



Rapport de
Projet IA et ML

Système de Détection de Fake News

Elaboré par :

Ameni ASKRI

Encadré par :

Mm Yassmine BEN DHIEB

Effectué à :

Etablissement : **ECOLE POLYTECHNIQUE PRIVEE "IBN KHALDOUN"**

Année universitaire : 2025/2026

Table des matières

1.Introduction Générale.....	1
2. Contexte général.....	2
3. Problématique.....	2
4. Solution proposée.....	2
5. Méthodologie et étapes de réalisation	3
5.0 Algorithmes et méthodes utilisés.....	3
5.1 Collecte et préparation des données	3
5.2 Vectorisation des textes (TF-IDF)	3
5.3 Modèle de classification : Multinomial Naive Bayes	4
5.4 Résumé automatique des textes.....	4
5.5 Visualisation avec WordCloud	4
5.6 Interface utilisateur avec Gradio	4
6. Modèles utilisés et résultats expérimentaux.....	5
6.1 Modèles de Machine Learning	5
a) Logistic Regression	5
b) Multinomial Naive Bayes	6
c) Linear Support Vector Machine (Linear SVM).....	6
d) K-Nearest Neighbors (KNN)	6
6.2 Modèles de Deep Learning.....	7
a) LSTM (Long Short-Term Memory).....	7
b) CNN (Convolutional Neural Network).....	7
6.3 Comparaison des modèles	7
7. Déploiement et réalisation.....	8
8. Conclusion générale	17
9.Références	18

Liste Des Figures

Figure 1: Histogramme comparatif de la longueur des articles Fake et Real	9
Figure 2: Mots les plus fréquents dans les Fake News	9
Figure 3: Mots les plus fréquents dans les Real News	10
Figure 4: Matrice de confusion obtenue avec la régression logistique	11
Figure 5: Matrice de confusion obtenue avec SVM.....	12
Figure 6: Matrice de confusion obtenue avec KNN.....	13
Figure 7: Comparaison des performances des modèles de classification (Accuracy)	13
Figure 8: Matrice de confusion obtenue avec LSTM.....	14
Figure 9: Matrice de confusion obtenue avec CNN	15
Figure 10: Déploiement sur Hugging Face Spaces	15
Figure 11: Interface Gradio déployée sur Hugging Face Spaces	16

1.Introduction Générale

À l'ère du numérique, Internet et les réseaux sociaux jouent un rôle central dans la diffusion et le partage de l'information. La rapidité de circulation des contenus permet un accès quasi instantané à l'actualité, mais elle a également favorisé la propagation massive de fausses informations, communément appelées Fake News. Ces informations trompeuses, souvent diffusées sans vérification préalable, peuvent influencer l'opinion publique, manipuler les comportements individuels, perturber la stabilité sociale et porter atteinte aux processus démocratiques.

Face à l'augmentation exponentielle du volume de données textuelles publiées chaque jour, les méthodes traditionnelles de vérification manuelle deviennent insuffisantes, coûteuses et difficilement généralisables. Dans ce contexte, l'Intelligence Artificielle (IA), et plus particulièrement le Machine Learning et le traitement automatique du langage naturel (Natural Language Processing – NLP), offre des solutions efficaces pour analyser, comprendre et classifier automatiquement les contenus textuels.

Ce projet s'inscrit dans cette dynamique et vise à concevoir et implémenter un système intelligent capable de détecter automatiquement les Fake News à partir de textes journalistiques. La solution proposée repose sur l'exploitation de différentes approches de Machine Learning et de Deep Learning, combinées à des techniques de prétraitement et de représentation des textes. En outre, le système intègre des outils d'analyse complémentaires tels que le résumé automatique et la visualisation des mots clés, afin d'améliorer l'interprétation des résultats et l'expérience utilisateur.

Dans cette optique, ce rapport présente les différentes étapes de réalisation du projet, depuis la préparation des données et la comparaison des modèles, jusqu'au déploiement d'une application interactive permettant une utilisation concrète du système en conditions réelles.

2. Contexte général

La diffusion croissante des Fake News constitue aujourd'hui un défi majeur pour les médias, les institutions publiques et les citoyens. Les plateformes numériques et les réseaux sociaux facilitent la circulation rapide et à grande échelle de contenus non vérifiés, rendant la distinction entre information fiable et information trompeuse de plus en plus complexe.

Ces fausses informations peuvent entraîner des conséquences importantes, telles que la désinformation du public, la manipulation de l'opinion et la perte de confiance envers les sources médiatiques. Les approches traditionnelles de vérification, essentiellement basées sur l'intervention humaine, deviennent insuffisantes face au volume et à la rapidité de diffusion des contenus.

Dans ce contexte, l'automatisation de la détection des Fake News à l'aide des techniques d'Intelligence Artificielle, notamment le Machine Learning et le traitement automatique du langage naturel, apparaît comme une solution essentielle pour assister les utilisateurs dans l'évaluation de la crédibilité des informations diffusées en ligne.

3. Problématique

La problématique principale de ce projet est la suivante :

Comment concevoir un système intelligent capable de classifier automatiquement un texte journalistique en Fake News ou Real News, tout en garantissant une bonne précision et une interprétation claire des résultats ?

Cette problématique soulève plusieurs défis, notamment le traitement du langage naturel, la représentation des textes sous forme numérique et le choix d'algorithmes de classification adaptés.

4. Solution proposée

La solution proposée repose sur une approche basée sur le Machine Learning et le NLP. Elle combine :

- La préparation et le nettoyage des données textuelles
- La vectorisation des textes à l'aide de la méthode TF-IDF
- L'utilisation de l'algorithme Multinomial Naive Bayes pour la classification
- L'intégration d'un modèle de résumé automatique basé sur Hugging Face

- La visualisation des mots clés à l'aide de WordCloud
- Une interface utilisateur interactive développée avec Gradio

5. Méthodologie et étapes de réalisation

5.0 Algorithmes et méthodes utilisés

Dans ce projet, plusieurs algorithmes et méthodes d'Intelligence Artificielle et de traitement automatique du langage naturel ont été utilisés afin de concevoir un système complet et performant de détection des Fake News. Chaque algorithme a été choisi en fonction de son efficacité, de sa simplicité et de son adéquation avec les données textuelles.

Les principaux algorithmes et méthodes utilisés sont :

- La méthode de vectorisation TF-IDF
- L'algorithme de classification Multinomial Naive Bayes
- Les modèles Transformers pour le résumé automatique
- Les techniques de visualisation textuelle (WordCloud)
- Une interface interactive basée sur Gradio

5.1 Collecte et préparation des données

Le projet utilise deux jeux de données[1] distincts :

- Un ensemble d'articles étiquetés comme *Fake News*
- Un ensemble d'articles étiquetés comme *Real News*

Chaque article est associé à une étiquette binaire (0 pour Fake, 1 pour Real). Les données sont ensuite fusionnées, mélangées aléatoirement et préparées pour l'entraînement du modèle.

5.2 Vectorisation des textes (TF-IDF)

Les textes bruts ne pouvant pas être directement exploités par les algorithmes de Machine Learning, la méthode TF-IDF (Term Frequency – Inverse Document Frequency) est utilisée afin de transformer les textes en vecteurs numériques.

Cette méthode permet de mesurer l'importance d'un mot dans un document par rapport à l'ensemble du corpus, tout en réduisant l'influence des mots fréquents peu informatifs.

5.3 Modèle de classification : Multinomial Naive Bayes

L'algorithme Multinomial Naive Bayes est choisi pour la classification des textes. Il est particulièrement adapté aux données textuelles représentées sous forme de fréquences ou de poids TF-IDF.

Ce modèle repose sur le théorème de Bayes et l'hypothèse d'indépendance conditionnelle entre les caractéristiques. Il offre de bonnes performances tout en étant simple et rapide à entraîner.

5.4 Résumé automatique des textes

Afin d'enrichir l'analyse, un modèle de résumé automatique basé sur la bibliothèque Hugging Face Transformers est intégré. Cette fonctionnalité permet de générer un résumé concis du texte analysé, facilitant la compréhension rapide de son contenu.

5.5 Visualisation avec WordCloud

La visualisation des mots clés est réalisée à l'aide de WordCloud. Cette représentation graphique permet de mettre en évidence les termes les plus fréquents dans le texte analysé.

5.6 Interface utilisateur avec Gradio

Une interface utilisateur interactive est développée à l'aide de Gradio. Elle permet à l'utilisateur de saisir un texte et d'obtenir :

- La prédiction Fake ou Real
- Un score de confiance
- Un résumé automatique
- Une visualisation WordCloud

6. Modèles utilisés et résultats expérimentaux

Pour évaluer les performances, plusieurs modèles de Machine Learning et Deep Learning ont été testés sur le même jeu de données. La métrique principale est l'accuracy, complétée par le classification report et la matrice de confusion.

6.1 Modèles de Machine Learning

Modèle	Accuracy
Logistic Regression	0.9915
Multinomial Naive Bayes	0.9359
Linear SVM	0.9989
K-Nearest Neighbors (K=5)	0.8804

Chaque modèle a été entraîné sur les vecteurs TF-IDF et évalué sur le jeu de test. Une matrice de confusion et un rapport de classification ont été générés pour analyser les performances.

a) Logistic Regression

La régression logistique [2] est un modèle de classification supervisée utilisé pour prédire la probabilité qu'un échantillon appartienne à une classe binaire (ici *Fake* ou *Real*). Le modèle estime la relation entre les variables d'entrée (vecteurs TF-IDF) et la probabilité d'appartenance à chaque classe en utilisant la fonction logistique (sigmoïde).

Usage pour la détection des Fake News :

Logistic Regression est rapide à entraîner et interprétable, idéal pour des textes vectorisés par TF-IDF, et permet d'obtenir un score de probabilité pour chaque prédiction.

b) Multinomial Naive Bayes

Naive Bayes est un modèle probabiliste basé sur le théorème de Bayes. Le Multinomial Naive Bayes [3] est adapté aux données textuelles représentant des fréquences ou poids de mots (TF-IDF).

Usage pour la détection des Fake News :

Naive Bayes est simple, rapide, efficace pour les textes longs, et fonctionne bien même avec un nombre élevé de caractéristiques TF-IDF.

c) Linear Support Vector Machine (Linear SVM)

Le SVM [4] linéaire est un modèle de classification qui cherche à séparer les classes par un hyperplan dans un espace à grande dimension. L'objectif est de maximiser la marge entre les points de classes différentes.

Usage pour la détection des Fake News :

Le Linear SVM est très performant sur des vecteurs TF-IDF de grande dimension et robuste face aux données bruitées.

d) K-Nearest Neighbors (KNN)

KNN [5] est un algorithme de classification basé sur la **proximité**. Un document est classé selon la majorité des k documents les plus proches dans l'espace des caractéristiques.

Principe :

1. Calculer la distance (ex : Euclidienne ou Cosine) entre le document à classer et tous les documents du jeu d'entraînement.
2. Identifier les k plus proches voisins.
3. Classer selon la majorité des classes des voisins.

Usage pour la détection des Fake News :

Simple et intuitif, mais moins efficace sur des jeux de données volumineux ou avec des vecteurs TF-IDF de grande dimension.

6.2 Modèles de Deep Learning

a) LSTM (Long Short-Term Memory)

LSTM [6] est un type de réseau de neurones récurrents (RNN) conçu pour capturer des dépendances à long terme dans les séquences, ici les textes.

LSTM utilise des portes (entrée, sortie, oubli) pour décider quelles informations conserver ou oublier à chaque pas de la séquence. Cela permet de modéliser le contexte des mots sur des textes longs.

Permet de capturer l'ordre des mots et les dépendances contextuelles, ce qui améliore la précision sur des textes complexes.

- Accuracy obtenue : 0.98

b) CNN (Convolutional Neural Network)

Appliqué au texte est un réseau de neurones capable d'extraire automatiquement des caractéristiques locales à partir des séquences de mots.

Principe :

1. Les mots sont transformés en vecteurs (*embeddings*).
2. Des filtres convolutifs glissent sur les séquences pour détecter des motifs locaux (ex : expressions, phrases types).
3. Une couche de pooling sélectionne les caractéristiques importantes pour la classification.

Usage pour la détection des Fake News :

Le CNN [7] capture des motifs et expressions clés qui peuvent être typiques des Fake News, offrant ainsi une performance élevée.

- Accuracy obtenue : 0.99

6.3 Comparaison des modèles

Les résultats montrent que les modèles linéaires comme Linear SVM et Logistic Regression offrent d'excellentes performances sur ce type de données. Les modèles de Deep Learning (LSTM et CNN) obtiennent également des résultats très élevés, mais nécessitent un temps d'entraînement plus important et davantage de ressources.

Malgré une accuracy légèrement inférieure, le modèle Naive Bayes a été retenu pour la réalisation finale en raison de sa simplicité, de sa rapidité d'exécution et de sa facilité d'intégration dans une application interactive.

7. Déploiement et réalisation

Après la phase d'expérimentation et de comparaison des modèles, le projet a été concrétisé par le déploiement d'une application fonctionnelle. Le modèle Multinomial Naive Bayes, combiné à la vectorisation TF-IDF, a été retenu pour l'application finale en raison de sa simplicité, de sa rapidité et de sa facilité d'intégration.

L'application dispose d'une interface interactive développée avec Gradio, qui permet à l'utilisateur de saisir un texte et d'obtenir instantanément :

- La prédiction (*Fake News* ou *Real News*)
- Un score de confiance associé à la prédiction
- Un résumé automatique du texte basé sur un modèle Transformer (Hugging Face)
- Une visualisation des mots clés sous forme de WordCloud

Cette interface a ensuite été déployée sur Hugging Face Spaces [8], offrant un accès en ligne simple et rapide, sans nécessiter d'installation ou de configuration locale. Le déploiement permet de transformer le projet en une véritable application IA utilisable en conditions réelles.

Captures d'écran de la réalisation

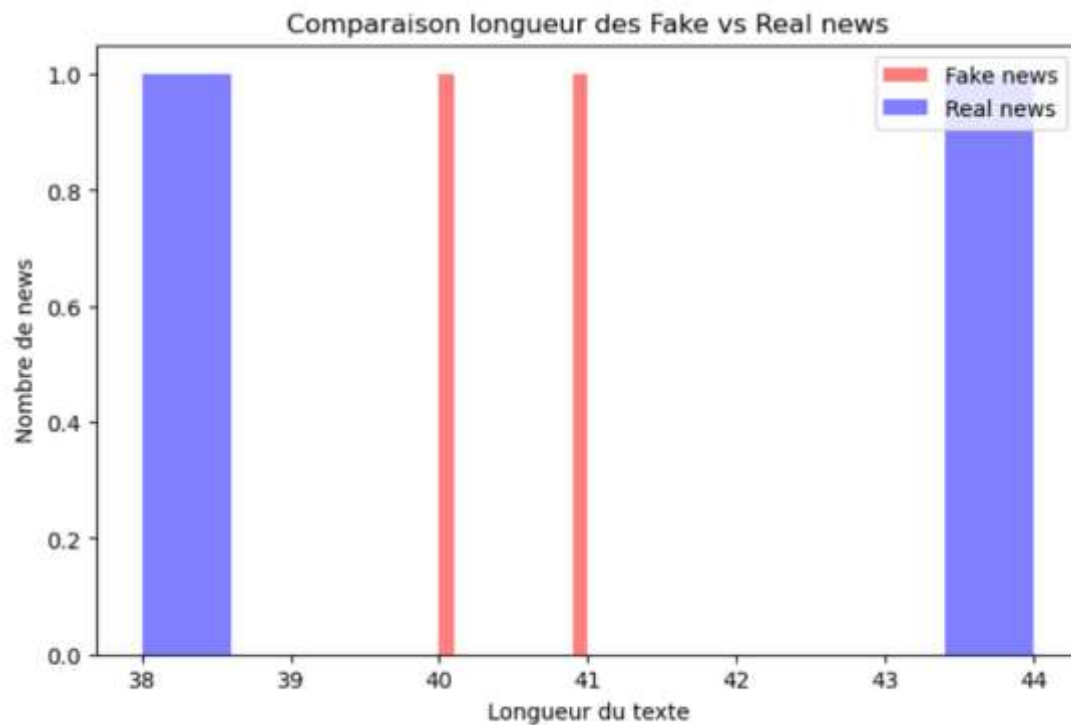


Figure 1: Histogramme comparatif de la longueur des articles Fake et Real

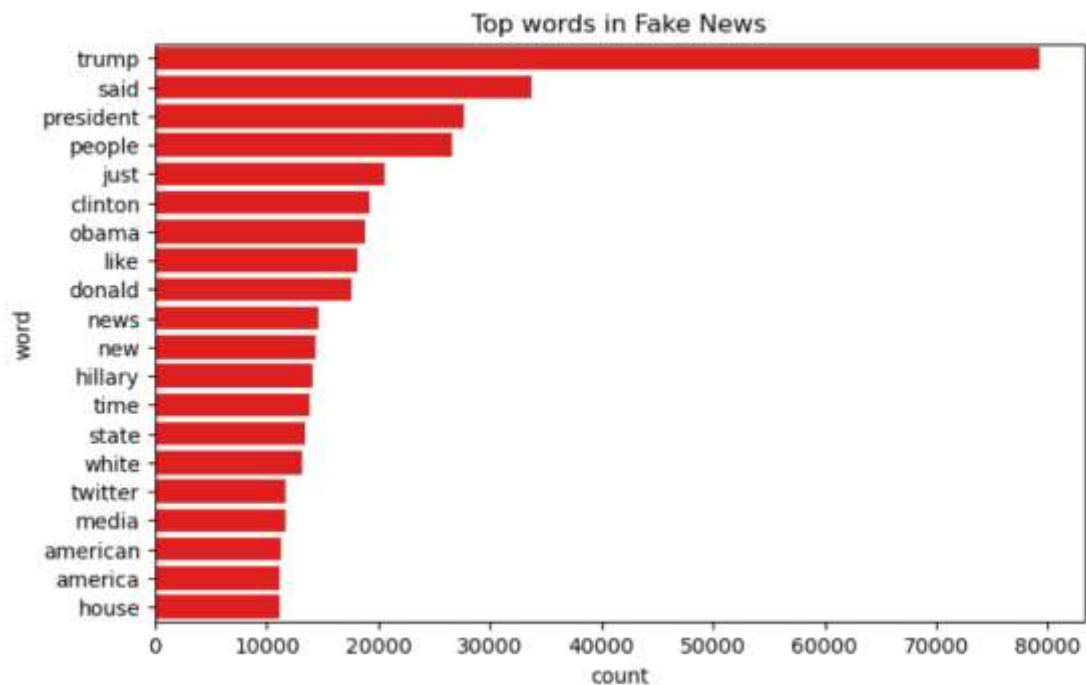


Figure 2: Mots les plus fréquents dans les Fake News

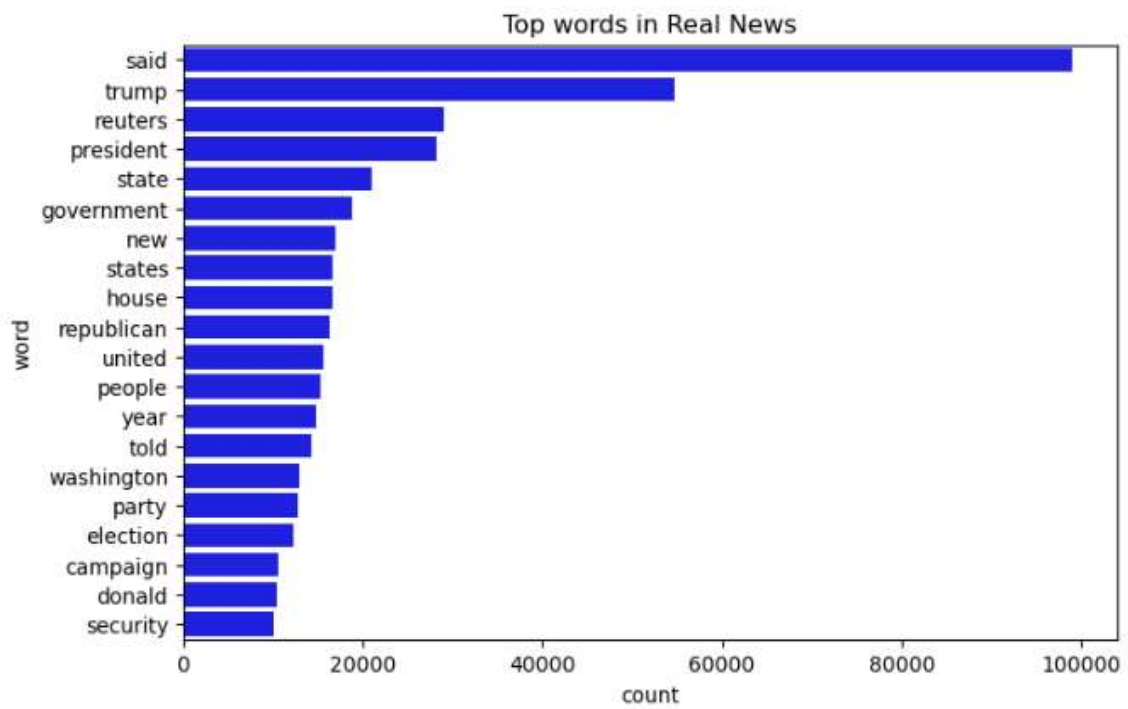


Figure 3: Mots les plus fréquents dans les Real News

[2]:

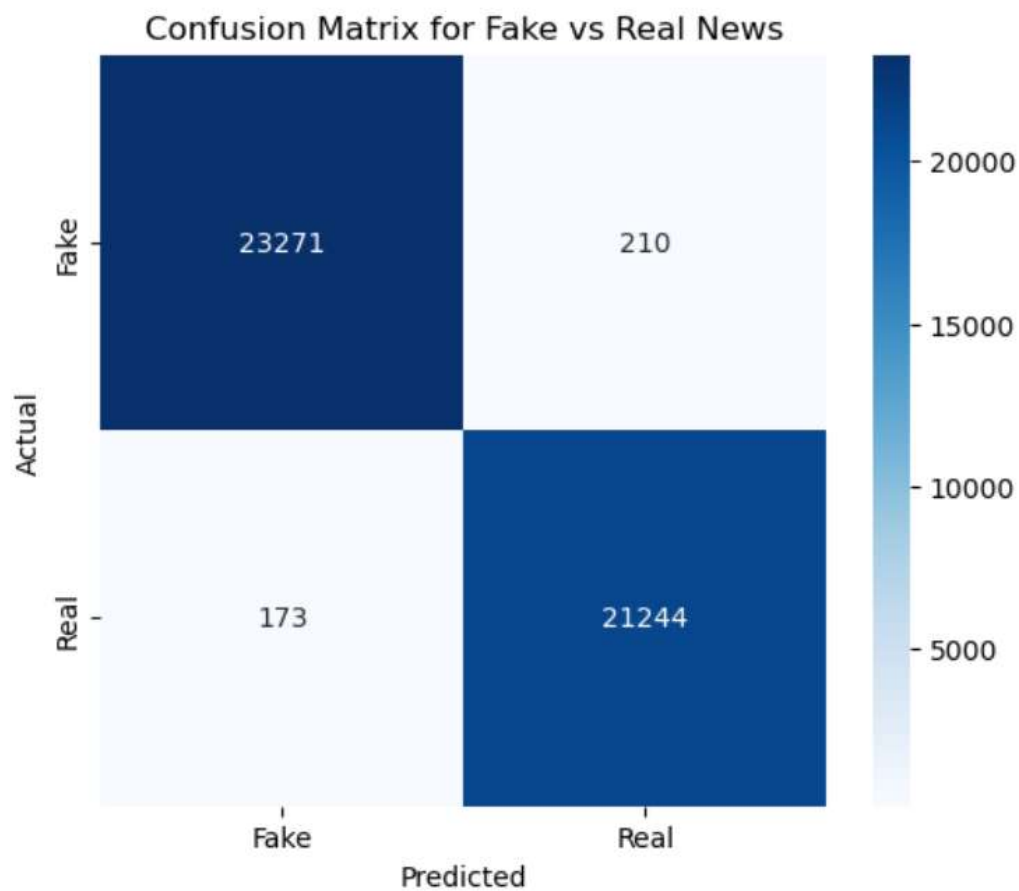


Figure 4: Matrice de confusion obtenue avec la régression logistique

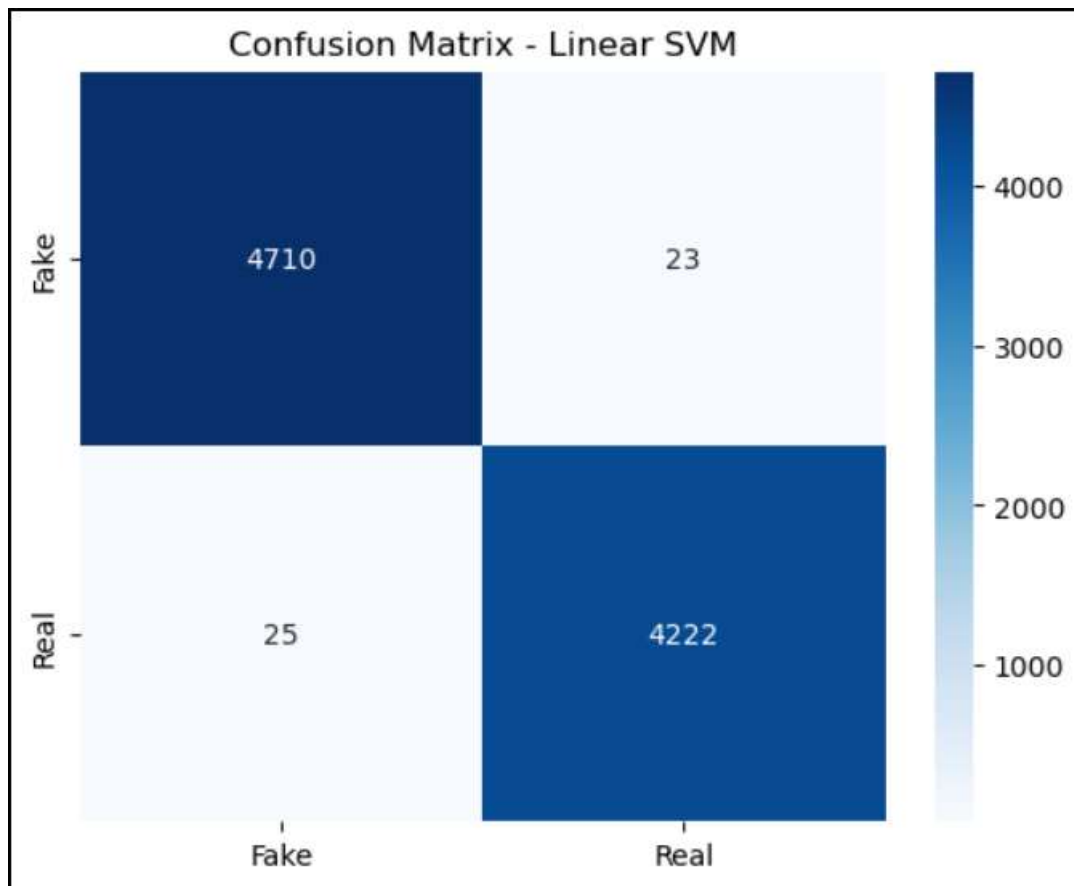


Figure 5: Matrice de confusion obtenue avec SVM

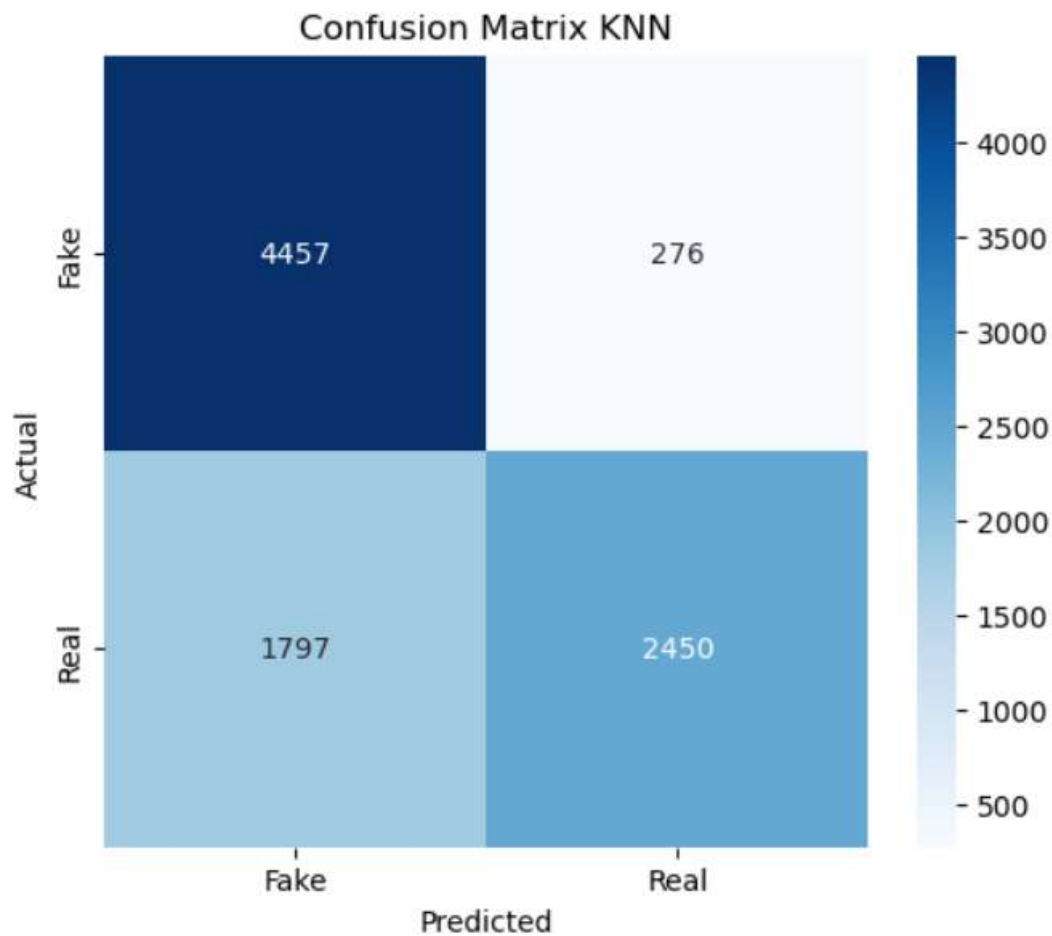


Figure 6: Matrice de confusion obtenue avec KNN

```

Logistic Regression → Accuracy : 0.9915
Naive Bayes → Accuracy : 0.9359
c:\Users\amani\anaconda3\Lib\site-packages\sk
warnings.warn(
Linear SVM → Accuracy : 0.9989
KNN → Accuracy : 0.8804

```

Figure 7: Comparaison des performances des modèles de classification (Accuracy)

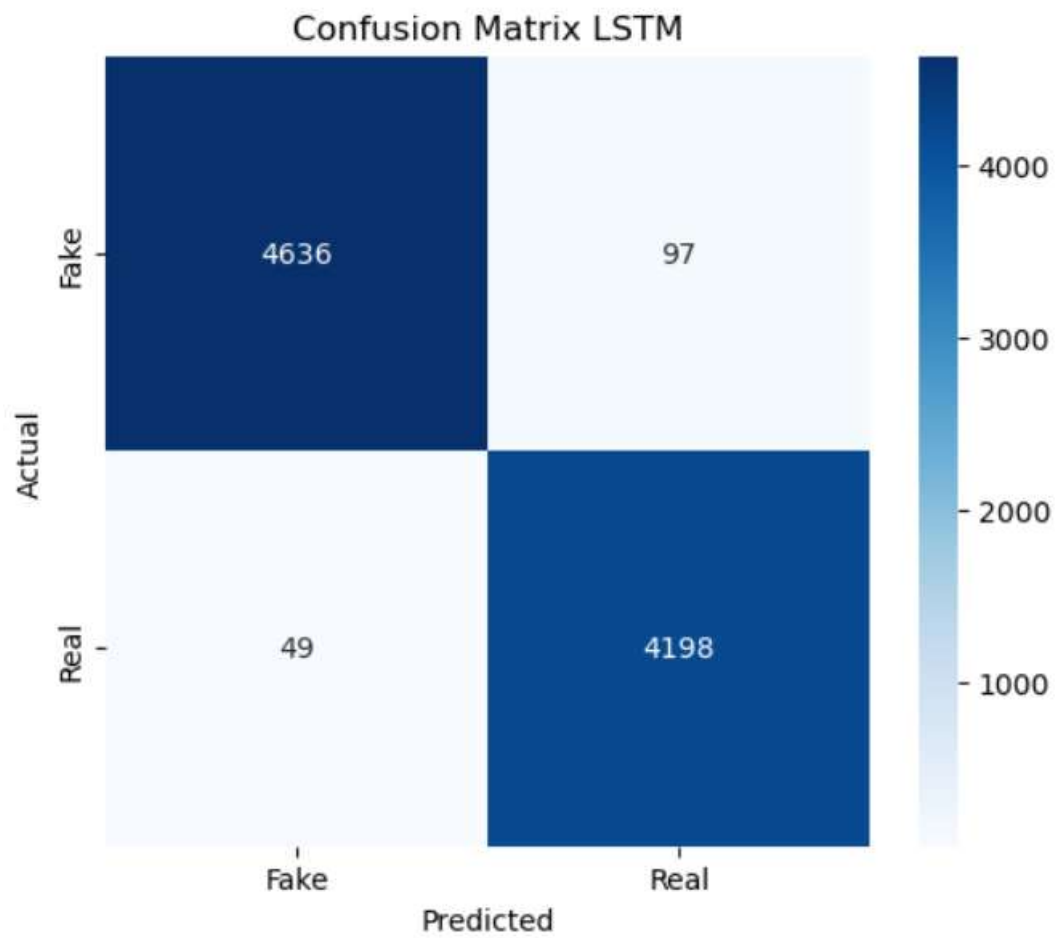


Figure 8: Matrice de confusion obtenue avec LSTM

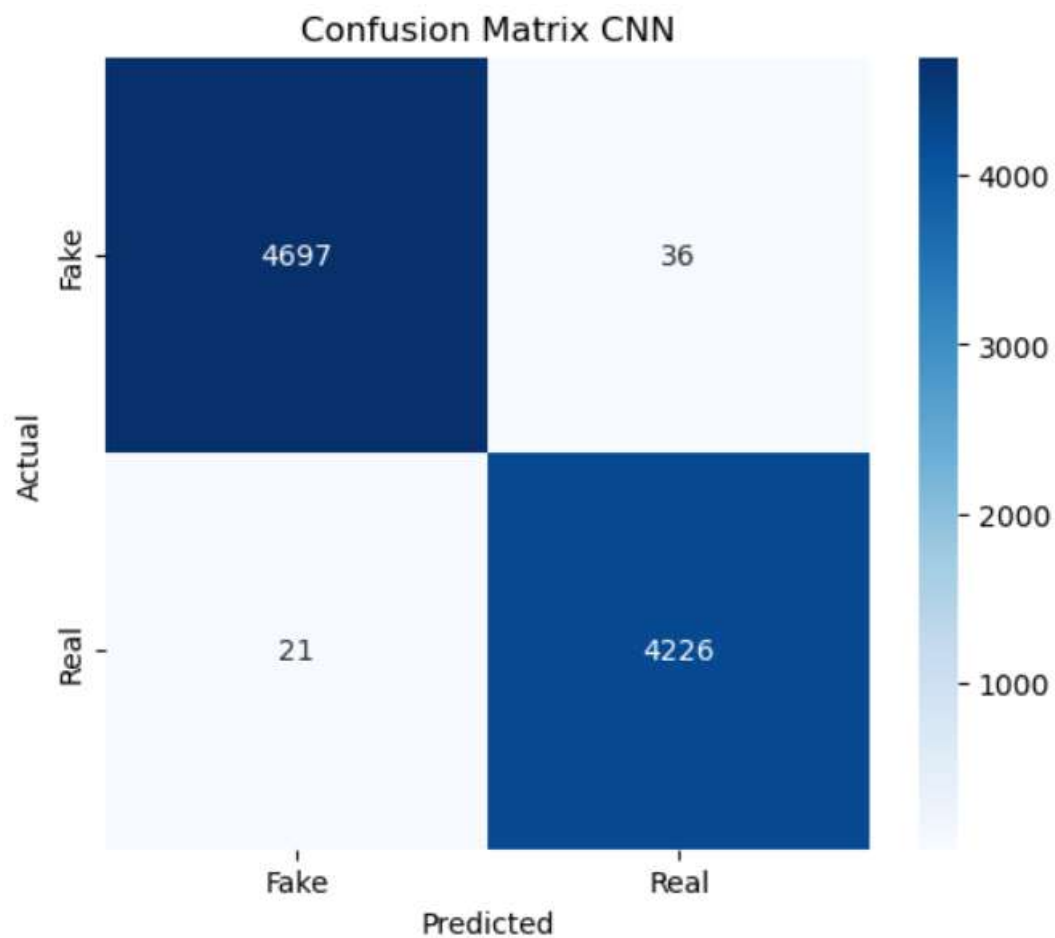


Figure 9: Matrice de confusion obtenue avec CNN

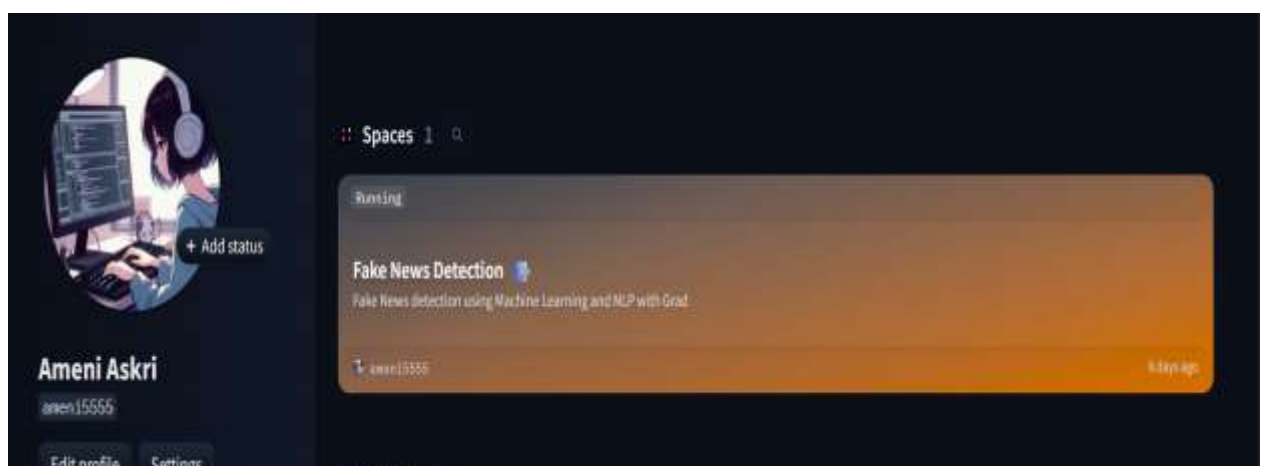


Figure 10: Déploiement sur Hugging Face Spaces



Figure 11: Interface Gradio déployée sur Hugging Face Spaces

8. Conclusion générale

Ce projet a permis de concevoir, développer et déployer une application intelligente de détection automatique des Fake News reposant sur les techniques de l'Intelligence Artificielle et du traitement automatique du langage naturel. L'approche adoptée, combinant la vectorisation TF-IDF et l'algorithme Multinomial Naive Bayes, s'est révélée efficace pour la classification des textes journalistiques, tout en offrant un bon compromis entre performance, simplicité et rapidité d'exécution.

L'étude comparative menée sur plusieurs modèles de Machine Learning et de Deep Learning a permis de mettre en évidence les avantages et les limites de chaque approche. Bien que certains modèles plus complexes aient présenté des performances élevées, le choix du modèle Naive Bayes pour la réalisation finale a été motivé par sa facilité d'intégration et son adéquation avec une application interactive en temps réel.

L'intégration de fonctionnalités complémentaires, telles que le résumé automatique basé sur les modèles Transformers et la visualisation des mots clés via WordCloud, a enrichi l'analyse des résultats et amélioré significativement l'expérience utilisateur. Le déploiement de l'application à l'aide de Gradio sur la plateforme Hugging Face Spaces a permis de transformer ce travail académique en une application concrète, accessible en ligne et exploitable dans des conditions réelles.

En perspective, ce projet pourrait être étendu par l'utilisation de modèles de Deep Learning et de Transformers plus avancés pour la classification, l'extension à une analyse multilingue, l'augmentation du volume et de la diversité des données, ainsi que l'intégration en temps réel sur des plateformes numériques et des réseaux sociaux. Ces améliorations permettraient de renforcer la robustesse et l'applicabilité du système face aux enjeux croissants de la désinformation.

9. Références

- [1] [Fake-News-Dataset](#)
- [2] [Logistic Regression - Recherche](#)
- [3] [Multinomial Naive Bayes - Recherche](#)
- [4] [Comprendre les Support Vector Machines \(SVM\)](#)
- [5] [K-Nearest Neighbor\(KNN\) Algorithm - GeeksforGeeks](#)
- [6] [LSTM \(Long Short-Term Memory\) - Recherche](#)
- [7] [Convolutional neural network - Wikipedia](#)
- [8] [ameni5555 \(Ameni Askri\)](#)