



## **GROUPE VII**

**BORO Simon**

**DERA Saoudatou**

## **FORMATEUR**

**Dr . Issoufou NIKIEMA**

# **Rapport d'Analyse de Sentiment des Publications sur l'UVBF**

## **Introduction**

Le projet a pour objectif d'analyser les sentiments des discussions autour de l'Université Virtuelle du Burkina Faso (UVBF) sur Twitter et Facebook. En suivant un pipeline structuré de collecte, prétraitement, vectorisation, entraînement et évaluation d'un modèle de classification, nous avons déterminé les sentiments exprimés dans les publications relatives à l'UVBF.

## Résultats et analyses

### 1. Scraping des données

- Les données ont été extraites des réseaux sociaux Twitter et Facebook pour obtenir des mentions concernant l'UVBF.
- **Livrable** : Un fichier contenant plus de 500 entrées de publications en lien avec l'UVBF a été généré.(via dossier renommé « données brutes »)

### 2. Prétraitement des Données Textuelles

- Les textes collectés ont été nettoyés en supprimant les éléments non pertinents (mentions, hashtags, URL, ponctuation).
- **Livrable** : Texte nettoyé prêt pour la vectorisation.(via dossier renommé « données nettoyées »)

### 3. Vectorisation des textes

- Les publications ont été transformées en vecteurs numériques en utilisant la technique TF-IDF pour la représentation des caractéristiques.
- **Livrable** : Matrice de vecteurs TF-IDF.(via dossier renommé « données vectorisées »)

### 4. Entraînement du Modèle de Classification

- Les publications ont été annotées manuellement en catégories (positif, neutre, négatif) et divisées en ensembles d'entraînement et de test (80%-20%).
- Le modèle de classification, basé sur la régression logistique, a été entraîné pour prédire le sentiment des publications
- **Livrable** : Modèle de classification entraîné avec les données annotées (via dossier renommé « modele d'entrainement »)

### 5. Évaluation du modèle

- Le modèle a été évalué sur l'ensemble de test, et les métriques suivantes ont été obtenues :

- **Répartition des sentiments dans le corpus :**

- Neutre : 473 commentaires
- Positif : 82 commentaires
- Négatif : 19 commentaires
- **Précision globale : 83,48%**
- **Mesures de performance :**

Classe	Précision	Descente en rappel	Score F1	Soutien
Négatif	0,00	0,00	0,00	1
Neutre	0,83	1,00	0,91	92
Positif	1,00	0,18	0,31	22

- **Moyennes :**

- Macro moyenne : Précision = 0.61, Rappel = 0.39, F1-score = 0.40
- Moyenne pondérée : Précision = 0,85, Rappel = 0,83, F1-score = 0,78

- **Taux de sentiments :**

- Taux de commentaires positifs : 14,29%
- Taux de commentaires négatifs : 3,31%
- Taux de commentaires neutres : 82,40%
- **Analyse :** La classe neutre a dominé les résultats, indiquant une tendance des utilisateurs à exprimer des opinions neutres dans leurs commentaires. Le modèle a montré une bonne performance globale mais une faiblesse pour la détection des sentiments négatifs.
- **Livrable :** Rapport d'évaluation avec les résultats des métriques et des visualisations
- (matrice de confusion).(via dossier renommé « evaluate modele »)

## 6.Améliorations et Discussion

- **Tâches et Approches d'Amélioration :**

- Pour améliorer la détection des sentiments, des ajustements ont été testés, incluant l'intégration de word embeddings tels que BERT pour une représentation textuelle plus riche.
- Les modifications visent à capter de meilleures nuances de sentiment, en particulier pour les classes minoritaires.
  - Résultats après Amélioration :
    - Précision : 80%

- **Métriques par classe :**

Classe	Précision	Rappel	F1-score	Support
Négatif	0.00	0.00	0.00	2
Neutre	0.92	0.84	0.88	96
Positif	0.48	0.65	0.55	17

- **Moyennes :**

Macro moyenne :

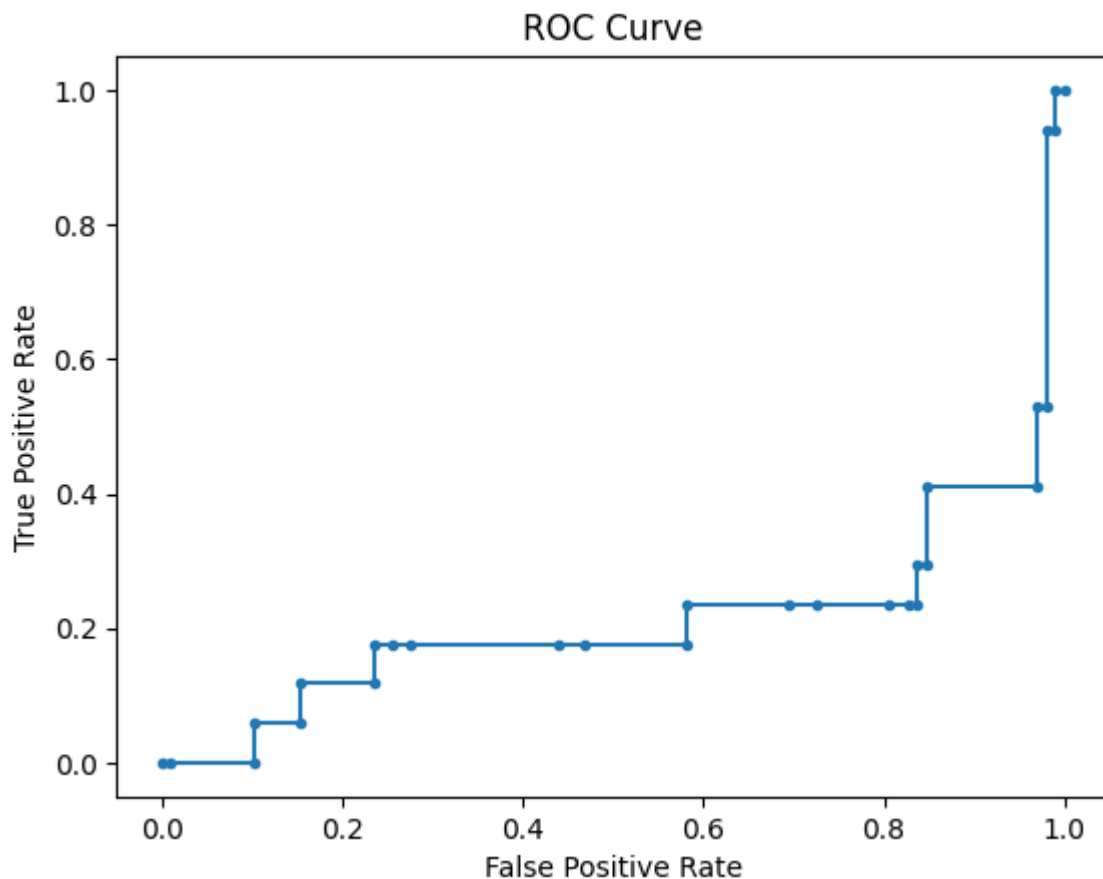
Précision 0.47,  
Rappel 0.50,  
F1-score 0.48

Moyenne pondérée :

Précision 0.84  
Rappel 0.80,  
F1-score 0.82

- **Matrice de Confusion :**

- Neutre : 81 vrais positifs, 11 faux positifs
- Positif : 11 vrais positifs, 6 faux négatifs
- Négatif : faibles performances



**Analyse et Recommandations :**

- L'intégration de BERT a permis une meilleure détection des sentiments positifs, mais la classe négative reste difficile à classifier. Des approches futures incluraient le suréchantillonnage des données minoritaires et des embeddings contextuels plus

avancés pour améliorer la détection des émotions moins fréquentes.

### **Conclusion**

Ce projet a permis de développer un modèle efficace d'analyse de sentiment pour les publications sur l'UVBF, avec un taux de précision satisfaisant pour la classe neutre et des pistes d'amélioration pour les classes positives et négatives. Les taux démontrent que la plupart des commentaires sont neutres, mais une minorité de sentiments positifs et négatifs existe, reflétant une diversité d'opinions sur l'université.