sociale des gens. Les personnes ayant des intentions malveillantes peuvent influencer les opinions du public par le biais de fausses nouvelles. La nature omniprésente des réseaux sociaux en ligne les rend vulnérables aux contenus malveillants ou à la désinformation. Le contenu automatisé, non pertinent et inapproprié est souvent qualifié de contenu de mauvaise qualité. Un schéma pour identifier le contenu de mauvaise qualité des tweets en arabe est proposé dans Alharthi et al. (2021) en utilisant des techniques d'apprentissage en profondeur. Il consiste en une extraction automatique des fonctionnalités des tweets et en l'identification des comptes de spam. Un schéma d'apprentissage en profondeur pour la détection de faux articles d'actualité à l'aide de techniques d'intégration de mots et de CNN est présenté dans Amine et al. (2019) en extrayant des caractéristiques basées sur du texte [2].

1.1.5 Système de recommandation

Un système de recommandation fait souvent partie d'un OSN pour recommander des personnes qu'un utilisateur peut connaître pour de futures connexions probables. La liste des personnes recommandées, par exemple, peut changer après que l'utilisateur a cliqué sur le profil d'un autre utilisateur. En outre, certains OSN peuvent suggérer des événements, des cours en ligne, des offres d'emploi, des publicités et des phrases instantanées en fonction du contexte lors de la saisie des messages.

Les listes de recommandations peuvent varier en fonction des clics de l'utilisateur pour d'autres éléments afin de fournir des services personnalisés en fonction des intérêts de l'utilisateur et d'autres facteurs associés [4].

1.1.6 Détection commnautaire

L'avènement des OSN a permis de former des communautés en ligne, et les rôles de chaque utilisateur dans la formation de ces communautés doivent être analysés. Une application de l'analyse des réseaux sociaux est la détection de communauté. pour détecter les événements. Les variations dans les communautés du réseau sont utilisées pour la détection des événements dans Aktunc et al. (2020).
Certaines méthodes utilisent une combinaison d'informations topologiques et de contenu de nœud pour la détection de communauté.

Cependant, ces méthodes ne révèlent pas les structures profondes du réseau ni n'attribuent de poids aux différentes sources de données. Un algorithme appelé représentation d'intégration profonde (DIR) est proposé dans Jin et al. (2017) basée sur la reconstruction articulaire profonde (DJR). Il apprend un équilibre approprié des poids pour les différentes composantes des données utilisées pour la détection de la communauté [2].

1.1.7 Analyse de la structure

L'analyse des propriétés structurelles d'un réseau social est appelée analyse structurale. Les propriétés structurelles d'un réseau social comprennent la distribution des degrés, les me-sures de centralité, les mesures de prestige, le coefficient de regroupement, la détection de communauté, etc. L'analyse des structures des réseaux sociaux dans les données de télécommunication constituées d'enregistrements détaillés des appels est réalisée dans Al-Molhem et al. (2019). Les utilisateurs qui influencent les autres sont identifiés à l'aide de mesures de centralité pour augmenter la croissance des campagnes marketing. Les utilisateurs disposant d'un module d'identité d'abonné multiple (SIM) sont identifiés à l'aide de mesures de similarité et de comportement. La structure d'un réseau social conduit à l'évolution d'une langue. Les interactions entre les membres d'un réseau social permettent un apprentissage décentralisé entre différents groupes d'utilisateurs. L'impact de la structure d'un réseau social sur les caractéristiques de communication entre ses membres est étudié dans Dubova et al. (2020) en utilisant l'apprentissage par renforcement.

1.1.8 Détection d'anomalies

Une application de l'analyse des réseaux sociaux est la détection d'anomalies. Les anomalies représentent un schéma irrégulier ou un comportement malveillant de certains utilisateurs d'un réseau social. Par exemple, les anomalies sont les spams inutiles, les fraudes, le harcèlement, la cyberintimidation, les comportements malveillants, etc. Une étude des schémas d'apprentissage automatique pour la détection d'anomalies dans un réseau social en ligne est présentée dans Savage et al. (2014). Dans ce document, il est suggéré que la détection d'anomalies consiste en deux étapes : (i) la sélection et le calcul des caractéristiques et (ii) la classification à l'aide de l'ensemble des caractéristiques.

Un schéma basé sur une combinaison de programmation par contraintes et d'appariement de graphes est proposé dans Graoui et al. (2016) pour la détection des valeurs aberrantes et des anomalies dans un réseau social en ligne.

Un modèle appelé DeepFriend pour la classification des nœuds malveillants dans un ARS est présenté dans Wanda et Jie (2021) en utilisant l'apprentissage profond dynamique. La cyberintimidation est un phénomène visant à harceler, embarrasser, menacer ou cibler une personne verbalement ou par le biais de messages sur un réseau social en ligne ou un média social. Des schémas d'apprentissage automatique et des algorithmes de traitement du langage naturel sont utilisés dans Altay et Alatas (2018) pour identifier la cyberintimidation dans les ARS[2].

1.2 conclusion

Dans ce chapitre, nous avons jeté quelques notions sur les réseaux sociaux et leurs applications. Il existe de nombreuses applications d'analyse des réseaux sociaux telles que l'analyse des sentiments, les systèmes de recommandation, la détection de communauté et la détection de fausses nouvelles. Une énorme quantité de données est générée dans le cas d'un réseau social Par conséquent, l'analyse à partir d'un réseau social est une tâche difficile

Chapitre 2

Filtrage de spam sur Youtube

2.1 Introduction

L'Internet offre plusieurs options aux utilisateurs pour créer et entretenir des relations, les sites de réseaux sociaux tels que Twitter, Facebook, YouTube et bien d'autres, facilitent encore plus cette tâche.

Malheureusement, par le temps, les sites de réseaux sociaux ouvrent des fenêtres d'opportunité pour les cybercriminels et les menaces en ligne.

Avec un public capté et divers moyens par lesquels les cybercriminels peuvent entrer en contact avec les utilisateurs, il n'est pas surprenant que les sites de réseaux sociaux soient des cibles constantes pour le spam, les escroqueries et autres attaques.

Dans ce chapitre, nous discutons le filtrage du spam dans les commentaires et nous représentons des travaux connexes qui ont été publiés pour résoudre ce problème.

2.2 Aperçu

Le processus de filtrage du spam est comme indiqué dans la figure 2.1 :

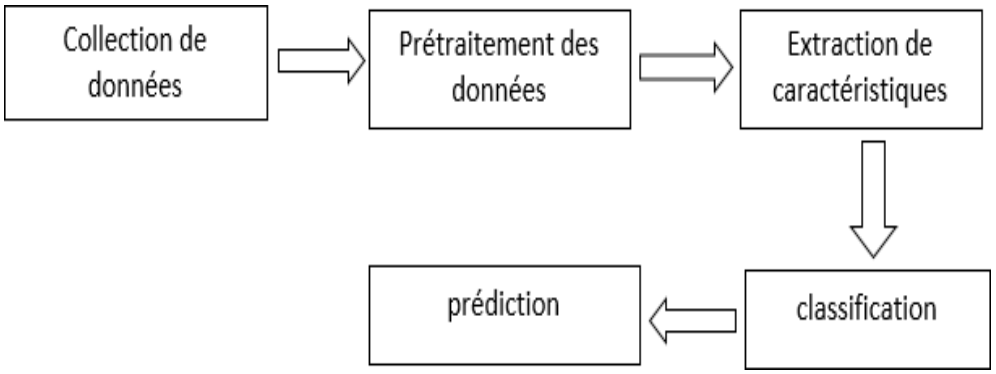

Figure 4 Le processus de filtrage du spam.

2.3 Etat de l'art (travaux connexes)

Le spam est généralement lié à des contenus avec des informations de faible valeur. Elles sont couramment trouvées sous forme d'images, de textes ou de vidéos, entraver la visualisation de contenu passionnant.
Il existe de nombreuses recherches liées au spam, comme le spam Web [6], le spam de blog [7], [8], e-mail spam [9], [10] and SMS spam [11], [12] filtrent. In social networks, undesired messages are known as social spam.

Alex Kantchelian et al. ont développé un Spam Technique de détection (SD) capable de calculer fonctionnalités inutiles et superflues dans les blogs, rendre les histoires significatives plus pratiques pour parties prenantes perpétuelles. Ils ont suggéré un l'extension de leurs travaux pour élargir la définition de spam tels que les URL, les messages courts l'enlèvement, etc. en plus de l'inclusion sensibilisation aux antagonistes, déploiement en ligne vers permettre la prédiction de commentaires futuristes et ainsi de suite [13].

Enhua Tan et al. conçu une SD d'exécution système appelé BARS : Blacklist-Assisted Runtime SD qui a construit une base de données de URL de spam contre quelle URL de chaque nouveau poste a été analysé pour déterminer si le poste était spam ou pas. Le clustering des ID utilisateur basé sur les URL partagées a également augmenté le efficacité de la détection. Cependant, l'efficacité de cette approche est une conséquence de dans quelle mesure la liste d'URL sur liste noire est-elle réussie ? favorisé [14].

L'article de HongyuGao mentionne que de nombreux les réseaux sociaux sont détectés sur l'Internet. Pour identifier le spam dans les réseaux sociaux, la méthode existante utilise les Publication sur le wall post dans Facebook. Les robots d'exploration sont utilisés pour collecter des messages sur le wall post dans Facebook exigeant utilisateurs. Puis ce message mural filtre et enfin recueille le post du wall post qui contient les URL.

Cette méthode différencie le wall post pour poster du texteet lien qui est mentionné dans le wall. Cette méthode collecte des groupes à partir d'une texture similaire contenu et le publie, y compris le même URL de destination. Graphique de similarité des publications algorithme de clustering est utilisé pour identifier les similitude entre la publication et l'URL. Basé sur cet utilisateur et ce post malveillants sont identifiés [15].

Seungwoo Choietal a expérimenté leur algorithme sur les vidéos Ted-Talks pour trouver le commentaire offrant une exposition et des informations sur le contenu vidéo.
Cependant,la méthode proposée s'est avérée inadéquate pour analyser les sentiments et opinions exprimés sur des plateformes telles que YouTube [16].

S. Lee et J. Kim, détails sur trois modules, Collecte de données, Extraction de caractéristiques, et Classement.
Sous Collecte de données, le système collecte les tweets avec URL en utilisant API Twitter Streaming qui est publiquement disponible pour obtenir des données de twitter. Dans Extraction de caractéristiques, les caractéristiques sont extraites de données existantes. L'URL redirige la longueur de la chaîne comme la fonctionnalité collecte le système car les attaquants utilisent longues chaînes de redirection d'URL pour faire des analyses difficile. L'URL suspecte sur Twitter est classé Basé sur la caractéristique [17].

Igor Santos et al. appliqué le concept de détection d'anomalie dans laquelle la divergence à partir d'e-mails authentiques a été utilisé comme métrique pour classer les e-mails comme spam ou ham. Meilleur précision a été obtenue en raison du nombre limité ensembles de formation comme on le voit dans les systèmes basés sur l'étiquetage [18].

2.4 Énoncé du problème

YouTube est devenu le site Web le plus populaire pour partager et visionner du contenu vidéo.

Cependant, cela s'est également transformé en une opportunité pour les utilisateurs malveillantsde partager du contenu promotionnel également appelé spam. Les commentaires de spam sontsouvent totalement sans rapport avec la vidéo donnée et sont généralement générés par des robots automatisés déguisés en utilisateur. Par conséquent, il est très important de trouver un moyen de détecter ces commentaires et de les signaler.

2.4.1 Spam sur Youtube

En 2016, la société a versé 2 milliards de dollars américains aux producteurs qui ont choisi de monétiser les revendications, depuis 2007. Après le lancement du système de monétisation, le site YouTube a été en proie à un contenu de très mauvaise qualité qui peut être considéré comme des vidéos de spam et des commentaires de spam. Le commentaire de spam peut être défini comme le commentaire qui n'est pas pertinent pour le contenu spécifique de la page Web. Les spams de commentaires ont été utilisés pour publier du contenu indésirable spécifique, déclarer des ventes, promouvoir du contenu pornographique, dégrader la position du site Web, rendre le site Web fiable en augmentant le nombre de vues. Le spam trouvé sur YouTube est directement lié au profit attractif offert par l'organisation de monétisation. Selon un communiqué de presse de Google, plus d'un million d'annonceurs utilisent les plateformes publicitaires Google, les bénéfices mobiles sur YouTube sont en hausse de 100% d'une année sur l'autre et le nombre d'heures que les gens regardent sur YouTube chaque mois est en hausse de 50% d'une année sur l'autre. Dans le même temps, selon Nexgate, une société de sécurité informatique, rien qu'au premier semestre 2013, le volume de spam social a augmenté de 55% [19].

Pour chaque spam trouvé sur n'importe quel réseau social, les 200 autres spams se trouvent sur Facebook et YouTube. Le problème est devenu si dangereux qu'il a motivé les utilisateursà créer une demande en 2012, lors de la 2e Conférence nationale sur les technologies émergentes, du 6 au 7 décembre 2016, Université d'Asie du Sud, Lahore, Pakistan 2, à laquelle ils demandent à YouTube de fournir des outils pour traiter les contenus indésirables [20].

En 2013, le blog officiel de YouTube fait état des efforts déployés pour traiter les remarques indésirables par la reconnaissance des liens malveillants, la détection d'illustrations ASCII et les modifications

Cependant, de nombreux utilisateurs ne sont toujours pas satisfaits de telles solutions. En effet, en 2014, l'utilisateur « PewDiePie », propriétaire de la chaîne la plus abonnée sur YouTube (près de 40 millions d'abonnés), a désactivé les commentaires sur ses vidéos, affirmant que la plupart des commentaires sont principalement du spam et qu'il n'existe aucun outil pour les traiter [22].

Le problème causé par le spam social a commencé à faire l'objet de discussions critiques à partir de 2010, mais un travail antérieur date de 2005 [23].

Cependant, les commentaires indésirables sur YouTube nuisent toujours à la communauté de la plate-forme ; la preuve d'un tel problème nécessite de l'attention et de la recherche. Les techniques établies pour le filtrage automatique du spam ont leurs performances dégradées lors du traitement des commentaires de YouTube. Cela est principalement dû au fait que ces messages sont généralement très courts et truffés d'idiomes, d'argots, de symboles, d'émoticônes et d'abréviations qui rendent même la tokenisation difficile.

apportées à l'affichage des longs commentaires [21].

2.4.2 Types de spam de commentaires sur Youtube

Nous pouvons classer la majorité des spams de commentaires sur Youtube dans l'un des types suivants. Le tableau 2.1 ci-dessous énumère le type de spam avec des exemples illustratives.

2.4.2.1 Spam basé sur des liens

Il s'agit d'une forme de spam très courante souvent vue sur Youtube. Les commentaires contiennent des liens hypertexte (HTTP) vers d'autres sites Web, généralement d'autres vidéos sur Youtube lui-même. De nombreux liens redirigent l'utilisateur vers des pages Web potentiellement malveillantes souvent à l'insu de l'utilisateur.

2.4.2.2 Spam promotionnel de la chaîne

C'est la forme de spam la plus répandue sur Youtube. Ces types de commentaires sont généralement composés d'utilisateurs qui tentent promouvoir leur propre chaîne en demandant des abonnés, en publiant des liens vers leurs vidéos, etc.

La nature	Exemple
basé sur des liens	Magnifique assortiment de bagues https ://www.ebay.com/b/Fashion-Jewelry/10968/bn_2408529 Le meilleur cadeau de la Saint-Valentin, je suis sûr qu'il vous impressionnera
Spam promotionnel	Je fais des vidéos drôles, abonnez-vous à ma chaîne

Tableau 1 Types de spam sur Youtube.

2.1 Conclusion

Il est très important de Développer des techniques de filtrage du spam dans les commentaires YouTube pour prot

Chapitre 3
approche propose

3.1 Introduction

Dans ce chapitre,nous introduisons d'abord la définition du ensemble de données utilisé avec la description, puis nous présentons le modèle utilisé n_gram avec les algorithmes de classification (naive bayes , random forest , svm) Pour évaluer les performances de ces modèles, nous avons utilisé F1-score, précision, recall. Enfin, nous présentons les résultats obtenus en comparant les algorithmes, et nous terminerons par une conclusion

3.2 processus de detection de spam

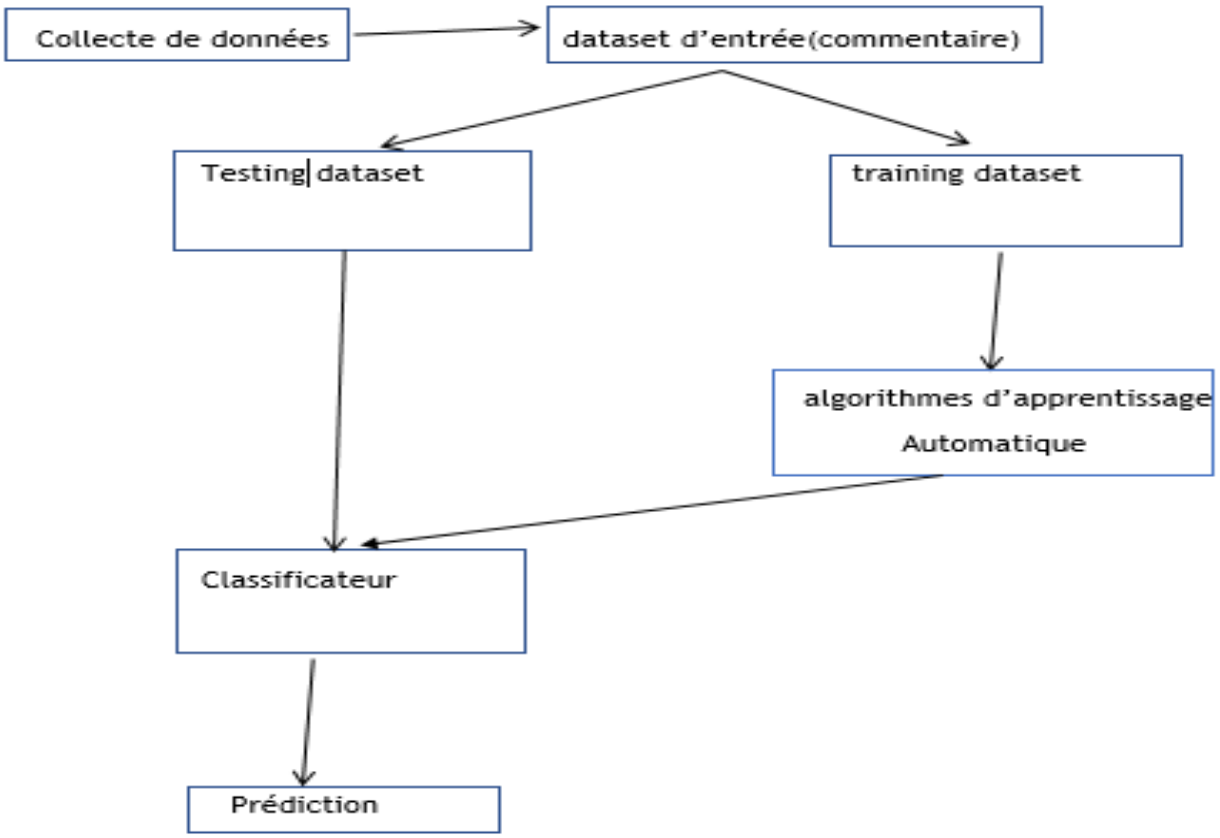

Figure 5 Le processus détaillé de filtrage du spam

3.2.1 Collection de données

Nous utilisons des ensembles de données ouverts.Ils se composent de données de commentaires sur cinq vidéos musicales populaires. Ils contiennent l'identifiant YouTube, l'auteur du commentaire, la date, le contenu du commentaire et la classe étiquetée (0 : Ham ou 1 : Spam). Nous utilisons uniquement le contenu des commentaires et la classe étiquetée. Chaque formation et test des cinq ensembles de données, comme illustré à la figure, peut entraîner un surajustement, où les cinq classificateurs fonctionnent bien uniquement sur ces données et ne s'appliquent pas bien aux données de commentaire dans d'autres vidéos.

3.2.2 Des informations sur Dataset

Le tableau ci-dessous représente le dataset, l'ID vidéo YouTube, le nombre d'échantillonsdans chaque classe et le nombre total commentaires par dataset. Ces échantillons ont été extraitsde la section des commentaires de cinq vidéos qui figuraient parmi les 10 vidéos les plus vuessur YouTube pendant la période de collecte.

Dataset	**Spam**	**Ham**	**Total**
Psy	175	175	350
KatyPerry	175	175	350
LMFAO	236	202	438
Eminem	245	203	448
Shakira	174	196	370
Total	1005	978	1983

Tableau 2 Des informations sur Dataset

3.3 Prétraitement des données

3.3.2 N_gramme

3.3.2.1 Définition N_gramme

N-gramme peut être défini comme la séquence continue de n éléments d'un échantillon donné de texte ou de parole. Les éléments peuvent être des lettres, des mots ou des paires de bases selon l'application. Les N-grammes sont généralement collectés à partir d'un corpus de texte ou de parole (un ensemble de données de texte long).

3.3.2.2 Modèle de langage N-gram

Un modèle de langage à N-grammes prédit la probabilité d'un N-gramme donné dans n'importe quelle séquence de mots du langage. Un bon modèle N-gramme peut prédire le mot suivant dans la phrase, c'est-à-dire la valeur de p(w|h)

Unigram (n = 1)

Bigram (n = 2)

Trigram (n = 3).

Exemple : I like to Play Cricket

N	Type de N-gramme	Résultat
1	**Ungram**	**"I", "like", "to", "Play", "Cricket"**
2	**Bigram**	**"I like", "Like to","to play", "play Cricket"**
3	**Trigram**	**"I like to","like to play", "to Play Cricket"**

Tableau 3 – Exemple de N_Gramme

3.4 Classification basé sur l'approntissage automatique

Après l'extraction des caractéristiques, le dataset transformé est introduit dans des techniques de classification : Support Vector Machine(SVM), Naïve Bayes(NB),Random Forest (RF) et N-Gram respectivement.

3.4.2 Support Vector Machine (SVM)

Une Support Vector Machine(SVM) est un classificateur discriminant qui peut être utilisé pour les deux problèmes de classification et de régression. L'objectif de la SVM est d'identifier une séparation optimale hyperplan qui maximise la marge entre les différentes classes des données d'entraînement. Dans en d'autres termes, étant donné les données d'entraînement étiquetées (apprentissage supervisé),

l'algorithme produit un hyperplan optimal qui catégorise de nouveaux exemples pour créer la plus grande distance possible à réduire une limite supérieure. Supports des vecteurs sont simplement les coordonnées des points de données qui sont les plus proches de l'hyperplan séparateur optimal fournissent les informations les plus utiles pour la SVM classification. En outre, une fonction de noyau appropriée est utilisée pour transformer les données en un haute dimension pour utiliser des fonctions discriminantes linéaires.

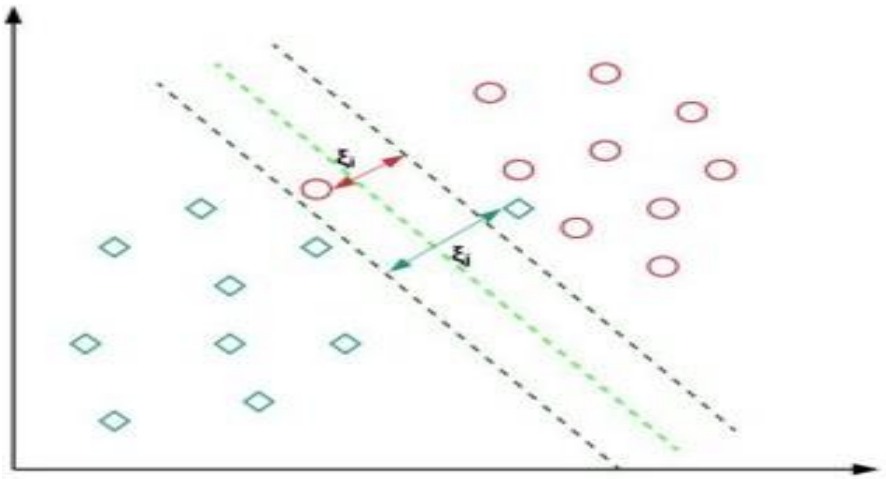

Figure 6 SVM with Kernel Trick

3.4.3 Naïve Bayes (NB)

Il s'agit d'une technique de classification basée sur le réseau bayésien avec une hypothèse de l'indépendance entre les prédicteurs. En d'autres termes, un classificateur Naive Bayes suppose que la présence d'une caractéristique particulière dans une classe n'est pas liée à la présence de toute autre caractéristique.Même si ces caractéristiques dépendent les unes des autres ou de l'existence d'autres caractéristiques, toutes ces propriétés contribuent l'indépendamment à la probabilité. NB est basé sur estimations de probabilité, appelée probabilité a posteriori.

$$P(a|b) = \frac{P(b|c) * P(a)}{P(b)}$$

3.4.4 Random Forest

Random forests es une méthode d'apprentissage d'ensemble pour classification, régression et autres tâches, qui opèrent en construisant une multitude de décisions arborescences au moment de l'apprentissage et sortie de la classe qui est le mode des classes (classification) ou prédiction moyenne (régression) des arbres individuels.

3.5 Les outils et les bibliothèques utilisés

3.5.2 Langage python

Python est un langage de programmation open source et multiplateforme, qui est devenu de plus en plus populaire au cours des dix dernières années. Il a été publié pour la première fois en 1991. La dernière version est la 3.7.0.CPython est l'implémentation de référence du langage de programmation Python.
Écrit en C, CPython est l'implémentation par défaut et la plus largement utilisée du langage. Python est un langage de programmation polyvalent (du fait de ses extensions nom-breuses),les exemples sont le calcul et les calculs scientifiques, les simulations, le développement Web (en utilisant, par exemple, le framework Web Django), etc.

3.5.3 Google colaboratory

Google Colab a été développé par Google pour fournir un accès gratuit aux GPU et aux TPU à quiconque en a besoin pour créer un modèle d'apprentis-sage automatique ou d'apprentissage en profondeur. Google Colab peut être défini comme une version améliorée de Jupyter Notebook. Comme tout autre produit de Google, vous pouvez accéder à Google Colab en vous connectant via votre compte Google.

3.5.4 NumPy

NumPy (Numerical Python) est une bibliothèque Python open source utilisée dans presque tous les domaines de la science et de l'ingénierie.

La bibliothèque NumPy contient des tableaux multidimensionnels et des structures de données matricielles. Il fournit ndarray, un objet de tableau homogène à n dimensions, avec des méthodes pour fonctionner efficacement dessus. NumPy peut être utilisé pour effectuer une grande variété d'opérations mathématiques sur des tableaux. Il ajoute de puissantes structures de données à Python qui garantissent des calculs efficaces avec des tableaux et des matrices et fournit une énorme bibliothèque de fonctions mathématiques de haut niveau qui fonctionnent sur ces tableaux et matrices.

3.5.5 Pandas

Pandas est une librairie python qui permet de manipuler facilement des données à analyser : manipuler des tableaux de données avec des étiquettes de variables (colonnes) et d'individus (lignes), on peut facilement lire et écrire ces dataframes à partir ou vers un fichier tabulé. on peut encore facilement tracer des graphes à partir de ces DataFrames grâce à matplotlib.

3.5.6 Matplotlib

Matplotlib est une bibliothèque Python open source, initialement développée par le neurobiologiste John Hunter en 2002. L'objectif était de visualiser les signaux électriques du cerveau de

personnes épileptiques. Pour y parvenir, il souhaitait répliquer les fonctionnalités de création graphique de MATLAB avec Python. Cette bibliothèque est particulièrement utile pour les personnes travaillant avec Python ou Num-Py. Elle est notamment utilisée sur des serveurs d'application web, des shells et des scripts Py-thon. Avec les APIs de matplotlib, il est aussi possible pour les développeurs d'intégrer des gra-phiques à des applications d'interface graphique.

3.5.7 wordcloud

Il s'agit d'une bibliothèque qui fournit des visualisations de données (telles que des graphiques, des infographies, etc.) L'utilisation d'un wordcloud peut donner vie à des données textuelles floues.

3.5.8 nltk

Le NLTK, ou Natural Language Toolkit, est une suite de bibliothèques logicielles et de programmes. Elle est conçue pour le traitement naturel symbolique et statistique du langage anglais en langage Python. C'est l'une des bibliothèques de traitement naturel du langage les plus puissantes.
Cette suite d'outils rassemble les algorithmes les plus communs du traitement naturel du langage comme le tokenizing, le part-of-speech tagging, le stemming, l'analyse de sentiment, la segmentation de topic ou la reconnaissance d'entité nommée.

3.5.9 sklearn

Scikit-learn (également connu sous le nom de sklearn) est une bibliothèque logicielle gratuite d'apprentissage automatique pour le langage de programmation Python.Il comporte divers algorithmes de classification, de régression et de clustering, notamment des support-vector ma-

chines, des random forests, gradient boosting, k-means et DBSCAN, elle est conçu pour interagir avec les bibliothèques numériques et scientifiques Python NumPy et SciPy.

3.5.10 seaborn

Seaborn est une bibliothèque Python de visualisation de données basée sur matplotlib. Il fournit une interface de haut niveau pour dessiner des graphiques statistiques attrayants et informatifs.

3.5.11 scikitplot

Scikit-Plot est une bibliothèque python qui fournit des visualisations pour de nombreuses métriques d'apprentissage automatique liées à la régression, à la classification et au clustering. Scikit-Plot est construit sur matplotlib.

3.6 Métriques d'évaluation utilisés

Quatre mesures sont utilisées pour l'évaluation : l'accuracy, la précision ,le F1 Score et recall. Chaque formule est basée sur la matrice de confusion.
La matrice de confusion donne la représentation matricielle, pour donner une image claire de la façon dont les classes cibles individuelles sont prédites par le modèle.

Matrice de confusion	Negative=0	Positive=1
Negative=0	True Negative : TN	Faux Negative : FN
Positive=1	Faux Positive :FP	True Positive :TP

Tableau 4 Matrice de confusion,Positive=spam,Negative=ham

1. **L'Accuracy**

   La précision est une mesure permettant d'évaluer les modèles de classification. De manière informelle, la précision est la fraction des prédictions que notre modèle a eu raison.

Formellement, la précision a la définition suivante :

$$Accuracy = \frac{Nombredeprdictionscorrectes}{Nombretotaldeprdictions}$$

2. La précision

La précision d'un modèle décrit combien d'éléments détectés sont vraiment pertinents. Il est calculé en divisant les vrais positifs par les positifs globaux.

Pour calculer la précision, nous utiliserons la formule ci-dessous :

$$Prcision = \frac{TP}{TP + FP}$$

3. Recall

Le rappel est l'une des métriques d'évaluation, en utilisant cette métrique, nous pouvons connaître le nombre de classes positives correctes parmi toutes les classes positives. Parallèlement à ces rappels d'informations, fournissez également des informations sur les classes positives mal classées.

La métrique d'évaluation de rappel peut être définie comme ci-dessous.

$$Recall = \frac{TP}{TP + FN}$$

4. F1 score

Le F1 score est la combinaison du score de précision et du score de recall. Nous pouvons définir le score F1 comme une simple moyenne pondérée de précision et de recall. Nous pouvons définir le F1 score comme ci-dessous (P=précision , R=recall) :

$$F1score = \frac{2 * P * R}{P + R}$$

3.7 Résultats

Nous présentons des ensembles de données de 80% pour les données d'entraînement et de 20% pour les données de test. Après cela, les techniques d'apprentissage automatique et de n_gram sont appliquées, comme indiqué dans le tableau Quatre échelles d'évaluation sont utilisées : acc taux de précision sc taux de détection de spam f1-score Chaque formule sera basée sur le tableau 3.3 .

Méthodes	**Acc**	**Recall**	**F1 score**	**précision**
N-Gram +NB	91,85%	0,98	0,93	0,88
N-Gram +RF	96,18%	0,94	0,96	0,99
N-Gram +KNN	80,15%	0,64	0 ,78	0,99
N-Gram +DT	93,38%	0,93	0,94	0,94
EMS-S	95,06%	0,95	0,95	0,86
N-Gram +SVM-L	**97,20%**	**0,96**	**0,97**	**0,99**

Tableau 5 résultats des méthodes de classification utilisées.

3.7.1 N_Gramme(2,2) :

Méthodes	ACC	F1-score	Recall	Precesion
Svm+n-gram	89%	0.90	0.82	0.99
RF+n-gram	91%	0.91	0.84	0,99
KNN+n-gram	71%	0.63	0.46	1.00
TR +n-gram	89%	0.90	0.85	0.96
NB+n-gram	82%	0.85	0.95	0.77
EMS_S	**95,06%**	**0,95**	**0,95**	**0,99**

Tableau 6 résultats des méthodes de n_gramme(2,2)

3.7.2 N_Gramme(2 ,3)

Méthodes	ACC	F1-score	recall	precesion
Svm+n-gram	**90%**	**0.90**	**0.82**	**0.99**
RF+n-gram	**90%**	**0.90**	**0.82**	**0.99**
KNN+n-gram	70%	0.61	0.44	1.00
TR +n-gram	89%	0.89	0.83	0.96
NB+n-gram	82%	0.85	0.96	0.77
EMS_S	**95,06%**	**0,95**	**0,95**	**0,99**

Tableau 7 résultats des méthodes de n_gramme(2,3)

3.8 Analyse des résultats

Comme indiqué dans le tableau 3.4, nous avons évalué les performances du modèle de N_-Gram avec quatre classificateurs d'apprentissage automatique sur les ensembles de données de différentes catégories de vidéoclips.

1. Les résultats expérimentaux ont montré que le modèle N_Gramme avec SVM Linéaire formait le meilleur résultat en termes de précision, de retour, de score f1 et de précession dans les vidéos par rapport au reste des classifications RF, naive bayes, knn, DT.
2. 2. Le modèle N_Gramme avec SVM Linéaire a également montré la meilleure évaluationpar rapport aux résultats expérimentaux précédents du modèle ESM-S.

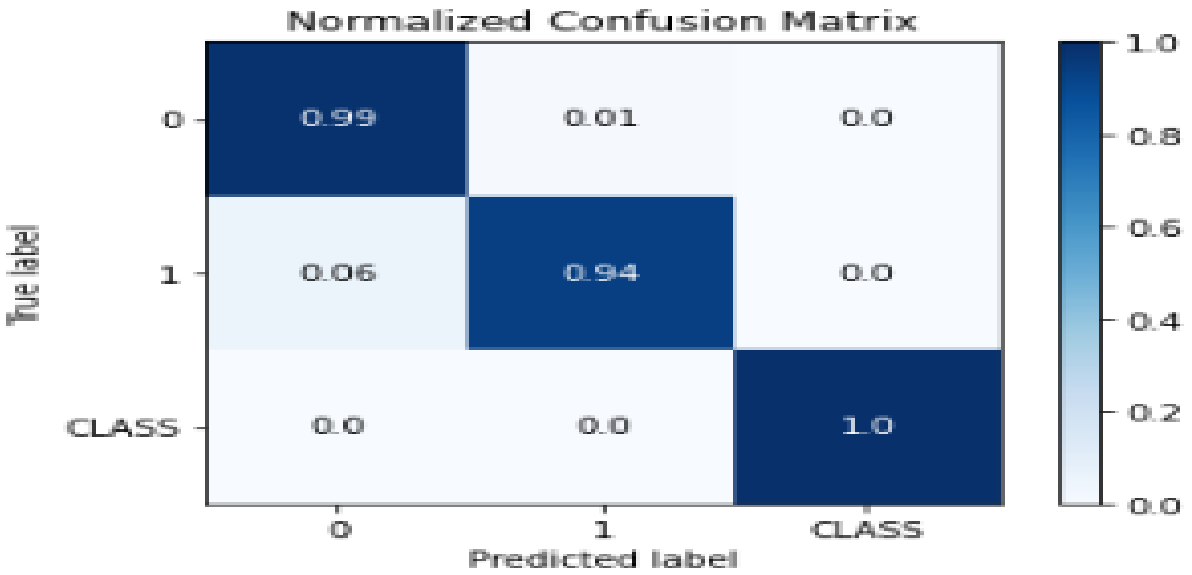

Figure 7 Confusion matrice n_gramme avec random forest

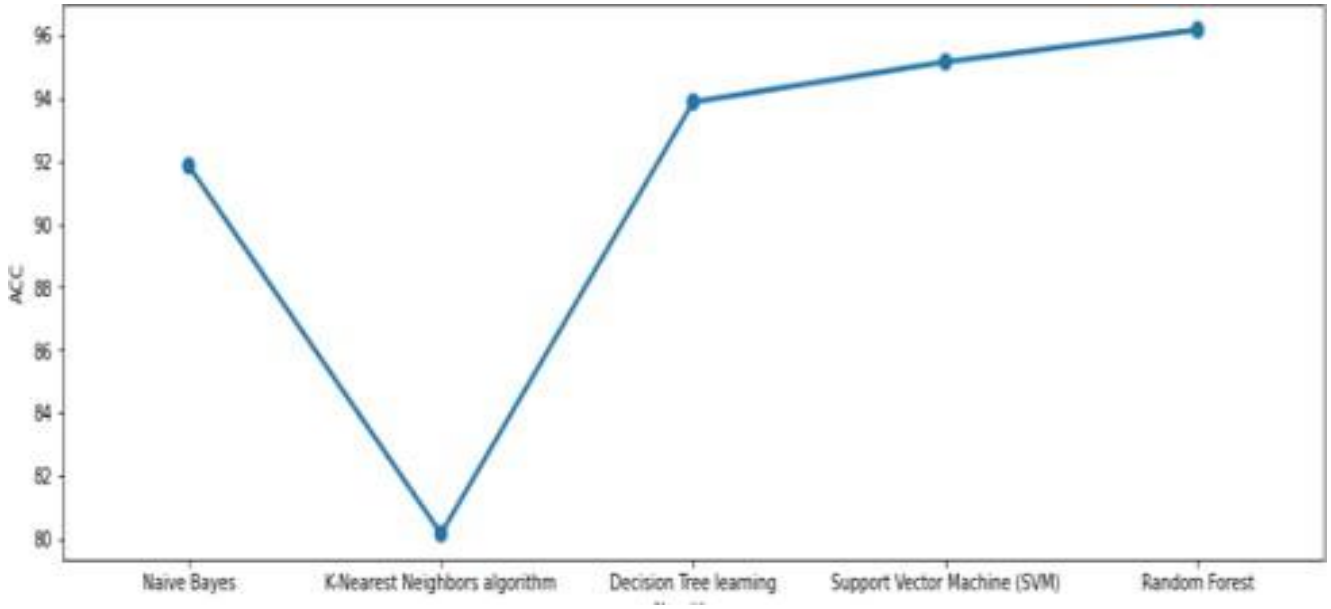

Figure 8 Courbe ACC pour comparer les classification proposés.

3.2 Conclusion

Dans ce chapitre, nous avons présenté la partie pratique de notre projet à travers le modèle n_Gram de détection des commentaires indésirables sur YouTube, qui a récemment connu une croissance fulgurante grâce au modèle d'approntissage automatique

Conclusion générale

Le domaine de la détection des spams a particulièrement progressé ces dernières années, grâce à l'introduction de techniques d'apprentissage automatique grandement améliorées pour le filtrage des spams, grâce aux progrès de la classification en spam et ham

Notre travail a étéorienté vers le développement d'une approche d'apprentissage automatique afin d'améliorer les performances du système de filtrage en utilisant un modèle n-gramme avec plusieurs étiquettesd'apprentissage automatique sur des vidéos de différentes catégories.

Le modèle n-gramme avec SVM-L a montré le meilleur en ACC, F1-SCORE et RECALL ,PRÉSION DANS LESJEUX DE DONNÉES

Dans les recherches futures, les performances devraient être meilleuressi la technologie d'apprentissage en profondeur TF-IDF ou RNN est ajoutée.

Bibliographie

[1] Omar, Normah, et al. "Understanding social network analysis (sna) in fraud detection." *Proceedings of the International Congress on Interdisciplinary Behaviour and Social Sciences*. 2014.

[2] Abbas, Ash Mohammad. "Social network analysis using deep learning: applications and schemes." *Social Network Analysis and Mining* 11.1 (2021): 1-21.

[3] Alkhereyf, Sakhar. *Text Classification: Exploiting the Social Network*. Columbia University, 2021

[4] Belfin, R. V., E. Grace Mary Kanaga, and Suman Kundu. "Application of Machine Learning in the Social Network." *Recent Advances in Hybrid Metaheuristics for Data Clustering* (2020): 61-83.

[5] Scott, John, and Peter J. Carrington. The SAGE handbook of social network analysis. SAGE publications, 2011

[6] Silva, Renato M., Tiago A. Almeida, and Akebo Yamakami. "Artificial neural networks for content-based web spam detection." *Proceedings on the International Conference on Artificial Intelligence (ICAI)*. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2012.

[7] Romero, Christian, Mario Garcia-Valdez, and Arnulfo Alanis. "A comparative study of blog comments spam filtering with machine learning techniques." *Soft Computing for Recognition Based on Biometrics*. Springer, Berlin, Heidelberg, 2010. 57-72.

[8] Henke, Márcia, et al. "Aprendizagem de máquina para segurança em redes de computadores: Métodos e aplicações." *Minicursos do XI Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais (SBSeg 2011)* 1 (2011): 53-103.

[9] Li, Ze, and Haiying Shen. "Soap: A social network aided personalized and effective spam filter to clean your e-mail box." *2011 Proceedings IEEE INFOCOM*. IEEE, 2011.

[10] Almeida, Tiago A., Jurandy Almeida, and Akebo Yamakami. "Spam filtering: how the dimensionality reduction affects the accuracy of Naive Bayes classifiers." *Journal of Internet Services and Applications* 1.3 (2011): 183-200.

[11] Hidalgo, José María Gómez, Tiago A. Almeida, and Akebo Yamakami. "On the validity of a new SMS spam collection." *2012 11th International Conference on Machine Learning and Applications*. Vol. 2. IEEE, 2012.

[12] Silva, Tiago P., et al. "Normalizaç ao Textual e Indexaç ao Semântica Aplicadas na Filtragem de SMS Spam." *Proc. of the 11st ENIAC, Sao Carlos, Brazil* (2014): 1-6.

[13] Kantchelian, Alex, et al. "Robust detection of comment spam using entropy rate." *Proceedings of the 5th ACM Workshop on Security and Artificial Intelligence*. 2012.

[14]Zhu, Tiantian, et al. "Beating the artificial chaos: Fighting OSN spam using its own templates." *IEEE/ACM Transactions on Networking* 24.6 (2016): 3856-3869.

[15]Gao, Hongyu, et al. "Detecting and characterizing social spam campaigns." *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. 2010.

[16]COMM conference on Internet measurementNovember, url = https ://doi.org/, doi = 10.1145/1879141.1879147

[17]Choi, Seungwoo, and Aviv Segev. "Finding informative comments for video viewing." *SN Computer Science* 1.1 (2020): 1-14.

[18]Survey Paper for WARNINGBIRD : Detecting Suspicious URLs in Twitter Stream,volume = 3

[19] Santos, Igor, et al. "Spam filtering through anomaly detection." *International Conference on E-Business and Telecommunications*. Springer, Berlin, Heidelberg, 2011.

[20] Hussain, Ahsan, and Bettahally N. Keshavamurthy. "Analyzing Online Location-Based Social Networks for Malicious User Detection." *Recent Findings in Intelligent Computing Techniques*. Springer, Singapore, 2019. 463-471.

[21] Alberto, Túlio C., Johannes V. Lochter, and Tiago A. Almeida. "Tubespam: Comment spam filtering on youtube." *2015 IEEE 14th international conference on machine learning and applications (ICMLA)*. IEEE, 2015.

[22] Madden, Amy, Ian Ruthven, and David McMenemy. "A classification scheme for content analyses of YouTube video comments." *Journal of documentation* (2013).

[23]Mishne, Gilad, David Carmel, and Ronny Lempel. "Blocking Blog Spam with Language Model Disagreement." AIRWeb. Vol. 5. 2005

[24]Aiyar, Shreyas, and Nisha P. Shetty. "N-gram assisted youtube spam comment detection." *Procedia computer science* 132 (2018): 174-182.

[25]Patil, Rahul C., and D. R. Patil. "Web spam detection using SVM classifier." *2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO)*. IEEE, 2015.