RAPPORT DU PROJET IA

Thème du projet : Fake news (le vrai du faux)



**Maxime Hou**

**Kongotehi Israël Dedieu M’BASSIE**

**Eddy El KHOURY**

**Jerry DOMBET**

M1-JV-DS2

15/05/2024

**TABLE DES MATIÈRES**

[**RÉSUMÉ**](#_9eat6k866g5f) **2**

[**INTRODUCTION**](#_yspy8tt3f0xe) **3**

[**HYPOTHÈSE**](#_75rf4vta81ax) **4**

[**DONNÉES**](#_lquiyrwpy6ke) **4**

[**PROBLÉMATIQUE**](#_kn5uvgo00ajj) **8**

[**RÉSULTATS**](#_o8rmzovhszmh) **9**

[**CONCLUSION**](#_chou9188p6co) **13**

[**RÉFÉRENCES**](#_upsdn5xevax7) **14**

# 

# 

# 

# 

# 

# 

# 

# 

# RÉSUMÉ

Le but de notre projet est d’identifier les fake news parmi une liste de News. Afin de résoudre ce problème nous avons pris une base de données contenant à la fois de vraies news et des fake news. Nous avons ensuite combiné ces deux bases puis nous avons essayé de retrouver les fake news en faisant une analyse sur cette base.

Tout d’abord, nous avons fait une analyse exploratoire des données afin d’avoir une idée de comment identifier les fake news. Nous avons aussi fait une étude des différentes catégories d’articles sous formes d’histogrammes, cela permet d’avoir une idée du domaine où les fake news sont le plus présentes.

Ensuite, une transformation TF-IDF (frequency-inverse document frequency) a été effectuée afin d’identifier les termes qui reviennent souvent (notamment dans les fake news). En effet, lors de notre exploration du jeu de données nous avons pu voir que les fake news étaient rédigées de manière différente à celle des faits réels, de plus elles sont plus ou moins écrites de la même façon. C’est pourquoi une transformation TF-IDF est utile dans ce cas.

Après cette transformation nous avons utilisé des techniques de machine learning afin de trier le vrai du faux, pour ce faire nous sommes partie sur l’algorithme de K-means afin de séparer les données en 2 clusters : les “real news” et les “fake news”.

# 

# INTRODUCTION

Nous vivons actuellement dans une ère où la technologie est forte et dominante, néanmoins cette évolution a des conséquences qui ne sont pas toujours idéales en particulier dans le domaine de l’information.

Avant la croissance exponentielle des nouvelles technologies, l’information était principalement distribuée par journaux, il était donc moins plausible de tomber dans la misinformation. Cependant, suite à la croissance technologique de nouveaux outils ont surgit : la télévision, les presses en ligne, les réseaux sociaux et actuellement l’intelligence artificielle.

Nous nous intéresserons principalement aux effets des réseaux sociaux. En effet, les réseaux sociaux dominent très fortement notre société, quasiment tout le monde y est accro et consomme les informations y circulant sans réellement y réfléchir, une étude a montré qu’en 2019, 42% des français utilisent les réseaux sociaux pour s’informer. Cela est source de manipulation et de mensonge avec les fake news, de plus avec l’apparition de l’intelligence artificielle, énormément de trucage ont lieu.

Ainsi, notre projet consiste à étudier le taux de fake news dans les réseaux sociaux par rapport aux sources d’informations traditionnelles, d’où notre problématique : les réseaux sociaux sont-ils réellement une source fiable d’informations ?

Nous répondrons à cette problématique à travers une analyse sur des news provenant de différentes sources que ce soit des réseaux sociaux ou des sites de presse. Nous allons comparer le taux de fake news par rapport aux news réelles et voir à quel point les réseaux sociaux sont fiables, notamment Twitter/X qui est souvent source de misinformation.

# 

# HYPOTHÈSE

Il est logique de penser que les réseaux sociaux sont les sources d’informations les moins fiables, la raison est que ce sont principalement des gens de tous les jours comme nous qui écrivons. De ce fait l’information n’est pas forcément recherchée et elle est souvent source de rumeur ou bien créée de toute pièce afin de faire le buzz.

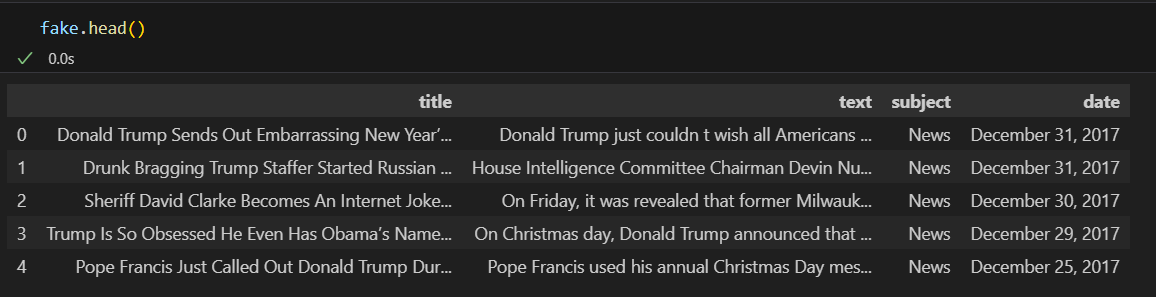
Ainsi, notre hypothèse est que les réseaux sociaux, en particulier Twitter ne sont pas des sources fiables quant à la récupération d’informations.

# DONNÉES

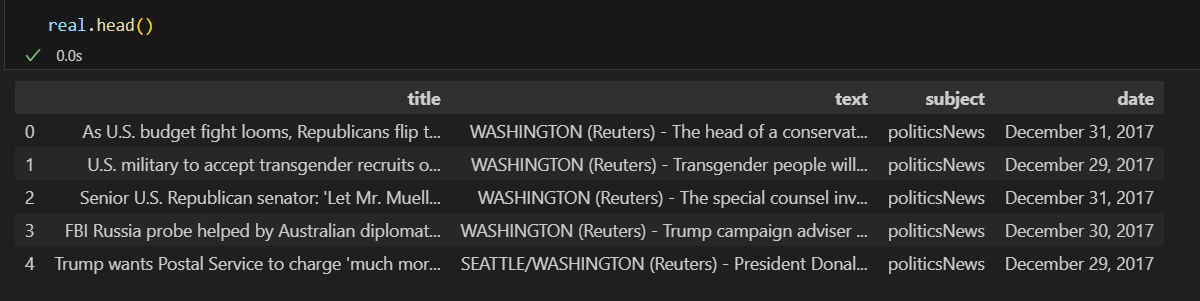
Initialement, nous avions prévu de récupérer une liste de News provenant directement de Twitter en utilisant Tweepy. Néanmoins nous avons rencontré des contraintes techniques : récupérer des tweets via Tweepy requiert un compte premium à Twitter API.

Ainsi pour notre projet, nous avons décidé de prendre des données provenant de deux fichiers issus de Kaggle: **Fake.csv** contenant 23 502 fausses informations et **True.csv** contenant 21 417 vraies informations, ces données traitent de diverses News datant de 2017.

Fake :

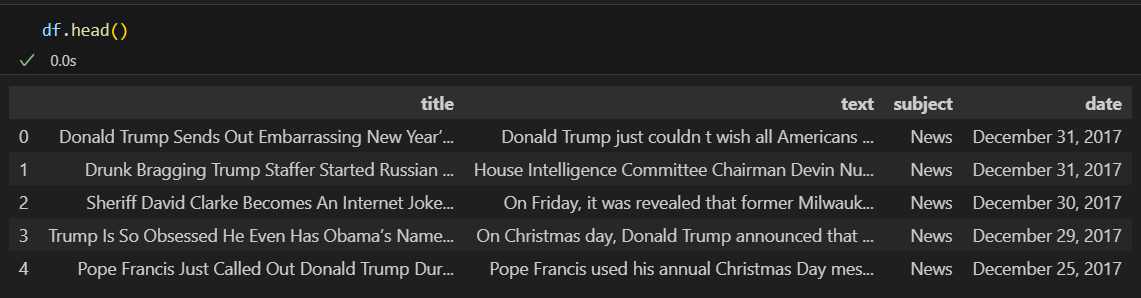
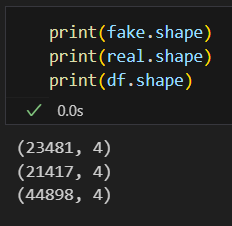


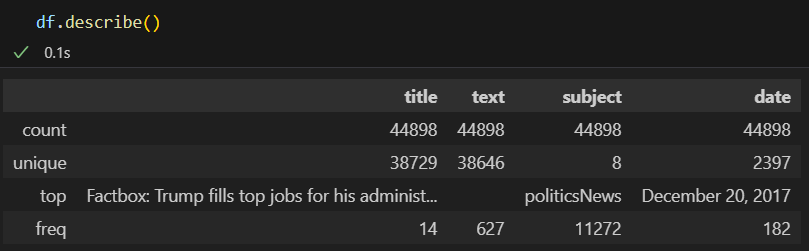
True :

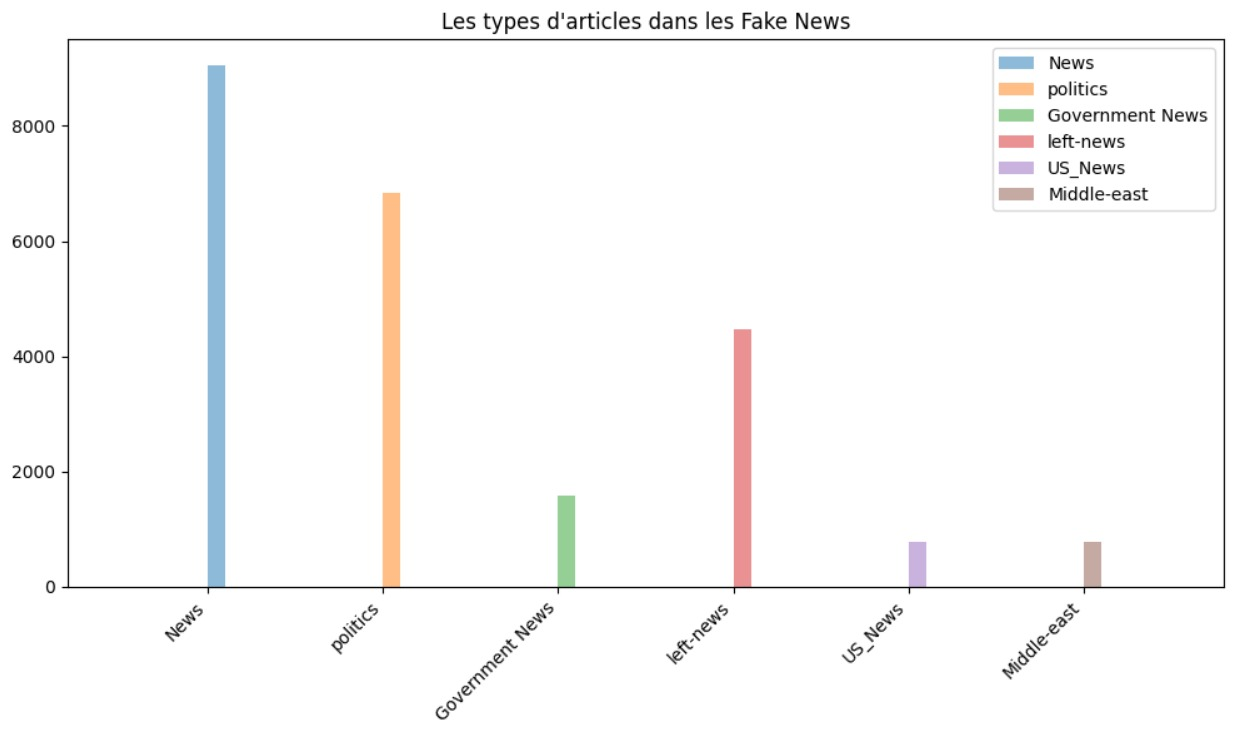


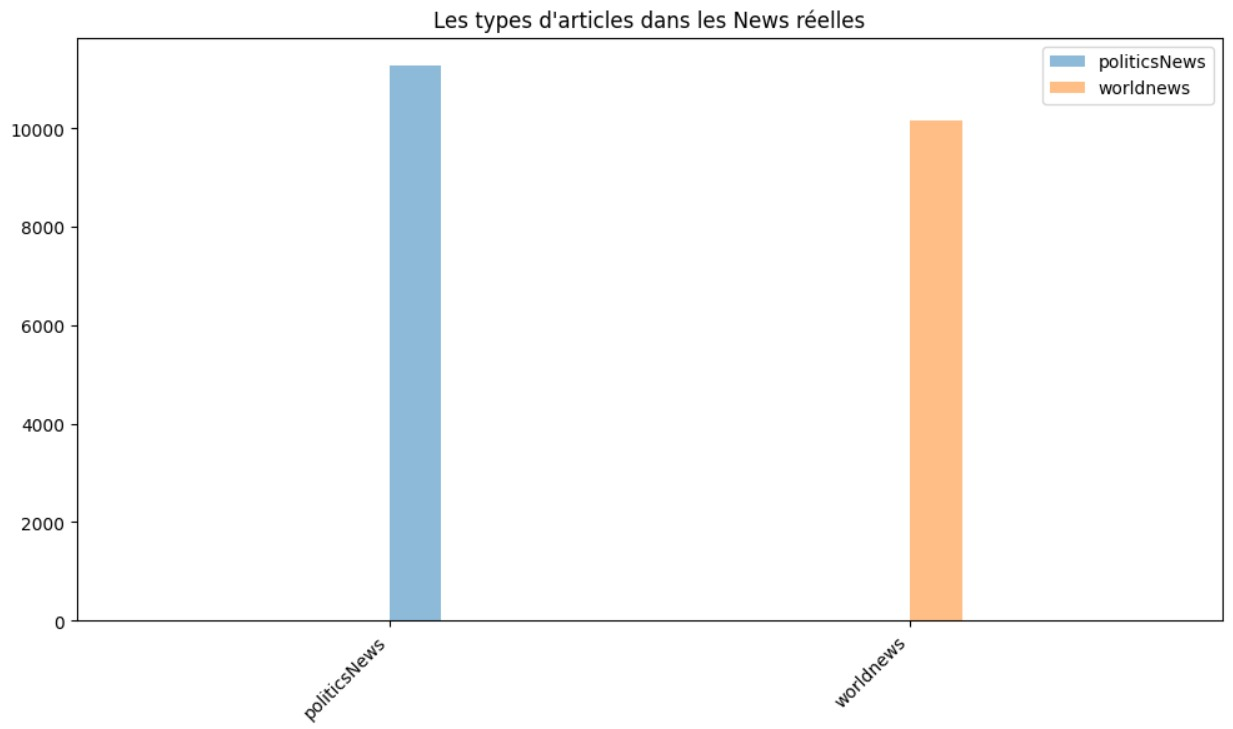
Avant même d’effectuer une analyse grâce aux outils d’intelligence artificielle, nous pouvons déjà voir une différence d’écriture entre les fake news et les faits réels.

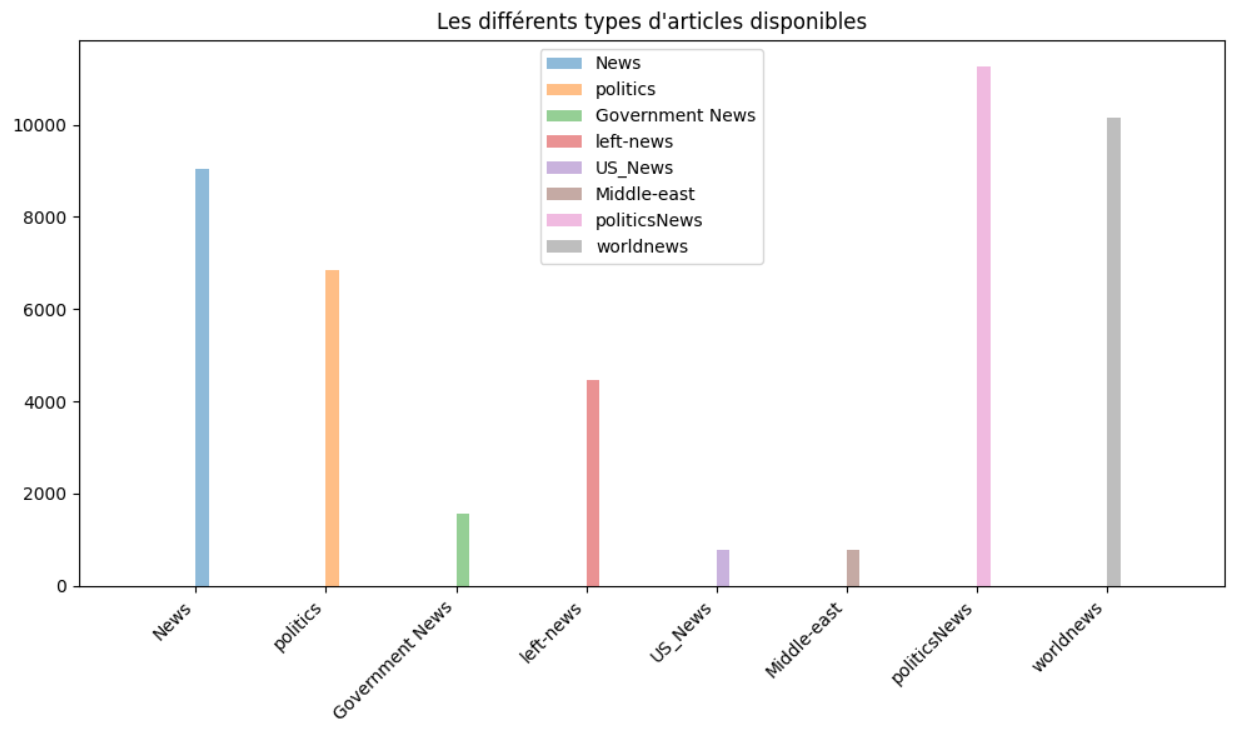
Chaque fichier comprend des colonnes pour le titre, le texte informatif, le sujet et la date de publication. Une exploration visuelle initiale montre une répartition variée des sujets, avec une prédominance pour la politique en général. C’est notamment en politique qu’il semble y avoir le plus de Fake News.











# PROBLÉMATIQUE

La problématique de la désinformation et des fake news sur les réseaux sociaux est un sujet de préoccupation croissant depuis plusieurs années. De nombreuses études ont été menées pour comprendre l'ampleur du phénomène et trouver des solutions pour y remédier.

Selon une étude menée par le Pew Research Center en 2019, environ deux tiers des Américains (64%) pensent que les fake news ont un impact important sur la confiance du public dans le gouvernement. De plus, une étude menée par l'Université de Stanford en 2016 a révélé que les élèves du secondaire et du collège ont du mal à distinguer les fake news des vraies informations en ligne.

Face à ce constat, de nombreuses initiatives ont été prises pour lutter contre la désinformation en ligne. Les géants des réseaux sociaux tels que Facebook et Twitter ont mis en place des mesures pour limiter la propagation des fake news sur leurs plateformes. Nous avons par exemple désormais des “notes communautaires” au sein de Twitter qui sont rédigés par des tiers afin de discréditer les fausses informations.

Cependant, malgré ces efforts, la désinformation reste un problème persistant sur les réseaux sociaux.

Conséquemment, le but central de notre projet est de déterminer la mesure dans laquelle les réseaux sociaux peuvent être considérés comme des sources fiables d’informations.

L'Intelligence Artificielle sera utilisée pour identifier les fausses informations des vraies informations en identifiant les similitudes entre chaque fake news, car en effet celles-ci semble suivre un certain pattern. De ce fait, grâce à une transformation TF-IDF, nous utiliserons de l’apprentissage non supervisé afin d’identifier ces similitudes et ainsi espérer retrouver les fake news en les séparant dans un groupe à part via un algorithme de clustering : celle du K-Means.

Une fois ces analyses effectuées par l’IA, nous pourrons identifier le nombre de fake news ainsi que celui des faits réels. Cela va nous permettre de calculer le taux de fiabilité (nombre de faits réels diviser par le nombre total) des sources à notre disposition, notamment celui des réseaux sociaux.

# RÉSULTATS

L’analyse de données pour ce projet à été concentrée sur l’utilisation de machine learning avancées pour classifier et comprendre la nature des informations diffusées sur les réseaux sociaux. En utilisant les datasets **Fake.csv** et **True.csv** de Kaggle, notre étude a employé le modèle K-means pour clusteriser les articles et la PCA pour réduire la dimensionnalité et visualiser les résultats.

1. **Description du modèle et de sa configuration**

Le **TfidfVectorizer** est configuré pour extraire 5000 termes les plus significatifs du corpus de textes, donnant ainsi un poids à chaque terme en fonction de sa fréquence dans un document et son inverse de fréquence dans l’ensemble du corpus. Cette démarche transforme le texte brute en un ensemble de caractéristiques, utilisable pour le clustering.

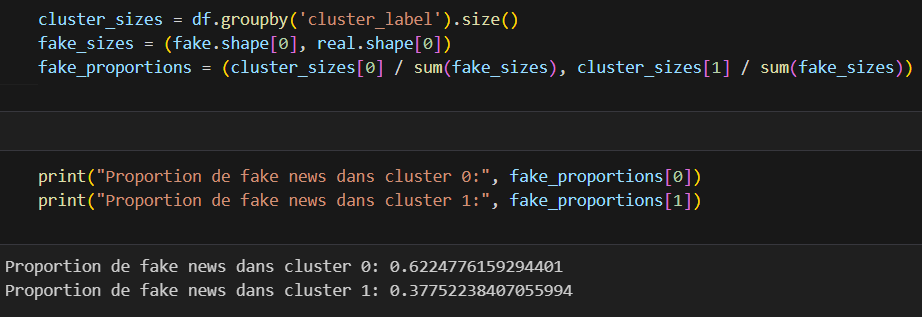
Le **clustering K-means** est une technique d’apprentissage automatique non supervisée utile pour regrouper des données en clusters selon leur similarité. L'algorithme K-means sélectionne initialement K points centraux (appelés centroïdes) au hasard dans l'ensemble de données, puis attribue chaque point de donnée au centroïde le plus proche.

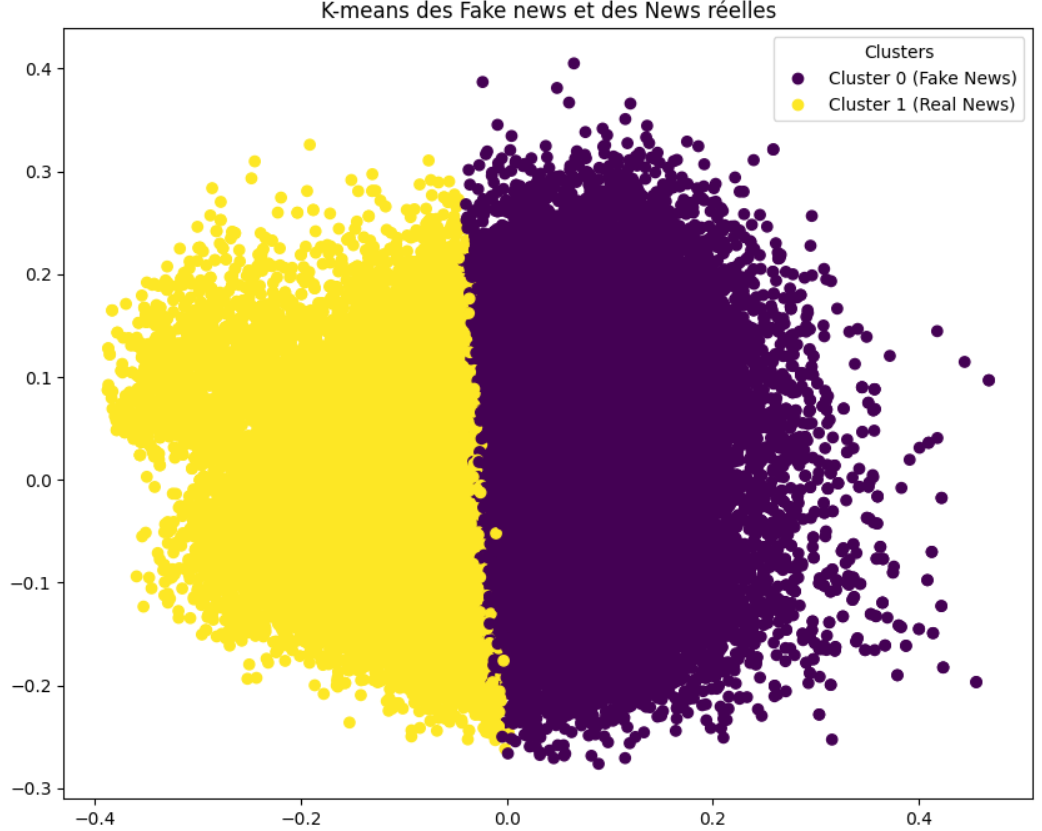
Dans notre étude, cette technique est appliquée avec un paramètre tel que le résultat donnera deux clusters reflétant nos catégories anticipées de “vraies” et “fausses” informations. Le choix de deux clusters est basé sur la première hypothèse qui stipule que les articles se divisent naturellement en ces deux groupes. L’algorithme a été initialisé avec un état aléatoire fixé pour garantir la reproductibilité des résultats.

1. **Résultats du clustering**

Après l’application du K-means , les articles ont été répartis en deux clusters principaux. Le premier cluster (Cluster 0) regroupe en grande majorité les articles identifiés comme fausses informations, alors que le second cluster (Cluster 1) contient en majorité les les articles identifiés comme vraies informations. Cependant une analyse plus détaillée à révélé une certaine contamination entre les clusters:

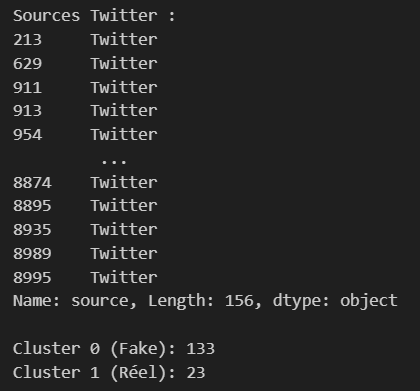
* Le **Cluster 0** à été trouvé pour contenir 62% de toutes les fausses informations mais également 38% de vraies informations.
* Le **Cluster 1** a montré une proportion inverse avec 37,7 % de fausses informations contre 62,3 % de vraies informations

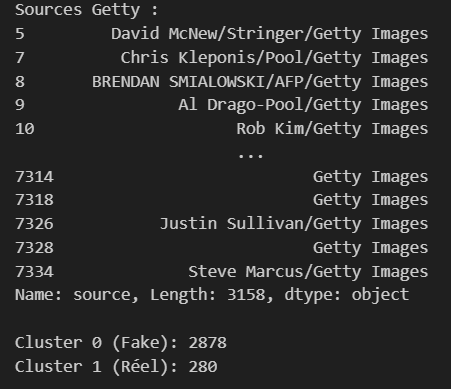


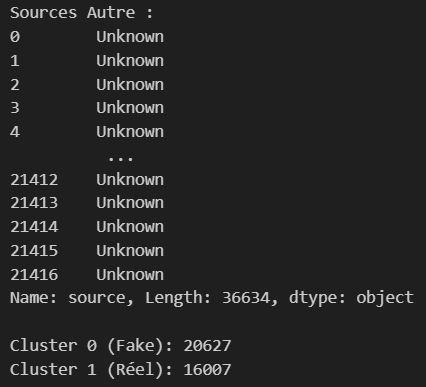


Cette distribution montre que, bien que le modèle puisse identifier des caractéristiques générales des fausses informations et vraies informations, il existe une zone de chevauchement significative où les fausses informations imitent stylistiquement ou thématiquement les vraies informations.

Après avoir obtenu nos clusters, nous avons regrouper les informations selon leurs sources (néanmoins certaines informations ne disposent pas de sources, ils seront donc catégorisé comme “Unknown”) :







Grâce à ces chiffres nous pouvons étudier le niveau de fiabilité :

* **Twitter** : réel/total => 23/(133+23) = 0.147 soit une fiabilité de 14.7%
* **Getty Images** : réel/total => 280/(2878+280) = 0.089 soit une fiabilité de 8.9%
* **Sources autres** : réel/total => 16007/(20627+16007) = 0.437 soit une fiabilité de 43.7%

Ainsi, nous pouvons voir que les réseaux sociaux ont un taux de fiabilité très faible, il est donc préférable d’éviter ces outils pour consommer de l’information. Néanmoins notre jeu de données ne dispose pas suffisamment de valeurs pour Twitter ou Getty, il est donc possible que les résultats soient erronés mais cela donne quand bien même un aperçu.

1. **Analyse du PCA et visualisation**

La PCA est utilisée pour réduire la dimensionnalité des données transformées par TfidfVectorizer à deux composantes principales, facilitant ainsi la visualisation des clusters. Les graphiques de dispersions générés montrent que, bien qu’il y ait une séparation notable entre les clusters, il existe une zone de mélange où les deux types d’informations se chevauchent.

1. **Comparatif avec d’autre modèle d’évaluation**

En comparaison avec d’autres modèles d’évaluation tels que la régression logistique et les réseaux de neurones, le K-means a montré des avantages en termes de simplicité et de vitesse d’exécution. Cependant, il a été moins efficace pour gérer les nuances et les subtilités des textes. Les modèles basés sur les réseaux de neurones, par exemple, ont montré une meilleure capacité à distinguer les nuances linguistiques qui caractérisent souvent les forces informations, bien que leur complexité et leurs exigences en ressources soient nettement plus élevées.

1. **Conclusion des résultats**

Les résultats indiquent que, si le clustering K-means est un outil utile pour une première analyse exploratoire de grandes quantités de données textuelles, il doit être complété par des analyses plus fines et des modèles plus sophistiqués pour améliorer la précision de la classification des informations en tant que vraies ou fausses. Cette étude souligne l'importance d'utiliser une combinaison de méthodes pour aborder la complexité de la détection des fausses informations sur les réseaux sociaux.

Nous pouvons aussi conclure (bien que la taille des échantillons n’est pas très grande) qu’en général, la source de fake news provient de réseaux sociaux ainsi l’utilisation de ceux-ci n’est pas idéal pour trouver des informations fiables, il vaut mieux utiliser des sources traditionnelles telles que des sites de presse officiels.

# 

# CONCLUSION

En conclusion, notre étude a mis en évidence la complexité de l'utilisation des réseaux sociaux en tant que sources fiables d'information. Bien que l'Intelligence Artificielle offre des outils puissants pour la détection des fake news, tels que le clustering K-means et la PCA, les taux d'erreur et les défis techniques persistent.

Les résultats de notre analyse montrent que les réseaux sociaux, en particulier Twitter, ont un taux de fiabilité très faible. Cela signifie que les utilisateurs doivent faire preuve de prudence lorsqu'ils consomment de l'information à partir de ces plateformes. Il est recommandé d'utiliser des sources traditionnelles d'information, telles que les sites de presse officiels, pour obtenir des informations fiables.

Cependant, il est important de noter que notre jeu de données ne disposait pas d'un nombre suffisant de valeurs pour Twitter ou Getty Images. Par conséquent, il est possible que les résultats soient erronés. Néanmoins, notre étude fournit un aperçu intéressant de la fiabilité des réseaux sociaux en tant que sources d'information.

De plus, notre étude recommande une approche multilatérale pour améliorer la fiabilité des informations sur les réseaux sociaux. Cette approche devrait combiner la technologie de l'IA et la vigilance humaine. Les utilisateurs doivent être éduqués sur la façon de repérer les fake news et de vérifier les informations avant de les partager. Les plateformes de réseaux sociaux devraient également prendre des mesures pour lutter contre la propagation des fake news, telles que l'utilisation de notes communautaires et la suppression des comptes qui propagent régulièrement de fausses informations.

Enfin, notre étude souligne l'importance d'utiliser une combinaison de méthodes pour aborder la complexité de la détection des fake news sur les réseaux sociaux. Les futures recherches devraient se concentrer sur le développement de modèles plus sophistiqués pour améliorer la précision de la classification des informations en tant que vraies ou fausses.

# RÉFÉRENCES

* <https://www.lesechos.fr/tech-medias/medias/aux-etats-unis-deux-tiers-des-adultes-sinforment-via-les-reseaux-sociaux-139220>
* <https://www.lemonde.fr/pixels/article/2016/11/23/fausses-informations-en-ligne-les-adolescents-facilement-dupes-selon-une-etude_5036468_4408996.html>
* <https://www.csa.fr/Informer/Collections-du-CSA/Focus-Toutes-les-etudes-et-les-comptes-rendus-synthetiques-proposant-un-zoom-sur-un-sujet-d-actualite/La-propagation-des-fausses-informations-sur-les-reseaux-sociaux-etude-de-la-plateforme-Twitter>