28/05/2024

Rapport projet IA

Detection de Fake News sur du texte

GROUPE LM- M1 - IAS - DA1

- AIT ABBAS Lydia
- MESSAADI Imene

2023/2024

Sommaire

Tables des figures
Liste des abréviations
Introduction Générale 1
Résumé 2
Introduction.....3
Données.....8
1. Visualisation des données.....8
Problématique.....11
1. Introduction.....11
2. A quoi l'IA va - t - elle répondre ?.....11
3. L'état de l'existant.....12
4. Conclusion.....13
Résultats.....14
1. Introduction.....14
2. L'environnement de travail.....14
3. Modèle proposé.....14
4. Fonctionnement du modèle.....15
5. Comparaison à d'autres modèles.....18
6. Conclusion 22
Conclusion.....23
Source.....24

Table des figures

Figure 1 : Types de Fake News 4
Figure 2 : L'exposition aux Fake News dans le monde 5
Figure 3 : Lois et articles sanctionnant les Fake News 6
Figure 4 : Distribution des news 8
Figure 5 : Longueur des news 9
Figure 6 : l'algorithme SVM 15
Figure 7 : Importation des bibliothèques 15
Figure 8 : Méthode de prétraitement du modèle 16
Figure 9 :Vecteurs de caractéristiques avec TF-IDF 17
Figure 10 :Explications d'un vecteur de caractéristiques TF-IDF obtenu 17
Figure 11 : Matrice de confusion: Matrice de confusion
.....18
Figure 12 : Exemple de la matrice de confusion 18
Figure 13 : Matrice de confusion du modèle Regression Logestic 19
Figure 14 : Matrice de confusion du modèle Decision Tree Classifier 19
Figure 15 : Matrice de confusion du modèle Support Vector Machin 19
Figure 16 : Indicateurs du modèle 'Decision Tree Classifier" 20
Figure 17 : Indicateurs du modèle "SVM" 20
Figure 18 : Taux d'erreur du modèle SVM 20
Figure 19 : Taux d'erreur du modèle SVM 21
Figure 20 : Courbe ROC pour un autre modèle 21

Liste des abréviations

IA : Intelligence Artificielle

SVM : Support Vector Machine

CSA : Conseil Supérieur de l'Audiovisuel

NLP : Natural Language Processing

POS : Part Of Speech

NLTK : Natural language Tool-kit

TF-IDF : Term Frequency - Invers Document Frequency

AUC : Air Under the Curve

Introduction Générale

À l'ère numérique actuelle , la recherche et la détections de fake news est encore au stade préliminaire de son développement . Plusieurs recherches et de nombreuses études sont en cours afin de pouvoir contrôler la propagation des fake news qui peuvent engendrer de graves conséquences dans la société . Mais qu'entendons - nous par une fake news ? Ou autrement dit 'fausses nouvelles' ?

Une fake news, c 'est tout simplement la création ou la diffusion d'un contenu dans le but manipuler ou tromper le public. Ce contenu peut avoir différents formats , allant des articles de presse aux vidéos, en passant par des images , messages vocaux... etc. et est diffusé à travers les médias. Nous donnerons plus de détails concernant cette définition plus bas.

Afin de résoudre ce défi , ou même - uniquement - surmonter le problème des fausses nouvelles , l'intelligence artificielle émerge comme une solution potentielle , tout comme l'apprentissage automatique . En combinant des techniques de traitement du texte avancées, visant la détection des fake news . Néanmoins, il convient de noter que la détections des fake news est classé comme étant une tâche ardue qui nécessite beaucoup d 'investissement, de temps et de discipline.

L'objectif de notre projet, exposé dans ce rapport, est de développer un modèle d'intelligence artificielle qui serait capable de détecter la véracité d'une nouvelle provenant de n'importe quel article. Pour cela nous utiliserons des techniques avancées de traitement de langage naturel et d'apprentissage et bien d'autres traitements que nous présenterons progressivement au long de ce rapport comprenant toutes les étapes suivies pour la mise en œuvre de notre modèle.

Ce rapport est organisé en trois sections principales, 'Données', 'Problématique' et 'Résultats'. Il débute par la présente partie qui constitue l'introduction générale incluant le contexte d'étude.

La première section "Données" portera sur la présentation de l'ensemble de données que nous avons à disposition, d'où proviennent - ils et puis sur la visualisation de données à l'aide de graphiques.

La deuxième section "problématique" se focalise sur la problématique de notre thématique, notamment à quoi l'IA va t–elle répondre, l'état de l'existant.

La troisième section "Résultats" présentera dans une première partie le modèle proposé et comment ce dernier fonctionne - t - il en décrivant également les étapes suivies pour la réalisation du modèle. Et en deuxième partie, la présentation des résultats que nous avons eu ainsi qu' un comparatif avec d'autres modèles.

Enfin, nous conclurons ce travail et nous synthétisons les compétences acquises tout au long de la réalisation de ce projet.

Résumé

Notre projet vise à développer une méthode de détection des fake news dans le domaine textuel, une problématique croissante dans notre société saturée d'informations. En utilisant les avancées en intelligence artificielle et en apprentissage automatique, nous cherchons à créer un modèle robuste capable d'identifier et de filtrer les informations trompeuses dans le flux d'actualités en ligne.

Après avoir analysé divers modèles de classification potentiels, tels que le Decision Tree Classifier, le Random Forest Classifier et le Support Vector Machine (SVM). sur ce dernier, nous avons effectué des évaluations approfondies. Le SVM s'est démarqué comme le modèle le plus performant, offrant des résultats prometteurs avec un taux d'erreur minimal de seulement 0,66%.

Ce rapport regroupe notre méthodologie, nos résultats et nos conclusions. Dans un environnement où la désinformation peut causer des dommages considérables, il est impératif de développer des solutions technologiques solides afin de préserver l'intégrité de l'information et de promouvoir une société éclairée et bien informée.

Introduction

Si le terme fake news a été popularisé durant la dernière campagne électorale américaine opposant Hillary Clinton et Donald Trump en 2016, l'utilisation de la désinformation et de la mésinformation ne date néanmoins pas d'hier mais bien présente depuis des siècles, voire des millénaires. Leur histoire remonte au Moyen Âge à partir du XIIIe siècle, où le terme "rumeur" imprégnait tous les aspects de la vie sociale, des rues aux cours des seigneurs : fausses nouvelles de la mort du roi, suspicion de complot ... Le mot « rumeur » à partir du XIIIe siècle, désigne un bruit informel, il reflétait les peurs populaires et les tensions entre les classes, tout en étant parfois instrumentalisée par les élites pour servir leurs intérêts.

En 1750, le terme "canard" désignant une fausse nouvelle apparaissait dans la presse, bien qu'au XVIe siècle, il faisait référence à une publication distribuée par un colporteur et traitant d'un fait divers.

Le 29 juillet 1881, la loi sur la liberté de la presse (article 27) a marqué l'introduction de la notion de "nouvelle fausse" dans le contexte juridique. Par la suite, cette terminologie a été incluse dans le code électoral, le code pénal et le code monétaire et financier, accompagnée des termes "fausse nouvelle", "information fausse" ou "fausse information".

Le 16 novembre 2016, le terme "post-vérité" a été désigné comme le mot international de l'année par le dictionnaire de l'université d'Oxford. Ce concept, né du postmodernisme des années 1980, a gagné en popularité avec les événements marquants de 2016 tels que le Brexit en juin et en juillet quand Donald Trump a obtenu l'investiture présidentielle du Parti républicain.

En 2016, ont vu le jour les "deepfakes", une abréviation de "Deep Learning" et "Fake", qui peut être traduit par "fausse profondeur", qui étendent les possibilités de manipulation grâce à des technologies numériques avancées. Ces outils compliquent également la tâche de vérification des informations.

Le terme "infox", composé des mots "information" et "intoxication". Il a été introduit en octobre 2018 par la Commission d'enrichissement de la langue française, qui le définit comme "une information mensongère, délibérément biaisée ou tronquée, diffusée par un média". Selon la linguiste Jeanne Bordeau, "infox" figure parmi les mots les plus utilisés par les médias en 2018. Le 22 décembre 2018, la France a promulgué une loi visant à lutter contre la manipulation de l'information, souvent appelée "loi fake news" ou "loi infox".

En 2019, les "shallow fakes" ont fait leur apparition, désignant des vidéos ou des images manipulées créées avec des techniques moins avancées que les deepfakes. Joe Biden a été ciblé par ces manipulations lors de la campagne présidentielle américaine, où une vidéo virale partagée par la Maison-Blanche le présentait faussement comme soutenant Donald Trump.

Avec l'avènement d'Internet, obtenir et partager de l'information est devenu extrêmement simple. Cependant, cette facilité ne garantit pas la fiabilité des informations. Il est difficile de trouver une source d'information qui soit à la fois fiable, gratuite, indépendante et sans publicité.

Une Fake News : qu’est-ce que c’est ?

L'expression "fake news", d'origine anglo-saxonne, est désormais utilisée en français pour décrire des informations mensongères ou délibérément biaisées. Ces fausses nouvelles sont souvent créées dans le but de favoriser un parti politique, de ternir la réputation d'une personne ou d'une entreprise, ou même de contredire des faits scientifiques établis, contribuant ainsi à la désinformation du public.

Un exemple célèbre de fake news est la prétendue déclaration du pape François en faveur de Donald Trump lors de sa campagne présidentielle.

Types de Fake News

Les fake news contextuels : Contenu réel avec des informations contextuelles fausses pour induire en erreur.
Les fake news manipulés : Contenu altéré via des techniques d'édition pour déformer les informations.
Les fake news trompeurs : Utilisation partielle d'un fait réel pour tromper l'opinion.
Connexion : Les titres et légendes utilisés ne correspondent pas au contenu réel.
Contenu fabriqué : est un contenu totalement faux, conçu pour tromper, parfois inconsciemment repris par les médias, utilisant parfois des deepfakes.
Le contenu imposteur : utilise des informations, photos ou vidéos manipulées portant des logos de médias célèbres pour induire en erreur.

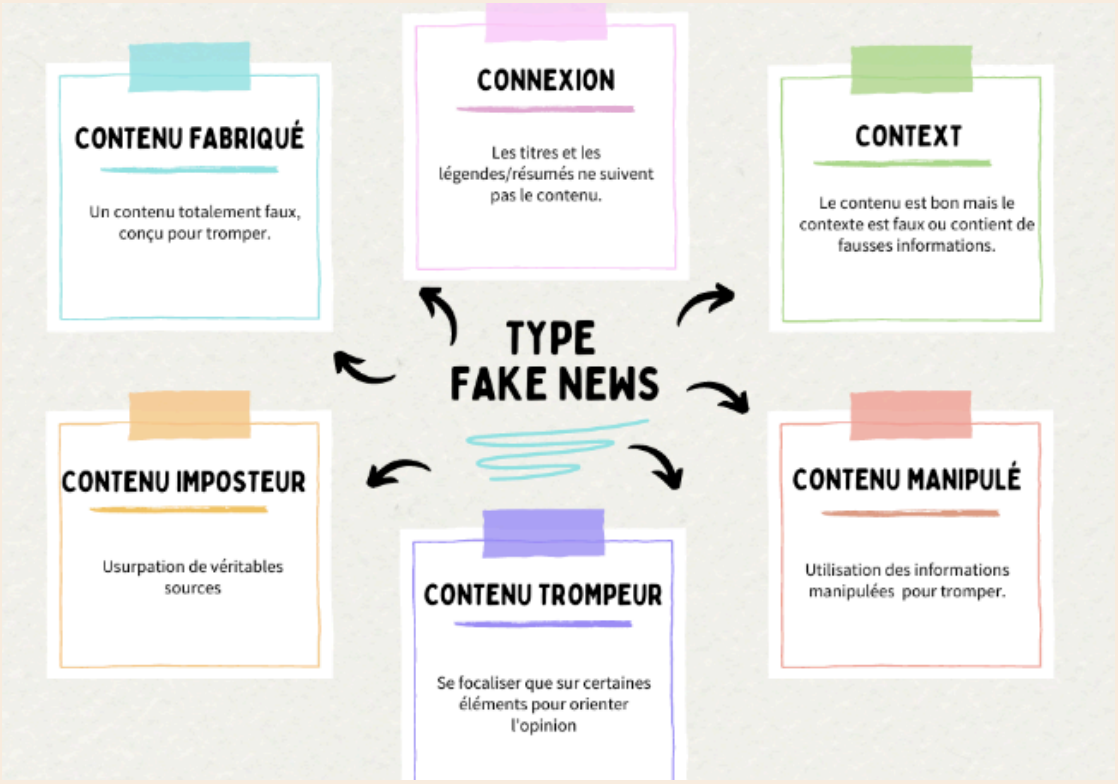

Figure 1 : Types de Fake News

Comment les fake news se propagent ?

La propagation des fake news repose sur plusieurs mécanismes. La plupart du temps, les fake news commencent sur des sites web anonymes. Leur visibilité est donc limitée. Elles sont ensuite amplifiées par les réseaux sociaux en utilisant des mots-clés et des contenus attrayants pour les algorithmes, où elles se propagent à grande vitesse grâce à notre tendance à partager rapidement ce qui nous semble intéressant ou choquant. Elles attirent l'attention par des titres accrocheurs et sensationnels, créant ainsi une demande pour ces contenus.

Les fake news cherchent à se fondre dans le paysage médiatique en imitant les formats et les styles des médias traditionnels, ce qui les rend encore plus difficiles à distinguer de l'information authentique. Leur diffusion est également facilitée par le manque d'éducation aux médias, qui rend les individus plus susceptibles de croire et de partager des informations erronées.

Enfin, les fake news peuvent également se propager en dehors d'internet, grâce au phénomène de la pensée de groupe et à la dépendance informationnelle au sein des communautés.

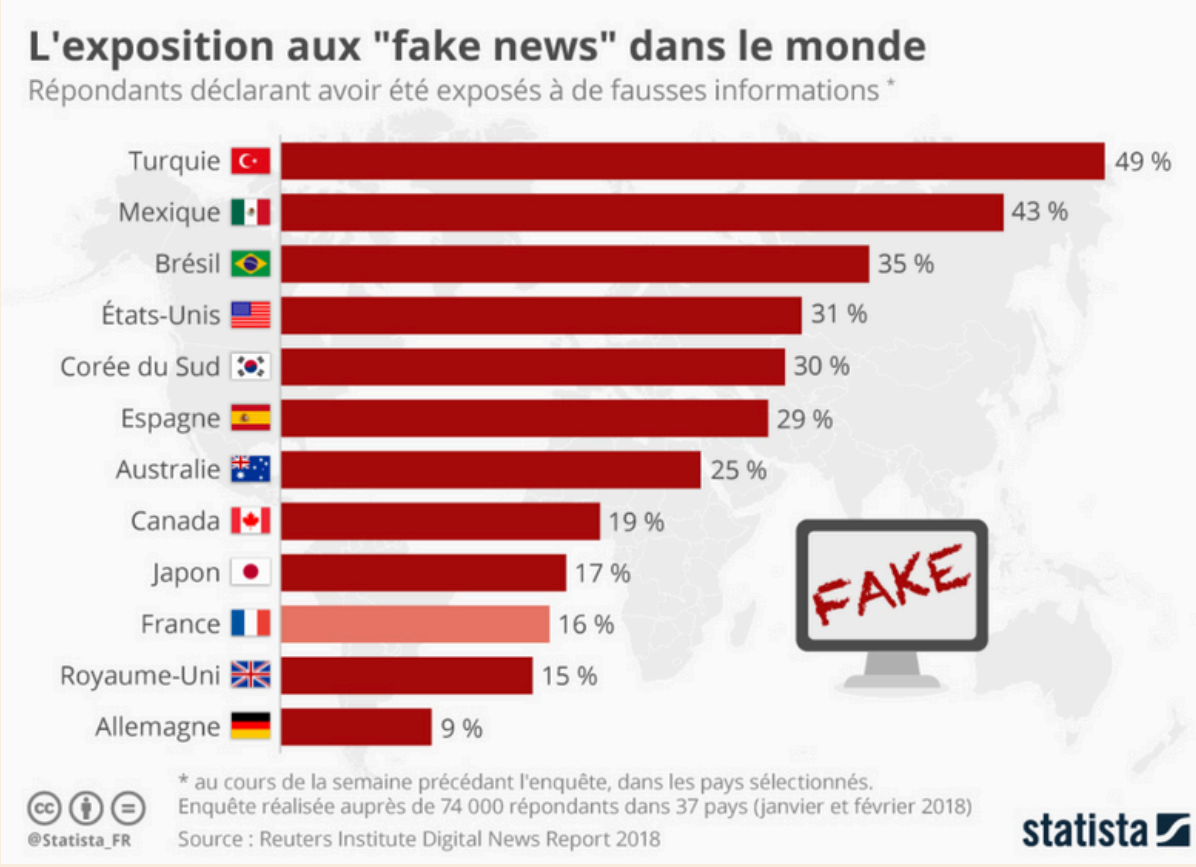

Figure 2 : L'exposition aux Fake News dans le monde

Que dit la loi contre les fake news ?

Ces textes restent difficiles à appliquer car il faut prouver la mauvaise foi et le trouble possible à l'ordre public. Depuis 2010, il y a eu trois condamnations pour fausses nouvelles.

Loi et Article	Sanction	Conditions pour la condamnation
Loi sur la presse du 27 juillet 1849	Emprisonnement d'un à deux ans , et d'une amende de 50 à 1000 francs	Publication ou reproduction, faite de mauvaise foi, de nouvelles fausses, de pièces fabriquées, falsifiées ou mensongèrement attribuées à des tiers, lorsque ces nouvelles ou pièces seront de nature à troubler la paix publique
Loi sur la liberté de la presse du 29 juillet 1881	Amende de 45 000 euros Amende jusqu'à 135 000 euros en cas de fausses nouvelles pouvant ébranler la discipline militaire ou entraver l'effort de guerre	Publication, la diffusion ou la reproduction, par quelque moyen que ce soit, de fausses nouvelles, de pièces fabriquées, falsifiées ou mensongèrement attribuées à des tiers lorsque, faite de mauvaise foi, elle aura troublé l'ordre public ou aura été susceptible de le troubler
Code électoral, Article L97	Emprisonnement d'un an et amende de 15 000 euros	Utilisation de fausses nouvelles, bruits calomnieux ou autres manœuvres frauduleuses pour surprendre ou détourner des suffrages

Figure 3 : Lois et articles sanctionnant les Fake News

Pendant les trois mois précédant le premier jour du mois d'élections générales ou européennes, les dispositions principales de la loi s'appliquent afin de lutter contre le phénomène.

- Un candidat ou un parti politique a la possibilité de saisir le juge des référés selon la loi. Ce juge est habilité à arrêter la diffusion de fausses informations. Il dispose d'un délai de 48 heures pour évaluer le contenu en question et, s'il est confirmé comme étant faux, ordonner sa suppression.

- Le CSA (Conseil Supérieur de l'Audiovisuel) peut prendre des mesures à l'encontre des médias "contrôlés par un État étranger, ou placés sous l'influence de cet État" si ceux-ci diffusent délibérément "de fausses informations de nature à altérer la sincérité du scrutin".

- Les plateformes en ligne sont tenues de divulguer l'identité des individus rémunérant la promotion de contenus liés à des sujets d'intérêt général et doivent spécifier le montant de la somme payée si elle dépasse 100 € HT.

Dans un monde où les informations affluent à une vitesse fulgurante et où les réseaux sociaux amplifient la diffusion, la distinction entre vérité et mensonge devient de plus en plus compliquée. Ainsi, la détection des fake news, notamment dans les textes, est devenue un défi difficile à relever.

Face à cette réalité, il est important de se demander : quel moyen pouvons-nous mettre en place pour détecter les fake news dans le flux incessant de textes diffusés sur Internet ?

Données

Dans le cadre de notre projet de détection de fake news, nous avons utilisé deux ensembles de données textuelles distincts, tous deux disponibles sur Kaggle, une plateforme en ligne qui propose des ensembles de données et des ressources pour l'apprentissage automatique et l'analyse de données. Le premier ensemble de données, **"Fake.csv"**, regroupe des exemples de fausses nouvelles, tandis que le second, **"True.csv"**, contient des textes provenant de sources authentiques. Ces données textuelles sont essentielles pour l'entraînement et l'évaluation de notre modèle de détection de fake news, contribuant ainsi à la préservation de l'intégrité de l'information en ligne.

Visualisation de données

Nombre de lignes et de colonnes des datasets

```
Entrée [15]: print("Nombre de ligne et de colonnes du dataset Fake.csv :",data_fake.shape)
             print("Nombre de ligne et de colonnes du dataset True.csv :" ,data_real.shape)

             Nombre de ligne et de colonnes du dataset Fake.csv : (23481, 4)
             Nombre de ligne et de colonnes du dataset True.csv : (21417, 4)
```

Distribution des news (en fonction du nombre d'articles)

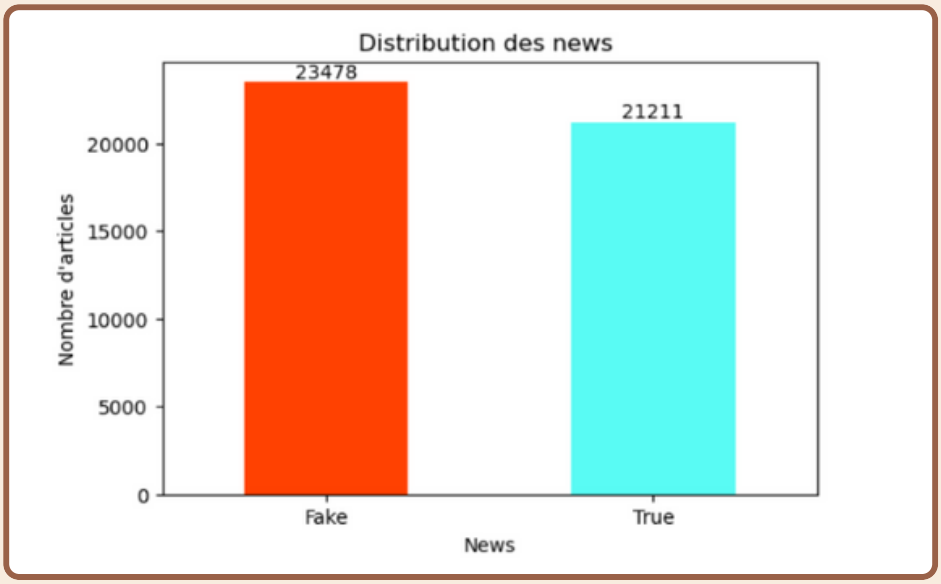

Figure 4 : Distribution des news

Longueur des news

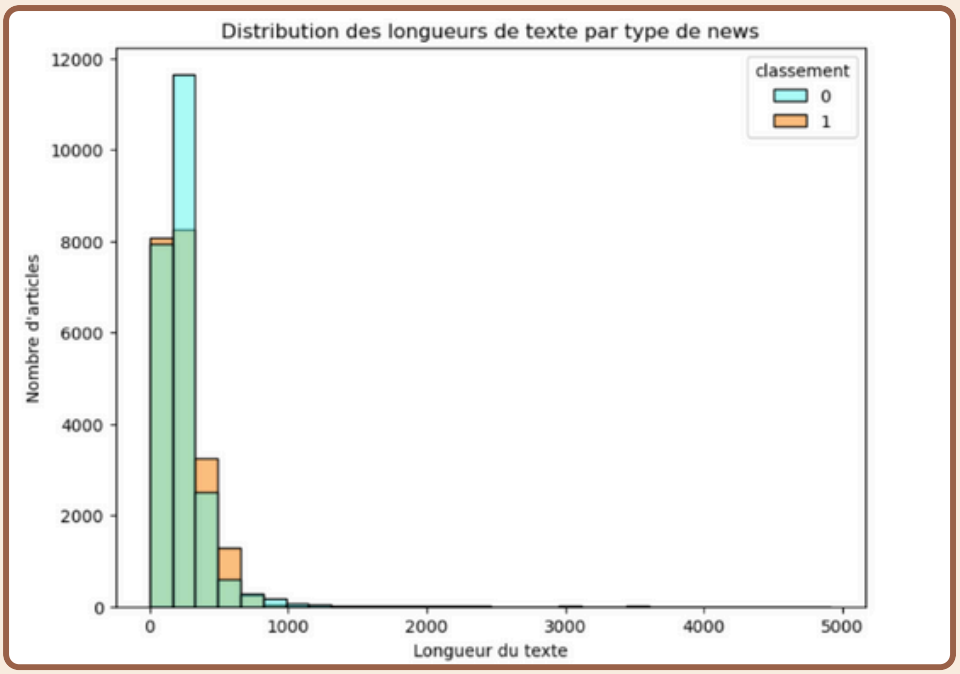

Figure 5 : Longueur des news

Afficher les premières lignes des datasets

```
Entrée [8]: data_fake.head()
```

Out[8]:

	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017

```
Entrée [9]: data_real.head()
```

Out[9]:

	title	text	subject	date
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017

Résumé statistique des données

```
Entrée [12]: data_fake.describe()
```

Out[12]:

	title	text	subject	date
count	23481	23481	23481	23481
unique	17903	17455	6	1681
top	MEDIA IGNORES Time That Bill Clinton FIRED His...		News	May 10, 2017
freq	6	626	9050	46

```
Entrée [13]: data_real.describe()
```

Out[13]:

	title	text	subject	date
count	21417	21417	21417	21417
unique	20826	21192	2	716
top	Factbox: Trump fills top jobs for his administ...	(Reuters) - Highlights for U.S. President Dona...	politicsNews	December 20, 2017
freq	14	8	11272	182

Dataset final

Après l'ajout d'une colonne "classement" pour identifier la véracité des textes et la fusion des deux ensembles de données en un seul, nous avons procédé au nettoyage des données.

```
Entrée [26]: # Fusionner les deux ensembles de données en un + nettoyage
both = pd.concat([data_fake, data_real], axis=0).dropna().drop_duplicates()
both
```

Out[26]:

	title	text	subject	date	classement
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017	0
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	0
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	0
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	0
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	0
...	...	...	...	...	...
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017	1
21413	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	August 22, 2017	1
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017	1
21415	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017	1
21416	Indonesia to buy $1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017	1

Problématique

Introduction

Dans cette section, nous abordons la problématique centrale de notre projet IA. Face à la rapidité dont l'information se propage dans l'environnement médiatique, se pose la question primordiale de la véracité des informations. Les fausses informations, de par leur diffusion et leurs graves conséquences, soulèvent d'importants défis. Ainsi nous nous pencherons sur la relation entre l'intelligence artificielle et la détection des fake news pour y faire face et limiter les conséquences.

A quoi l'IA va-t-elle répondre ?

De nos jours, les technologies de l'information et des communications représentent un élément fondamental de la société, spécialement, l'environnement médiatique qui est saturé de contenus et d'informations en ligne. L 'évolution dans le domaine de l'informatique et de l'intelligence artificielle ouvre également des opportunités pour l'innovation et la compétitivité. La communication et la diffusion de l'information sont alors devenues des éléments essentiels à la réussite des organisations dans la société . Néanmoins, il est pareillement compliqué de vérifier la véracité de ces informations ou du contenu en ligne. Indépendamment du domaine concerné, une fake news peut avoir des conséquences néfastes sur la société.

Dans ce contexte, l'IA, comme elle permet d'accélérer la propagation des fake news à travers la diffusion, que ce soit en créant des textes, reproduire des voix ou même générer de fausses vidéos de personnes existantes, peut également jouer un rôle crucial et constituer un outil puissant en les identifiant et démasquant. Au cours de notre projet, nous nous concentrons spécifiquement sur ***la détection de fake news sur du texte***. Face à cette problématique, nous cherchons à développer un modèle qui sera capable de détecter les informations authentiques des informations

> **Dans l'ère numérique, l'unique important alliée contre la désinformation est l'intelligence artificielle**

trompeuses.

Pour automatiser et améliorer le processus de la détection, l'utilisation de l'intelligence artificielle est indispensable afin de pouvoir développer un modèle performant, efficace et précis en utilisant des techniques avancées d'apprentissage automatique. On retrouve à titre d'exemple :

- L'analyse du texte (linguistiques, contextuelle ou autres), une étape ou l'IA est capable d'analyser de vaste quantité de volume en peu de temps et de détecter les signaux de désinformation ;
- la détection des similitudes ou les divergences qui pourraient révéler la nature frauduleuse ;
- la classification des textes selon leur véracité également (Vrai ou faux).

Par ailleurs, nous détaillerons plus de fonctionnalités dans la section des résultats.

Voir quoi rajouter.

L'état de l'existant

Concernant la détection des fake news sur de texte, l'état de l'existant est en progression et évolution constante. Il y a eu plusieurs recherches qui ont été faites dans ce domaine ainsi que beaucoup de développements. Néanmoins il existe toujours des difficultés à vérifier la véracité des news, d'où plusieurs approches et techniques sont abordées pour répondre à cette problématique par exemple : approches supervisées, approches basées sur des langage naturel et vérifications des faits.

Les approches supervisées ça consiste en l'utilisation des modèles d'apprentissage supervisées tels que les réseaux des neurones profonds, les arbres de décision ou d'autres modèle d'apprentissage selon la complexité des données ainsi que des performances souhaitées dans le but de classer les textes selon leur véracité. Il est crucial d'entrainer le modèle à partir des données dont la véracité est connue. Quand l'objectif est de détecter les fake news il est primordial que cet ensemble de données comporte des exemples sur différents textes en précisant leur véracité, vrai ou faux, ce qui est appelé étiquette de véracité. Pendant l'entrainement, le modèle est optimisé afin de minimiser la perte et les erreurs. La classification est binaire (0 pour les fake news et 1 pour les non-fake).

Les caractéristiques du texte extraites pour entraîner le modèle peuvent varier en fonction de l'approche spécifique. Cela peut inclure des fonctionnalités linguistiques telles que le nombre de mots qu'un article compte, la fréquence des mots dans un ensemble de données, les caractéristiques syntaxiques, sémantiques ou contextuelles, selon ce qui est pertinent pour la tâche de détection des fausses informations. Une fois l'entrainement est fait, il est évalué sur un ensemble de données.

Conclusion :

Dans cette section, nous avons défini la problématique du projet qui est une étape cruciale dans la mesure où toute la suite de projet en dépend. Tout au long de cette section, nous avons présenté la problématique et sa relation avec l'intelligence artificielle.

Résultats

Introduction

A ce stade de la réalisation du projet, nous avons fait une présentation détaillée des données que nous avons à disposition et la définition de la problématique est achevée. Par conséquent, la dernière phase du processus s'agit de développement de notre modèle et ensuite la présentation des résultats obtenus. Dans cette section, nous présentons l'environnement de développement ainsi que les outils que nous avons utilisés, les étapes suivies et nous donnons un aperçu des résultats.

Environnement de travail

Jupyter est une application web open source utilisée pour programmer. Jupyter permet de réaliser des calepins ou notebooks, c'est-à-dire des programmes contenant à la fois du texte, simple ou enrichi typographiquement et sémantiquement grâce au langage à balise simplifié 'Markdown', et du code, lignes sources et résultats d'exécution [15].

Modèle proposé :

Après une recherche approfondie, nous avons identifié divers modèles de classification pouvant potentiellement être appliqués afin de répondre à notre problématique. Certains d'entre les modèles identifiés ont été rapidement écartés, soit car ça ne répondait pas pleinement à notre besoin, soit en raison de leur temps de charge qui est très lent, où bien parce qu'ils sont plus performants sur d'autres types de modèles. Par exemple : le modèle ANN (les réseaux des neurones artificielles). Ensuite, nous avons expérimenté d'autres modèles qui se sont révélés efficaces, offrant des performances de classification satisfaisantes.

Parmi les modèles que nous avons pu tester , l'un d'entre eux n'a pas donné de bons résultats , il s'agit du modèle de régression logistique , un modèle performant pour la classification de simples phrases et facile à interpréter car les coefficients sont directement associés aux caractéristiques du modèle. Néanmoins , ce modèle a montré des lacunes lorsqu'il est appliqué à de longs textes où les résultats n'étaient pas tout le temps corrects .

Le ' *Decision Tree Classifier', le* ' *Random Forest Classifier'* et le *'Support Vector Machine'* quant à eux, ont donné des résultats plus corrects c'est pourquoi nous avons décidé de poursuivre les tests sur ces trois modèles. À la fin de notre programme, pour évaluer la précision et l'exactitude de chaque modèle (à l'aide de graphiques), le meilleur modèle avec plus de précision et un faible taux d'erreur - uniquement 0,69% - était le modèle SVM (Support Vector Machine), donc nous avons décidé de ne garder que ce modèle . En seconde position se trouvait le modèle ' *Decision Tree Classifier*'.

Définition de chacun des trois modèles :

- ***Decision Tree Classifier***: Un arbre de décision peut servir à bâtir des modèles prédictifs automatisés, dont les applications peuvent concerner l'apprentissage automatique. Dans ces arbres de décision, les nœuds représentent les données. Chaque branche contient un ensemble d'attributs (règles de classification), associés à une étiquette de classe spécifique que

l'on retrouve à l'extrémité de la branche [7].

.

- ***Random Forest Classifier*** : aussi un modèle d'arbre de décision mais il consiste en de multiples arbres conçus pour accroître le taux de classification [7].
- ***Support Vector Machine*** : les SVMs sont une famille d'algorithmes d'apprentissage automatique qui permettent de résoudre des problèmes tant de classification que de régression ou de détection d'anomalie. Ils sont connus pour leurs solides garanties théoriques, leur grande flexibilité ainsi que leur simplicité d'utilisation même sans grande connaissance de data mining [8].

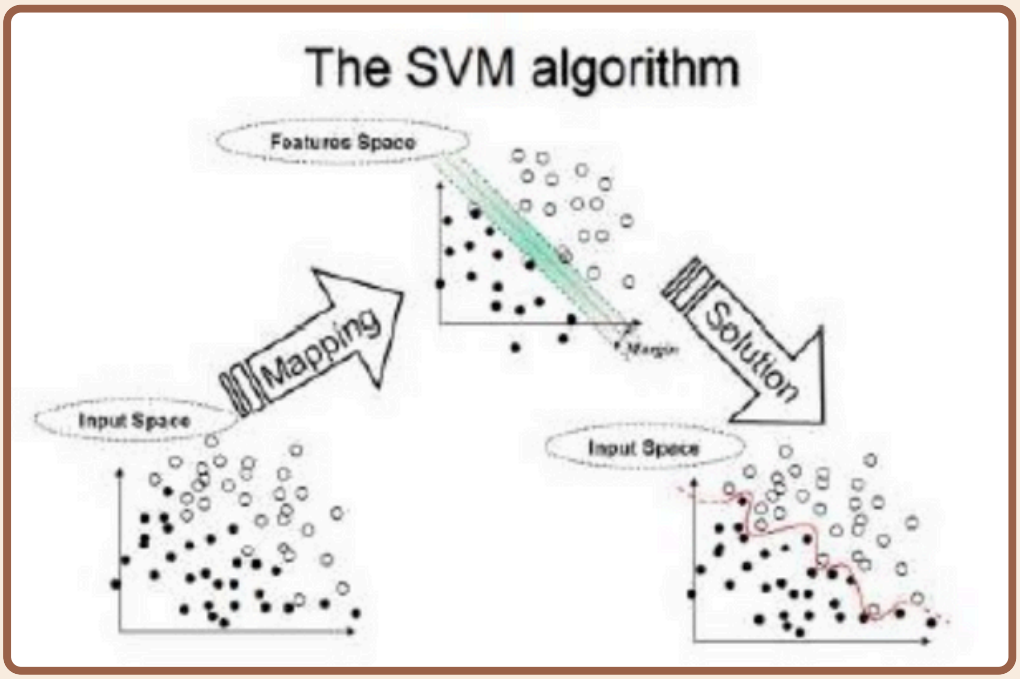

Figure 6 : l'algorithme SVM

Fonctionnement du modèle :

Afin de développer notre modèle, nous avons suivi les étapes ci-dessous :

L'importation des bibliothèques : nous avons importé différentes bibliothèques, dont des bibliothèques pour optimiser les calculs et les parcours, telles que pendas et numpy. Pour la visualisation des données, nous avons intégré matplotlib et seaborn. Pour la manipulation de chaîne de caractères et les expressions régulières, notamment le nettoyage et l'analyse des données, nous avons utilisé string et re. En outre, la bibliothèque sklearn (autrement dit scikit-learn) pour l'apprentissage automatique et NLTK pour le traitement de langage naturel (NLP) qui inclue de fonctionnalités comme la tokenisation, la lemmatisation et la racinisation (cf la figure 7).

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import re
import string

from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split, cross_val_predict
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from sklearn.model_selection import GridSearchCV
from sklearn.svm import SVC

from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
from nltk import pos_tag
import nltk
```

Figure 7 : Importation des bibliothèques

Prétraitement des données : Avant l'étape de prétraitement, nous avons ajouté une colonne "classement" assignant la valeur 0 à tous les articles "fake" et la valeur 1 à tous les articles considérés comme "non-fake". Ensuite, nous avons fusionnés les deux datasets en un seul, nommé "both" en éliminant toutes les valeurs manquantes ainsi que les doublons. Enfin, nous avons créé une méthode pour le prétraitement de texte qui consistait à transformer le texte des articles en minuscule, à supprimer les caractères spéciaux, les liens, les expressions entre crochets ou entre parenthèses et à remplacer les sauts de lignes par un espace.

La figure ci-dessous illustre le contenu de la méthode de prétraitement:

```
def preprocess_text(text):
    text = text.lower() # Convertir le texte en minuscules


    text = re.sub(r'[^a-zA-Z0-9\s]', '', text) # Supprimer les caractères spéciaux et les espaces supplémentaires
    text = re.sub(r'\s+', ' ', text) # remplacer plusieurs espacesc par un seul espace
    text = re.sub(r'https?://\S+|www\.\S+', '', text) #supp les liens
    text = re.sub(r'\n', ' ', text)  # Remplacer les sauts de ligne par un espace
    #text = re.sub(r'<.*?>+', '', text)  # Supprimer les balises HTML
    text = re.sub(r'\[.*?\]', '', text)  # Supprimer les expressions entre crochets
    text = re.sub(r'\(.*?\)', '', text)

    # Tokenization
    tokens = word_tokenize(text)
    tokens_and_pos = pos_tag(tokens)

    # Supprimer les mots vides
    stop_words = set(stopwords.words('english'))
    tokens_and_pos = [word for word in tokens_and_pos if word[0] not in stop_words]

    # Lemmatization (stemming)
    lemmatizer = WordNetLemmatizer()
    tokens = [lemmatizer.lemmatize(word[0], pos=pos_tag_to_pos_lemmas(word[1])) for word in tokens_and_pos]

    # Rejoindre les tokens en une seule chaîne de texte
    preprocessed_text = ' '.join(tokens)

    return preprocessed_text
```

Figure 8 : Méthode de prétraitement du modèle

Cette méthode inclue également :

- *La tokenisation* : qui consiste à diviser le texte en plusieurs unités plus petites, ces unités s'appellent 'tokens'. Cette étape permet la segmentation du texte pour une analyse approfondie.
- *La suppression des Stopwords* : réside en la suppression de tous les mots vides qui ne porte pas de sens spécifique, exemple : by, the, and, we, as, but, if, be, can … etc.
- *La lemmatisation* : Pour la lemmatisation, nous avons utilisé une librairie python de traitement de langage appelée NLTK (Natural language Tool-kit). Cette librairie offre une méthode "lemmatize" permettant de trouver la lemme de chaque mot. Afin de tirer un maximum d'avantage de celle-ci, il est important de lui indiquer pour chaque mot que nous voulons lemmatiser son rôle -composant de langage- (ou pos : Part Of Speech) dans la phrase (verbe, nom, adjectif…). Cette méthode implique de réduire des mots en leurs formes de base afin de faciliter la comparaison et l'analyse.
- "*pos_tag*" : NLTK offre aussi une méthode "pos_tag" permettant de trouver le rôle d'un mot. Cela dit, cette dernière renvoie un rôle beaucoup trop précis pour chaque mot en plus d'avoir un encodage des rôles différent. À titre d'exemple : Si nous voulons déterminer le lemme d'un mot qui serait un adjectif, la méthode "pos_tag" nous indiquera comme rôle "ADJ", or pour indiquer à la méthode lemmatize que le mot en question est un adjectif, il faudra positionner dans le paramètre "pos" la lettre "à".

Pour palier ça, nous avons créé la méthode "pos_tag_to_pos_lemma".

- "*pos_tag_to_pos_lemma*": permet de faire la transcodification entre l'encodage utilisé par "pos_tag" et celui utilisé par "lemmatize".

L'extraction des caractéristiques avec TF-IDF (Term Frequency - Invers Document Frequency): consiste à la vectorisation du texte, transformant ainsi chacun des articles en un vecteur numérique, ce qui permet d'évaluer l'importance d'un mot dans un article par rapport à un ensemble de données (comprenant un ensemble d'articles). Les vecteurs de caractéristiques obtenus sont utilisés comme entrée pour l'algorithme de classification.
Le calcul s'effectue comme suit : dans la liste de "features names" récupérée, nous cherchons la fréquence de chacun des mots appartenant à cette liste dans chaque article du dataset. Ensuite le résultat obtenu est multiplié par le logarithme du rapport entre le nombre total des articles et le nombre d'articles ou un mot apparait. Les figures [9] et [10] illustrent le résultat des "features names" ainsi que les vecteurs de caractéristiques obtenues, dans notre cas nous avons limité la liste à 100 mots.

```
print(text_to_vector.get_feature_names_out())
print(X_vectorized)
```

```
['accord' 'administration' 'also' 'america' 'american' 'ask' 'attack'
 'back' 'bill' 'call' 'campaign' 'clinton' 'come' 'could' 'country'
 'court' 'day' 'democrat' 'democratic' 'donald' 'election' 'even' 'first'
 'force' 'former' 'get' 'give' 'go' 'government' 'group' 'help' 'hillary'
 'house' 'image' 'include' 'issue' 'know' 'last' 'law' 'leader' 'like'
 'make' 'many' 'may' 'medium' 'member' 'million' 'month' 'national' 'need'
 'new' 'news' 'obama' 'office' 'official' 'one' 'party' 'people' 'percent'
 'plan' 'police' 'policy' 'political' 'president' 'presidential' 'report'
 'republican' 'reuters' 'right' 'russia' 'say' 'security' 'see' 'senate'
 'show' 'since' 'state' 'statement' 'support' 'take' 'tax' 'tell' 'think'
 'time' 'trump' 'try' 'two' 'united' 'use' 'vote' 'want' 'washington'
 'way' 'week' 'well' 'white' 'woman' 'work' 'would' 'year']
  (0, 33)       0.1150892095195854
  (0, 75)       0.11220705270225907
  (0, 7)        0.1094954772760088
  (0, 78)       0.11245212923916233
  (0, 42)       0.11061570529580397
  (0, 76)       0.23908163723992204
  (0, 21)       0.10678781653040118
  (0, 9)        0.09647609055111672
```

Figure 9 :Vecteurs de caractéristiques avec TF-IDF

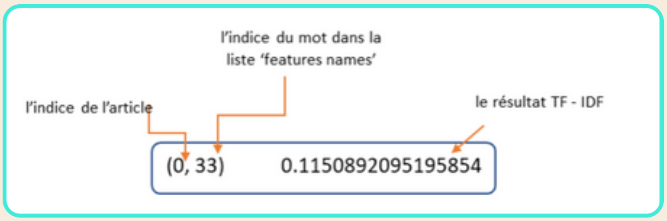

Figure 10 :Explications d'un vecteur de caractéristiques TF-IDF obtenu.

Entrainement du modèle : pour la partie entrainement nous avons opté pour le modèle SVM, en exploitant les données d'entrainement 'X_train' (le résultat de la vectorisation dans l'étape précédente) et 'y_train' (qui correspond à la colonne classement de notre dataset) afin que nous puissions à la fin distinguer les fausses informations des informations authentiques.

Evaluation du modèle :Ici nous testons la performance de notre modèle à l'aide des indicateurs.

Phase de test : dans cette partie nous avons pris des articles sur internet et tester si la classification de notre modèle se faisait correctement ou pas.

Comparaison à d'autres modèles:

Cette partie est consacré à la comparaison du modèle SVM avec les deux modèle "'Logistic Regression" et le modèle "Decision Tree Classifier".D'abord nous vous présentons la différence entre les trois modèles via la matrice de confusion. Ensuite nous vous illustrant les résultats que nous avons eu sur le modèle que nous avons choisi "SVM".

La matrice de confusion est un outil de mesure de la performance des modèles de classification à deux classes ou plus. Dans le cas binaire (i.e. à deux classes, le cas le plus simple), la matrice de confusion est un tableau à quatre valeurs représentant les différentes combinaisons de valeurs réelles et valeurs prédites comme dans la figure ci-dessous[16].

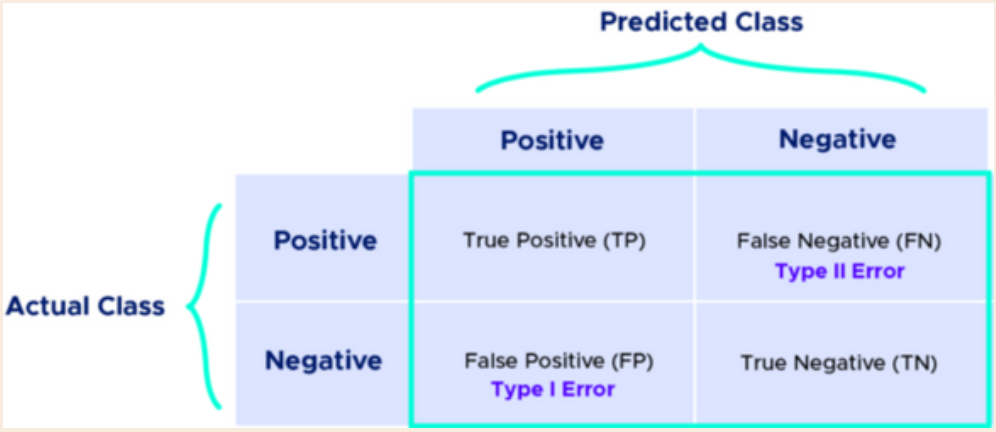

Figure 11 : Matrice de confusion: Matrice de confusion

Pour visualiser les performances, Elle compare les données réelles pour une variable cible à celles prédite par un modèle. Les prédictions vraies et fausses sont révélées et réparties par classe, ce qui permet de les comparer avec des valeurs définies [9].

Les quartes cas possible lors d'une prédiction sont les suivants (cf la figure 12 [10]):

- Vrai négatif : quand la prédiction et la réalité sont positives.
- Vrai positif : quand la prédiction et la réalité sont négatives.
- Faux positif : quand la prédiction est positives mais la réalités est négative.
- Faux négatif : quand la prédiction est négatives mais la réalité est positive.

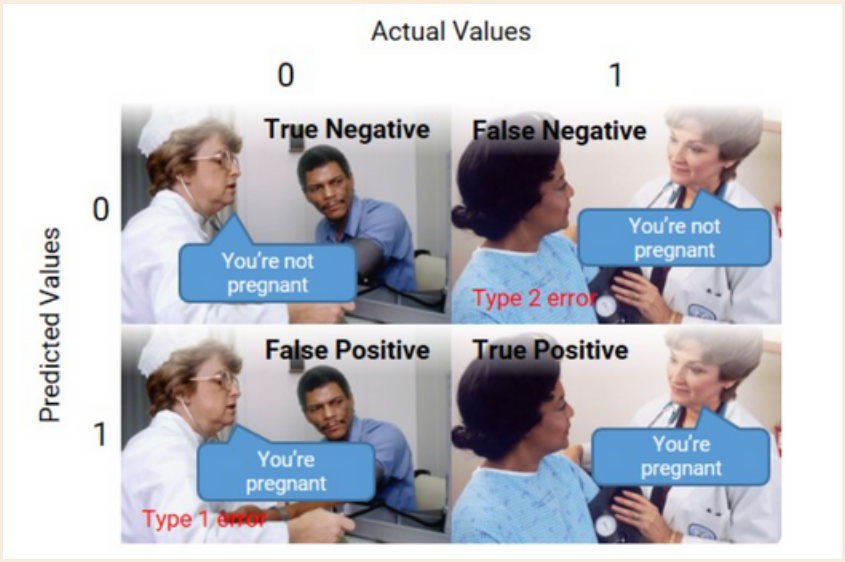

Figure 12 : Exemple de la matrice de confusion

Ci-dessous, nous avons choisi de partager le résultat de la matrice de confusion de quelques modèles que nous avons testé.
Le nom de chaque modèle est indiqué au dessus de chaque matrice, notamment surligné en jaune
Plus la couleur est foncé plus le nombre d'article est petit.

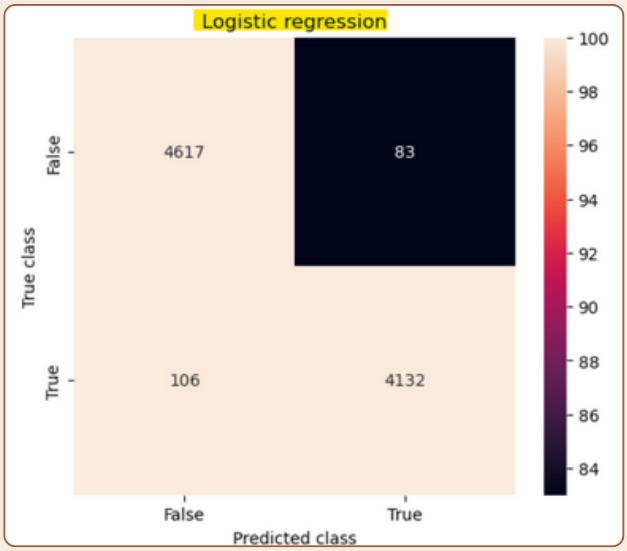

Figure 13 : Matrice de confusion du modèle Regression Logestic

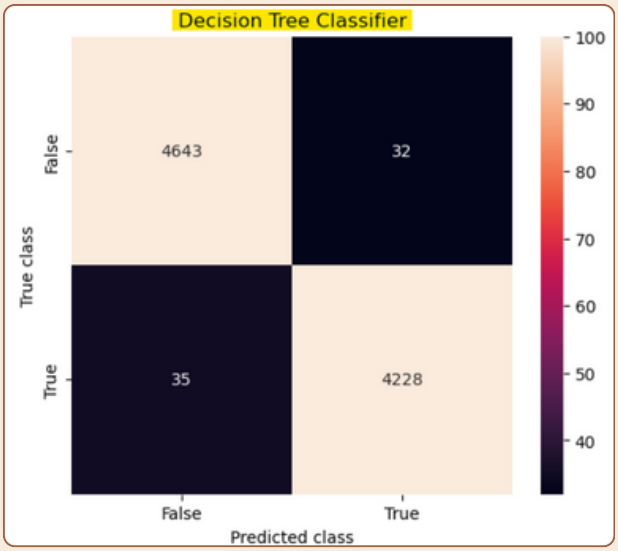

Figure 14 : Matrice de confusion du modèle Decision Tree Classifier

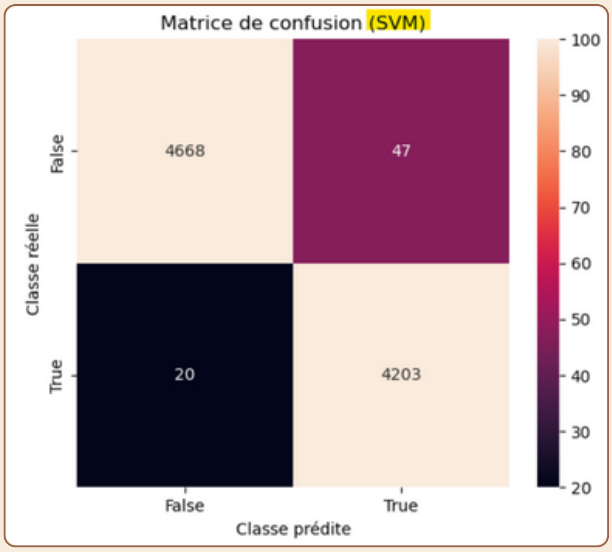

Figure 15 : Matrice de confusion du modèle Support Vector Machin

Ce qui nous intéresse le plus sur la matrice de confusion ce sont les deux colonnes "Faux positif" & "Faux négatif".

Pour la colonne "Faux positif" les résultats sont 106, 35 et 20 pour les modèles "Regression Logistic", "Decision Tree Classifier" et "Support Vector Machine" dans l'ordre consécutif. On peut conclure alors que le modèle qui renvoie le moins d'erreur est le modèle SVM avec uniquement 20 articles "faux positif".

Pour la colonne "Faux négatifs", les résultats sont : 83 pour "Regression Logistic", 32 pour "Decision Tree Classifier" et 47 pour "SVM".

Les méthodes les plus utilisée pour tirer des informations interessantes de ce genre de matrice est sont les indicateurs ou également appelé “métriques”.

Les figures X et X illustrent les détails des deux meilleurs modèles que nous avons testé

```
Accuracy:  0.9838890132020586
Precision:  0.9828226555246053
Recall:  0.9837360594795539
F1-Score:  0.9832791453785416
Matthews correlation coefficient:  0.9677354990694023
```

Figure 16 : Indicateurs du modèle ‘Decision Tree Classifier”

```
Accuracy:  0.9925039158648468
Precision:  0.9889411764705882
Recall:  0.995264030310206
F1-Score:  0.9920925292104331
Matthews correlation coefficient:  0.9849853553483161
```

Figure 17 : Indicateurs du modèle “SVM”

Interprétation :
Accuracy : qui représente le pourcentage des prédictions correctes. La valeur est élevé ce qui veut dire que le modèle classe correctement la plus part des articles.

Précision : les articles positifs qui sont correctement prédits. plus la valeur est élevée plus le modèle à un faible taux d’erreur(peu de faux positifs). Dans notre cas, le modèle Decision Tree Classifier a une plus grande précision.

Recall : le nombre d’échantillons positifs qui sont correctement prédit parmi tous les échantillons réellement positifs (différemment de précision, parmi tous les échantillons prédits positifs). Une valeur élevé signifie que le taux d’erreur est petit (pour les faux négatifs).

F1-Score : c’est à la fois une précision et la capacité d’un modèle a identifier correctement tous les échantillons positifs. Plus le résultat est élevé plus la précision et la capacité sont meilleures.

Mathews correlation coefficient : prends en compte les 4 valeurs de la matrice de confusion et mesure la qualité de classification. une valeur proche de 1 veut dire que la classification est parfaite.

```
diff = 0
for i, v in enumerate(y_test):
    #print("hello")
    if prediction[i] != v:
        diff = diff + 1
    #print(f"prediction {prediction[i]} actual {v}")
print(f"taux d'erreur : {diff/len(y_test)*100} %")
```

```
taux d'erreur : 0.6601029313045425 %
```

Figure 18 : Taux d’erreur du modèle SVM

La figure ci-dessus indique le taux d’erreur de notre modèle.

Ci-dessous le graphique montrant la courbe ROC de notre modèle. On peut voir que la performance de notre modèle est parfaite.

La courbe ROC aide à illustrer la performance du modèle et représente la sensibilité en fonction de 1 où l'axe des abscisses représente les taux de faux positifs, l'axe des ordonnées représente les taux des vrais positifs.
La courbe ROC (en vert), plus elle est proche du coin supérieur gauche plus le modèle est meilleur, ce qui veut dire un taux élevé de vrai positifs.

AUC (Air Under the Curve) : plus l'air est proche de 1 meilleur est la performance du modèle.

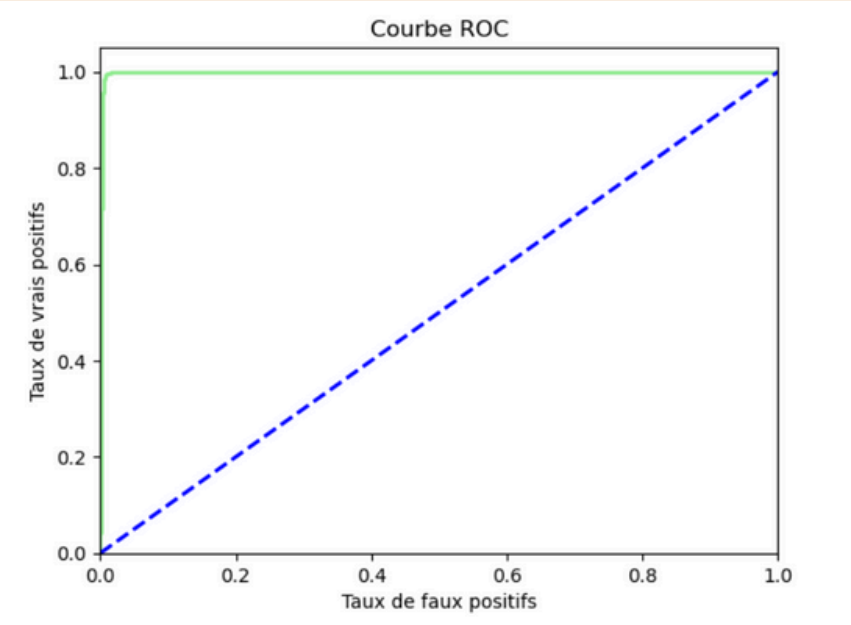

Figure 19 : Taux d'erreur du modèle SVM

Voici un exemple d'une courbe ROC d'un autre modèle pris sur un article. Quand une AUC à 50% cela ne donne pas beaucoup d'informations[17].

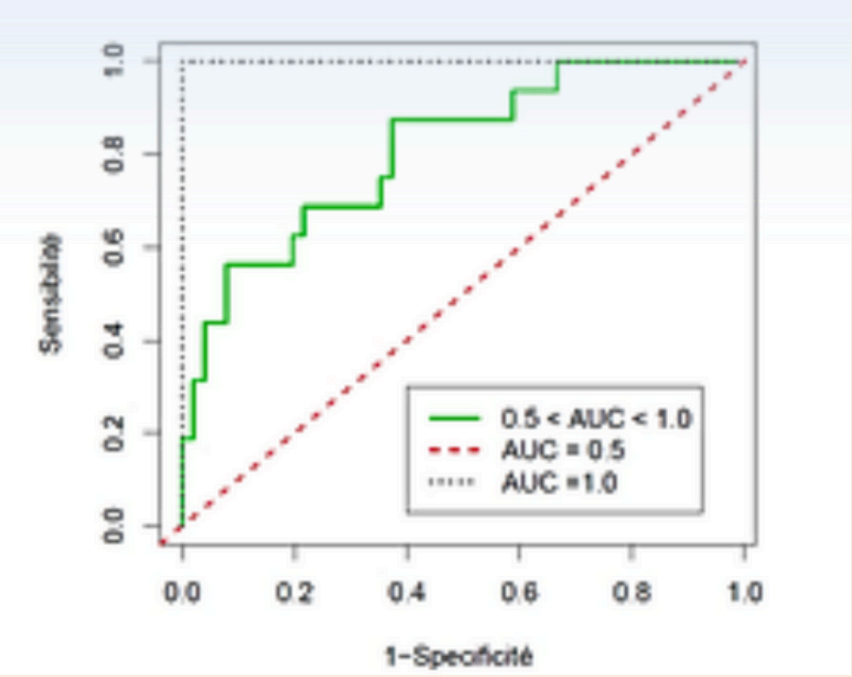

Figure 20 : Courbe ROC pour un autre modèle

Conclusion :

Dans cette section, nous avons présenté l'environnement de développement de notre modèle, suivie d'une explication détaillée des différentes étapes suivies pour le développement de notre projet. Cela dans le but de fournir une meilleure compréhension du travail effectué.

Conclusion

Ce travail s’inscrit dans le cadre de la création de notre projet d’intelligence artificielle "détection des fake news sur du texte". Il est décomposé en trois sections principales. La première a été consacrée à l'introduction de quelques généralités afin de mieux comprendre le contexte général du projet ainsi que des détails sur les données utilisées. L'étape suivante a été réservée à l’expression et à la présentation de la problématique où nous avons montré la relation entre l’intelligence artificielle et les fake news. Dans la dernière section, nous nous sommes penchés sur la programmation du modèle.

Pour ce faire, nous avons choisi de travailler sur l'environnement Jupyter. Lors du développement de notre modèle, nous avons utilisé des techniques avancées en intelligence artificielle et en apprentissage automatique. Après l'analyse de différents modèles de classification, tels que le Decision Tree Classifier, le Random Forest Classifier et le Support Vector Machine (SVM), nous avons choisi le modèle qui a donné un meilleur résultat en termes de performance.

Nous espérons que notre modèle apportera une contribution à la lutte contre la propagation des fake news. Nous nous engageons à continuer d’affiner et à améliorer notre approche afin de fournir un modèle plus performant et précis.

Sources

[1] https://larevuedesmedias.ina.fr/la-rumeur-au-moyen-age-media-des-elites-et-voix-du-peuple

[2] https://www.legifrance.gouv.fr/loda/id/ LEGITEXT000006070722

[3] Tandoc, E. C., & Lim, Z. W., & Ling, R.(2018). Defining “fake news.”Digital Journalism, 6, 137-153.

[4] Rapport de fin de cycle “Conception et réalisation d'une application web python pour la gestion du programme TUP-HIMO”

[5] https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000037847559

[6] https://fr.wikipedia.org/wiki/Loi_contre_la_manipulation_de_l%27information#:~:text=La%20loi%20contre%20la%20manipulation,et%20promulgu%C3%A9e%20le%2022%20d%C3%A9cembre%20.

[7] https://www.lucidchart.com/pages/fr/arbre-dedecision#

[8] https://dataanalyticspost.com/Lexique/svm/

[9] https://www.jedha.co/formation-ia/matrice-confusion#

[10] https://kobia.fr/classification-metrics-matrice-de-confusion/

[11] https://www.nltk.org/api/nltk.stem.wordnet.html

[12] https://en.wikipedia.org/wiki/Tf%E2%80%93idf

[13] https://github.com/scikit-learn/scikit-learn/blob/872124551/sklearn/feature_extraction/text.py#L2112

[14] https://www.sciencespo.fr/executive-education/fr/actualites/nouveau-mot-du-dictionnaire-infox-quevoque-t-il-de-notre-societe/#

[15] https://fr.wikipedia.org/wiki/Jupyter

[16) https://datascientest.com/matrice-de-confusion

[17] https://www.idbc.fr/tutoriel-comment-lire-une-courbe-roc-et-interpreter-son-auc/