



**FMI** Faculté des  
Mathématiques et  
d'Informatique



République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
Université de Bordj Bou Arreridj Mohammed El Bachir El Ibrahimi  
Faculté des Mathématiques et d'Informatique  
Département d'Informatique

---

# **Mémoire**

En vue de l'obtention du diplôme de Master

Domaine : Mathématiques et Informatique

Filière : Informatique

Option : Ingénierie de l'Informatique Décisionnelle (IID)

## **Intitulé**

**Propagation de la rumeur sur les médias  
sociaux**

**Présenté par :**

- RIGHI Mohamed el Manar
- BOUSSAHEL Djalel Eddine

**Proposé et dirigé par :**

- Djamila MOHDEB

**Devant le jury composé de :**

-

Président  
Examinatrice

Année Universitaire 2020/2021



# *Dédicaces*

*Je remercie tout d'abord le bon dieu de m'avoir donné le courage et la patience  
pour accomplir ce travail*

*Je dédie ce modeste travail :*

*A mes chers parent, tous les mots du monde ne soudaient exprimer l'immense  
amour que je vous porte, ni la profonde gratitude que je vous témoigne pour  
tous les efforts et sacrifices que vous n'avez jamais cessé de consentir pour mon  
instruction et mon bien-être. J'espère avoir répondu aux espoirs que vous avez  
fondé en moi.*

*Que dieu tout puissant vous garde et vous procure santé, bonheur et longue vie  
Inchallah.*

*A moi-même*

*A mes chères sœurs, ma grand-mère. Je vous souhaite une vie pleine de bonheur  
et de succès*

*A tous mes amis inébranlables et merci pour tous les professeurs du  
département d'Informatique.*

**MANAR**

# *Dédicaces*

*Je dédie ce travail à mes parents qui ont fait en sorte que l'amour du savoir soit un véritable crédo pour leurs enfants.*

*A ma sœur Imène et mon frère Houssam qui m'ont encouragé sans cesse et qui ont cru en moi.*

*Je tiens à remercier particulièrement Madame Mohdeb pour ses énormes qualités d'femme de sciences, cela suscite respect et admiration.*

*Je dédie aussi ce travail à mes défunts grands-parents, que Dieu leur accorde son vaste paradis.*

**DJALEL**

# ***Remerciements***

*Nous remercions tout d'abord Dieu le tout puissant de nous avoir donné la volonté et la persévérance pour réaliser ce travail.*

*Nous tenons à saisir cette occasion et adresser nos profonds remerciements à Madame MOHDEB DJAMILA, notre encadrant de mémoire, pour ses précieux conseils et son orientation ficelée et surtout pour sa disponibilité tout au long de notre recherche.*

*Nos vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils ont porté à notre recherche en acceptant d'examiner notre travail et de l'enrichir par leurs propositions.*

*Nous souhaitons adresser nos remerciements les plus sincères aux personnes qui nous ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire ainsi qu'à la réussite de cette formidable année universitaire.*

*Nous remercions tous nos enseignants pour leur dévouement, leur patience et leur contribution à notre formation.*

*Enfin, nous adressons nos plus sincères remerciements à tous nos proches et amis, qui nous ont toujours encouragés au cours de la réalisation de ce mémoire.*

*Merci à tous et à toutes.*

# ملخص

تعتبر منصات وسائل التواصل الاجتماعي مصادر إخبارية واسعة الانتشار تقدم تحديثات بشكل أسرع مقارنة بالوسائط التقليدية. ومع ذلك، فإن عدم وجود آليات للتحقق من المعلومات المنشورة يؤدي إلى انتشار الإشاعات التي يمكن أن يكون لها عواقب سلبية على المجتمع.

يعالج هذا المشروع بشكل خاص هذه المشكلة من خلال دراسة حالة واقعية ميزت الحياة السياسية الجزائرية لبضعة أشهر بين نهاية عام 2020 وبداية عام 2021. ويتعلق الأمر بالإشاعات المنتشرة حول صحة رئيس الجمهورية الجزائرية عبد المجيد تبون التي تم تداولها عبر شبكات التواصل الاجتماعي ونذكر منها يوتوب.

مساهمتنا هي نهج أكاديمي ذو هدفين: نمذجة انتشار هذه الإشاعات باستخدام النماذج الرياضية وتصنيف مواقف الجزائريين تجاهها باتباع نهج التعلم الانتقالي. تؤكد النتائج المتحصل عليها قدرة النموذج الوبائي SIR على تتبع عملية انتشار الإشاعات والأداء الجيد للمحولات أحادية اللغة مقارنة بالمحولات متعددة اللغات في مهمة تصنيف المواقف.

**كلمات مفتاحية:** المعلومات الخاطئة، انتشار الإشاعات، نماذج الأوبئة، التعلم الآلي، نقل التعلم، تصنيف المواقف، تصنيف النص، المعالجة الآلية للغة الطبيعية.

# Résumé

---

Les plateformes des médias sociaux sont omniprésentes en tant que sources d'informations fournissant souvent des mises à jour plus rapides que les médias traditionnels. Néanmoins, l'absence des mécanismes de vérification pour ces informations publiées conduit à une propagation rapide des rumeurs qui peuvent avoir des conséquences négatives de grande envergure sur la société,

Ce projet s'attaque justement à cette problématique en examinant un cas réel qui a marqué la vie politique algérienne pendant quelques mois entre la fin de l'an 2020 et le début de l'an 2021. Il s'agit des rumeurs qui ont circulé sur la plateforme sociale YouTube et les autres réseaux sociaux à propos de la santé du président de la République algérienne Abdelmadjid Tebboune.

Notre contribution propose une démarche académique pour deux objectifs : modéliser la propagation de ces rumeurs à l'aide des modèles mathématiques et classer les attitudes des algériens à son égard en suivant une approche d'apprentissage par transfert. Les résultats obtenus confirment la capacité du modèle épidémique SIR à retracer le processus de diffusion des rumeurs et la bonne performance des transformateurs monolingues par rapport aux transformateurs multilingues dans la tâche de classification des attitudes.

**Mots-clés** : fausses informations, rumeurs, propagation des rumeurs, modèles épidémiques, apprentissage par transfert, classification des attitudes, classification de texte, TALN.

# Abstract

---

Social media platforms are ubiquitous as news sources that often provide updates faster than traditional media. Nevertheless, the absence of verification mechanisms for posted information leads to a rapid spread of rumors which can have negative consequences on the society.

This project particularly tackles this problem by examining a real case that marked the Algerian political life for a few months between the end of 2020 and the beginning of 2021. It concerns the rumors that have circulated on the social platform YouTube and other social networks about the health of the President of the Algerian Republic Abdelmadjid Tebboune.

Our contribution is an academic approach that aims to two objectives: to model the propagation of these rumors using mathematical models and to classify the stances of Algerians towards it by following a transfer learning approach. The obtained results confirm the ability of the SIR epidemic model to trace the rumor spreading process and the good performance of monolingual transformers compared to multilingual transformers in the task of stance classification.

**Keywords:** false information, rumor propagation, epidemic models, machine learning, transfer learning, stance classification, text classification, NLP.

# Table des Matières

---

Liste des figures

Liste des tableaux

Liste des algorithmes

**Introduction générale**

**Chapitre 1 : La rumeur : Concepts de base**

Introduction .....	4
1. Définition.....	4
2. Les types des rumeurs .....	5
3. Caractéristiques de la rumeur .....	5
4. Diffuseurs des rumeurs en ligne .....	6
5. Dynamique et cycle de vie d'une rumeur sur les médias sociaux .....	6
6. Motivations de diffusion des rumeurs .....	7
7. Les rumeurs les plus répandues sur le Vaccin contre Covid-19.....	8
Conclusion.....	9

**Chapitre 2 : Propagation des rumeurs en ligne**

Introduction .....	11
<b>I. Diffusion de l'information et de la rumeur .....</b>	<b>11</b>
1. Les modèles épidémiques.....	11
1.1. Le modèle SI .....	11
1.2. Le modèle SIR.....	12
1.3. Le modèle SIS .....	13
1.4. Le modèle SIRS.....	14
2. Les modèles mathématiques.....	15
2.1. Le modèle à seuil linéaire (Linear Threshold Model) .....	15
2.2. Le modèle en cascade indépendant (Independent Cascade Model) .....	16
2.3. Le modèle stochastique (Stochastic model) .....	16
<b>II. Classification des attitudes (Stance Classification) .....</b>	<b>17</b>
1. Définition de la classification des attitudes envers les rumeurs (Rumor Stance Classification) ...	17
2. Approches de classification des attitudes .....	18
2.1. Approche d'apprentissage automatique.....	18
2.2. Approche d'apprentissage approfondi.....	18
2.3. Approche d'apprentissage ensemblistes.....	18
2.4. Approche d'apprentissage par transfert.....	19

Conclusion.....	19
<b>Chapitre 3 : Propagation de la rumeur sur les réseaux sociaux : étude de cas</b>	
Introduction .....	21
1. Description du projet .....	21
2. Description du cas d'étude .....	23
3. Préparation des données d'étude .....	23
3.1. Collection et nettoyage de données .....	23
3.2. Annotation de données .....	27
3.3. Prétraitement de données.....	27
4. Extraction et généralisation du modèle de propagation des rumeurs .....	28
5. Apprentissage par transfert pour la classification du texte.....	28
5.1. Définition.....	28
5.2. Les transformateurs .....	29
6. Apprentissage par transfert pour le traitement automatique du langage naturel (NLP) .....	30
6.1. Modèles de base .....	30
6.2. Modèles multilingues .....	30
6.3. Modèles monolingues.....	31
Conclusion.....	31
<b>Chapitre 04 : Implémentation &amp; Résultats</b>	
Introduction .....	33
1. Environnement de travail et outils d'implémentation .....	33
1.1. Matériel .....	33
1.2. Langage de programmation Python.....	33
1.3. Environnement de programmation .....	33
1.4. Principaux packages Python utilisés .....	34
2. Analyse exploratoire de données.....	35
2.1. Informations générales .....	35
2.2. Distribution générale de classes .....	35
2.4. Fréquence des termes (N-Gramme).....	37
2.5. Style linguistique du texte .....	38
2.6. Perspective temporelle.....	39
3. Propagation des rumeurs basée sur le modèle SIR.....	40
3.1. Variables et paramétrage .....	40
3.2. Modèle de propagation des rumeurs basé sur $\beta$ et $\gamma$ personnalisées.....	41
3.3. Modèle de propagation des rumeurs basé sur les valeurs déduites $\beta$ et $\gamma$ .....	43
3.4. Discussion du modèle de propagation résultat .....	43

4.	Classification des attitudes envers les rumeurs .....	44
4.1.	Annotation de données .....	44
4.2.	Nettoyage et prétraitement .....	44
4.3.	Implémentation des transformateurs .....	45
4.4.	Discussion des résultats .....	47
	<b>Conclusion générale .....</b>	<b>48</b>
	<b>Bibliographie</b>	

# Liste des figures

---

<b>Figure 1.1 :</b> Cycle de vie d'une rumeur sur les médias sociaux.....	7
<b>Figure 2.1 :</b> Le modèle SI.....	12
<b>Figure 2.2 :</b> Le modèle SIR .....	13
<b>Figure 2.3 :</b> Le modèle SIS .....	13
<b>Figure 2.4 :</b> Le Modèle SIRS .....	14
<b>Figure 2.5 :</b> Le modèle stochastique.....	16
<b>Figure 2.6 :</b> Représentation schématique de la procédure de détection des positions .....	17
<b>Figure 3.1:</b> Architecture de système proposé.....	22
<b>Figure 3.2 :</b> Représentation de l'architecture d'un Transformateur .....	30
<b>Figure 4.1 :</b> Jupyter et Google Colaboratory .....	35
<b>Figure 4.2 :</b> Distribution de classes de l'ensemble de données « Rumeurs_Commentaires »	37
<b>Figure 4.3 :</b> Les mots les plus fréquents dans l'ensemble de données « Rumeurs_Commentaires » .....	38
<b>Figure 4.4 :</b> Les bi-grammes et les tri-grammes les plus fréquents dans l'ensemble de données « Rumeurs_Commentaires » .....	39
<b>Figure 4.5 :</b> Distribution de la longueur des textes .....	40
<b>Figure 4.6 :</b> Densité des éléments textuelles des commentaires.....	40
<b>Figure 4.7 :</b> Distribution des commentaires par mois .....	41
<b>Figure 4.8 :</b> Courbe de propagation de la rumeur selon le modèle SIR avec $\beta$ et $\gamma$ personnalisés et les périodes d'infection a, b, et c .....	44
<b>Figure 4.9 :</b> Courbe de propagation de la rumeur selon le modèle SIR avec $\beta$ et $\gamma$ déduites .	45

# Liste des tableaux

---

<b>Tableau 3.1:</b> Les rumeurs analysées .....	24
<b>Tableau 3.2:</b> Les chaines YouTube choisies .....	25
<b>Tableau 3.3:</b> Les vidéos YouTube choisies .....	27
<b>Tableau 4.1:</b> Caractéristiques du matériel utilisé .....	34
<b>Tableau 4.2:</b> Caractéristiques de l'ensemble de données « Rumeurs_Commentaires » .....	36
<b>Tableau 4.3:</b> Distribution de classes de l'ensemble de données « Rumeurs_Commentaires » .....	36
<b>Tableau 4.4:</b> Distribution de classes selon le diffuseur de la rumeur .....	38
<b>Tableau 4.5:</b> Quelques exemples des commentaires annotés .....	46
<b>Tableau 4.6:</b> Exemples des commentaires après la segmentation .....	46
<b>Tableau 4.7:</b> Exemples des commentaires après la suppression des mots vides .....	46
<b>Tableau 4.8:</b> Exemples des commentaires après le nettoyage .....	46
<b>Tableau 4.9:</b> Les résultats obtenus lors de l'exécution des transformateurs .....	48

# Liste des algorithmes

---

<b>Algorithme 4.1</b> : Algorithme de classification des attitudes .....	47
--	----

# Introduction générale

---

### 1. Contexte

Les plates-formes et les réseaux sociaux ont réussi petit à petit à remplacer les médias de masse traditionnels tels que les journaux, la télévision et le radio. Ces nouvelles technologies à grande communauté nous offrent l'information en temps réel sans effort et sans coût, mais au détriment de sa crédibilité, véracité et certitude, ce qui peut causer des crises politiques, économiques, ou culturelles entre individus et même au sein des communautés et des sociétés. Parmi les problèmes les plus répandus qui sont liés au sujet de l'expansion des informations non vérifiées sur le Web, on souligne le phénomène de propagation des rumeurs sur les médias sociaux populaires tels que Facebook, Twitter et YouTube.

### 2. Problématique

Les rumeurs, qui peuvent se représenter comme des informations dont on n'est pas sûre, est devenu un vrai sujet d'attention au cours de ces dernières années. Leur propagation à grande échelle et à grande vitesse sur le Web et surtout sur les réseaux sociaux, multipliée par leur capacité à toucher un grand public dans des périodes de temps très réduites, a impliqué un intérêt pour l'étude de ses processus de diffusion et d'évolution de la part de la communauté scientifique en informatique mais aussi les chercheurs dans d'autres disciplines à l'image de la sociologie, la physique, les mathématiques, la psychologie, ...etc.

### 3. Objectif et contribution

L'analyse et la classification des attitudes des individus à l'égard des rumeurs est l'une des approches les plus utilisées pour étudier la prévalence des rumeurs auprès d'un public cible, voire leur niveau de certitude et de véracité.

Dans ce projet, nous nous sommes intéressés par l'étude du processus de diffusion d'un ensemble de rumeurs qui s'étaient déclenchées sur le Web et les réseaux sociaux durant la période d'hospitalisation hors pays du président algérien Abdelmadjid Tebboune (d'octobre 2020 à mars 2021) pour des complications liées à sa contamination par le nouveau coronavirus. Notre but était d'explorer le modèle de propagation de ces rumeurs et également les positions prises à son égard par l'audience algérienne sur la plateforme sociale YouTube en particulier. Pour ces deux objectifs, nous avons exploité les modèles épidémiologiques mathématiques pour modéliser et retracer le processus de propagation de ces rumeurs depuis son début jusqu'à sa déclinaison, ainsi que les techniques multilingues et monolingues de l'apprentissage par

transfert pour développer un modèle efficace de détection et de classification des attitudes tout en respectant la méthodologie de la classification du texte liée au domaine du traitement automatique du langage naturel.

### **4. Plan de mémoire**

Notre contribution dans ce projet de fin d'études est détaillée dans ce mémoire. Ce dernier est composé de quatre chapitres qui sont organisés comme suit :

- Le premier chapitre est un tour d'horizon sur les rumeurs, leurs types et caractéristiques ainsi que ses diffuseurs et leurs motivations.
- Le deuxième chapitre présente une revue des travaux académiques qui sont intéressés par la modélisation de la propagation d'information, et spécialement la propagation de la rumeur.
- Le troisième chapitre explique la démarche de notre propre contribution qui vise la modélisation de processus de diffusion des rumeurs ainsi que la classification des attitudes des individus envers eux.
- Le dernier chapitre présente les différentes phases suivies lors de déploiement de l'approche proposée afin de mettre en œuvre, valider et interpréter les résultats obtenus par les solutions proposées.
- Enfin, ce mémoire se termine par une conclusion et une vision claire sur quelques caractéristiques techniques et théoriques utilisées lors du déploiement et de l'implémentation de notre projet.

# **Chapitre I**

## **La rumeur : concepts de base**

---

### Introduction

L'internet et les réseaux sociaux ont rapidement et largement modifié la manière dont nous produisons, diffusons et consommons l'information. Néanmoins, cette transformation sociologique et économique compliquée vers le cyberspace a aggravé le phénomène de propagation des fausses nouvelles et informations. En particulier, la problématique de la diffusion des rumeurs est devenue un champ actif de recherche et d'intérêt en raison de son influence inconsciente majeure sur la vie quotidienne moderne.

Le but de ce chapitre est de donner une vision claire sur ce phénomène en présentant dans un premier temps, tout ce qui concerne les rumeurs, ses catégories, ses promoteurs, la façon dont elles se propagent, et enfin, des exemples d'actualité qui retracent son impact considérable sur la société d'aujourd'hui.

### 1. Définition

La rumeur est une information incertaine non vérifiée circulant de personne à personne et portant sur un objet, un événement ou une question d'intérêt public [1]. Elle est destinée être crue, se rapportant à l'actualité, transmise et répandue par un vecteur formel ou informel, humain et technique (individus, organisations, presse, Web, réseaux et média sociaux...etc.) sans vérification officielle [2].

Les rumeurs, qui ne sont pas nécessairement fausses, « naissent des questions spontanées que se pose le public et auxquelles on ne fournit pas de réponse. Elles satisfont le besoin de comprendre l'événement si celui-ci n'est pas clair » [3]. Par conséquence, elles apparaissent souvent d'une source non déterminée dans une situation de crise, transmises dans un groupe social ou dans une institution pour agir sur les comportements, les croyances et les opinions d'un individu.

La rumeur est très nocive quand elle propage de fausses nouvelles pour manipuler les gens, voire des pays entiers (la rumeur est utilisée comme une arme entre les nations hostiles). La rumeur fait naître la peur au sein de la population (rumeur d'une épidémie, d'une catastrophe, etc.) et peut servir aussi à influencer l'opinion, à promouvoir un produit de manière frauduleuse, à détruire un concurrent (en économie, en politique ou au milieu professionnel) ...etc.

### 2. Les types des rumeurs

Autre que la manipulation du public à des fins politiques ou commerciales avec des informations intentionnellement déformées, les rumeurs peuvent également servir à d'autres

objectifs et être de différentes formes. Certaines d'entre elles visent à fournir de l'information alors que d'autres sont principalement motivées par l'expression d'une émotion. Ce qui nous amène à distinguer trois types de rumeurs selon la motivation du transmetteur [4] :

### 2.1. Les rumeurs de souhait (wish rumours)

Ce sont des rumeurs qui peuvent être répandues par une seule personne ou un groupe de personnes pour l'intérêt personnel. Ces rumeurs sont moins dangereuses parce qu'elles diffusent des récits qui aident à rassurer un public en supportant ses aspirations et ses attentes [2].

### 2.2. Les rumeurs de peur (the Bogie rumours)

Ce sont des rumeurs qui visent à augmenter la peur et l'anxiété d'un auditoire. Elles expriment les peurs et les angoisses d'un groupe en donnant une illusion de savoir et de contrôler le danger.

### 2.3. Les rumeurs agressives ou hostiles (wedge-driving aggression rumours)

Des rumeurs qui sont généralement motivées par un désir de division et génèrent de la haine. Elles sont principalement tournées contre un groupe pour semer la discorde entre ses membres et détruire les loyautés.

## 3. Les caractéristiques de la rumeur

Les rumeurs peuvent être perçues comme un phénomène collectif se transmet oralement ou par les médias. Bien qu'elles portent un caractère de non-vérification de l'information et proviennent des sources non officielles, elles sont censées être informatifs qu'elles soient vraies ou fausses.

Les différents types de rumeurs qui circulent *off-line* ou bien *on-line* sur le web et les réseaux sociaux ont plusieurs propriétés communes, on cite notamment :

- La source reste cachée, mystérieuse aucun responsable.
- Elle échappe à tout contrôle.
- Son résultat est aléatoire.
- Sa propagation est rapide.
- Elle peut même se retourner contre ses émetteurs.

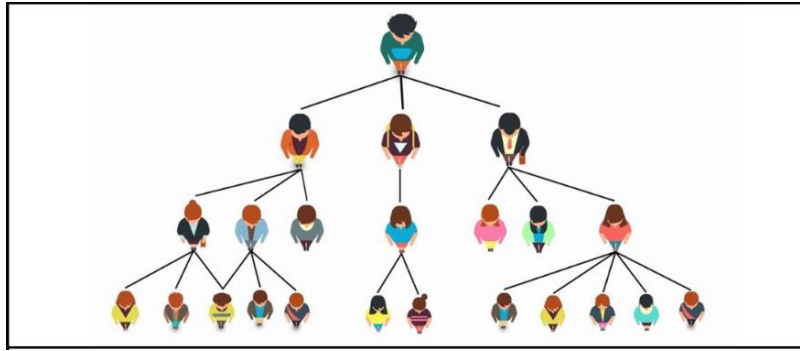
### 4. Les diffuseurs des rumeurs en ligne

Différentes catégories des utilisateurs des médias sociaux contribuent à la manipulation des communautés en ligne par la diffusion intentionnelle ou accidentelle des rumeurs. On peut citer entre autres [5] :

- **Les journalistes :** le journalisme joue un rôle très important dans ce domaine, principalement à cause de son impact lourd sur la société. Ce qui entraîne que les journalistes soient les premiers responsables de diffusion des rumeurs sur les médias et les réseaux sociaux pour des fins d'attraction du public et augmentation de leur statut et popularité au sein de la société.
- **Les gouvernements :** les politiciens utilisent les médias et les réseaux sociaux principalement en période électorale, afin de diffuser des rumeurs ou des informations mensongères dans le but de manipuler l'opinion publique et faire pencher la balance électorale en leur faveur.
- **Les bots :** ensemble de programmes malveillants reliés à des faux comptes automatisés qui interagissent sur les réseaux sociaux en tant que personnes réelles pour participer à la publication et la diffusion des rumeurs dans le but de manipuler les gens.
- **Les « Trolls » :** sont les utilisateurs qui partagent les rumeurs uniquement pour l'amusement personnel. Ils s'amusent à semer la confusion et la discorde dans les réseaux sociaux, en provoquant d'autres utilisateurs et créant le conflit à travers des publications et des commentaires agressifs et haineux.
- **Les idiots utiles :** les idiots utiles sont les internautes ordinaires qui peuvent partager enthousiasment une fausse information sans vérifier son contenu ou sa source. Ils sont facilement manipulés par les personnes ou les organisations qui contribuent à la propagation des rumeurs à leur entourage.
- ...etc.

### 5. La dynamique de la rumeur sur les médias sociaux

La rumeur se propage dans les réseaux sociaux sous forme d'une **cascade**. Elle débute toujours par un promoteur ou un premier diffuseur. D'autres personnes qui l'ont réceptionné vont la diffuser à leurs tours en reflétant ainsi l'effet boule de neige dont le volume s'augmente en descendant une montagne.



**Figure 1.1.** Cycle de vie d'une rumeur sur les médias sociaux

Une cascade d'information est une représentation formelle sous forme d'un arbre ou d'une structure arborescente qui capture directement la propagation d'information. Le nœud racine d'une cascade représente l'utilisateur qui a partagé l'information pour la première fois. Les autres nœuds représentent les utilisateurs qui ont ensuite propagé ladite information en la transférant après sa publication par leurs nœuds parents, auxquels ils sont connectés via des arêtes.

Ainsi, suivant la classification adoptée par Zafarani et al. [6], on distingue trois rôles essentiels dans le processus de diffusion de la rumeur :

- a) **Les expéditeurs** lancent le processus de la diffusion de la rumeur.
- b) **Les diffuseurs** transmettent la rumeur à leurs abonnés.
- c) **Les destinataires** (utilisateurs finaux) reçoivent et consomment les rumeurs diffusées.

## 6. Les motivations de la diffusion des rumeurs en ligne

Face au grand flux d'informations partagées sur les réseaux sociaux, les utilisateurs profitent de l'indépendance et de la liberté d'expression pour propager les rumeurs. Les raisons varient selon l'objectif du diffuseur. Parmi les motifs de diffusion en ligne les plus fréquents [14] :

- 6.1. Intention politique :** l'action de dénigrer l'image publique de l'opposant ou de promouvoir une personne ou un groupe.
- 6.2. Bénéfice financier :** les rumeurs déclenchent la motivation d'investissements à grande échelle et affectent les prix des produits et des services. Les rumeurs économiques sont intentionnellement partagées dans le but d'augmenter les profits des entreprises.

**6.3. Passion pour l'influence ou la promotion d'une idéologie :** certains utilisateurs sont en particulier passionnés par une organisation, une idéologie, une personne ou une philosophie. Ils participent ardemment à répandre leurs passions et leurs idées par tous les moyens.

**6.4. Raisons technologiques :** les algorithmes des médias sociaux en ligne sont structurés en fonction de la popularité, non pas de la précision [7]. Les moteurs de recherche [8] font partie des flux algorithmiques. En conséquence, la circulation et la propagation des rumeurs peuvent être intentionnellement générées pour attirer l'attention des utilisateurs.

**6.5. Manipuler l'opinion publique dans une économie de consommation :** l'opinion publique concernant une entreprise, un service, un produit ou une personne est d'une importance significative, car dans une économie de consommation, c'est le client qui décide du sort des actions, des ventes, des résultats des élections, de tous types d'entreprises et bien d'autres.

**6.6. Générer les revenus de la publicité :** les créateurs des rumeurs peuvent réaliser un profit considérable via les moteurs de publicité automatisés tel que AppNexus, Facebook Ads et Google AdSense.

## 7. Exemples des rumeurs les plus répandues en l'an 2020/2021

La crise sanitaire du Covid-19 était accompagnée par une épidémie de fausses informations sur le virus, les traitements, les statistiques, les vaccins...etc. Les réseaux sociaux ont considérablement gagné du terrain comme source principale de l'information non vérifiée et de la rumeur. Parmi les rumeurs les plus répandues sur les vaccins anti-Covid-19 :

- **Les vaccins à "ARN messenger" vont modifier l'ADN humain :** cette rumeur a été partagée dès le lendemain de l'annonce du vaccin Pfizer en novembre 2020. Elle prétend que les vaccins dits à "ARN messenger" vont modifier l'ADN humain et "transformer nos gènes" [9][10].
- **Bill Gates veut implanter ses puces électroniques via des vaccins :** entre février et juillet 2020 cette rumeur été la plus partagée sur les médias sociaux, selon laquelle le vaccin a pour but de surveiller et contrôler la population mondiale [11].
- **Les vaccins contre la Covid-19 provoqueraient des infections au VIH :** Des messages relayés sur Twitter puis Facebook affirment, à tort, que des vaccins anti-Covid provoqueraient des infections au VIH [12]. Le site médicale *NEWS MEDICAL Science*

## Chapitre 01 : La rumeur - concepts de base

a reporté après que le vaccin anti-Covid qui est de famille adénovirus est généralement sûr et ne provoque que très peu d'effets secondaires [13].

### Conclusion

De nos jours, une partie considérable des informations qu'on rencontre sur le web et les réseaux sociaux sont des fausses nouvelles ou des rumeurs. Dans ce chapitre, nous avons présenté des notions liées au concept de la rumeur. Nous avons illustré la différence entre ses divers types et la nature de ses diffuseurs. Une attention particulière a été portée sur la dynamique de la dissémination en ligne de la rumeur et les motivations de ses diffuseurs. Le chapitre suivant porte plus de détails sur les modèles de propagation correspondants à notre objet de recherche et leurs contextes d'application.

## **Chapitre II**

# **Propagation des rumeurs en ligne**

---

### Introduction

Avec cette formidable croissance de l'usage des réseaux sociaux pour le partage d'informations et de ressources, l'efficacité et la rapidité du processus lié à la circulation et la diffusion de ces derniers ont été également avancées. Le problème de la dissémination de la rumeur en ligne qui peut avoir un impact majeur avec des conséquences indésirables sur une cyber-audience visée a été largement étudié par la communauté académique et scientifique partout dans le monde.

L'objet de ce chapitre est d'exposer l'état de l'art des modèles développés pour la poursuite et le retracement de la propagation des rumeurs sur le Web et les réseaux sociaux.

### I. La diffusion de l'information et de la rumeur

Les modèles classiques et nouvelles de propagation de l'information servent également à identifier les processus de la transmission et la réception de la rumeur entre les utilisateurs dans un réseau social en-ligne ou hors-ligne.

Une rumeur est un élément informatif qui exhibe les mêmes propriétés de dissémination que l'information ordinaire. Cependant, il faut noter qu'elle a tendance à se propager d'autant plus rapidement que chaque individu est amené à rencontrer un certain nombre d'individus liés à son cercle social.

Il existe différents modèles conçus pour cerner la diffusion de l'information et de la rumeur au sein d'un groupe social. Les plus connus sont les suivants [14] :

#### 1. Les modèles épidémiques

Les modèles épidémiques décrivent le processus de diffusion de l'information dans un réseau comme étant une propagation de maladie infectieuse dans une communauté.

##### 1.1. Le modèle SI

Le modèle **SI** (*Susceptible, Infecté*) est le plus simple modèle dans la catégorie des modèles épidémiques. Dans sa représentation, les nœuds d'un réseau sont généralement arrangés en deux classes :

- i. « **S** » : décrit la susceptibilité d'un individu à être infecté ;
- ii. « **I** » : décrit l'individu infecté, qui est dans un état actif pour infecter les autres individus ;

## Chapitre 02 : Propagation des rumeurs en ligne

Le seul évènement pouvant survenir ici est l'infection d'un individu 'susceptible'. C'est-à-d, le passage de l'état « S » (susceptible à être infecté) à l'état « I » (infecté).

Dans le contexte de la diffusion de l'information sur un réseau social, les nœuds '**infectés**' peuvent être considérés comme ceux qui ont déjà été informés de la rumeur, et sont prêts à la transmettre à leurs voisins. Les nœuds '**susceptibles**' sont les utilisateurs qui ne sont pas informés mais qui peuvent être informés par d'autres.

On considère que la probabilité qu'un individu 'susceptible' devienne 'infecté' est proportionnelle au nombre d'individus actuellement infectés et que le coefficient de proportionnalité soit  $\beta \geq 0$ .

Etant donné un nombre précis d'individus susceptibles S, on peut s'attendre à ce que  $(\beta \cdot I \cdot S)$  d'entre eux seront infectieux chaque jour. Sous forme d'équations différentielles, on obtient le système suivant [15] :

$$\frac{dS}{dt} = -\beta \cdot I \cdot S$$

$$\frac{dI}{dt} = \beta \cdot I \cdot S$$

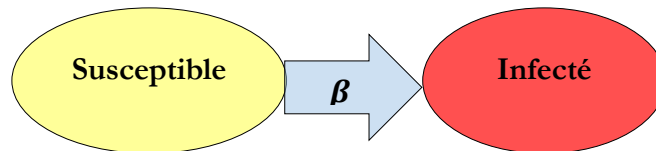


Figure 2.1. Le modèle SI

### 1.2. Le modèle SIR

Le modèle SIR (*Susceptible*, *Infecté*, *Récupéré*) est également l'un des modèles de diffusion épidémique les plus courants. C'est une variation du modèle SI (*Susceptible*, *Infecté*) auquel est ajouté le troisième état **R** (*Recovered*). Un individu « récupéré » est celui qui s'est rétabli et est vacciné contre la maladie ; c-à-d, le nœud qui est sorti d'un état d'infection et ne peut pas être de nouveau susceptible ou infecté.

Dans notre contexte, les nœuds récupérés sont ceux qui ont développé une immunité contre la rumeur ou l'information en circulation. Ils peuvent être informés par la rumeur mais ils n'ont pas intérêt à transmettre l'information aux voisins.



**Figure 2.2.** Le modèle SIR

Le modèle SIR est composé principalement de trois compartiments comme suivant :

- $S(t)$  Nombre d'individus susceptibles ;
- $I(t)$  Nombre d'individus infectés ;
- $R(t)$  Nombre d'individus récupérés.

Les individus infectés ont un paramètre global  $\gamma$  qui représente le taux de guérison ou la probabilité que des nœuds infectés soient récupérés.

Avec ces trois états différents nous pouvons associer les trois équations suivantes [15] :

$$\frac{dS}{dt} = S'(t) = -\beta \cdot I(t) \cdot S(t)$$

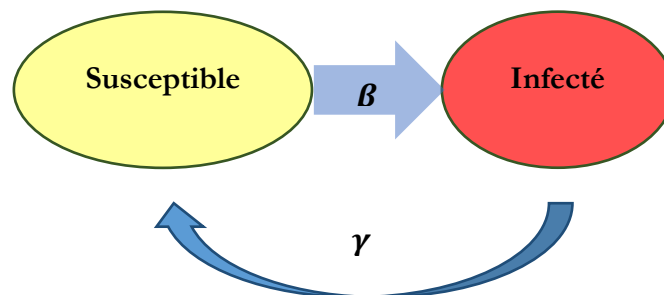
$$\frac{dI}{dt} = I'(t) = \beta \cdot I(t) \cdot S(t) - \gamma \cdot I(t)$$

$$\frac{dR}{dt} = R'(t) = \gamma \cdot I(t)$$

Avec  $N$  : la population totale,  $N = S(t) + I(t) + R(t)$  et  $S, I > 0$

### 1.3. Le modèle SIS

Le modèle SIS (*Susceptible, Infecté, Susceptible*) est une autre variation du modèle SI (*Susceptible, Infecté*) avec la possibilité de la réinfection des individus ; c-à-d au bout d'un certain temps, les individus 'infectés' peuvent redevenir 'susceptibles'. C'est un modèle qui ne se procure pas d'immunité contre la maladie.



**Figure 2.3.** Le modèle SIS

## Chapitre 02 : Propagation des rumeurs en ligne

Le système d'équations devient donc [15] :

$$\frac{dS}{dt} = -\beta \cdot I \cdot S + \beta \cdot I$$

$$\frac{dI}{dt} = \beta \cdot I \cdot S - \gamma \cdot I$$

### 1.4. Le modèle SIRS

Le modèle SIRS (*Susceptible, Infecté, Récupéré, Susceptible*) combine les deux modèles SIS et SIR en autorisant les individus récupérés de perdre leur immunité à un taux  $\Gamma$  pour retourner dans le compartiment des susceptibles  $S$ .

Le système d'équations devient donc [15] :

$$\frac{dS}{dt} = -\beta \cdot I \cdot S + \Gamma \cdot R$$

$$\frac{dI}{dt} = \beta \cdot I \cdot S - \gamma \cdot I$$

$$\frac{dR}{dt} = \gamma \cdot I - \Gamma \cdot R$$

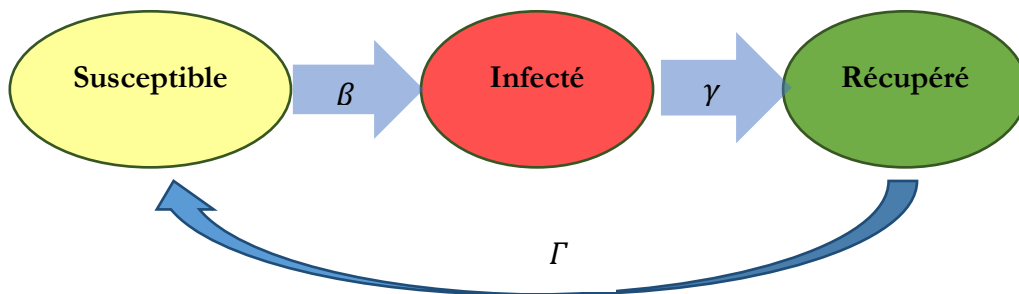


Figure 2.4. Le Modèle SIRS

## 2. Les modèles mathématiques

Les modèles de diffusion mathématiques représentent le réseau social par un graphe  $G = (V, E)$  ; avec  $V$  (les nœuds) est l'ensemble des individus constituant le réseau et  $E$  (les arêtes) est l'ensemble de tous les liens qui existent entre eux.

Dans ce cas, étudier le processus de diffusion d'information consiste à retracer le chemin dont la rumeur suit depuis sa source promotrice aux nœuds du niveau inférieur.

### 2.1. Le modèle à seuil linéaire (Linear Threshold Model)

Dans ce modèle, le risque pour un nœud  $v$  d'être infecté par la rumeur d'un voisin  $w$  est déterminée en fonction du poids  $b_{v,w}$  qui satisfait l'équation (1) pour tout  $w$  dans le voisinage  $N(v)$  de  $v$  [14] :

$$\sum_{w \in N(v)} b_{v,w} \leq 1 \quad (1)$$

Ce modèle [16], utilise un seuil  $\theta_v \in [0,1]$  pour déterminer à partir de quel instant un nœud est infecté par une rumeur. Dans un premier temps, chaque nœud  $v \in V$  sélectionne indépendamment un seuil  $\theta_v \in [0,1]$  uniformément en aléatoire. L'objectif de ce seuil est de représenter la fraction pondérée des voisins actifs de  $v$  qui doivent adopter la rumeur pour que  $v$  devienne actif. Un nœud est « **actif** » s'il est infecté par la rumeur (soit initialement ou séquentiellement) de l'un de ses voisins, et « **inactif** » autrement.

Maintenant, étant donné les seuils choisis pour tous les nœuds  $v$  dans  $V$ , la propagation progresse à partir d'un ensemble initial de nœuds infectés  $I$  comme suit :

À l'étape  $t$ , tous les nœuds actifs de l'étape  $(t - 1)$  restent actifs, et tout nœud inactif  $v$  pour lequel la somme des poids des liens venant de ses voisins actifs (infectés par la rumeur) est supérieure au seuil  $\theta_v$  (2),  $v$  devient donc infecté :

$$\sum_{w \in N_{Actifs}(v)} b_{v,w} \geq \theta_v \quad (2)$$

### 2.2. Le modèle en cascade indépendant (Independent Cascade Model)

Le modèle en cascade indépendant [17] est la forme généralisée du modèle épidémique de type SIR. Tout nœud  $v$  infecté par la rumeur n'aura qu'une seule fois la chance d'activer (infecter) ses voisins  $N(v)$  actuellement inactifs (susceptibles).

Un nœud  $v$  peut infecter un nœud voisin  $w$  avec une fonction probabiliste  $p(u, v)$ . Si le nœud  $w$  est liés à plusieurs voisins nouvellement activés, ils essaieront de l'activer séquentiellement dans un ordre arbitraire. Si l'une de ces tentatives réussit,  $w$  est activé au temps  $t$  et la même procédure se poursuit plus loin par  $w$ , c'est-à-dire que  $w$  va essayer de transmettre la rumeur à son tour à sa voisinage inactive au temps  $t + 1$  [14].

### 2.3. Le modèle stochastique (Stochastic model)

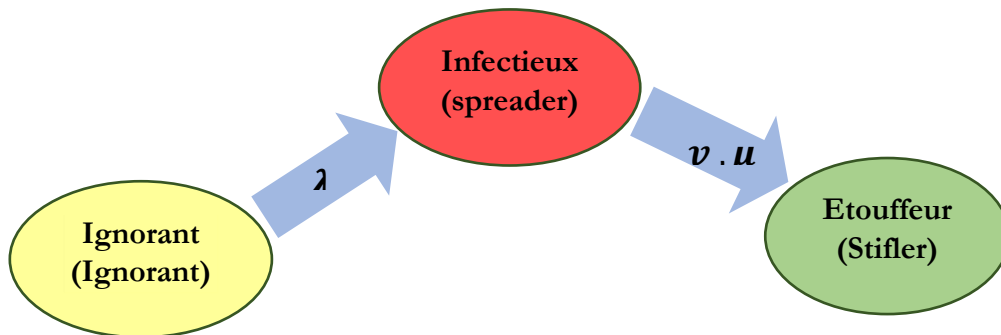
Le premier modèle stochastique a été proposé par McKendrick en 1926 [18], et développé ensuite par Barlette en 1949 [19] et par Kendall en 1956 [20]. Ces modèles non

## Chapitre 02 : Propagation des rumeurs en ligne

déterministes sont basés sur les probabilités, les estimations et les variables aléatoires pour la réalisation des événements, tel que le changement d'état d'un individu entre les trois classes décrits ci-après [15].

Le modèle stochastique est composé de trois compartiments :

- **Les ignorants** (*ignorants*) ceux qui ne sont pas infectés par la rumeur ;
- **Les infectieux** (*spreaders*) ceux qui partagent et diffusent la rumeur ;
- **Les étouffeurs** (*stiflers*) ceux qui sont déjà infectés mais ne partagent pas la rumeur [21].



**Figure 2.5.** Le modèle stochastique

Les paramètres du modèle stochastique comportent principalement :

- **La probabilité d'infection  $\lambda$**  ; l'« ignorant » peut devenir un « infectieux » s'il entre en contact avec un autre « infectieux ».
- **La probabilité d'immunité  $\mu$**  ; un « infectieux » peut devenir un « étouffeur » s'il entre en contact avec un autre « infectieux » ou un « étouffeur ».
- **La probabilité de perdre l'intérêt  $\nu$**  ; l'« infectieux » peut après une période de temps, perdre l'intérêt à partager la rumeur et devenir donc un « étouffeur ».

## II. La classification des attitudes (Stance Classification)

La classification des attitudes est un des sujets récents les plus actifs dans le traitement automatique du langage naturel. Elle est généralement considérée comme un sous-problème de l'analyse des sentiments qui vise à identifier la position de l'auteur du texte envers une cible (une entité, un concept, un événement, une idée, une opinion, un sujet...etc.). Le but étant de mieux comprendre les positions prises par les gens (avec, contre ou neutre) et la façon duquel ils réagissent à une information [22].

### 1. Définition de la classification des attitudes envers les rumeurs (Rumor Stance Classification)

La classification des attitudes envers les rumeurs est un domaine d'application très important qui aide avec ces processus de vérification à déterminer la véracité de la rumeur dans les différents réseaux sociaux en utilisant le système **SDQC** (Support, Deny, Request, Comment) qui classifie les réactions des gens vis-à-vis une rumeur en quatre catégories :

- **Supporter** (Support) : l'auteur du texte à examiner soutient clairement la véracité de la rumeur.
- **Refuser** (Deny) : l'auteur du texte à examiner nie clairement la véracité de la rumeur.
- **Requête** (Request) : l'auteur du texte à examiner n'a pas pu confirmer la véracité de la rumeur et cherche des preuves supplémentaires.
- **Commentaire** (Comment) : l'auteur du texte à examiner fait son propre commentaire sans juger clairement la véracité de la rumeur.

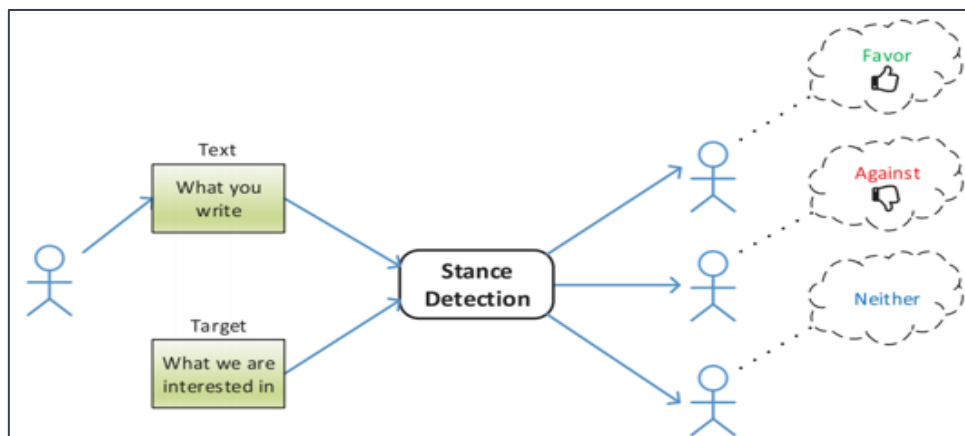


Figure 2.6. Représentation schématique de la procédure de détection des positions [22]

## 2. Approches de classification des attitudes

### 2.1. Approche d'apprentissage automatique

La plupart des approches de classification automatique des positions reposent sur des techniques d'apprentissage automatique (Machine Learning) avec ses deux types ; l'apprentissage supervisé et l'apprentissage non supervisé. Parmi les méthodes de base les plus utilisées : les Machines à Vecteurs de Support (SVM), Naïve Bayes et les Arbres de Décision.

### 2.2. Approche d'apprentissage profond

L'apprentissage profond « Deep Learning » fait partie des méthodes d'apprentissage supervisés qui utilisent les réseaux de neurones multicouches composés de milliers d'unités (les

## Chapitre 02 : Propagation des rumeurs en ligne

neurones) pour effectuer des opérations et des calculs à travers plusieurs couches afin de produire une ou des sorties qui représentent les classes finales des données traitées [22].

Il existe un grand nombre d'architectures profondes utilisés pour la classification des attitudes [23], prenant comme exemple : les Réseaux Neuronaux Convolutifs (CNN), les Réseaux de Neurones LSTM (Long Short Term Memory), les Réseaux de Neurones Récurents à portes GRU.

### 2.3. Approche d'apprentissage ensembliste

Les méthodes ensemblistes combinent plusieurs algorithmes et classificateurs dans la phase d'apprentissage de données afin d'améliorer la performance et obtenir de meilleures prédictions pour résoudre une tâche de classification donnée [24].

Il existe plusieurs architectures ensemblistes utilisés pour la classification des attitudes telles que : le Bagging (Bootstrap Aggregation), le Boosting et les forêts aléatoires (Random Forest).

### 2.4. Approche d'apprentissage par transfert

L'apprentissage par transfert « Transfer Learning » consiste à utiliser des outils permettant d'entraîner plus efficacement un modèle d'une architecture profonde sur un grand ensemble de données. Le modèle final pré-entraîné qui résulte, peut être utilisé directement avec quelques réglages de paramètres sur des nouveaux ensemble de données réelles et différents. Ce processus facilite l'apprentissage et booste la performance de classification. Cette technique est actuellement considérée comme étant l'approche la plus performante dans le domaine du traitement automatique du langage naturel et dans le domaine de classification des attitudes [25].

De plus amples informations sur l'apprentissage par transfert avec en particulier, les modèles pré-entraînés mBERT, araBERT, XLM-Roberta, seront présentées dans le chapitre suivant.

## Conclusion

Le phénomène de propagation de la rumeur a été étudié à travers deux approches : celui des modèles épidémiques qui définit la rumeur en tant qu'une maladie infectieuse, et celui des modèles mathématiques à l'aide des équations différentielles et modèles stochastiques. Tous ces modèles procèdent à attribuer des rôles aux utilisateurs d'un réseau social afin de déterminer les éléments et les événements clés du processus de propagation. Pour bien déterminer ces rôles,

## Chapitre 02 : Propagation des rumeurs en ligne

plusieurs approches existent. Notre intérêt dans ce mémoire a porté sur le paradigme de classification des attitudes qui propose une variété des techniques d'apprentissage automatique, d'apprentissage profond et d'apprentissage par transfert pour la résolution de ce problème. Ces modèles ont montré leur efficacité à travers plusieurs études et applications.

Dans le reste de ce mémoire, nous exploiterons certaines de ces approches pour explorer la propagation des rumeurs sur les réseaux sociaux avec un cas d'étude récent et important.

## **Chapitre III**

# **Propagation de la rumeur sur les réseaux sociaux : étude de cas**

---

### Introduction

Nous revenons dans ce chapitre à un cas d'étude réel et récent qui a marqué la vie politique algérienne pendant quelques mois entre fin de l'an 2020 et début de l'an 2021. Nous exposons notre méthodologie pour examiner ce cas d'étude en illustrant les étapes du projet, les méthodes, et les techniques que nous avons exploité pour retracer, en premier lieu, le processus de propagation que les rumeurs en question ont suivi et pour détecter, en deuxième lieu, la réaction des algériens à ces rumeurs sur les réseaux et les médias sociaux.

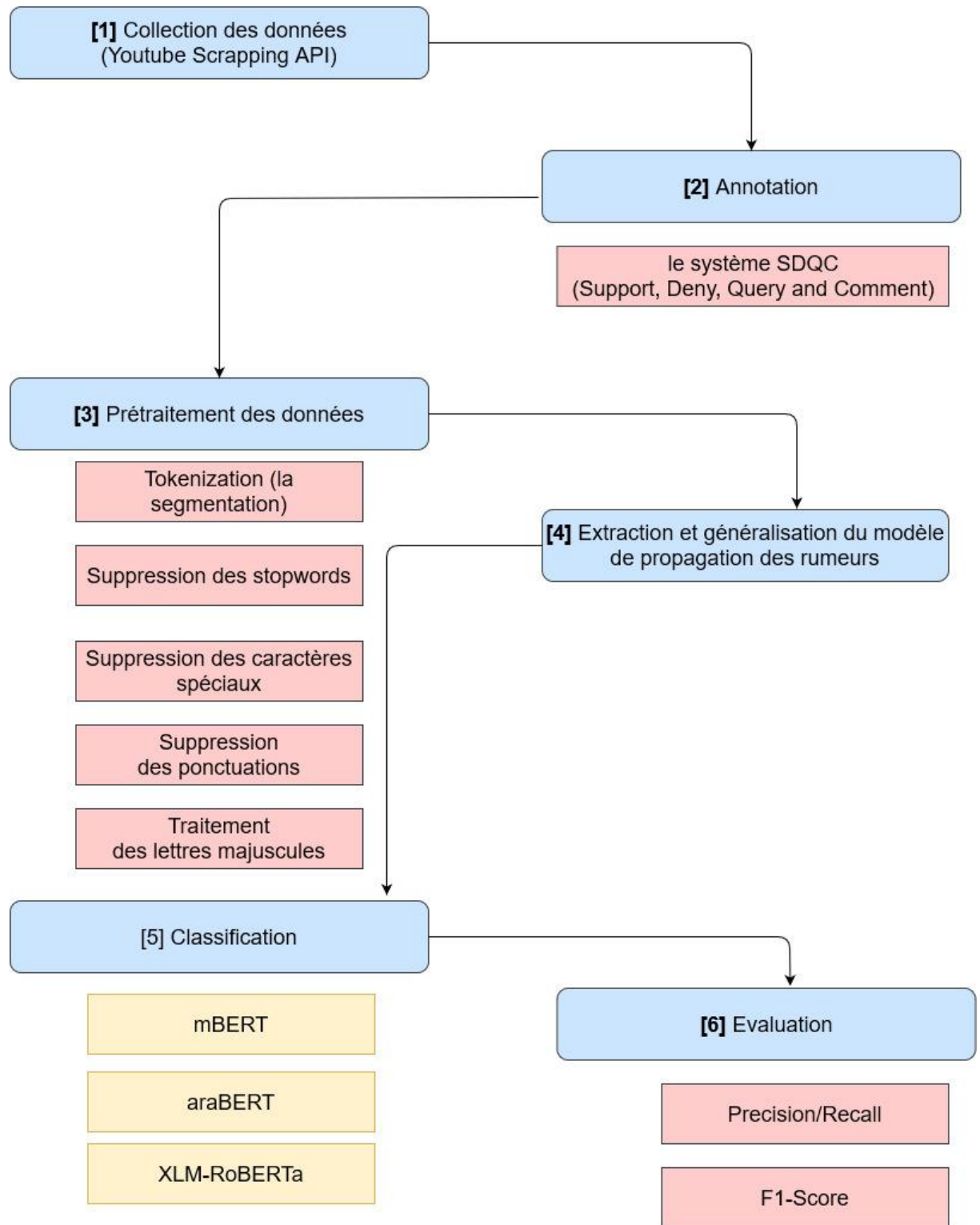
### 1. Description du projet

Dans le but d'examiner le processus de diffusion des rumeurs et de détecter les attitudes du public visé par ces derniers sur les médias sociaux, nous avons proposé ce projet qui aborde le cas des rumeurs qui ont circulé sur la plateforme sociale YouTube à propos de la santé du président de la République algérienne Abdelmadjid Tebboune durant la période entre octobre 2020 et mars 2021.

Notre méthodologie de réalisation est illustrée dans la Figure 3.1. En résumé, nous avons suivi les étapes décrites ci-dessous :

1. Collection de données à l'aide de YouTube Scrapping API.
2. Nettoyage de données collectées.
3. Annotation (étiquetage) manuelle des données collectées.
4. Extraction et généralisation du modèle approximatif de propagation des rumeurs à l'aide du modèle épidémique SIR.
5. Prétraitement et préparation de données collectées pour la tâche de détection des attitudes envers les rumeurs étudiées.
6. Classification des attitudes à l'aide des modèles d'apprentissage par transfert.
7. Evaluation des résultats obtenus.

## Chapitre 03 : Propagation de la rumeur : étude de cas



**Figure 3.1.** Architecture de système proposé

### 2. Description du cas d'étude

Durant la période entre octobre 2020 et mars 2021, l'absence du nouvel président Abdelmadjid Tebboune qui était atteint du Covid 19 a alimenté beaucoup d'interrogations, spéculations et de rumeurs auprès de la population algérienne. Cette absence a été tombé en pleine crise ; une crise politique marquée par les protestations d'après-Hirak et une grave crise sanitaire et économique liée aux répercussions de la propagation de la pandémie de Coronavirus. Nous listons ci-après les dates clés de ce cas d'étude :

- **24 octobre 2020** : M. Abdelmadjid Tebboune s'était mis volontairement à l'isolement (date de son dernier tweet) après avoir été en contact avec de hauts responsables de la présidence et du gouvernement contaminés par Covid 19, puis admis dans une unité de soins spécialisés de l'hôpital militaire d'Ain Naâdja, à Alger.
- **28 octobre 2020** : M. Abdelmadjid Tebboune est hospitalisé en Allemagne.
- **8 novembre 2020** : M. Abdelmadjid Tebboune avait tenu personnellement à rassurer le peuple algérien qui attend son retour au pays, affirmant que son état de santé s'améliore.
- **30 novembre 2020** : la présidence de la République algérienne a indiqué, dans un communiqué que le président Abdelmadjid Tebboune serait de retour dans son pays.
- **13 décembre 2020** : M. Abdelmadjid Tebboune est réapparu publiquement pour la première fois depuis le début de son hospitalisation en Allemagne.
- **29 décembre 2020** : retour du président Abdelmadjid Tebboune en Algérie après sa 1<sup>ère</sup> hospitalisation en Allemagne.
- **10 janvier 2021** : l'information a été confirmée par la présidence algérienne : en raison de complications au pied dues au Covid-19, M. Abdelmadjid Tebboune est reparti pour l'Allemagne.
- **20 janvier 2021** : la présidence algérienne a déclaré que M. Abdelmadjid Tebboune a été opéré avec succès.
- **12 février 2021** : retour à Alger du président après une nouvelle hospitalisation en Allemagne

### 3. Préparation des données d'étude

#### 3.1. Collection et nettoyage de données

En utilisant YouTube API, nous avons collecté un ensemble de 7 500 commentaires à partir de 11 publications visuelles. Les rumeurs contenues dans ces publications sont illustrées dans le Tableau 3.1.

## Chapitre 03 : Propagation de la rumeur : étude de cas

ID_Vidéo		Rumeurs
Vidéo 1	Amir DZ	<ul style="list-style-type: none"> <li>- Le président Tebboune est allé en Allemagne à cause de sa maladie thrombose veineuse profonde et du cancer du poumon.</li> <li>- Le gouvernement essaye de manipuler l'opinion publique à travers les médias de masse publiques et quelques pages sur les réseaux sociaux.</li> </ul>
Vidéo 2	Hichem Aboud	<ul style="list-style-type: none"> <li>- Le président Tebboune n'est pas allé en Allemagne pour des examens réguliers mais pour un traitement médical à cause de sa maladie compliquée.</li> <li>- Le président Tebboune est en coma artificiel.</li> </ul>
Vidéo 3	Hichem Aboud	<ul style="list-style-type: none"> <li>- Le président Tebboune a une maladie neuronale et une paralysie dans le côté droit de son visage.</li> </ul>
Vidéo 4	Hichem Aboud	<ul style="list-style-type: none"> <li>- Une équipe de journalistes sont allés pour enregistrer le discours du président Tebboune. Vu son mauvais état de santé, les généraux ont interdit cet enregistrement d'être diffusé sur la télévision, ce qui a poussé le président à le publier sur sa page Twitter.</li> </ul>
Vidéo 5	Amir DZ	<ul style="list-style-type: none"> <li>- La disparition du président Tebboune avec sa famille.</li> <li>- Quelques membres de la famille du président Tebboune ont été tués.</li> </ul>
Vidéo 6	Amir DZ	<ul style="list-style-type: none"> <li>- La vidéo du président Tebboune publiée sur Twitter confirme qu'il a une thrombose veineuse profonde et qu'il a perdu sa conscience.</li> </ul>
Vidéo 7	Mohamed Larbi Zitout	<ul style="list-style-type: none"> <li>- Le président Tebboune et quelques membres de la présidence ont attrapé le Covid-19.</li> </ul>
Vidéo 8	Mohamed Larbi Zitout	<ul style="list-style-type: none"> <li>- Les médecins allemands ont coupé les doigts du pied droit du président Tebboune à cause de sa maladie du diabète.</li> </ul>
Vidéo 9	Mohamed Larbi Zitout	<ul style="list-style-type: none"> <li>- Le président Tebboune était dans l'hôpital militaire Ain Naâdja depuis 14 jours et il a été transféré vers un hôpital allemand parce qu'il était en mauvaise santé.</li> </ul>
Vidéo 10	Mohamed Larbi Zitout	<ul style="list-style-type: none"> <li>- Le président a une gangrène au niveau du pied droit. Ce dernier risque d'être coupé jusqu'au niveau de genou.</li> </ul>
Vidéo 11	Mohamed Larbi Zitout	<ul style="list-style-type: none"> <li>- Le président Tebboune est dans le coma à l'hôpital Ain Nadja parce qu'il est malade du Covid-19.</li> <li>- Les reins et les poumons du président sont en mauvaise santé.</li> </ul>

**Tableau 3.1.** Les rumeurs analysées

Après le nettoyage et la suppression des commentaires inutiles et hors-sujet, le nombre total de données a diminué jusqu'à 3 300 commentaires (c-à-d, la moitié des commentaires publiés sont des commentaires hors sujet).

Ci-dessous, nous résumons nos choix dans cette première étape du projet.

**Choix de plateforme sociale :** nous avons choisi la plateforme YouTube pour deux raisons. Premièrement, parce qu'elle assemble une grande communauté algérienne qui est très active,

## Chapitre 03 : Propagation de la rumeur : étude de cas

très réactive et qui s'intéressent bien aux sujets d'intérêt commun. Deuxièmement, parce que YouTube permet à ces utilisateurs inscrits de scraper le contenu publié sur les chaînes YouTube publiques contrairement aux autres réseaux sociaux comme Twitter et Facebook qui posent des restrictions sur l'exploration et la collection de données, ce qui peut influencer la quantité et la qualité de données à examiner.

**Choix de chaînes YouTube et des vidéos :** nous avons choisi une playlist de 11 vidéos (Tableau 3.3) qui ont abordé le sujet de la maladie du président Abdelmadjid Tebboune entre octobre 2020 et mars 2021. Ces vidéos ont été publiées sur 03 chaînes YouTube principales (Tableau 3.2) dont les propriétaires sont accusés régulièrement par les autorités algériennes de désinformation pour des fins de manipulation de l'opinion publique.

Nom de chaîne	Date de création	Description	Pays	Nbr des abonnés	Nbr de vidéos
Amir DZ	2018-10-26	الصحفي الاستقصائي الجزائري الذي كشف النقاب على عدة ملفات و مستندات تخص النظام الجزائري، سياسيين و جنرالات عسكريين. متابع بتسعة مذكرات اعتقال دولية من طرف النظام الجزائري بسبب الملفات السرية التي يكشفها	GB	986000	934
Mohamed Larbi Zitout	2018-10-07	هذه هي القناة الرسمية الرئيسية للمقاوم السياسي الجزائري محمد العربي زيتوت، عضو مؤسس لحركة رشاد وديبلوماسي جزائري سابق	GB	751000	5496
Hichem Aboud	2018-05-10	قناة الصحفي هشام عبود. اخبار و تحاليل و تعليقات حول الأحداث السياسية و الاقتصادية و الثقافية و الرياضية	FR	324000	1970

**Tableau 3.2.** Les chaînes YouTube choisies

### 🚩 Chaîne 01 : AMIR DZ

Le propriétaire de cette chaîne s'appelle Amir Boukhors. Un blogueur algérien de 37 ans, qui se présente comme un journaliste d'investigation et activiste politique. Il réside actuellement en Allemagne et est très suivi en Algérie. Sa page Facebook compte presque 500 000 abonnés et sa chaîne YouTube compte 986 000. Recherché pour neuf mandats d'arrêt internationaux, ses publications sur les réseaux sociaux suscitent un tollé de réactions notamment à cause de son atteinte à la vie privée des politiciens et des personnalités publiques. [26]

### 🚩 Chaîne 02 : Mohamed Larbi Zitout

## Chapitre 03 : Propagation de la rumeur : étude de cas

Le propriétaire de cette chaîne s'appelle **Mohamed Larbi Zitout**, ancien diplomate algérien et un des membres fondateurs du mouvement Rachad créé en 2007.

Mohamed Larbi Zitout obtient un diplôme de l'École Nationale d'Administration d'Alger en 1986 et il est désigné comme premier secrétaire de l'ambassade d'Algérie en Libye en 1991. Il démissionne de son poste en 1995 et s'exile à Londres sous le statut de réfugié.

Très présent sur les réseaux sociaux lors du soulèvement populaire du Hirak, il est condamné à 20 ans de prison par contumace le 17 décembre 2019 par le tribunal d'Oran pour « intelligence avec une puissance étrangère, atteinte à l'autorité de l'armée, insulte et diffamation ». [27]

### **Chaîne 03 : Hichem Aboud**

Hichem Aboud est un journaliste algérien. Il est le fondateur et ancien directeur du quotidien 'Mon Journal' et 'Djaridati'. En 1975, il s'est engagé dans l'armée algérienne en qualité d'élève officier. Alors qu'il poursuivait ses études à l'Institut des Sciences Politiques et de l'Information de l'Université d'Alger, il obtient sa licence en sciences journalistiques en 1978.

À l'issue de ses études universitaires, il rejoint la direction du commissariat politique de l'armée algérienne où il a eu à diriger la rédaction de l'édition francophone du magazine mensuel El-Djeïch. En 2019, il fonde la chaîne de télévision suisse Amel TV.

Le 17 février 2021, il est condamné à sept ans de prison, avec émission d'un mandat d'arrêt international, pour avoir diffusé des informations confidentielles sur la famille d'une personnalité militaire et pour 'atteinte à l'unité nationale et participation à une entreprise de démoralisation de l'armée'. [28]

ID_Vidéo	Chaîne	Date de publication	Nbr de vues	Nbr de likes	Nbr de dislikes	Nbr de commentaires
Vidéo 1	Amir DZ	2020-12-01	944559	18433	3613	2547
Vidéo 2	Hichem Aboud	2020-10-30	188451	5378	348	226
Vidéo 3	Hichem Aboud	2020-12-14	518776	7180	1451	60
Vidéo 4	Hichem Aboud	2020-12-17	198869	5132	417	114
Vidéo 5	Amir DZ	2020-11-24	559312	13644	1298	1659
Vidéo 6	Amir DZ	2020-12-21	55	3	0	30

## Chapitre 03 : Propagation de la rumeur : étude de cas

Vidéo 7	Mohamed Larbi Zitout	2020-10-27	187918	3315	453	185
Vidéo 8	Mohamed Larbi Zitout	2021-02-26	105639	1982	242	299
Vidéo 9	Mohamed Larbi Zitout	2020-10-30	163944	2703	384	247
Vidéo 10	Mohamed Larbi Zitout	2021-03-02	136531	2688	241	322
Vidéo 11	Mohamed Larbi Zitout	2020-10-28	394848	4221	898	952

**Tableau 3.3.** Les vidéos YouTube choisies

### 3.2. Annotation de données

Cette étape vise à étiqueter chaque commentaire par rapport à la position exprimé du commentateur envers le contenu de la vidéo et envers le diffuseur de la vidéo. Nous avons suivi le système SDQC expliqué dans le chapitre précédent pour obtenir les quatre libellés suivants :

- **Supporter**
- **Refuser**
- **Requête**
- **Commentaire** : l'auteur de la réponse fait son propre commentaire sans contribuer clairement à apprécier la véracité de la rumeur.

Pour enrichir cette dernière classe, nous avons ajouté deux autres étiquettes :

- **Commentaire positif** : l'auteur n'apprécie pas la véracité de la rumeur mais soutient le diffuseur de la rumeur en personne ou attaque à ses détracteurs.
- **Commentaire négatif** : l'auteur n'apprécie pas la véracité de la rumeur mais critique le diffuseur de la rumeur en personne ou attaque ses fans.

### 3.3. Prétraitement de données

Le prétraitement de données est une étape très importante qui a pour but de dégager tous ce qui est bruit du texte étudié (les mots inutiles et redondants, les chiffres, les suffixes et préfixes...). Un bon nettoyage améliore le résultat de la classification peu importe la capacité du classificateur appliqué. Le prétraitement consiste principalement à :

## Chapitre 03 : Propagation de la rumeur : étude de cas

- Éliminer les mots vides (stop words) : cette étape consiste à éliminer les déterminants, adjectifs, prépositions, conjonctions... du texte en se basant sur des coups spéciaux pour chaque langue, exemple : et, a, ou, alors...
- La suppression des caractères spéciaux : consiste à éliminer les caractères spéciaux du texte comme : +, @, #...
- La suppression de la ponctuation.
- Le traitement des lettres majuscules pour le texte non arabe : les lettres majuscules sont très utilisées dans les paragraphes. On les trouve aux débuts des phrases, dans les mots propres et même dans les abréviations. Une façon qui est la plus évidente de les traiter est de, tout simplement, les convertir en minuscules.
- La segmentation (Tokenization) : découpage d'un flux de caractères en phrases, symboles ou mots. Nous nous intéressons plus particulièrement aux mots.

### 4. Extraction et généralisation du modèle de propagation des rumeurs

Une fois les données nettoyées et étiquetées par le système SDQC, on peut procéder à construire les ensembles des acteurs « Susceptibles », « Infectés » et « Récupérés » dont le modèle épidémique SIR que nous avons choisi en a besoin dans le but de tracer la propagation des rumeurs étudiées sur le média sociale YouTube. Après l'extraction des paramètres de diffusion, la courbe de la propagation obtenue est ensuite généralisée afin de donner une vision approximative et plus claire sur le processus de diffusion réel.

On va revenir sur cette étape avec plus d'explications dans le chapitre suivant.

### 5. Apprentissage par transfert pour la classification du texte

L'apprentissage par transfert est une méthode très populaire d'apprentissage automatique supervisé qui consiste à transférer des connaissances d'un modèle neuronal déjà entraîné sur une tâche pour améliorer les performances et réduire le temps d'exécution et par conséquent économiser le cout d'une nouvelle tâche dans le même domaine [29].

En raison de ses résultats incroyablement précis dans diverses tâches de programmation en langage naturel comme la classification de texte, il est devenu très important et a donné aussi un grand aide au développement du domaine de l'apprentissage automatique.

#### 5.1. Définition

La définition mathématique de l'apprentissage par transfert est représentée en termes de tâches et des domaines [30]

- **Domaine** : le domaine  $D$  est composé de deux parties,  $X$  : un espace de description (Feature Space), et  $P(X)$  : la distribution marginale alors  $D = \{X, P(X)\}$  avec  $X = \{x|x_i \in X, i = 1, \dots, n\}$
- **Tâche** : la tâche  $T$  est composée de deux parties,  $Y$  : un espace d'étiquettes (Label Space), et  $f_s(.)$  : une fonction de prédiction alors  $T = \{Y, f\}$

Vue que le domaine source est plus riche en données annotées que le domaine cible, donc à partir des connaissances d'une source  $P_s(D_s, T_s)$  et une cible  $P_c(D_c, T_c)$  tels que  $T_s = T_c$  et  $D_s \neq D_c$  l'apprentissage par transfert sera capable à améliorer l'entraînement de la fonction de prédiction cible  $f_c(.)$  dans  $D_c$  [31].

### 5.2. Les transformateurs

Les modèles d'apprentissage par transfert se basent sur le concept du transformateur. Un transformateur est un modèle séquence à séquence développée par une équipe de Google en 2017. Il propose une nouvelle manière de construire un modèle de langue fondé sur des « mécanismes d'attentions » et capable de prédire les mots les plus fréquents et prendre la décision en fonction du contexte pour faciliter la distribution des modèles pré-entraînés.

Le transformateur utilise une architecture Encodeur-Décodeur (Figure 3.2) pour transformer des vecteurs réels en entrées vers des vecteurs réels en sorties, ce que lui permet de garder une grande quantité de données en mémoire sans perte dans le but d'améliorer significativement les performances dans les différentes tâches [32].

- **Un Encodeur** : chaque encodeur se compose de deux sous-couches sous-couches d'attention « Attention sub-layer » suivie d'un FeedForward Network (FFN) complètement connectées. En plus, chaque sous-couche possède en sortie une couche qui ajoute, additionne, les sorties de la couche et du raccord à une connexion dite résiduelle « qui connecte directement les valeurs d'entrée de la couche à la sortie de la couche » et qui normalise l'ensemble après la lecture [33].
- **Un Décodeur** : le décodeur est également composé d'un bloc d'attention « Attention sub-layer » et d'un Feed-Forward, mais il contient en plus une couche « Encoder-Decoder Attention » qui a pour but de permettre au décodeur de réaliser le mécanisme d'attention entre la séquence d'entrée « la sortie de l'encodeur » et la séquence de sortie « en train d'être décodée » pour créer la prédiction de la tâche [33].

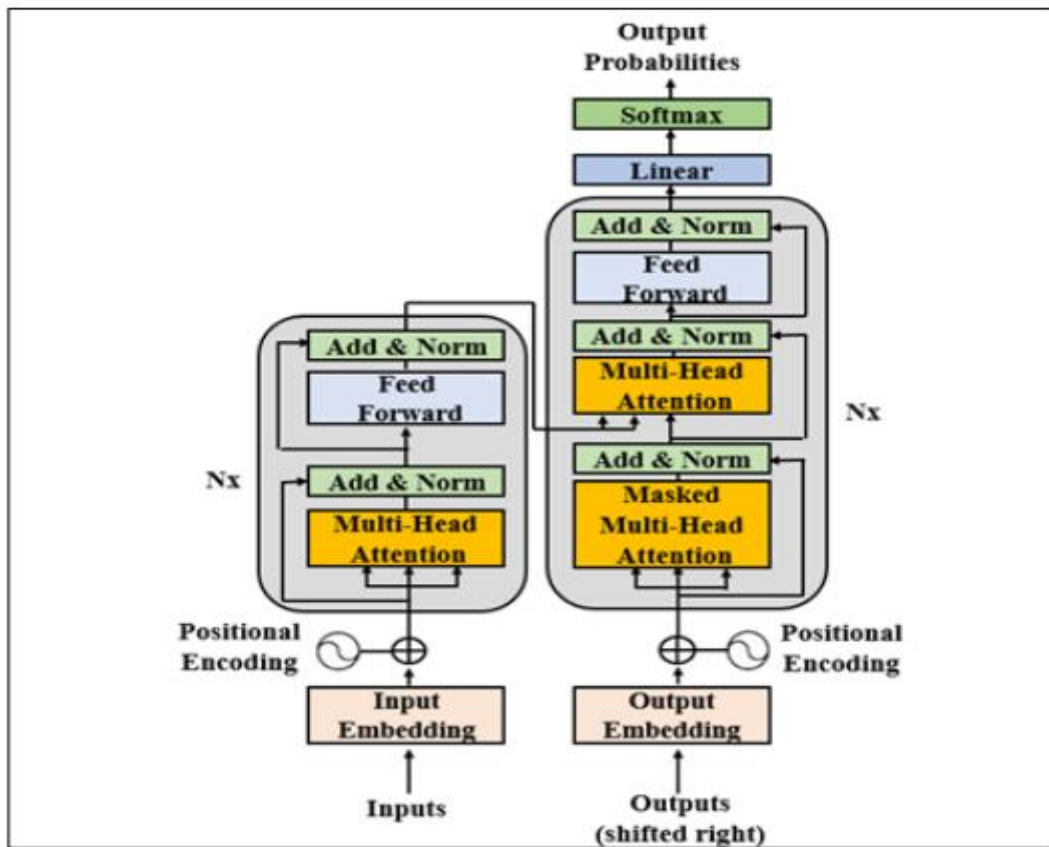


Figure 3.2. Représentation de l'architecture d'un Transformateur

## 6. Apprentissage par transfert pour le traitement automatique du langage naturel (NLP)

Malgré que la création la discipline d'apprentissage par transfert soit encore récente, plusieurs modèles sont apparus pour le traitement automatique de la langue naturelle. Ce qui se suit sont les plus connues :

### 6.1. Modèles de base

- **Modèle BERT** : BERT est l'abréviation de **B**idirectional **E**ncoder **R**epresentations from **T**ransformers. Publié en fin 2018, BERT est le premier modèle créé pour l'apprentissage par transfert. Il a été développé par un groupe de chercheurs travaillant à Google et a été entraîné sur des données textuelles en langue anglaise extraites de Wikipedia pour les tâches de classification du texte, des systèmes de Questions-Réponses (QA), de traduction automatique ...etc. [34]
- **Modèle XLM** : XLM, est une version améliorée de BERT présenté par Facebook IA. Ce modèle utilise une technique de pré-traitement connue (BPE) et un mécanisme d'apprentissage bilingue avec BERT afin d'apprendre les relations entre les mots dans

## Chapitre 03 : Propagation de la rumeur : étude de cas

différentes langues. Le modèle surpasse les autres modèles dans une tâche de classification multilingue (implication de phrases dans 15 langues) et améliore considérablement la traduction automatique lorsqu'un modèle pré-entraîné est utilisé pour l'initialisation du modèle de traduction [35].

- **Modèle GPT** : GPT c'est le modèle de traitement automatique du langage développé par Open AI. Sa dernière version GPT-3 annoncé le 28 mai 2020 et ouvert aux utilisateurs via l'API d'Open AI en juillet 2020. Il se distingue par une capacité de généralisation singulière. Il atteint des performances étonnantes avec un nombre très réduit d'exemples [36].

### 6.2. Modèles multilingues

Comme le titre indique, ce sont des modèles plus généralisés, destinés aux tâches NLP qui traitent plusieurs langues. Dans notre projet, nous avons exploité deux modèles :

- **Modèle mBERT** : c'est le BERT multilingue (mBERT). Il fournit des représentations de phrases pour 104 langues, qui sont utiles pour de nombreuses tâches multilingues en utilisant l'apprentissage par transfert sur des tâches morphologiques et syntaxiques. Beaucoup de travaux de recherche ont confirmé l'efficacité et la bonne performance de mBERT pour l'apprentissage transversal entre langues [37].
- **Modèle XLM-Roberta** : le modèle XLM-Roberta est la mise à jour du XLM-100 qui a été publié en novembre 2019 par Facebook AI. Il est aussi basé sur le transformateur avec une capacité de gérer des textes dans 100 langues différentes, et avec un grand ensemble de données d'entraînement dont la taille peut arriver jusqu'à plus de 2 To.

### 6.3. Modèles monolingues

Il existe plusieurs variantes de modèles transformateurs BERT qui ont été pré-entraînés sur des données textuelles écrites dans une langue autre que la langue anglaise. Comme la langue de nos données collectées est la langue arabe, nous avons exploité un modèle BERT destiné à l'arabe. Ce modèle s'appelle **araBERT**.

araBERT est un transformateur publié par Antoun et al. [38] en 2020. Il est conçu sur la base du modèle BERT et entraîné sur un très grand corpus de textes écrits en langue arabe. Le BERT de la langue arabe ou le araBERT a atteint des performances remarquables sur la plupart des tâches de NLP arabe testées, qui a évité plusieurs limitations de l'original BERT qui sont liés

## Chapitre 03 : Propagation de la rumeur : étude de cas

au manque de ressources textuelles dans la langue arabe par rapport à la langue anglaise et aux autres langues occidentales.

### Conclusion

Les rumeurs au sujet de la santé du chef d'état algérien Abdelmadjid Tebboune a causé une incertitude au sein de la communauté algérienne sur le Web et les réseaux sociaux pendant une période de cinq mois. Ce projet a été réalisé dans le but d'analyser la propagation de ces rumeurs et de détecter les attitudes du public algérien envers elles. Nous avons choisi le modèle épidémique SIR pour le premier objectif et l'apprentissage par transfert pour le deuxième objectif. Dans le chapitre suivant on va revenir aux détails techniques de la réalisation et aux résultats obtenus avec leurs perspectives.

# **Chapitre IV**

## **Implémentation et résultats**

---

### Introduction

Après avoir illustré notre approche de conception dans le chapitre précédent, nous présentons dans ce chapitre les résultats de différentes phases et procédures de réalisation et déploiement de ce projet de recherche qui exploite un des modèles de propagation des rumeurs et l'apprentissage par transfert pour la classification des attitudes envers les rumeurs liées à notre cas d'étude expliqué auparavant.

### 1. Environnement de travail et outils d'implémentation

#### 1.1. Matériel

Caractéristiques	Poste de travail N°01	Poste de travail N°02
PC	DESKTOP	HP
Système d'exploitation	Windows10 Professionnel	Windows10 Professionnel
Processeur	Intel® Core™ i5-8400 CPU @ 2.80GHz	Intel® Core™ i3-5200 CPU @ 2.80GHz
RAM	16 GB	6 GB
Type de système	SE 64 bits	SE 64 bits

Tableau 4.1. Caractéristiques du matériel utilisé

#### 1.2. Langage de programmation Python

Pour l'implémentation, nous avons choisi le langage Python, version 3.8.8 (Février 2021). Python est un langage de programmation *open source* qui est très populaire, très pratique pour les projets liés à la science de données et qui dispose d'une grande communauté de développeurs tout autour du monde.

Python est un langage interprété, multiparadigme et multiplateformes, placé sous une licence libre, qui présente des avantages et des facilitations grâce à sa flexibilité et les nombreuses outils, bibliothèques et packages qu'il offre pour la collection, le nettoyage, l'exploration et la modélisation de données [39].

#### 1.3. Environnement de programmation

Ce projet a été développé sur deux environnements de programmation :

- **Jupyter Notebook** : est une application web interactive *open-source* utilisée en science de données pour créer et partager en temps réel des documents constitués d'une liste ordonnée de cellules d'entrées et de sorties. Ce dernier peut supporter plus de 40 langages de programmation et aide au nettoyage et au visualisation des données [40].

- **Google Colaboratory** : est un environnement Google simple et gratuit particulièrement adapté au Machine Learning pour créer et exécuter les algorithmes d'apprentissage automatique dans des documents de type Jupyter Notebooks en utilisant des bibliothèques populaires telles que TensorFlow, PyTorch [40].

Nous avons utilisé Google Colab principalement parce que les modèles d'apprentissage par transfert requièrent le type « **Runtime GPU** » en tant qu'environnement d'exécution pour plus d'efficacité et de rapidité.



**Figure 4.1.** Jupyter et Google Colaboratory

### 1.4. Principaux packages Python utilisés

En plus de packages de base Pandas, Numpy et Matplotlib, la réalisation de ce projet nécessite l'utilisation de trois bibliothèques importantes :

- **NLTK « Natural Language ToolKit »** : est une suite très puissante d'outils de traitement automatique du langage naturel. Elle se base sur le langage de programmation Python et offre un ensemble des bibliothèques de prétraitement de données, de tokenization, stemming, lemmatisation et beaucoup plus [41].
- **PyArabic** : est une bibliothèque Python dédiée à la langue arabe. Elle offre des fonctions de bases pour manipuler le texte arabe, comme la détection et le groupement des lettres, la suppression des signes diacritiques, ou la conversion des nombres en lettres arabes [42].
- **Transformers** : est une librairie d'apprentissage profond (Deep Learning) utilisé principalement pour l'apprentissage par transfert dans le domaine de traitement du langage naturel. Il a été introduit en 2017 par « Hugging Face » qui contient plus de 30 modèles pré-entraînés capables à gérer 100 langues différentes [43].

### 2. Analyse exploratoire de données

#### 2.1. Informations générales

La première étape de notre projet nous a permis, en utilisant YouTube Scraper API avec les fonctions de nettoyage, d'extraire un ensemble de 3147 commentaires à partir de 11 vidéos appartenant à 03 chaînes YouTube dont les acteurs sont très actifs à la diffusion des rumeurs de notre cas d'étude (comme nous l'avons déjà décrit dans le chapitre 3). Les caractéristiques de notre base de données finale sont présentées dans le tableau ci-dessous :

Colonne	Type	Description
<b>ID_Commentaire</b>	String	Identifiant du commentaire
<b>ID_Video</b>	String	Identifiant de la vidéo
<b>ID_Chaine</b>	String	Identifiant de la chaîne YouTube
<b>Name</b>	String	Nom d'utilisateur
<b>Comment</b>	String	Texte du commentaire
<b>Label</b>	String	Attitude exprimée envers les rumeurs : <ul style="list-style-type: none"> <li>- Supporter</li> <li>- Refuser</li> <li>- Requête</li> <li>- Commentaire</li> <li>- Commentaire Positif</li> <li>- Commentaire Négatif</li> </ul>
<b>Time</b>	DateTime	La date et l'heure de la publication
<b>Likes</b>	Float64	Le nombre de mentions « J'aime »
<b>Reply Count</b>	Float64	Le nombre de réponses

**Tableau 4.2.** Caractéristiques de l'ensemble de données « Rumeurs\_Commentaires »

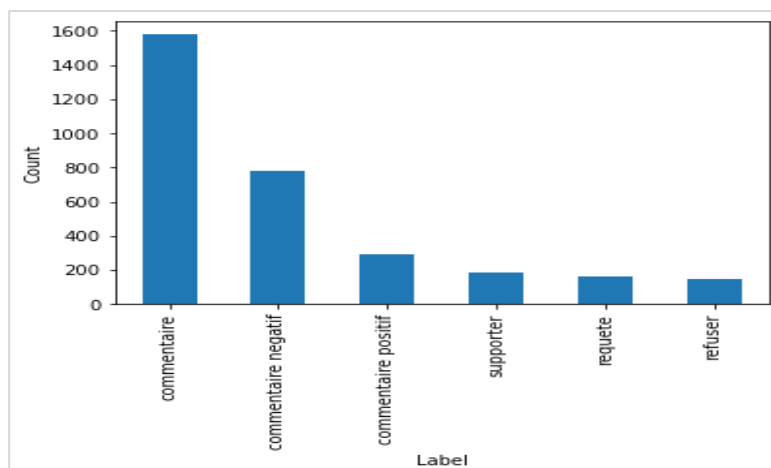
#### 2.2. Distribution générale de classes

L'ensemble de données « Rumeurs\_Commentaires » contient 3147 commentaires dont les classes « Supporter », « Refuser », « Requête », « Commentaire », « Commentaire Positif », « Commentaire Négatif » sont distribuées comme illustré dans le Tableau 4.3 et la Figure 4.2.

Label	Nombre
Commentaire	1580
Commentaire Négatif	781
Commentaire Positif	295
Supporter	186
Requête	160
Refuser	145
<b>Total</b>	<b>3147</b>

**Tableau 4.3.** Distribution de classes de l'ensemble de données « Rumeurs\_Commentaires »

Label	Nombre
Commentaire	1580
Commentaire Négatif	781
Commentaire Positif	295
Supporter	186
Requête	160
Refuser	145
<b>Total</b>	<b>3147</b>



**Figure 4.2.** Distribution de classes de l'ensemble de données « Rumeurs\_Commentaires »

### 2.3. Distribution de classes selon la chaine qui diffuse les rumeurs

D'après le tableau ci-dessous (Tableau 4.4), la distribution totale des commentaires n'est pas uniforme, que ce soit selon la chaine ou bien les classes qui définissent les positions des commentateurs. Nous remarquons que :

- Les visiteurs des trois chaines étudiées ne tendent pas à exprimer clairement leurs positions envers les rumeurs diffusées (la plupart des commentaires sont de classe « commentaire »).
- À l'exception de la chaine 01 (Hichem Aboud), les visiteurs prennent une position négative de diffuseur de la rumeur.
- L'écart entre la classe « Supporter » et la classe « Refuser » sur chaque chaine est étroit. En plus que la plupart des visiteurs ne réagissent pas à la rumeur elle-même, la petite fraction de ceux qui croient aux rumeurs diffusées et très proche à la fraction de ceux qui les refusent.

Label	<b>Chaine 01 : Hichem Aboud</b>	<b>Chaine 02 : Mohamed Larbi Zitout</b>	<b>Chaine 03 : Amir DZ</b>	<b>Total</b>
<b>Supporter</b>	21 (09,7%)	87 (08,9%)	78 (03,9%)	<b>186</b>
<b>Refuser</b>	17 (07,9%)	80 (08,2%)	48 (02,4%)	<b>145</b>
<b>Requête</b>	12	14	134	<b>160</b>
<b>Commentaire</b>	111	577	892	<b>1580</b>
<b>Commentaire Positif</b>	50	24	221	<b>295</b>
<b>Commentaire Négatif</b>	4	185	592	<b>781</b>
<b>Total</b>	<b>215</b>	<b>967</b>	<b>1965</b>	<b>3147</b>

**Tableau 4.4.** Distribution de classes selon le diffuseur de la rumeur

## 2.4. Fréquence des termes (N-Gramme)

Les figures 4.3 et 4.4 présentent les mots, les bi-grammes et les tri-grammes les plus fréquents dans notre corpus de commentaires. Nous distinguons la fréquence remarquable des structures reliées principalement au sujet de débat : «مات», «تبون», «الشعب», «يشفيه», mais aussi des structures que nous pouvons l'interpréter par le fait que la communauté algérienne sur YouTube a accusé, à une certaine période le gouvernement d'être complotiste dans une manipulation politique par la mention des différents acteurs politiques comme : «شنقرية», «دولة مدنية ماشي عسكرية», ou revendications politiques comme «الجنيرالات», «العصابة», «بوتقليقة», ou «فرنسا», «المغرب», qui accusent des entités extérieures de semer la discorde et de désinformation médiatique contre le public algérien en exploitant l'état instable du pays qui souffre d'une crise politique et sanitaire sans précédent.

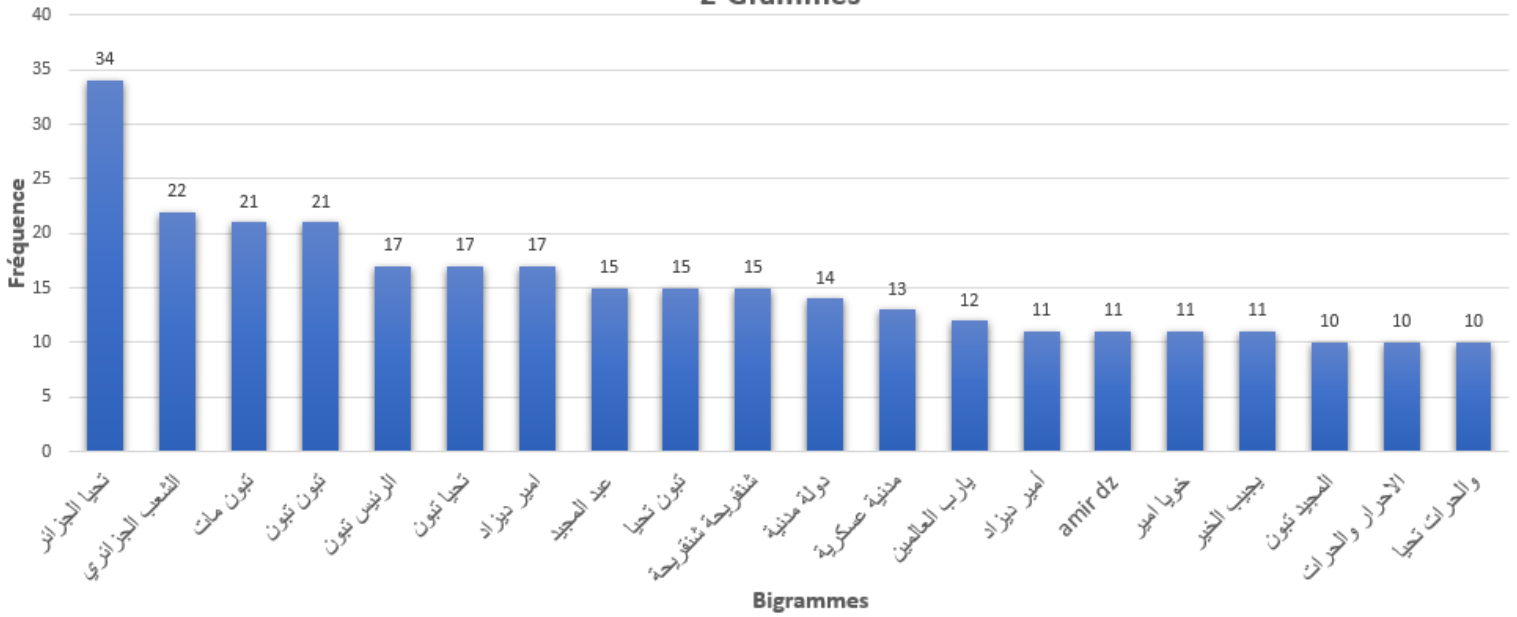


**Figure 4.3.** Les mots les plus fréquents dans l'ensemble de données

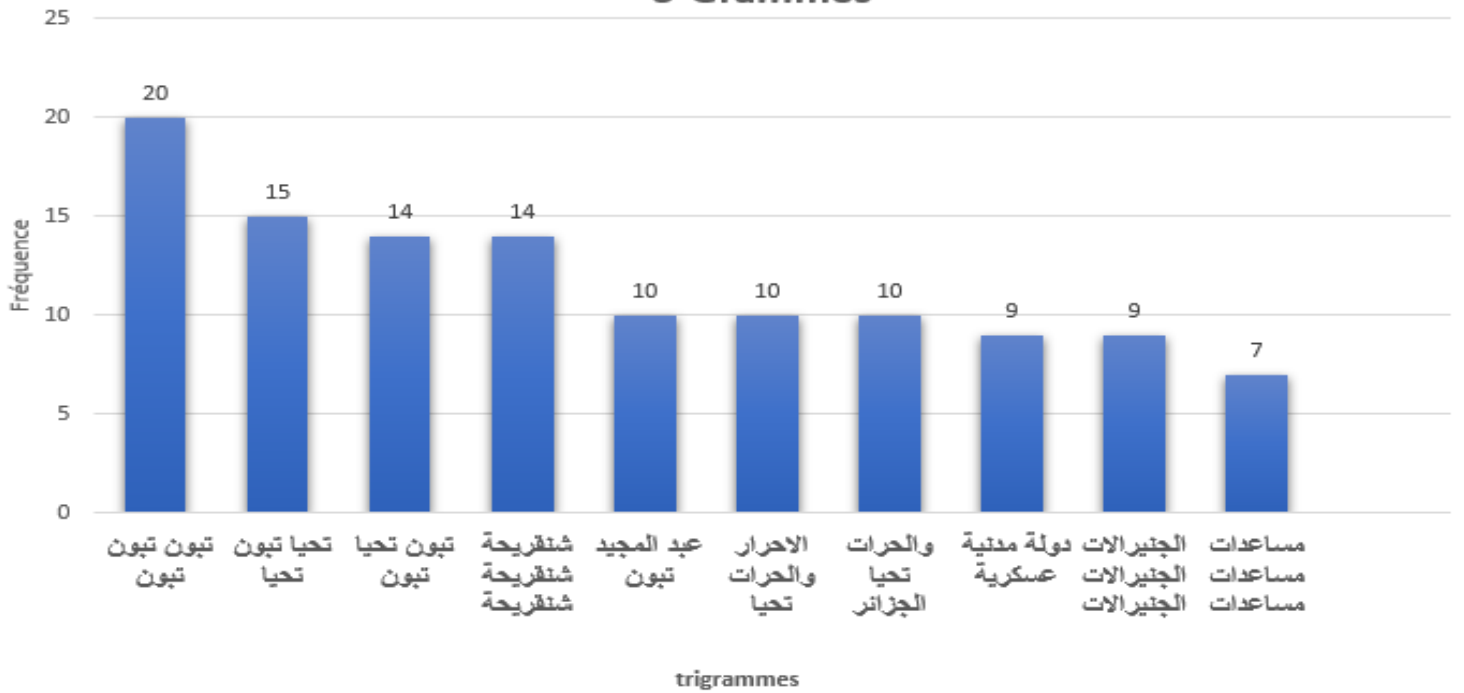
« Rumeurs\_Commentaires »

## Chapitre 04 : Implémentation et résultats

### 2-Grammes



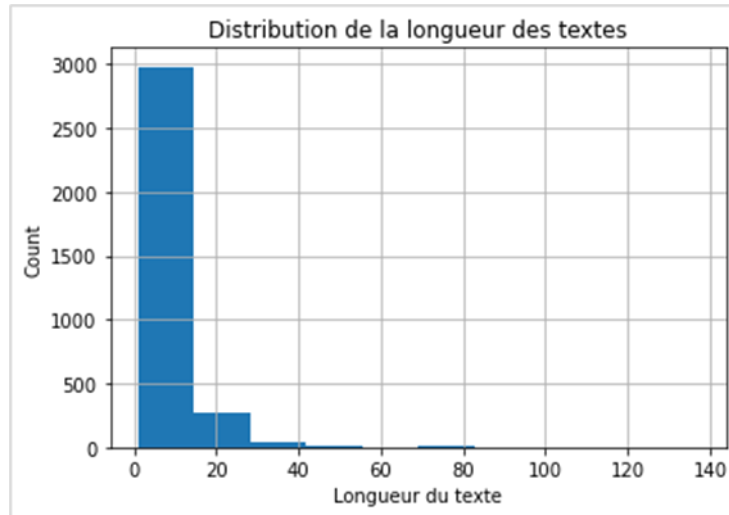
### 3-Grammes



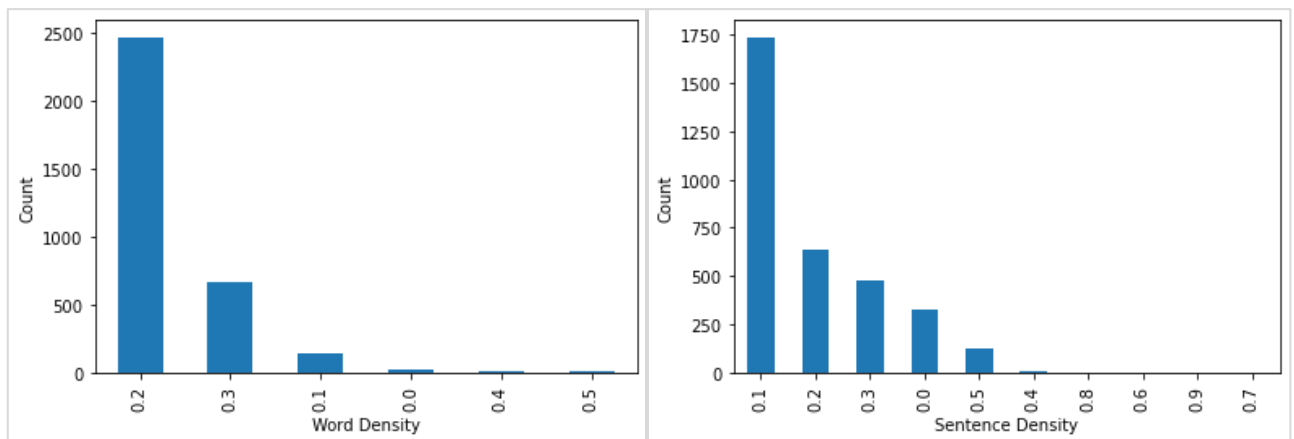
**Figure 4.4.** Les bi-grammes et les tri-grammes les plus fréquents dans l'ensemble de données  
« Rumeurs\_Commentaires »

### 2.5. Style linguistique du texte

Les figures suivantes illustrent le style linguistique des commentaires écrits par les *followers* des trois chaînes YouTube. La longueur du texte (nombre moyen de mots dans chaque commentaire) d'une grande partie des commentaires est située entre 0 et 25, sachant que la densité des mots (longueur moyenne des mots) atteint un niveau moyen de 20% et la densité de phrases (nombre de phrases par rapport au nombre de mots) et de 10% à 30%.



**Figure 4.5.** Distribution de la longueur des textes

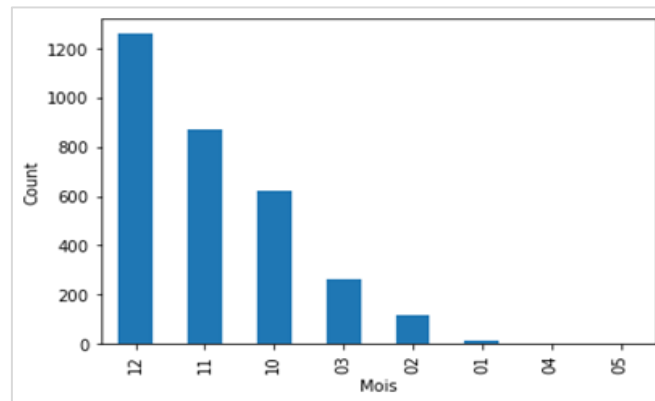


**Figure 4.6.** Densité des éléments textuelles des commentaires

### 2.6. Perspective temporelle

La figure 4.7 présente la distribution des commentaires collectées par rapport au mois de l'année appartenant à l'intervalle de temps [octobre 2020, mai 2021]. Nous remarquons que la majorité des commentaires ont été publiés dans les premiers mois de l'hospitalisation du président algérien Abdelmadjid Tebboune (octobre, novembre, décembre). Après une période

de quelques mois la rumeur a perdu son intérêt et a commencé à se déclinier jusqu'à sa disparition en mai 2021.



**Figure 4.7.** Distribution des commentaires par mois

### 3. Propagation des rumeurs basée sur le modèle SIR

Dans cette étape, nous nous concentrons sur l'extraction du modèle de propagation des rumeurs en se basant sur le modèle épidémique SIR que nous avons déjà présenté dans le deuxième chapitre. Nous avons étudié deux cas possibles de propagation :

- **1<sup>er</sup> cas** : le modèle de propagation SIR avec des paramètres personnalisés extraits de l'observation de la dynamique de la rumeur durant la période d'étude.
- **2<sup>ème</sup> cas** : le modèle de propagation SIR avec les paramètres principaux extraits directement à partir de données collectées.

#### 3.1. Variables et paramétrage

Etant donné le modèle de propagation SIR expliqué dans le Chapitre 02, nous définissons ci-après ses paramètres selon la nature des données collectées de la plateforme YouTube.

- $S = \text{nbr\_commentaires de type (Commentaire + Requête)} (t)$
- $I = \text{nbr\_commentaires de type (Supporter + Commentaire Positif)} (t)$
- $R = \text{nbr\_commentaires de type (Refuser + Commentaire Négatif)} (t)$
- $N = \text{Nbr. Total de population} = \text{Nbr\_Total des commentaires}$
- Les fractions des populations susceptibles, infectés et récupérés :  $\# \text{Susceptible} = S/N$ ,  $\# \text{Infected} = I/N$ ,  $\# \text{Recovered} = R/N$ ,  $\text{Total} = N$
- Intervalle temporel d'étude = 206 jours, de 27/10/2020 à 20/05/2021.

Rappelant que :

- $\beta$  détermine le **taux de transmission** : taux de contacts qu'un individu « Susceptible » peut avoir avec des personnes « Infectés ». Il est égal au nombre de contacts par jour divisé par le nombre total de la population.

**Exemple** : supposant qu'un individu a en moyen une probabilité de contacter 2 personnes « Infectés » par jour,  $\beta = 2/nbr\_population$ .

- $\gamma$  détermine le **taux de guérison**. Il est égal à :  $1/période\ d'infection$ .

**Exemple** : si la période d'infection d'un individu « Infecté » = 07 jours,  $\gamma = 1/7$ .

### 3.2. Modèle de propagation des rumeurs basé sur $\beta$ et $\gamma$ personnalisées

Pour tracer la courbe du 1<sup>er</sup> cas (Figure 4.8), nous avons proposé les valeurs suivantes pour les paramètres du modèle de propagation SIR :

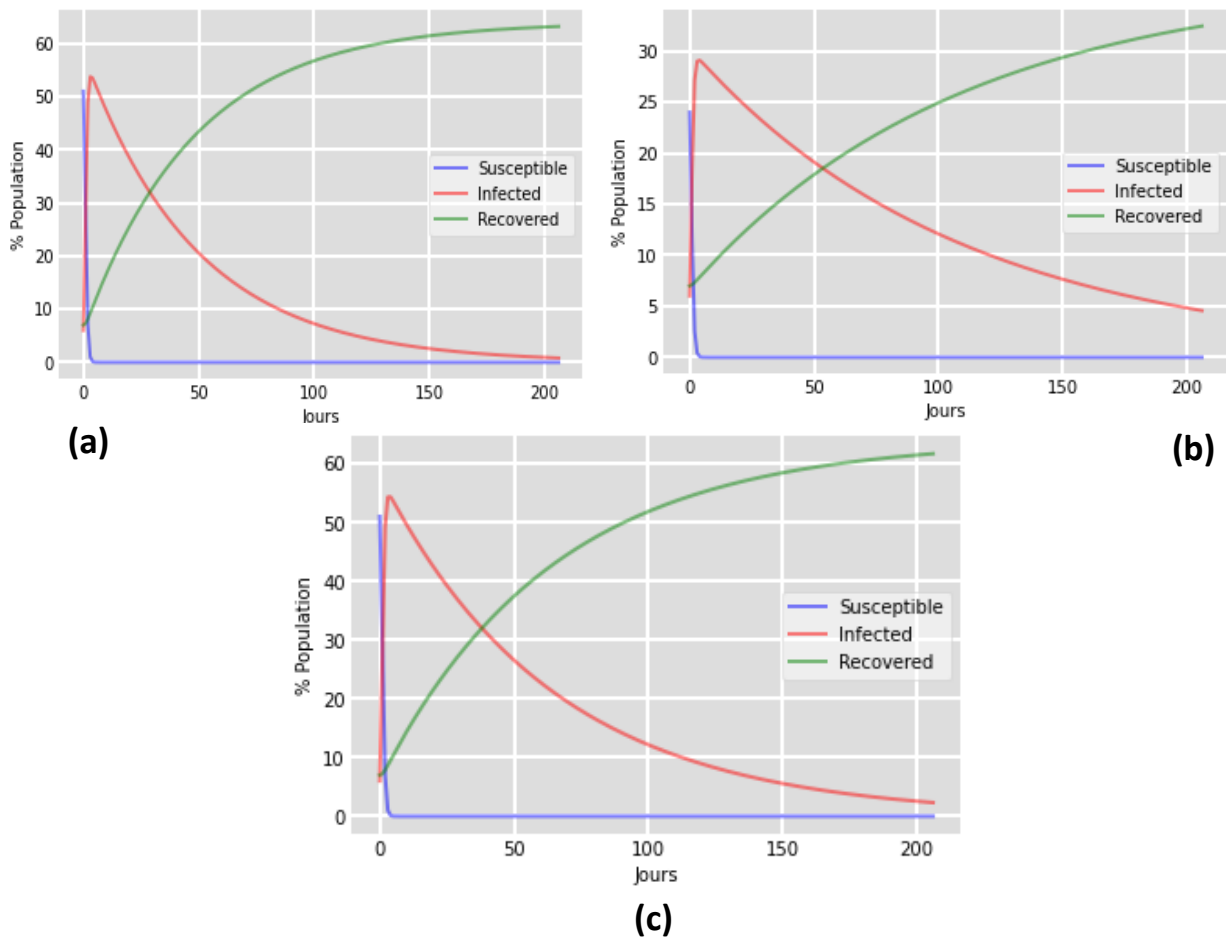
- $\beta = nbr\ moyen\ de\ commentaires\ de\ type\ (Supporter + Commentaire\ Positif)\ par\ jour / nbr\_total\_population$
- Nbr moyen de commentaires de type (« Supporter » + « Commentaire Positif ») par jour = Somme de commentaires (« Supporter » + « Commentaire positif ») / Nbr\_total\_jours.
- $\gamma = 1/Période\ d'infection$
- **Période d'infection**
  - ✓ Période d'infection 01 = **de 27/10/2020 à 12/02/2021** = 03 mois et 16 jours = 108 jours,  $\gamma = 1/108$ . (Figure 4.8 (a))
  - ✓ Période d'infection 02 = **de 27/10/2020 à 29/12/2020** = 02 mois et 02 jours = 63 jours,  $\gamma = 1/63$ . (Figure 4.8 (b))
  - ✓ Période d'infection 03 = **de 27/10/2020 à 13/12/2020** = 01 mois et 17 jours = 47 jours,  $\gamma = 1/47$ . (Figure 4.8 (c))

Rappelant que :

- La date **13/12/2020** correspond à la première apparition du président algérien Abdelmadjid Tebboune après son hospitalisation. Il s'était exprimé dans un discours à destination des algériens, promettant un retour 'dans les plus brefs délais'.
- La date **29/12/2020** correspond au retour du président Abdelmadjid Tebboune en Algérie après sa 1<sup>ère</sup> hospitalisation en Allemagne.
- La date **12/02/2021** correspond au retour final du président Abdelmadjid Tebboune en Algérie après sa 2<sup>ème</sup> hospitalisation en Allemagne.

## Chapitre 04 : Implémentation et résultats

Considérons les paramètres personnalisés mentionnés ci-dessus, nous remarquons une grande ressemblance entre les figures 4.8 (a, b, c) malgré la différence significative entre les périodes d'infection. En fait, les courbes montrent une décroissance rapide de l'ensemble « Infectés », une croissance rapide de l'ensemble « Récupérés » et un taux proche au nul de l'ensemble « Susceptibles » après une courte période qui marque le début la de propagation des rumeurs. Ces résultats suggèrent que les rumeurs au sujet de la santé du président n'étaient pas une préoccupation pour le public algérien. Or, cela ne peut pas être confirmé ou admis vu la sensibilité de ce sujet à cette période de crise politique et sanitaire. Nous pouvons donc constater que la distribution réelle des ensembles « Infectés », « Récupérés », et « Susceptibles » n'est pas liée aux deux facteurs que nous avons proposé qui sont les trois dates clés mentionnées précédemment et le nombre moyen de commentaires de type « Supporter » et « Commentaire Positif » par rapport au nombre total des commentaires.

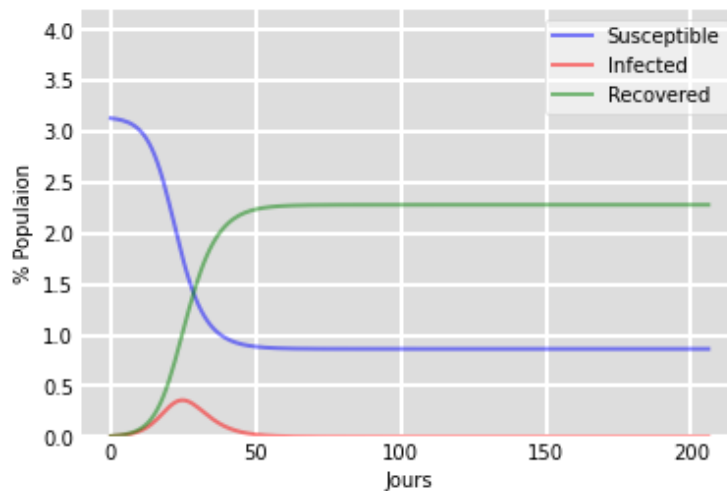


**Figure 4.8.** Courbe de propagation de la rumeur selon le modèle SIR avec  $\beta$  et  $\gamma$  personnalisés et les périodes d'infection a, b, et c

### 3.3. Modèle de propagation des rumeurs basé sur les valeurs déduites $\beta$ et $\gamma$

Notre objectif dans ce cas est de tracer la courbe de propagation des rumeurs étant donné les paramètres clés  $\beta$  et  $\gamma$  (Figure 4.8). Nous avons déduit la valeur moyenne de  $\beta$  et de  $\gamma$  à partir de la distribution et de la dynamique temporelle des ensembles  $S$  (Susceptibles),  $I$  (Infectés) et  $R$  (Récupérés) par rapport à l'intervalle temporel d'étude (206 jours).

Nous avons trouvé que  $\beta \sim 0.53$  et  $\gamma \sim 0.30$ .



**Figure 4.9.** Courbe de propagation de la rumeur selon le modèle SIR avec  $\beta$  et  $\gamma$  déduites

### 3.4. Discussion de modèle de propagation résultat

A partir de ce que nous avons comme résultats (modèle SIR basée sur des données réelles avec les paramètres  $\beta = 0.53$  et  $\gamma = 0.30$ ), nous pouvons observer que la partie décroissante « Susceptibles » est symétrique par rapport à la partie croissante « Récupérés » alors que la courbe représentant le nombre d'individus « Infectés » qu'ils étaient en contact avec la rumeur est en forme de cloche. Ce dernier ne cesse d'augmenter depuis le premier jour de publication de la rumeur « le 27 octobre 2020, le transfert de président algérien Abdelmadjid Tebboun en Allemagne » pour atteindre une valeur maximale vers le 25<sup>ème</sup> jour, puis il commence à se diminuer.

En fait, la population susceptible de supporter la rumeur selon laquelle le président algérien serait mort, décroît après son apparition dans un discours qui a été diffusé sur les réseaux sociaux (le 13/12/2020). C-à-d que le temps que la rumeur est contredite par une source officielle, la dynamique de propagation de la rumeur s'affaiblit progressivement. Au bout d'un temps suffisamment long, on remarque un très grand nombre de personnes récupérées (Recovered) et un nombre très réduit de personnes susceptibles.

## 4. Classification des attitudes à l'égard des rumeurs

Nous présentons dans cette section les étapes d'implémentation des modèles d'apprentissage par transfert pour la tâche de classification des positions de la communauté algérienne sur YouTube à l'égard des rumeurs à propos de la santé du président de la république.

### 4.1. Annotation de données

L'annotation de données a été effectuée manuellement suivant le système SDQC qui a été expliqué dans les chapitres précédents. Nous présentons ci-dessous quelques exemples :

Texte	Annotation
بارك الله فيك استاذ هشام على مصداقيتك	Supporter
اسسس جيت هنا حبيت نقلك تبون ماماتش	Refuser
بصح قولنا وشار تبون باش دارو عليه	Requête
الله. برحم. بوك. يرجع. ولا. ميرجعش اصلا. راه. خضرا. فوق طعام	Commentaire
صح خاطبه السياسة و حتى كلامه سوقي و لي يتابعوه 60 بالمئة منهم مراركة	Commentaire Négatif
أمير صدق ومصادقية عقدهم وذباب يدخل بيرر تحت ذبابة واحدة فاتح منتي كونت	Commentaire Positif

Tableau 4.5. Quelques exemples des commentaires annotés

### 4.2. Nettoyage et prétraitement

Le nettoyage et le prétraitement de données sont des tâches très importantes qui doivent intervenir pour corriger la qualité de données ou ignorer les données manquantes et erronées afin d'établir un modèle performant. Nous avons appliqué un ensemble des fonctions pour préparer et nettoyer les données brutes dans cet ordre :

#### ✓ Segmentation (Tokenization)

Texte brute	Texte après la segmentation
أني صدق و مصداقية	['أني', 'صدق', 'و', 'مصداقية']
بصح واين راه رائيس لي شافوه مشفنا والو	['بصح', 'واين', 'راه', 'رائيس', 'لي', 'شافوه', 'مشفنا', 'والو']
الله. برحم. بوك. يرجع. ولا. ميرجعش اصلا. راه. خضرا. فوق طعام	['الله', 'برحم', 'بوك', 'يرجع', 'ولا', 'ميرجعش', 'اصلا', 'راه', 'خضرا', 'فوق', 'طعام']
والله غير تحشم بهذا الاسلوب نتاع الكلام نتاع خاطيك السياسة	['والله', 'غير', 'تحشم', 'بهذا', 'الاسلوب', 'نتاع', 'الكلام', 'نتاع', 'خاطيك', 'السياسة']
Lah yarhmah	['lah', 'yarhmah']

Tableau 4.6. Exemples des commentaires après la segmentation

#### ✓ Suppression des mots vides (Stopwords)

Texte brute	Texte après la suppression des mots vides
['انسان', 'متقف', 'ماشاء', 'الله']	['انسان', 'متقف', 'ماشاء', 'الله']
['يجب', 'خروج', 'لي', 'عاصمة']	['خروج', 'عاصمة']

['واين', 'رائيس', 'شافوه', 'مشفنا', 'واو']	['واين', 'رائيس', 'شافوه', 'مشفنا']
['language', 'de', 'rue', 'essayes', 'd', 'être', 'un', 'vrai', 'journaliste']	['language', 'rue', 'essayes', 'vrai', 'journaliste']
['لا', 'حول', 'و', 'لا', 'قوة', 'لا', 'بالله']	['قوة', 'بالله']

**Tableau 4.7.** Exemples des commentaires après la suppression des mots vides

- ✓ Autres fonctions de nettoyage : conversion des lettres non arabe en minuscules, suppression de la ponctuation, des caractères spéciaux, des émojis et des signes diacritiques

Texte brute	Texte après la suppression des caractères spéciaux
ظهر الحق وزهق الباطل 🤖🤖	['ظهر', 'الحق', 'وزهق', 'الباطل']
['حسبنا', 'الله', 'ونعم', 'الوكيل', 'في', 'هاذو', 'المخلوقات']	['حسبنا', 'هادو', 'المخلوقات']
تبون، راه، غي، مستشفي، المانيا، ياس، ي، محمد، العربي، زيوط	['تبون', 'غي', 'مستشفي', 'المانيا', 'ياس', 'ي', 'محمد', 'العربي', 'زيوط']
INSPECTEUR CLOUSOU / MOUHARIJ / CLOWN	['inspecteur', 'clousou', 'mouharij', 'clown']

**Tableau 4.8.** Exemples des commentaires après le nettoyage

### 4.3. Implémentation des transformateurs

L'algorithme 4.1 décrit le processus de classification des attitudes à l'égard des rumeurs étudiées. L'implémentation a été effectuée sous Google Colab à l'aide du package « Transformers ».

---

#### Algorithme 4.1 : Algorithme de classification des attitudes

---

**Entrées :** commentaires\_à\_classifier, TransformateurPré-entraîné

**Objectif :** classer les commentaires en (*Supporter, Refuser, Commentaire, Commentaire Négatif, Commentaire Positif*)

1. Configurer l'environnement d'exécution sur « Google Colab » : confirmer l'existence du GPU pour charger le jeu de données.
  2. Nettoyage, pré-traitement et préparation du jeu de données.
  3. Décomposer le jeu de données en deux sous-ensembles : ensemble d'entraînement, ensemble de test.
  4. Télécharger le modèle transformateur pré-entraîné.
  5. Encoder les commentaires du jeu de données avec le « Tokenizer » du transformateur appliqué.
  6. Convertir les données de chaque sous-ensemble en type TensorDataset.
  7. Déterminer l'architecture du transformateur à appliquer.
  8. Entraîner le modèle.
  9. Évaluer le modèle sur l'ensemble de test.
  10. Évaluer la performance du modèle avec les métriques : Precision, Recall, F1\_Score.
-

## Chapitre 04 : Implémentation et résultats

Notant que nous avons défini les mêmes paramètres principaux (paramètres recommandés par les développeurs du BERT) pour tous les modèles pré-entraînés : mBERT (version : bert-base-multilingual-cased), XLM-R (version : xlm-roberta-base) et araBERT (version : aubmindlab/bert-base-arabertv02).

- ✓ Max Length = 256 (taille maximale d'une séquence textuelle à traiter).
- ✓ Batch Size = 16
- ✓ Epochs = 8
- ✓ Learning Rate = 5e-5

Nous avons évalué la performance de ces modèles à l'aide des métriques suivantes :

- **La précision** : (en anglais *precision*) : est définie comme le nombre de prédictions faites qui sont réellement correctes ou pertinent parmi toutes les prédictions basées sur la classe positive. Ceci est également connu comme **valeur prédictive positive** et peut être représentée par la formule [44] :

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

- **Le rappel** : (en anglais *recall*) : est défini comme le nombre d'instances de la classe positive qui étaient correctement prédit. Ceci est également connu sous le nom de **couverture** ou de **sensibilité** et peut être représenté par la formule [44] :

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

- **F1-Score** : est une autre mesure de précision qui est calculée en prenant la moyenne harmonique de la précision et du rappel et peut être représentée comme suit [44] :

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

### 4.4. Discussion des résultats

Modèles	Précision	Rappel	F1-Score
mBERT (bert-base-multilingual-cased)	0.45	0.45	0.45
XLM-Roberta (xlm-roberta-base)	0.39	0.39	0.38
araBERT (aubmindlab/bert-base-arabertv02)	<b>0.55</b>	<b>0.52</b>	<b>0.53</b>

**Tableau 4.9.** Les résultats obtenus lors de l'exécution des transformateurs

A partir de ce que nous avons comme résultats (Tableau 4.9), nous remarquons que le modèle araBERT donne des performances élevées par rapport aux autres modèles mBERT et

## Chapitre 04 : Implémentation et résultats

XLM-Roberta. Cela est due essentiellement à deux facteurs. En premier lieu, à son corpus linguistique purement arabe qui englobe plus de 24 GB de textes avec 64K comme taille de vocabulaire alors que BERT par exemple utilise 4.3 GB avec moins de vocabulaire [45]. Le deuxième facteur est le fait que araBERT a été entraîné pour résoudre des différentes tâches de traitement de la langue naturelle (NLP) comme l'analyse des sentiments (Arabic Sentiment Analysis) qui est très proche à notre problématique de classification des attitudes.

Comme conclusion préliminaire à notre expérimentation, nous constatons que la performance des modèles monolingues peut surpasser celle des modèles multilingues selon la langue traitée et le contexte d'étude. AraBERT est un choix pertinent pour les tâches de traitement automatique de la langue arabe en comparaison avec les modèles multilingues.

### Conclusion

Dans ce chapitre, nous avons montré les étapes d'implémentation et de réalisation de notre projet de fin d'étude qui consiste à l'étude de la diffusion des rumeurs au sujet de la maladie du président algérien Abdelmadjid Tebboune. Nous avons appliqué le modèle épidémiologique SIR sur nos données réelles collectées à l'aide de YouTube Scrapping API pour mieux comprendre comment les rumeurs ont propagé sur la plateforme YouTube. Nous avons ensuite appliqué l'apprentissage par transfert pour classifier les attitudes des utilisateurs algériens de YouTube à l'égard des rumeurs diffusées en moyennant deux transformateurs multilingues et un transformateur monolingue. L'évaluation de ces modèles a montré la capacité optimale du araBERT, le transformateur monolingue, à traiter les données textuelles pour produire des meilleurs résultats.

# **Conclusion générale**

---

La propagation des rumeurs sur les réseaux sociaux est un sujet sensible qui a attiré l'attention des chercheurs, compte tenu de son importance et de son influence ainsi que sa complexité.

Dans ce présent travail, nous avons étudié la diffusion de la rumeur politique sur les réseaux sociaux. Contrairement à beaucoup des études existantes qui se concentrent sur des simulations mathématiques à l'aide des différents types des modèles de diffusion d'information pour comprendre ce phénomène, nous avons abordé ici le problème d'un autre point de vue, en étudiant la diffusion à partir d'un cas d'étude réel qui s'inscrit dans le contexte politique et social algérien.

Dans ce cadre, nous avons entamé ce travail par une prospection des concepts de base qui sont liés à la propagation des rumeurs en ligne en passant par un état de l'art sur ce qu'a été déjà fait dans ce sens et en montrant ensuite notre système proposé pour l'étude de la diffusion des rumeurs et les attitudes de l'auditoire cible à son égard.

Comme le sujet du phénomène de propagation des rumeurs sur les réseaux sociaux est relativement récent, l'apport académique tenu par cette étude peut être considéré comme limité. Cependant, nous avons pu exploiter un ensemble de travaux existants sur la modélisation de diffusion de l'information et spécialement de la rumeur, que nous avons jugé utiles pour notre contribution. Nous avons également exploité l'apprentissage par transfert qui est une approche très récente dans le domaine de traitement automatique du langage naturel, pour étudier et redessiner les différentes réactions du public des médias sociaux envers ce phénomène. Nos résultats soulignent la robustesse du modèle épidémique de propagation SIR et la bonne performance du nouvel transformateur monolingue AraBERT par rapport aux modèles multilingues pré-entraînés mBERT qui était développé par Google, et XLM-Roberta qui était développé par Facebook.

Les défis majeurs rencontrés dans ce projet sont divers. Le défi le plus important concerne l'étape d'annotation de données. Cette étape a été effectuée manuellement à cause de l'absence des techniques efficaces et pratiques pour l'étiquetage automatique des données textuelles en langue arabe avec ses variations dialectales. En plus, le style linguistique incompréhensible ou non clair de beaucoup des commentaires sur YouTube rendent la tâche d'annotation plus laborieuse. Un autre problème concerne les variations linguistiques dans les commentaires collectés. En fait, le commentateur algérien sur YouTube écrit ses textes en arabe classique, dialecte arabe algérien, tamazight, français, anglais, arabizi (franco-arabe), et parfois en

d'autres dialectes arabes maghrébins ou du moyen-orient. Ces variations ont influencé la performance des transformateurs linguistiques qui ont été utilisés pour la classification automatique des positions des commentateurs à l'égard des rumeurs. La performance des modèles appliqués a été affectée aussi par la nature des données des médias sociaux qui rend déséquilibrée tout ensemble de données collectées à partir de ces plates-formes. Le résultat est que notre base de données a inclut un nombre très élevé des textes non pertinents de type « Commentaire » par rapport aux textes informatifs qui expriment clairement la position du commentateur (commentaires de type « Supporter », « Requête » ou « Refuser »).

En expliquant les difficultés d'étude du phénomène de propagation des rumeurs sur les médias sociaux dans le contexte algérien, nous considérons que ce travail ouvre d'autres perspectives intéressantes pour développer d'autres approches automatiques plus performantes pour appréhender cette problématique.

On espère aujourd'hui que nos objectifs ont été atteints à un niveau raisonnable, et que notre étude a pu aider le lecteur de notre mémoire à comprendre le phénomène des rumeurs en ligne avec sa complexité et ses possibilités d'être traité automatiquement à l'aide des techniques de modélisation mathématique et d'apprentissage automatique.

# Bibliographie

---

- [1] Peterson W., Gist N. (1951). Rumor and Public Opinion, *American Journal of Sociology*, 57, p. 159-167.
- [2] Knapp R. (1944). A Psychology of Rumor», *Public Opinion Quarterly*, 8 (1), 1944, p. 22-37.
- [3] Kapferer, J. N. (1987). Rumeurs : le plus vieux média du monde. Editions du seuil.
- [4] Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. (2018). Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys (CSUR)*, 51(2), 1-36.
- [5] S Zannettou, M Sirivianos, J Blackburn, N Kourtellis .(2019).The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans.- *Journal of Data and Information Quality(JDIQ)*,vol.11,no 3, p. 1-22.
- [6] Zafarani, R., Abbasi, M. A., and Liu, H. (2014). Social Media Mining: An Introduction. Cambridge University Press, New York, NY, USA.p. 1-2
- [6] Zafarani, R., Zhou, X., Shu, K., & Liu, H. (2019, July). Fake news research: Theories, detection strategies, and open problems. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 3207-3208).
- [7] Reilly, I. (2018). F for fake: Propaganda! Hoaxing! Hacking! Partisanship! and Activism! in the fake news ecology. *The Journal of American Culture*, 41 (2), 139–152. doi:10.1111/jacc.12834.
- [8] Mohseni, S., Ragan, E., & Hu, X. (2019). Open issues in combating fake news: Interpretability as an opportunity. *arXiv preprint arXiv:1904.03016*.
- [9] Thomas Saint-Cricq.(2020,10 Décembre).Covid-19 vaccines don't change your DNA.*Service fact-checking de l'AFP*. Repéré à <https://factcheck.afp.com/vaccines-dont-change-your-dna>
- [10] mlive Michigan.(2021,14 Avril).No, the COVID-19 vaccine will not change your DNA.*Journal électronique MLive Michigan*. Repéré à <https://www.mlive.com/public-interest/2021/04/no-the-covid-19-vaccine-will-not-change-your-dna.html>

- [11] Jack Goodman and Flora Carmichael. (2020,30 Mai),Coronavirus: Bill Gates ‘microchip’ conspiracy theory and other vaccine claims fact-checked. *BBC Reality Check*. Repéré à <https://www.bbc.com/news/52847648>
- [12] GreatgameIndia(2020, 21 Décembre).Researchers Warn COVID-19 Vaccines May Cause HIV .*GreatGameIndiaInfection.Journal on Geopolitics & International Relations*. Repéré à <https://greatgameindia.com/covid-19-vaccines-hiv/>
- [13] Dr. Sanchari Sinha Dutta.(2021,10 Mars).Quels sont les vaccins basés sur adénovirus ?.*News Medical Life Sciences*. Repéré à [https://www.news-medical.net/health/What-are-Adenovirus-Based-Vaccines-\(french\).aspx](https://www.news-medical.net/health/What-are-Adenovirus-Based-Vaccines-(french).aspx)
- [14] P Meel, DK Vishwakarma.(2020).Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities, *Expert Systems with Applications*,p8-11.
- [13] Nicolas Turenne (2018). The rumour spectrum. *Plos one*, p9-10
- [13] M Li, X Wang, K Gao, S Zhang.(2017).A survey on information diffusion in online social networks: Models and methods.*Information*,p4-7
- [14] NP Nguyen, G Yan, MT Thai. (2013). Analysis of misinformation containment in online social networks with Aid of celebrities. *Computer Networks*, p4
- [15] JJ Cheng, Y Liu, B Shen, WG Yuan. (2013).An epidemic model of rumor diffusion in online social networks.*The European Physical Journal B*,p3
- [16] Granovetter M. (1978). Threshold models of collective behavior. *Am J Sociol* 83(6),p1420–1443
- [17] Goldenberg J, Libai B, Muller E (2001b) Talk of the network: a complex system look at the underlying process of word-of-mouth. *Mark Lett* 12(3):211–223
- [18] Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772), 700-721.
- [19] Bartlett, M. S. (1949). Some evolutionary stochastic processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2), 211-229.
- [20] Kendall, D. (1956). Deterministic and Stochastic Epidemics in Closed Populations.

- [21] W Huang.(2011).On rumour spreading with skepticism and denial, *Working paper*,p1
- [22] D Küçük, Fazli Can.(2020).Stance Detection : A Survey, *ACM Computing Surveys (CSUR)*,p2-18.
- [23] Yiwei Zhou, Alexandra I. Cristea, and Lei Shi. (2017). Connecting targets to tweets: Semantic attention-based model for target-specific stance detection. *In Proceedings of the International Conference on Web Information Systems Engineering*. p18–32.
- [24] Bifet, Albert, et al. (2009). New ensemble methods for evolving data streams. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, p1-2.
- [25] Zarrella, Guido, and Amy Marsh. (2016). Mitre at semeval-2016 task 6: Transfer learning for stance detection. *arXiv preprint arXiv:1606.03784*, p1-2
- [26] Pétition : Libérez Amir DZ. Repéré à <https://www.mesopinions.com/petition/droits-homme/liberez-amir-dz/95596>
- [27] Wikipédia, Mohamed Larbi Zitout. Repéré à [https://fr.wikipedia.org/wiki/Mohamed\\_Larbi\\_Zitout](https://fr.wikipedia.org/wiki/Mohamed_Larbi_Zitout)
- [28] Wikipédia, Hichem Aboud. Repéré à [https://fr.wikipedia.org/wiki/Hichem\\_Aboud](https://fr.wikipedia.org/wiki/Hichem_Aboud)
- [29] Zhuang, Fuzhen, et al. (2020). A comprehensive survey on transfer learning. *Proceedings of the IEEE* 109.1,p1,3
- [30] Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), p1345-1359.
- [31] Meftah, Sara, et al. (2019). Exploration de l'apprentissage par transfert pour l'analyse de textes des réseaux sociaux. *26e Conférence sur le Traitement Automatique des Langues Naturelles*. ATALA, p3
- [32] Denis Rothman. (2021). Transformers for Natural Language Processing: Build innovative deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and more, *Packt Publishing Ltd* .p5
- [33] Olivier. (2020). À la découverte du Transformer. leDataScientist.IA. Repéré à <https://ledatascientist.com/a-la-decouverte-du-transformer/>

- [34] Wikipédia. BERT (language model). Repéré à [https://en.wikipedia.org/wiki/BERT\\_\(language\\_model\)](https://en.wikipedia.org/wiki/BERT_(language_model))
- [35] Rani Horev (2019).XLM — Enhancing BERT for Cross-lingual Language Model. Cross-lingual Language Model Pretraining. Repéré à <https://towardsdatascience.com/xlm-enhancing-bert-for-cross-lingual-language-model-5aeed9e6f14b>
- [36] Pierre-Jean Larpin,A-S.et Tom St.(2021).GPT-3, la dernière révolution du NLP ?.Quant Blog des Experts Data & IA. Repéré à <https://www.quantmetry.com/blog/analyse-gt3/>
- [37] J Libovický, R Rosa, A Fraser.(2019).How language-neutral is multilingual BERT? », *arXiv preprint arXiv:1911.03310*,p1
- [38] W Antoun, F Baly, H Hajj,(2020).Arabert: Transformer-based model for arabic language understanding», *arXiv preprint arXiv:2003.00104*
- [39] Wikipédia,Python. Repéré à [https://fr.wikipedia.org/wiki/Python\\_\(langage\)](https://fr.wikipedia.org/wiki/Python_(langage))
- [40] Wikipédia. Jupyter. Repéré à <https://fr.wikipedia.org/wiki/Jupyter>
- [41] Natural Language Toolkit.(2009). APA Style. Repéré à <https://www.nltk.org/>
- [42] Taha Zerrouki. (2010). PyArabic. Repéré à <https://pypi.org/project/PyArabic>
- [43] Nagesh Singh Chauhan. (2021). Hugging Face Transformers Package – What Is It and How To Use It.*KD Nuggets*. Repéré à <https://www.kdnuggets.com/2021/02/hugging-face-transformer-basics.html>
- [44] Cheyma TAGUIA,Zahra SID.(Septembre 2019). Détection de fausses informations sur le web et les réseaux sociaux, *Université Mohamed El Bachir El Ibrahimi de Bordj Bou Arréridj*,
- [45] W Antoun, F Baly, H Hajj.(2020).«AraBERT: Transformer-based Model for Arabic Language Understanding»,*arXiv preprint arXiv:2003.00104*,p5