****

**MASTER II : Fouille de données et intelligence artificielle**

**Construction de modèles et leur déploiement**

**Modèle de Détection de Tweet Suspect**

**Projet réalisé par Groupe 9:**

**SOGO Armel Emmanuel**

**OUEDRAOGO Relwindé Guillaume Victorien**

**Enseignant:**

**Dr. Abdoul Kader KABORE**

**Introduction :**

La régulation des médias sociaux, en particulier X, est essentielle pour maintenir un espace numérique sûr et respectueux. Étant donné la rapidité avec laquelle les informations circulent sur cette plateforme, les tweets malveillants qui incluent discours haineux, désinformation, harcèlement, et manipulation peuvent avoir des répercussions immédiates et graves sur les individus et les communautés.

La détection de tels tweets repose sur des algorithmes de traitement de langage naturel et sur des systèmes de modération humaine.

Ce projet vise à développer un modèle de Machine Learning capable de classifier les tweets en deux catégories : "suspect" (menaces, terrorisme, intimidation) et "non suspect". L'objectif est d'automatiser l'identification des discours dangereux sur les plateformes sociales pour prévenir des comportements à risque.

**Méthodologie :**

**1- Exploration et Prétraitement des Données**

**Importation et Exploration des données**

* Le jeu de données tweets\_suspect.csv a été importé depuis notre stockage Google Drive.
* Les tweets sont stockés dans la colonne message, et les labels (0 pour non suspect, 1 pour suspect) dans la colonne label.
* Un fort déséquilibre entre les classes a été observé, avec une prédominance de tweets suspects.
* Analyse de la polarité et des sentiments des tweets à l'aide de TextBlob.

**Prétraitement des Données :**

* Les tweets ont été nettoyés en supprimant les balises HTML, les espaces multiples et les caractères spéciaux, la conversion en minuscules, et la suppression des mots vides (stop words). Ces étapes sont essentielles pour simplifier les données et réduire le bruit.
* Une lemmatisation a été appliquée pour réduire les mots à leur forme de base.

**Représentation Vectorielle (Embedding) :**

* Le modèle DistilBERT a été utilisé pour générer des embeddings (représentations vectorielles) des tweets.
* DistilBERT est un choix judicieux pour les environnements avec des ressources limitées.

**2- Construction du Modèle**

**Équilibrage des classes :** Le dataset étant déséquilibré, la technique de sur-échantillonnage SMOTE a été appliquée pour corriger ce déséquilibre.

**Division des Données :**

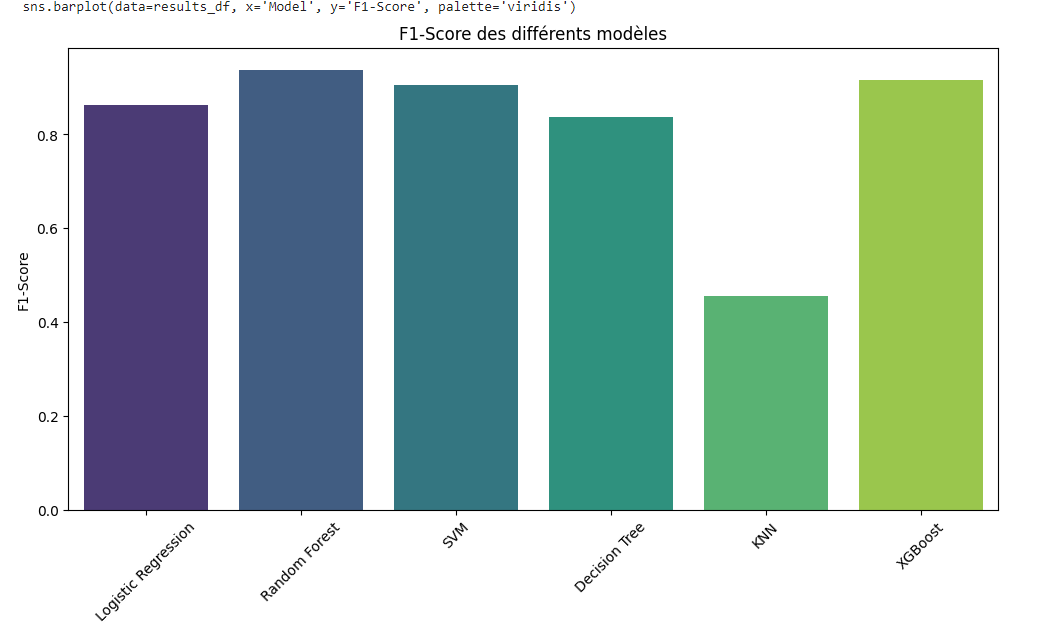
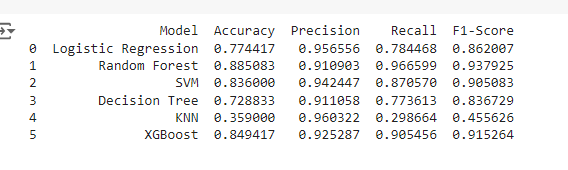
Les données ont été divisées en ensembles d'entraînement (80%) et de test (20%).

**Entraînement et Évaluation des Modèles :**

* Plusieurs modèles de classification ont été entraînés et évalués avec scikit-learn :
* Régression Logistique
* Random Forest
* SVM
* Arbre de Décision
* KNN
* XGBoost
* Les performances des modèles ont été mesurées en utilisant l'accuracy, la précision, le rappel, le F1-score, les matrices de confusion et la courbe ROC

**3- Résultats :**

* Le modèle Radom Forest s'est avéré le plus performant, avec une accuracy de 88,5% et un excellent rappel de 96,7%.
* XGBoost a également obtenu de bons résultats, mais légèrement inférieurs à ceux de Random Forest.
* Les autres modèles ont montré des performances correctes, mais moins efficaces.



**4- Réflexions sur le Projet et Améliorations Futures**

### Réflexions

* Le nettoyage et la préparation des données ont joué un rôle crucial dans l'obtention de bonnes performances.
* L'utilisation de DistilBERT pour générer des embeddings a permis de capturer des informations sémantiques importantes des tweets.
* Le traitement du déséquilibre de classes a été essentiel pour éviter un biais du modèle vers la classe majoritaire.
* L'optimisation des hyperparamètres du modèle SVM a permis d'améliorer encore ses performances.

### Améliorations possibles

* Explorer d'autres techniques de prétraitement du texte, comme la stemming.
* Expérimenter avec d'autres modèles de langage pour les embeddings (ex: BERT, RoBERTa).
* Ajuster davantage les hyperparamètres du modèle en utilisant des techniques plus avancées.
* Collecter plus de données pour améliorer la robustesse du modèle.
* Intégrer des informations contextuelles supplémentaires, comme les informations sur l'utilisateur

**Conclusion :**

Ce projet a démontré la faisabilité de la classification de tweets suspects avec un haut niveau de précision en utilisant un modèle SVM. La démarche rigoureuse de préparation des données, d'entraînement et d'optimisation du modèle a permis d'obtenir des résultats satisfaisants. Il existe cependant des pistes d'amélioration pour rendre le modèle encore plus performant et robuste

Lien du projet : <https://github.com/Sogo95/Detection-de-tweet>