****

**RAPPORT DE PROJET**

Etudiant :

**- GORO Issouf Ousmane**

**- KONSEIBO Abdel Aziz**

Enseignant(s) :

* **Dr. Abdoul Kader KABORE** [abdoulkader.kabore@protonmail.com]

**Détection de tweets suspects**

Lien du repo git :

**1. Introduction**

Le développement des médias sociaux est en pleine expansion de nos jours. Ces médias sociaux constituent des sources d’information et permettent aussi de lier des relations avec la communauté des utilisateurs de ces médias. Parmi ces médias il y’a le réseau social X anciennement appelé Twitter qui permet d’échanger avec ses abonnés de courts messages informatifs (tweets).

Notre étude consiste à mettre en place un modèle de machine learning qui permet de dire si un message tweet est suspect ou non suspect.

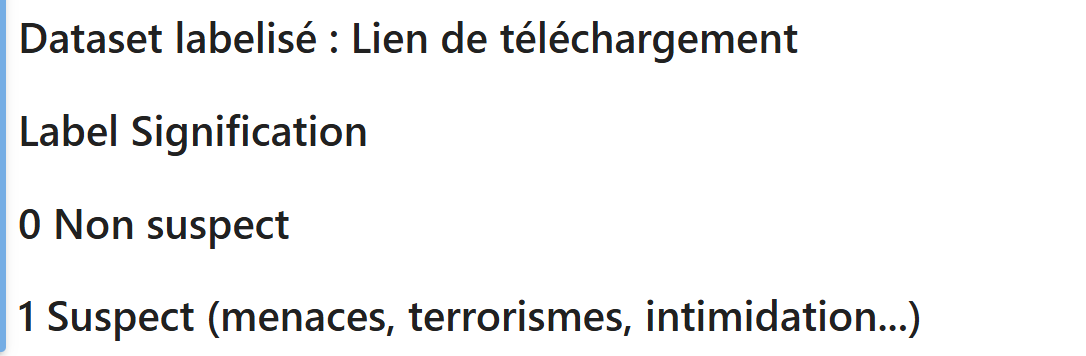
introduction, une méthodologie, des résultats, une discussion sur les limites et des suggestions d’améliorations.

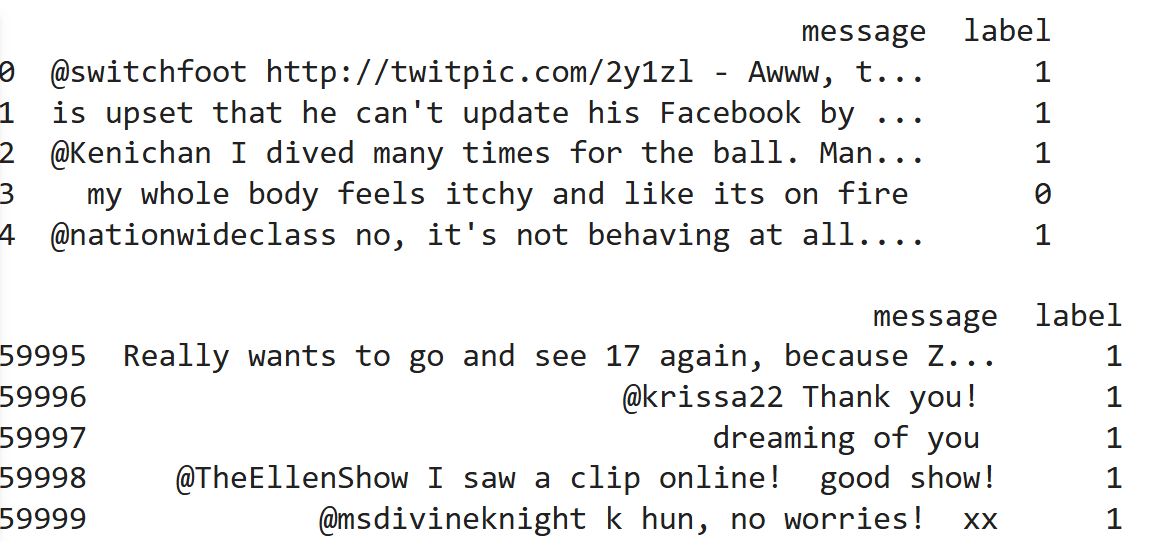
**2. Méthodologie**

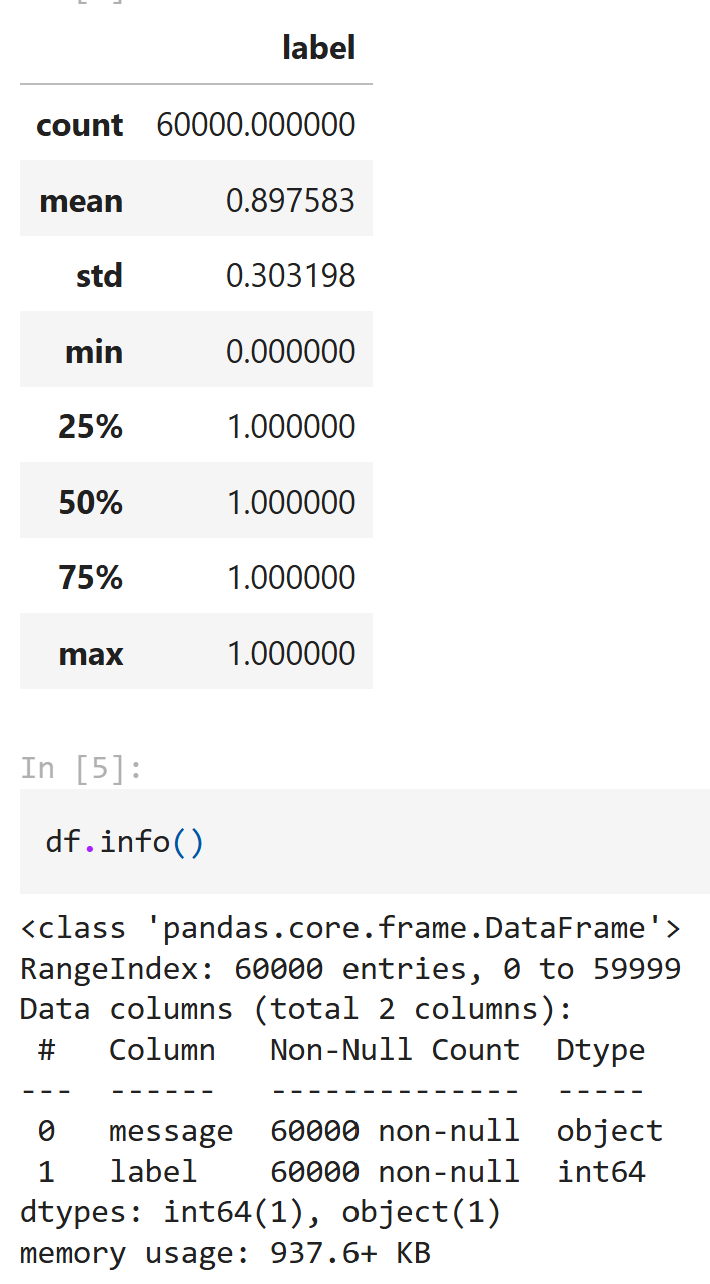
Notre méthodologie consiste à utiliser dans un premier temps le modèle pré entrainé RoBERTa (Robustly Optimized BERT Approach) qui est une version améliorée du modèle BERT (Bidirectional Encoder Representations from Transformers) développé par Google. Ce modèle a été mis en place pour améliorer les performances du modèle BERT en modifiant certains hyperparamètres clés et en utilisant des techniques d'entraînement plus robustes. Dans un second temp nous avons utilisé également les modèles classiques de régression logistique, d’arbres de décision et de KNN pour analyser les tweets.

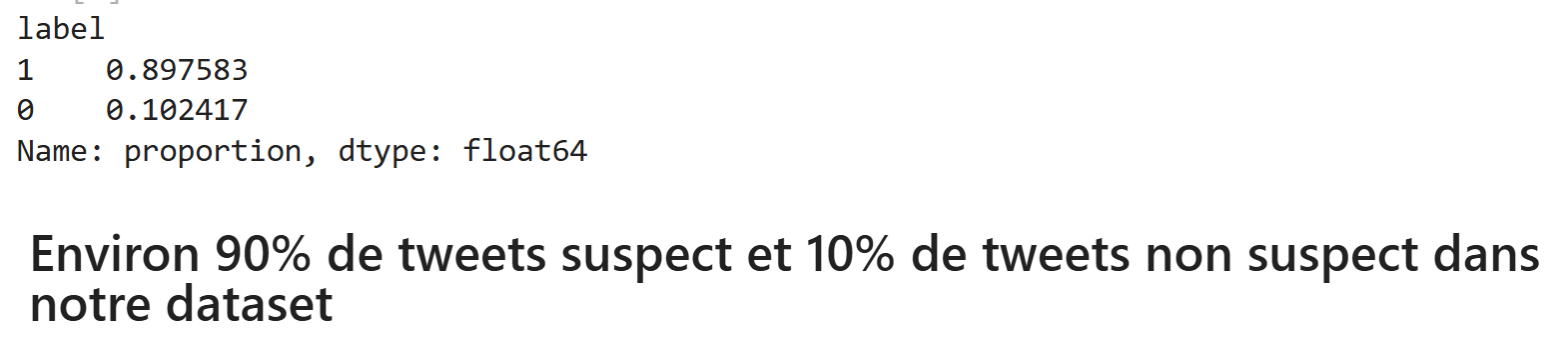
**2.1 Exploration et prétraitement des données**

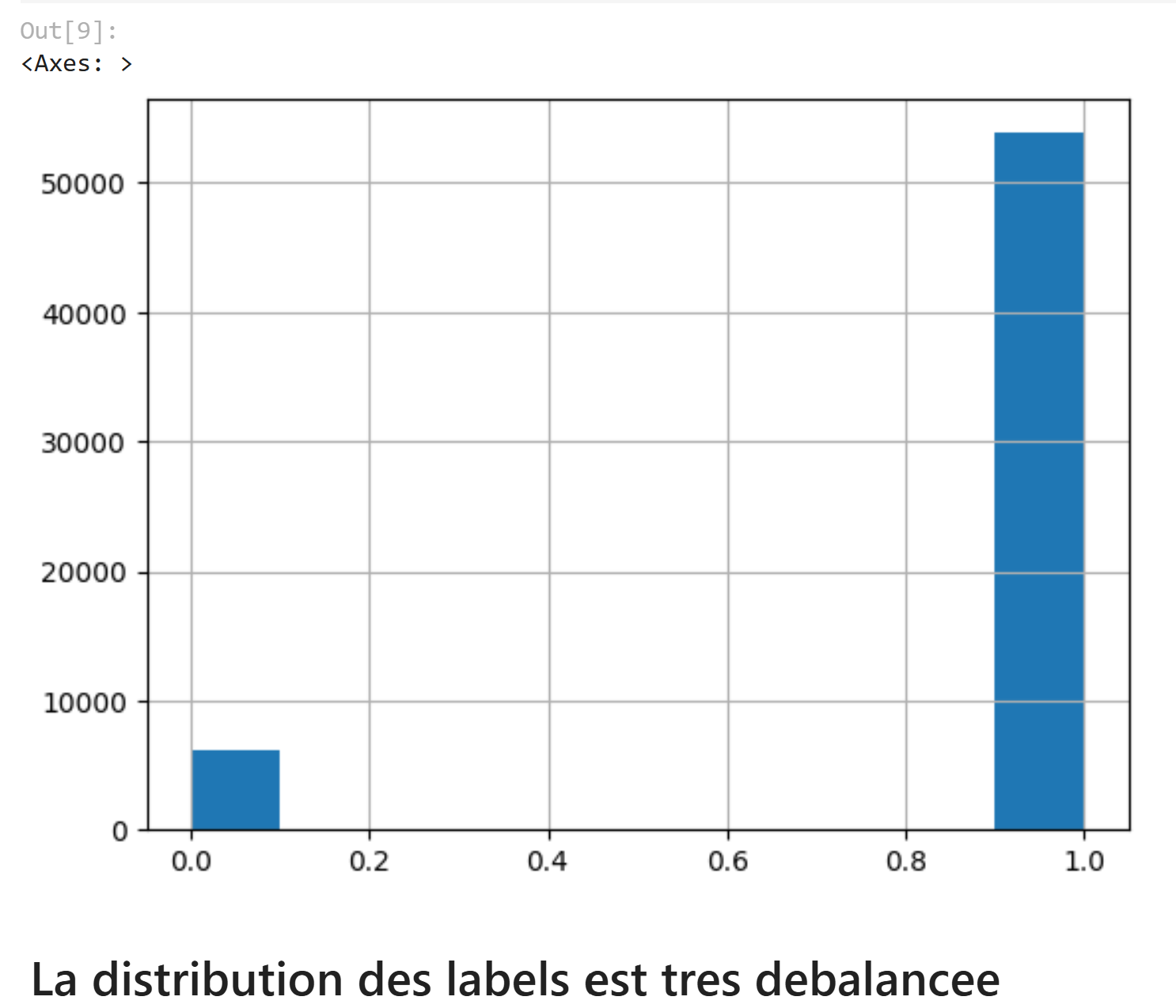
Nous avons utilisé une base de données de 60.000 tweets labellisés en tweet suspects et non suspect (0 ou 1 comme label). L’analyse exploratoire du dataset nous donne ce qui suit :











La distribution des labels est très débalancée

Lorsque la distribution des données est déséquilibrée, les mesures de performance traditionnelles comme l’exactitude (Accuracy) peuvent être trompeuses. Voici quelques métriques plus appropriées pour évaluer les modèles dans ce contexte que nous allons utiliser pour notre cas :

Matrice de confusion:

Fournit une vue d’ensemble des prédictions correctes et incorrectes, en distinguant les vrais positifs (TP), vrais négatifs (TN), faux positifs (FP) et faux négatifs (FN).

Précision (Precision):

Preˊcision=TP+FPTP​

Indique la proportion de prédictions positives correctes parmi toutes les prédictions positives.

Rappel (Recall) ou Sensibilité:

Rappel=TP+FNTP​

Mesure la capacité du modèle à identifier correctement les instances positives.

Score F1:

Score F1=2×Preˊcision+RappelPreˊcision×Rappel​

Harmonie moyenne de la précision et du rappel, utile lorsque vous avez besoin d’un équilibre entre les deux.

Courbe ROC et AUC (Area Under the Curve):

La courbe ROC trace le taux de vrais positifs (TPR) contre le taux de faux positifs (FPR) à différents seuils de classification.

L’AUC mesure la capacité du modèle à distinguer entre les classes.

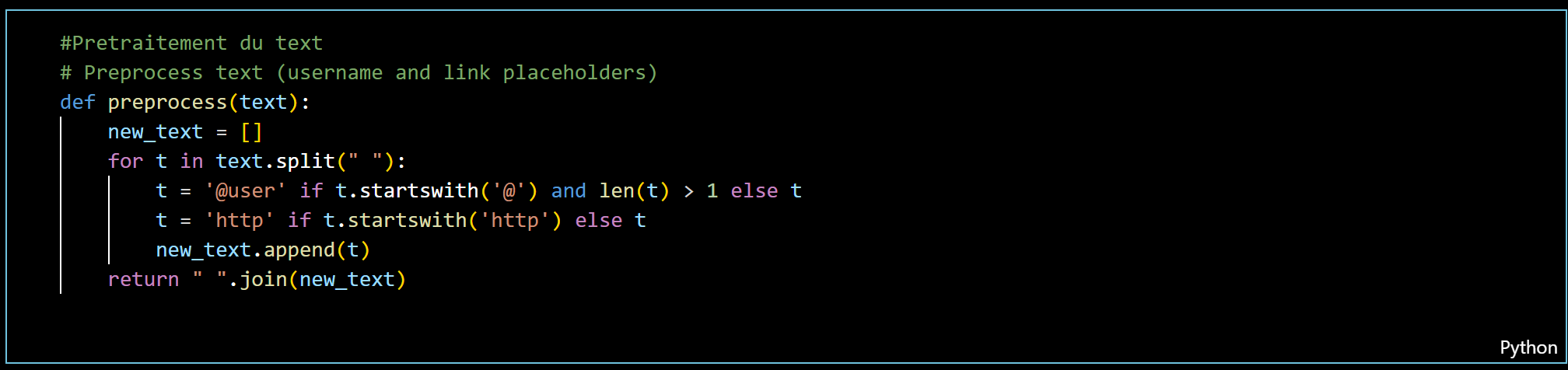
Métrique de G-Mean:

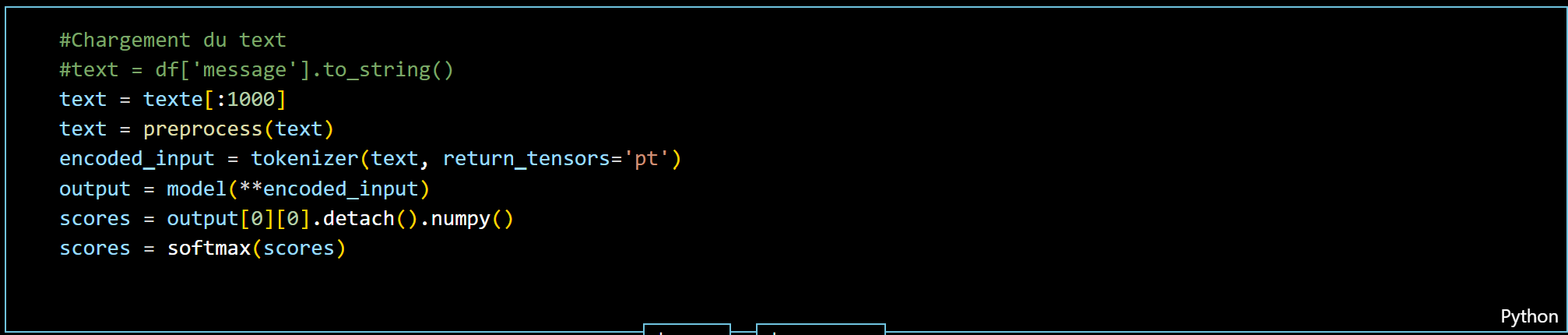
G-Mean=Sensibiliteˊ×Speˊcificiteˊ​

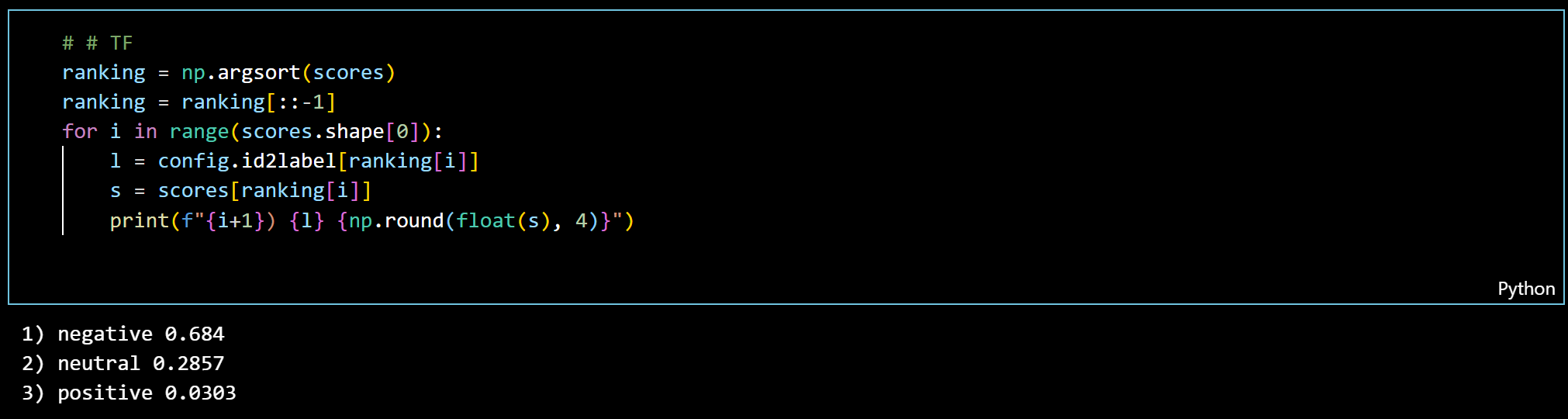
Évalue l’équilibre entre la sensibilité et la spécificité.

Ces métriques fournissent une image plus claire des performances du modèle dans toutes les classes, surtout lorsque les classes sont déséquilibrées.

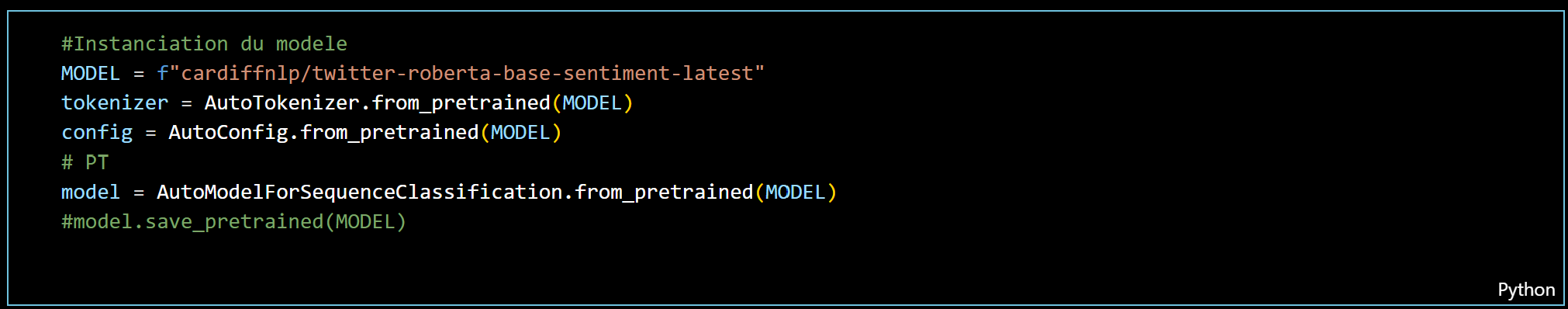
**2.2 Embeddings et représentation des données**



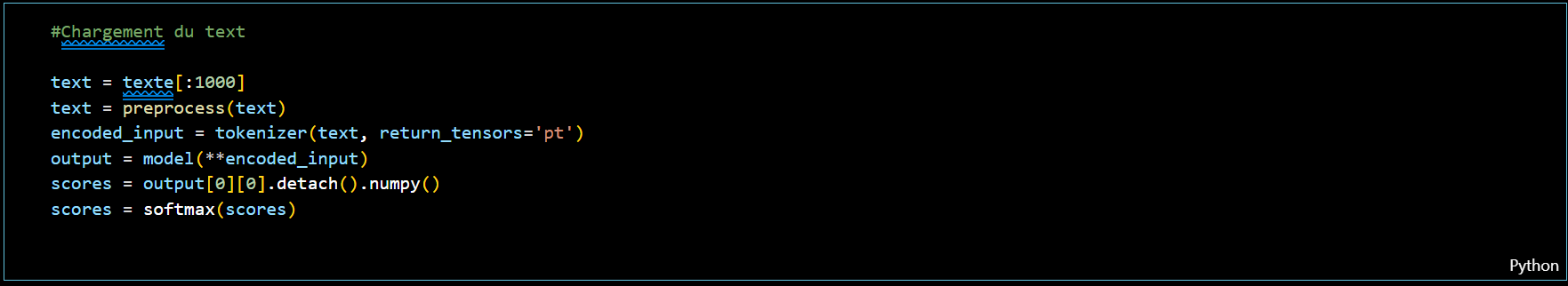




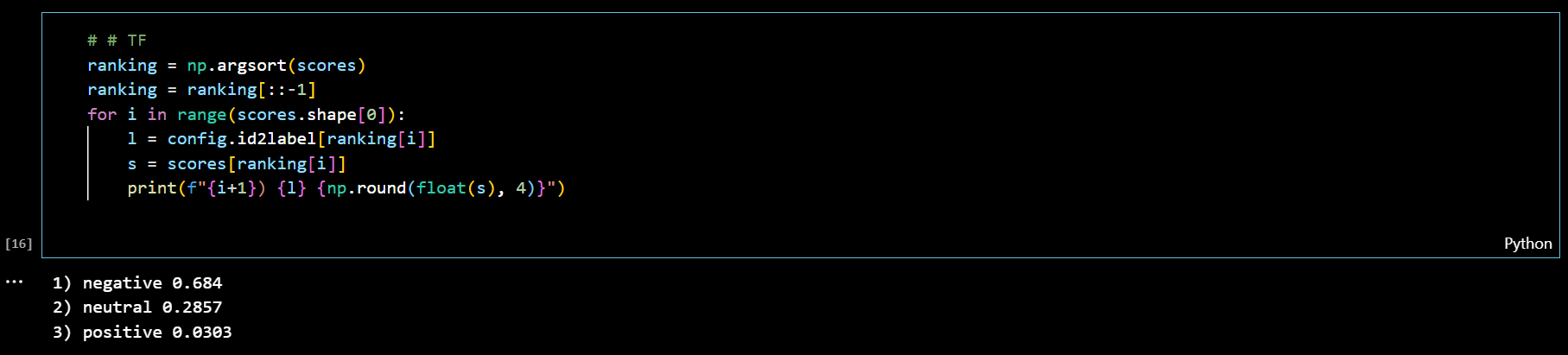
**2.3 Construction du modèle**

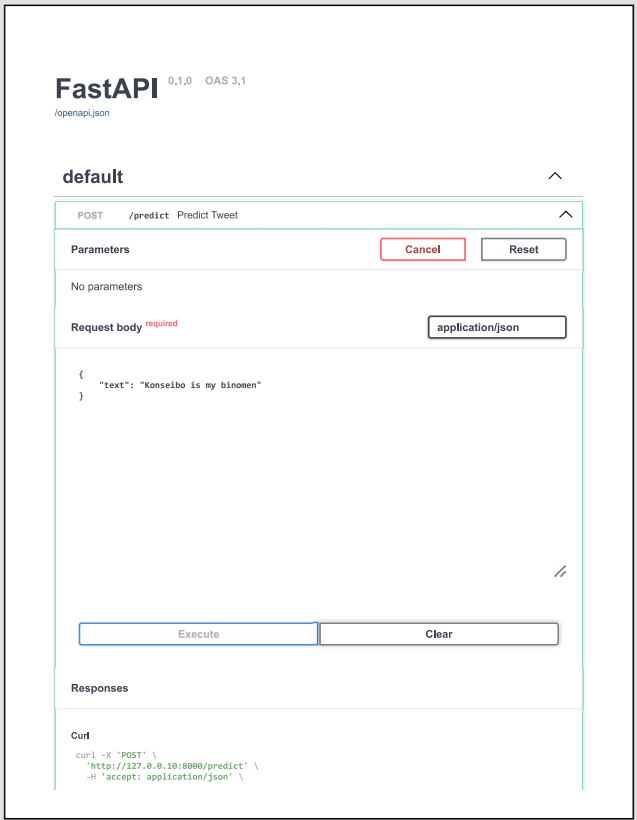


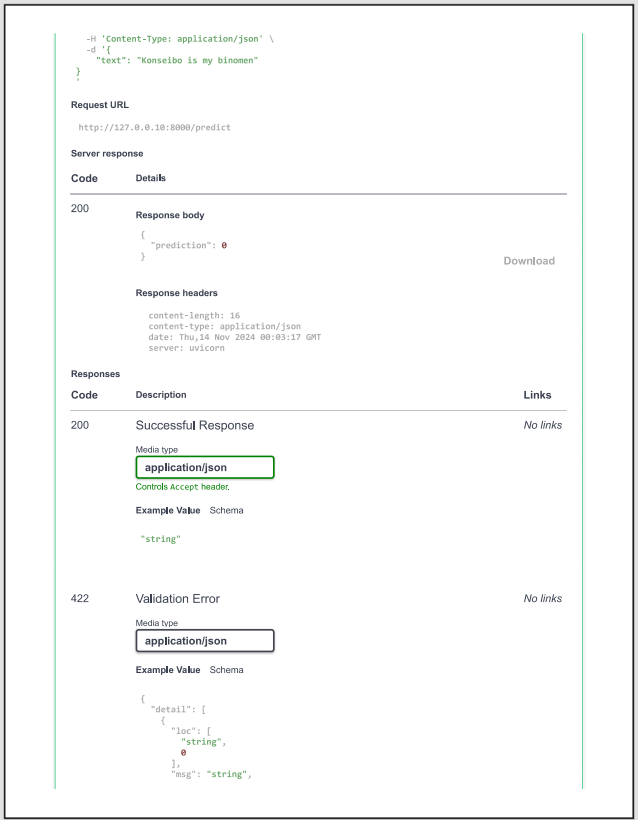
**2.4 Entrainement et validation**

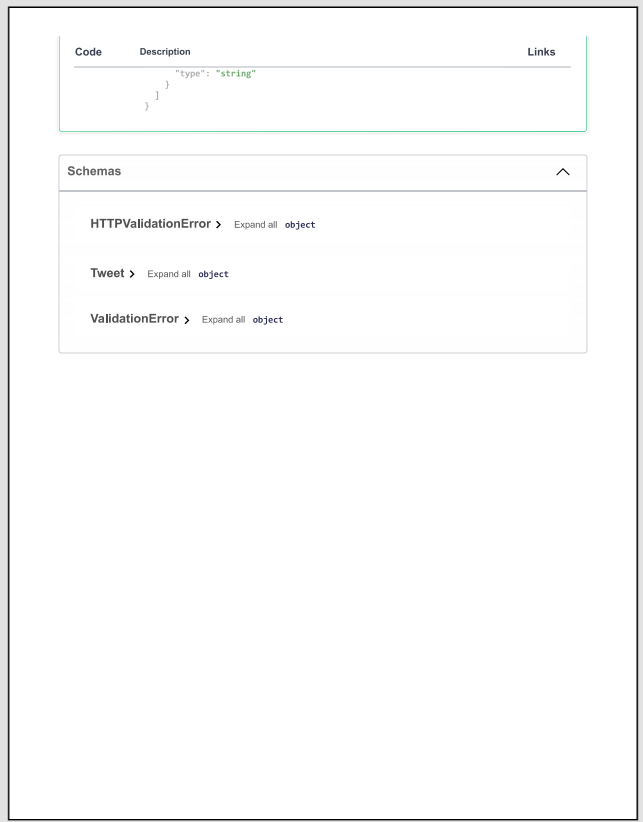


**2.4 Évaluation et ajustement**

**2.5 Déploiement**

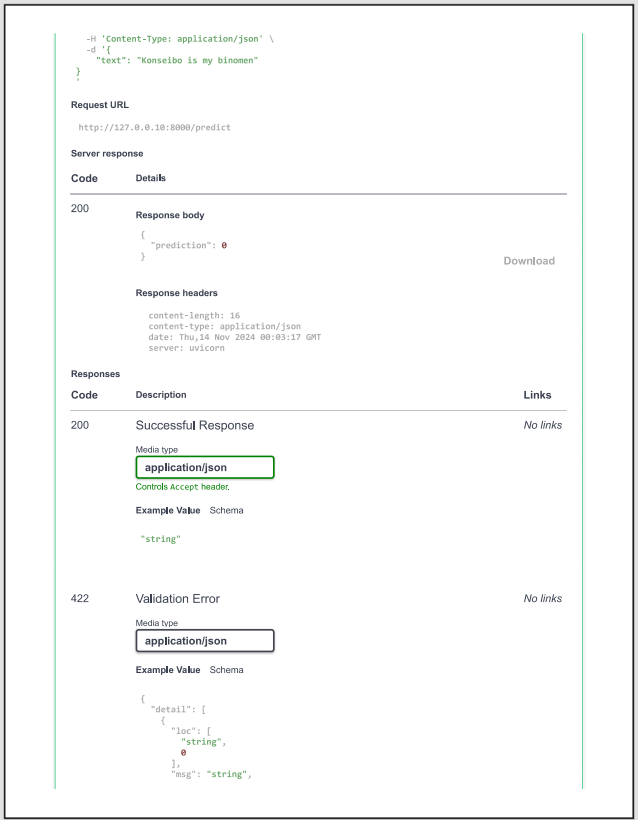
****

****

****

**3. Résultats**

Le résultat donne la prédiction 0 du texte saisi sous forme de tweet c’est-à-dire que le tweet est non suspect.

****

**4. Discussion sur les limites et améliorations**

Notre modèle a été entraine sur une partie de nos données (1000 phrases) prise de façon aléatoire. Ce choix a été fait pour des raisons de contraintes de disponibilité de ressources machine. Pour lever cette limitation il nous faut disposer de ressources machines suffisantes pour entrainer le modèle sur l’ensemble des données d’entrainement ce qui nous donnera de meilleures prédictions.

**5. Conclusion**

Notre étude qui a porté sur la classification des tweets en tweet suspect et non suspect a été fait selon deux approches ;

* Une première approche basée sur le modèle preentraine **twitter-roberta-base-sentiment-latest** que nous avons customisé.
* Une seconde approche basée sur la régression logistique, les arbres de décision et le KNN pour classifier les tweets.

Nous avons utilisé un échantillon aléatoire de la base de données originale compte tenu de la limitation des ressources machines dont nous possédons. Malgré ces limitation nos modèles ont obtenu une bonne capacitée de prédiction.

Les sont donc perfectible pour la suite de notre étude.