

# TP Final

## Développer une API d'Analyse de Sentiments

Frédéric  
Dabadie

### Analyse des matrices de confusion

#### Matrice de confusion (Model Positif)

Classe	Prédit 0 (Haineux)	Prédit 1 (Non Haineux)
Réel 0 (Haineux)	5	2
Réel 1 (Non Haineux)	3	3

#### Matrice de confusion (Model Négatif)

Classe	Prédit 0 (Non Haineux)	Prédit 1 (Haineux)
Réel 0 (Non Haineux)	3	3
Réel 1 (Haineux)	2	5

### Les biais potentiels

- Biais de données:**
  - Le jeu de données est petit et peut ne pas être représentatif de la diversité des tweets.
  - Certaines catégories de langage (argot, abréviations) peuvent ne pas être bien capturées par le modèle.
- Biais de Classe:**
  - Un déséquilibre entre les classes pourrait fausser les prédictions. Si une classe est sous-représentée, le modèle pourrait avoir du mal à bien la reconnaître.
- Biais de Prétraitement:**
  - L'élimination des mots vides (stopwords) peut parfois supprimer des mots avec des informations importantes pour l'interprétation du texte.

# Explication des Métriques de Performance

**Précision :** Proportion des prédictions correctes parmi celles faites pour une classe donnée. Elle mesure combien des tweets classés comme haineux/non haineux le sont réellement.

**Rappel :** Proportion des exemples réels d'une classe correctement identifiés. Il indique la capacité du modèle à trouver tous les tweets haineux ou non haineux.

**F1-score :** Moyenne harmonique entre précision et rappel. Une valeur élevée indique un bon équilibre entre ces deux métriques.

**Support :** Nombre d'exemples réels pour chaque classe dans l'ensemble de test.

## Analyse des Performances

### Modèle Positif

Classe	Précision	Rappel	F1-score	Support
0 (Haineux)	0.62	0.71	0.67	7
1 (Non Haineux)	0.60	0.50	0.55	6
Moyenne pondérée	0.61	0.62	0.61	13

### Modèle Négatif

Classe	Précision	Rappel	F1-score	Support
0 (Haineux)	0.60	0.50	0.55	6
1 (Non Haineux)	0.62	0.71	0.67	7
Moyenne pondérée	0.61	0.62	0.61	13

## Observations

- Le score F1 moyen est de 0.61, indiquant une performance modérée.
- Le rappel pour la classe 1 (non haineux) dans le modèle positif est faible (0.50), ce qui signifie que de nombreux tweets non haineux sont mal classés.
- Le modèle négatif a un rappel plus élevé pour la classe haineuse (0.71), mais souffre de faux positifs (3 tweets mal classés).

## Recommandations pour Améliorer le Modèle

- **Augmenter la taille du dataset:**
  - Le jeu de données actuel est très limité (50 tweets positifs/négatifs). L'ajout de plus de données réduirait les généralisations.
- **Améliorer la préparation des données:**
  - Tester d'autres méthodes de nettoyage de texte (Suppression des mots rares, etc).

## Exemple d'un appel API

```
1 {  
2   "tweets": [  
3     "Un leader qui n'a pas peur de dire ce qu'il pense, ça fait du bien !",  
4     "Son dernier discours était rempli de contradictions...",  
5     "Il a vraiment transformé l'économie, les chiffres parlent d'eux-mêmes.",  
6     "Son influence sur la politique est une catastrophe.",  
7     "Il est le seul à s'opposer aux élites corrompues.",  
8     "Encore une déclaration absurde, il vit dans son propre monde.",  
9     "Treupe 2024, c'est notre seul espoir pour l'avenir du pays !",  
10    "Comment peut-on encore soutenir une telle personne après tout ça ?",  
11    "Il ose affronter le politiquement correct, c'est rare aujourd'hui.",  
12    "Une gestion des crises totalement désastreuse, un vrai fiasco."  
13  ]  
14 }
```

Body Cookies Headers (5) Test Results ↻

{ } JSON ▾

▶ Preview

🔗 Visualize ▾

```
1 {  
2   "Comment peut-on encore soutenir une telle personne après tout ça ?": -0.2,  
3   "Encore une déclaration absurde, il vit dans son propre monde.": -0.498,  
4   "Il a vraiment transformé l'économie, les chiffres parlent d'eux-mêmes.": 0.294,  
5   "Il est le seul à s'opposer aux élites corrompues.": 0.509,  
6   "Il ose affronter le politiquement correct, c'est rare aujourd'hui.": 0.282,  
7   "Son dernier discours était rempli de contradictions...": -0.467,  
8   "Son influence sur la politique est une catastrophe.": -0.485,  
9   "Treupe 2024, c'est notre seul espoir pour l'avenir du pays !": 0.503,  
10  "Un leader qui n'a pas peur de dire ce qu'il pense, ça fait du bien !": 0.614,  
11  "Une gestion des crises totalement désastreuse, un vrai fiasco.": 0.026  
12 }
```