

# **Titre du Projet : Analyse de Sentiment des Publications sur l'UVBF à partir des Données de Twitter et Facebook**

**Objectif :** Les étudiants vont collecter et analyser des données provenant des réseaux sociaux Twitter (X) et Facebook pour mesurer les sentiments des discussions autour de leur université, UVBF. Le projet inclut les étapes de scraping, prétraitement, vectorisation, entraînement et évaluation d'un modèle de classification.

## **Étapes du Projet :**

### **1. Scraping de Données sur Twitter et Facebook**

- **Tâches :**
  - Utiliser l'API Twitter (ou Tweepy) pour récupérer des tweets contenant des mots-clés associés à **UVBF** (ex. : "Université UVBF", "UVBF", hashtag #UVBF).
  - Utiliser l'API de Facebook (ou une technique de scraping) pour extraire des publications publiques mentionnant **UVBF**.
  - Sauvegarder les données dans un fichier CSV/JSON structuré, avec le texte des publications, la date, et l'auteur (si disponible).
- **Livrable :** Fichier contenant des tweets et des publications Facebook, avec au moins 500 entrées relatives à **UVBF**.

### **2. Prétraitement des Données Textuelles**

- **Tâches :**
  - Nettoyer les tweets et publications : suppression des mentions (@), des hashtags, des URL, de la ponctuation, des emojis et tout texte non pertinent.
  - Transformer le texte en minuscules, tokenisation, suppression des stop words et lemmatisation/stemming (avec spaCy ou NLTK).
- **Livrable :** Texte nettoyé et prêt à être utilisé pour la vectorisation.

### 3. Vectorisation des Textes

- **Tâches :**
  - Transformer les textes en vecteurs numériques en utilisant TF-IDF.
  - Créer une matrice de caractéristiques représentant les tweets et publications Facebook.
- **Livrable :** Matrice de vecteurs TF-IDF prête pour l'entraînement du modèle.

### 4. Entraînement d'un Modèle de Classification de Sentiments

- **Tâches :**
  - Annoter les tweets et publications en définissant manuellement des labels (positif, neutre, négatif).
  - Diviser les données en ensembles d'entraînement (80%) et de test (20%).
  - Entraîner un modèle de classification tel que Naive Bayes, Logistic Regression ou SVM pour prédire le sentiment des publications.
  - Donner le taux commentaire positif et négatif sur l'uvbf sur les réseaux sociaux.
- **Livrable :** Modèle de classification entraîné avec les données annotées.

### 5. Évaluation du Modèle

- **Tâches :**
  - Évaluer le modèle sur l'ensemble de test avec des métriques telles que la précision, le rappel, la F1-score, et l'exactitude.
  - Générer une matrice de confusion et analyser les faux positifs/négatifs.
- **Livrable :** Rapport d'évaluation avec les résultats des métriques et des visualisations (matrice de confusion).

## 6. Améliorations et Discussion

- **Tâches :**
  - Proposer des améliorations telles que l'intégration de word embeddings (ex. : BERT) ou d'autres ajustements au pipeline de prétraitement.
  - Tester les nouvelles approches et comparer les performances.
- **Livrable :** Rapport final contenant une analyse des résultats et des recommandations pour améliorer les performances du modèle.

### Technologies et Bibliothèques :

- **Scraping :** API Twitter (Tweepy), API Facebook ou scraping avec BeautifulSoup.
- **Prétraitement :** NLTK, spaCy.
- **Vectorisation :** TF-IDF (via sklearn).
- **Machine Learning :** Naive Bayes, Logistic Regression, SVM (via sklearn).
- **Évaluation :** sklearn.metrics.

### Délivrables Finaux :

1. **Code complet :** pour le scraping, prétraitement, vectorisation, entraînement, et évaluation du modèle.
2. **Rapport :** décrivant les approches utilisées, les résultats obtenus et les améliorations proposées.
3. Délai : 30 octobre 2024 à 23h 59
4. Vous pouvez déposer vos livrables dans ce dossier :  
[https://drive.google.com/drive/folders/1PK6\\_ujyrQy7ggydF\\_73MqUQFG95eAqko?usp=s\\_haring](https://drive.google.com/drive/folders/1PK6_ujyrQy7ggydF_73MqUQFG95eAqko?usp=s_haring)

**NB : tous les livrables doivent être mis dans un seul dossier et le dossier doit être renommé au nom du groupe (Groupe1).**