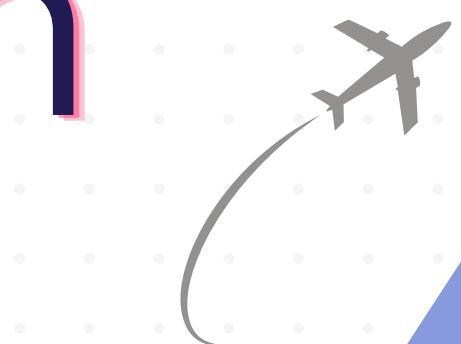


Project NLP:

Edinburgh Airport



Moad HADMEN
Angela GRIMAUD
Silya MOUSSOUS
Anastasiia SAVELKOVA



UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Sommaire

1. Pre-process des données

2. Analyse de sentiments : Vader et RoBERTa

3. Prediction de la satisfaction

4. Prediction des notes

5. Topic Modeling

6. Conclusion



1. Preprocess des données

customer_responses_2025.xlsx

Renaming and mapping columns

Function : cleaning text

Function : lemming

Function : token

DataFrame : df_clean

```
columns = ['How likely are you to recommend Edinburgh Airport to a friend or colleague?',
           'Overall, were you satisfied with your most recent experience at Edinburgh Airport?',
           'Are there any premium services that you would like to see introduced at Edinburgh?',
           'Do you have any other feedback or suggestions?']
```

```
column_mapping = {columns[0] : 'Rates',
                  columns[1] : 'Satisfaction',
                  columns[2] : 'Services_suggestions',
                  columns[3] : 'Feedbacks'}
```

```
satisfaction_mapping = {'Yes' : 1, 'No':0}
```

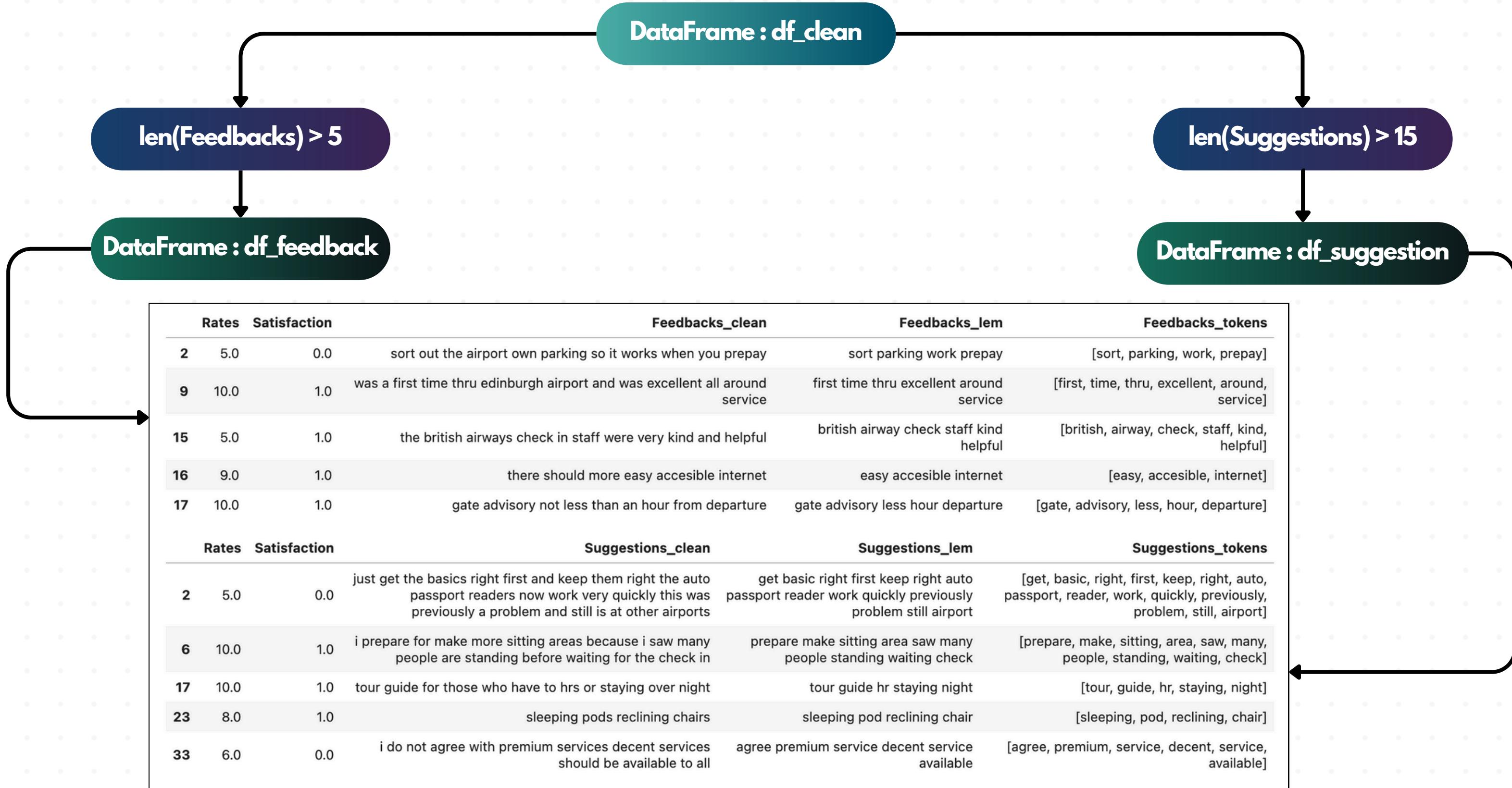
	Rates	Satisfaction	Feedbacks_clean	Feedbacks_lem	Feedbacks_tokens	Suggestions_clean	Suggestions_lem	Suggestions_tokens
0	9.0	1.0		no	[]	no		[]
1	10.0	1.0		no	[]	no		[]
2	5.0	0.0	sort out the airport own parking so it works when you prepay	sort parking work prepay	[sort, parking, work, prepay]	just get the basics right first and keep them right the auto passport readers now work very quickly this was previously a problem and still is at other airports	get basic right first keep right auto passport reader work quickly previously problem still airport	[get, basic, right, first, keep, right, auto, passport, reader, work, quickly, previously, problem, still, airport]
3	9.0	1.0		no	[]	no		[]
4	10.0	1.0		no	[]	no		[]



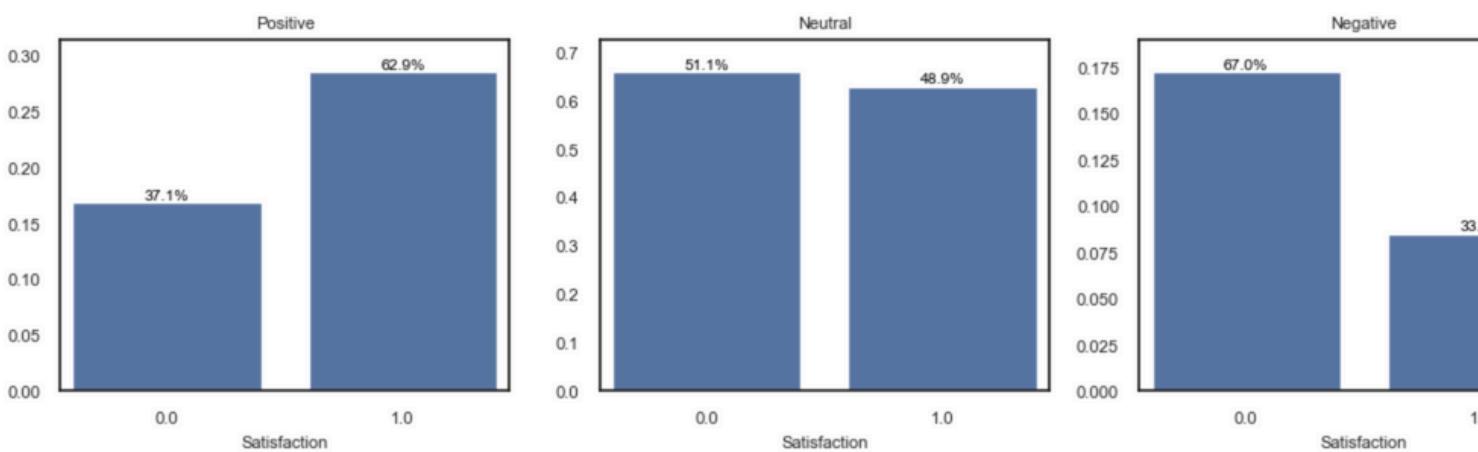
UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

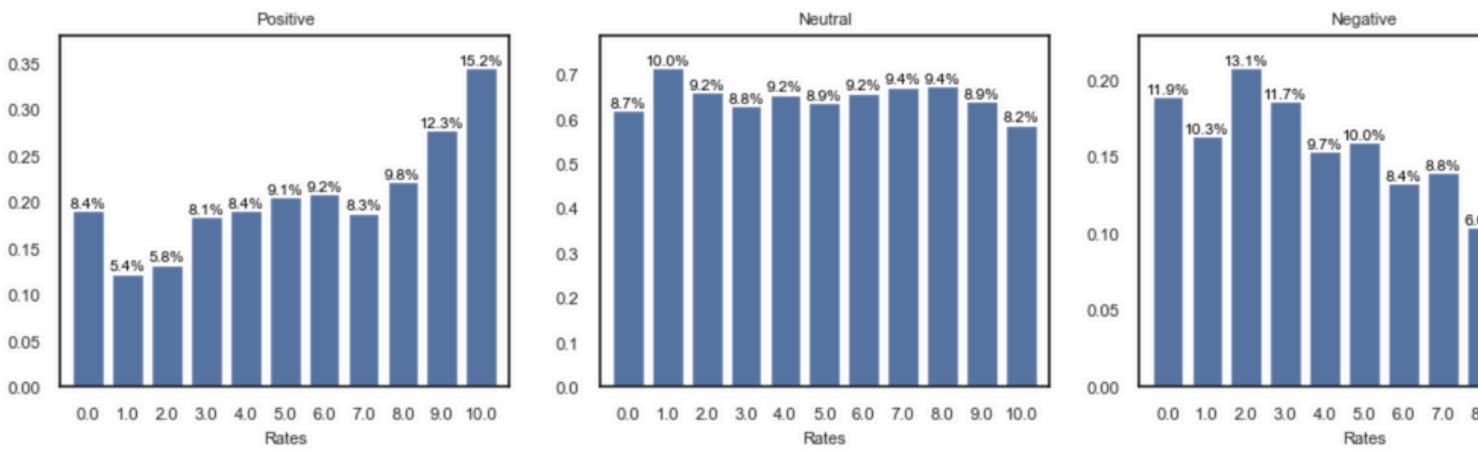
1. Preprocess des données



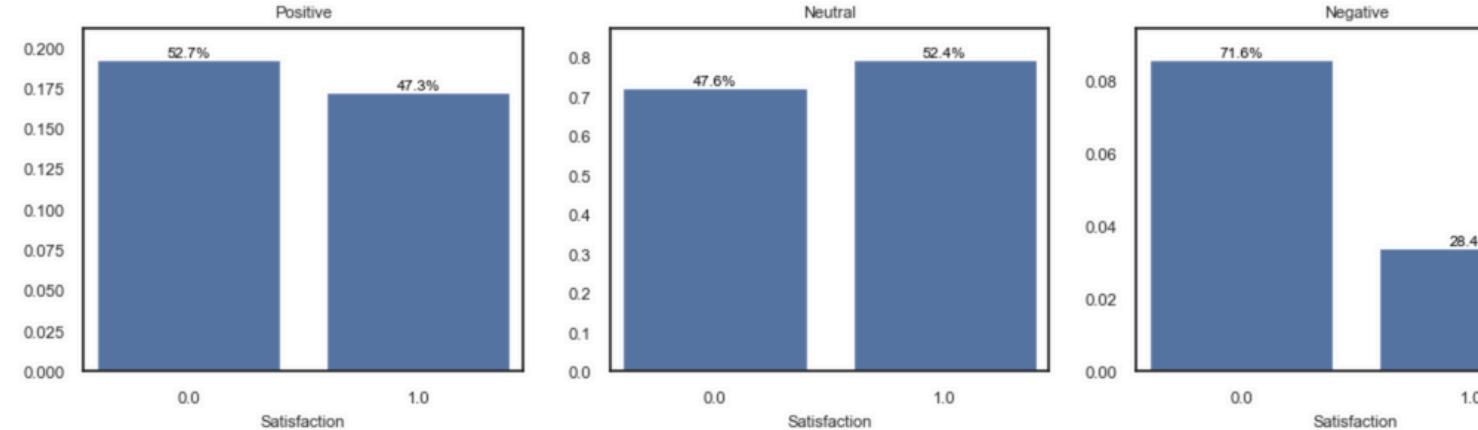
Sentiment Analysis - Target variable: 'Satisfaction' (Dataset: Feedbacks)



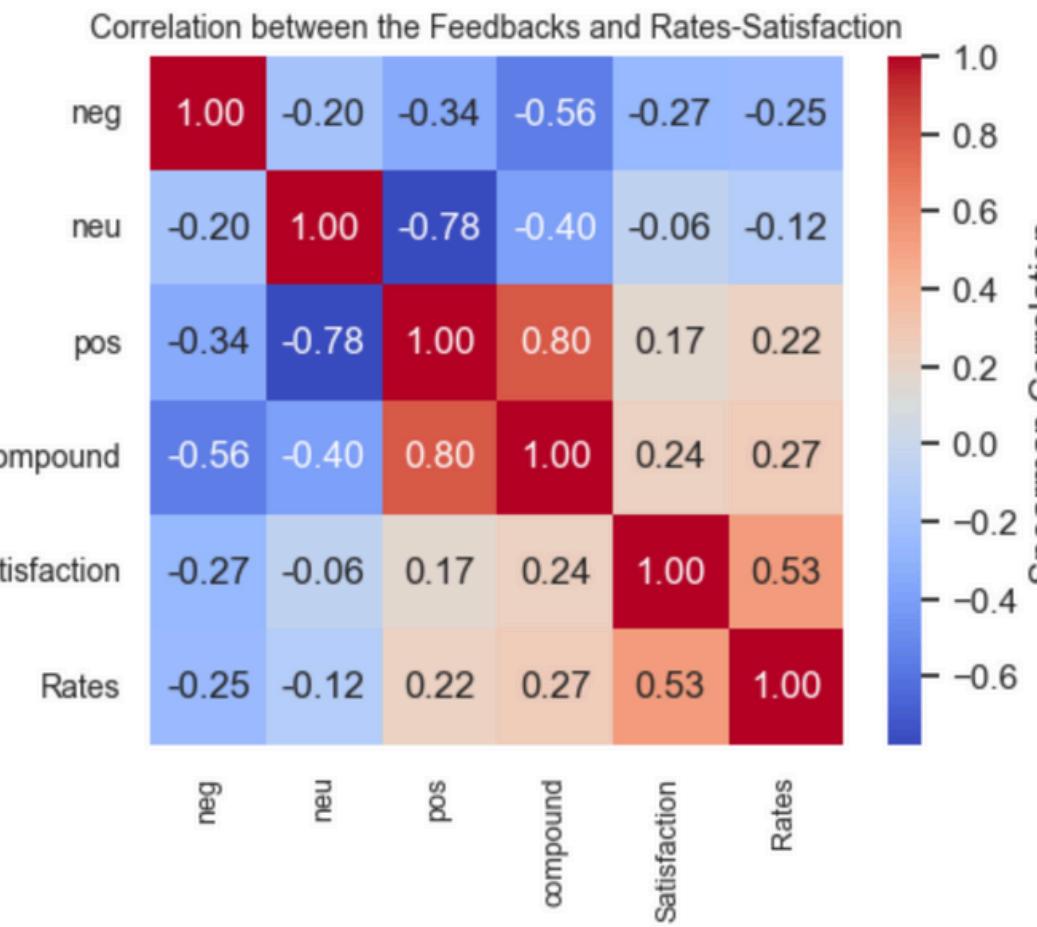
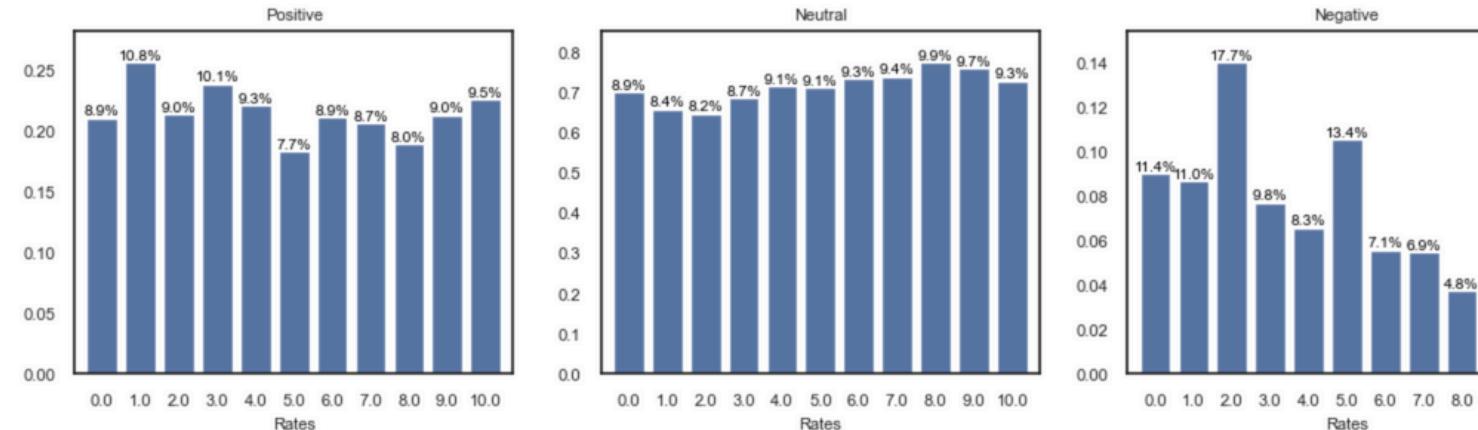
Sentiment Analysis - Target variable: 'Rates' (Dataset: Feedbacks)



Sentiment Analysis - Target variable: 'Satisfaction' (Dataset: Suggestions)



Sentiment Analysis - Target variable: 'Rates' (Dataset: Suggestions)



2. Analyse de sentiments : Vader

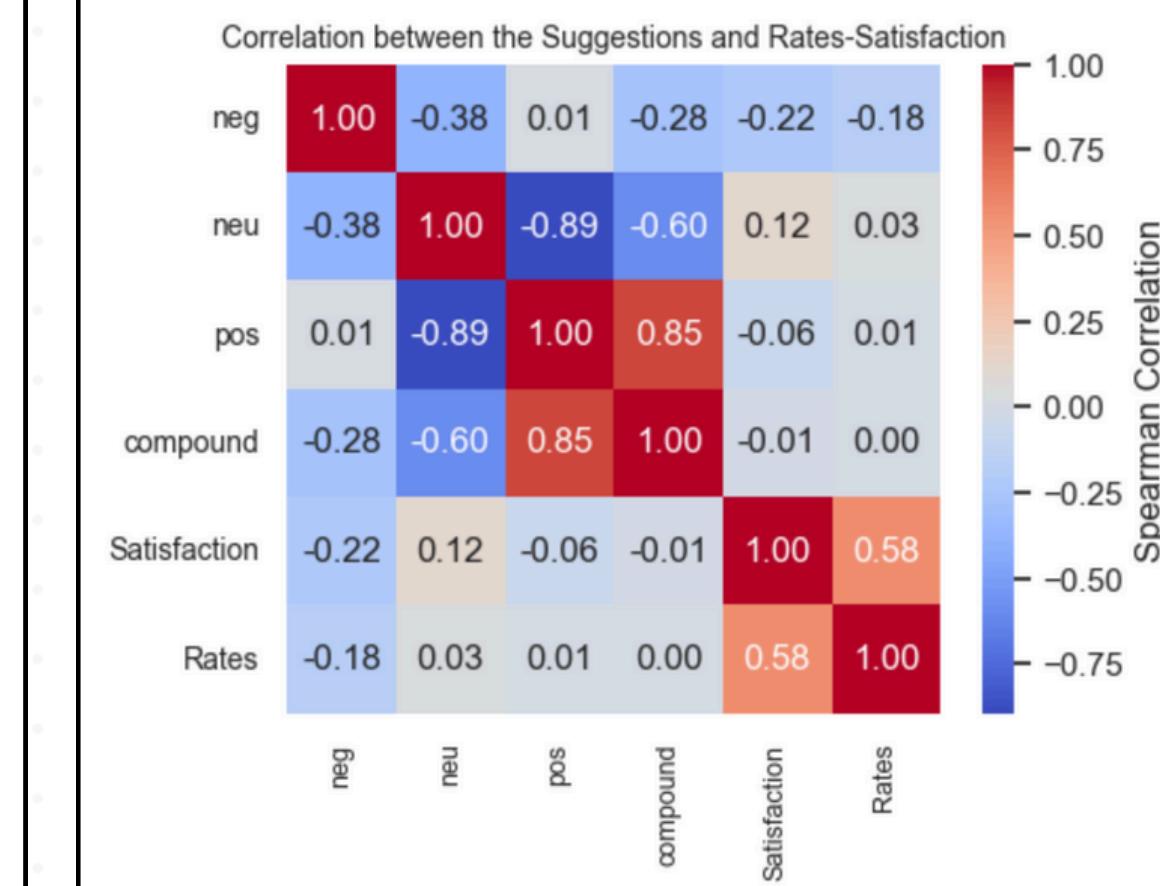
DataFrame : df_feedback

DataFrame : df_suggestion

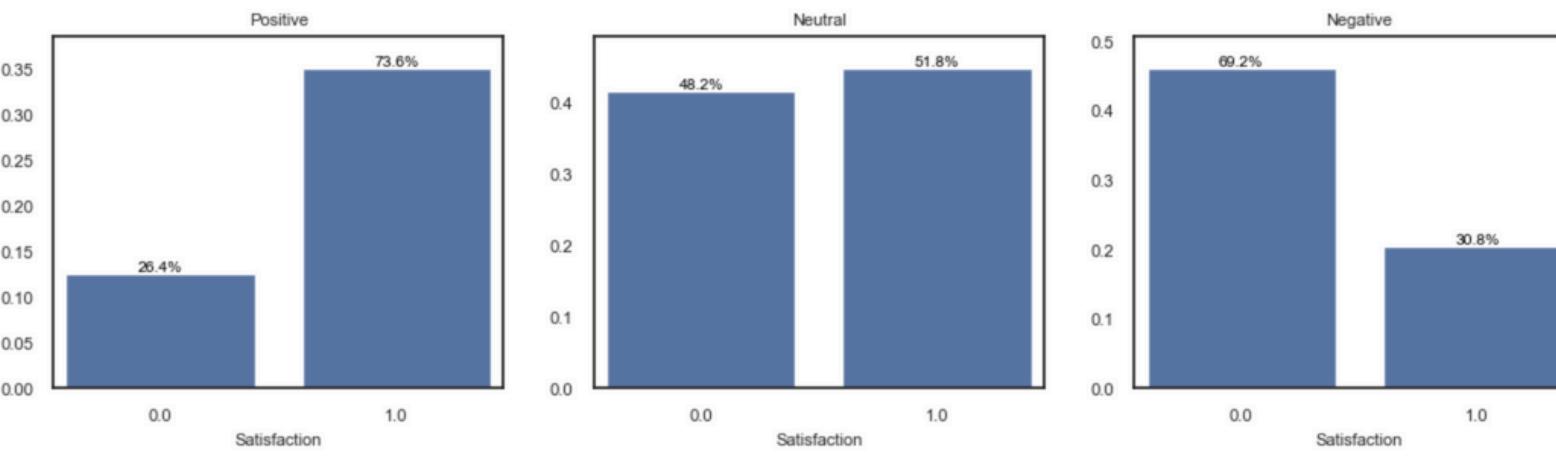
Function : Scoring_vader

Function : plot_sentiment

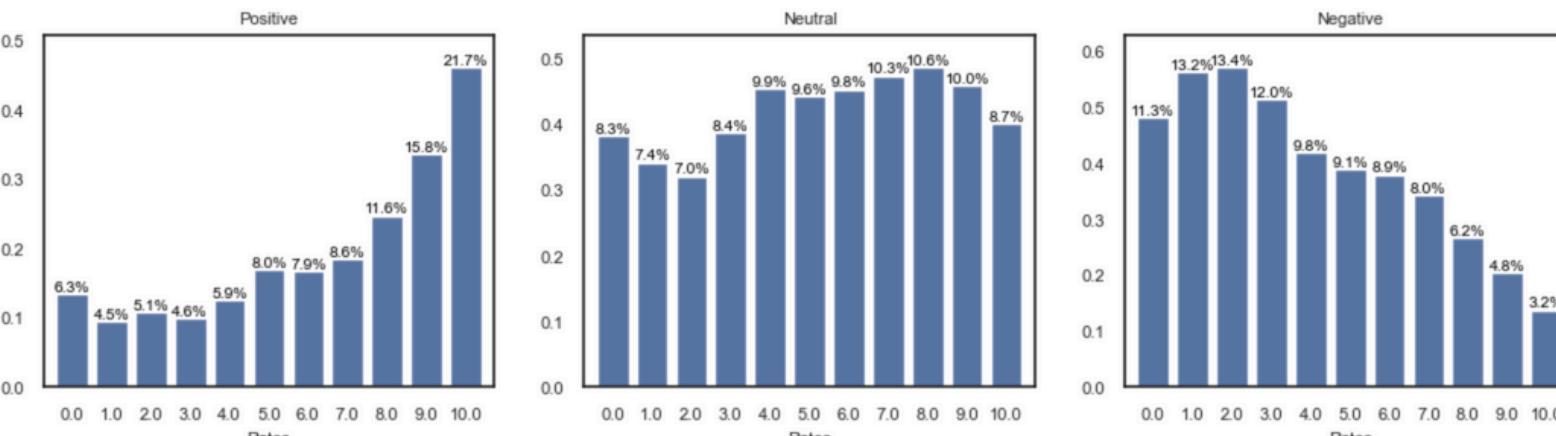
Function : heatmap



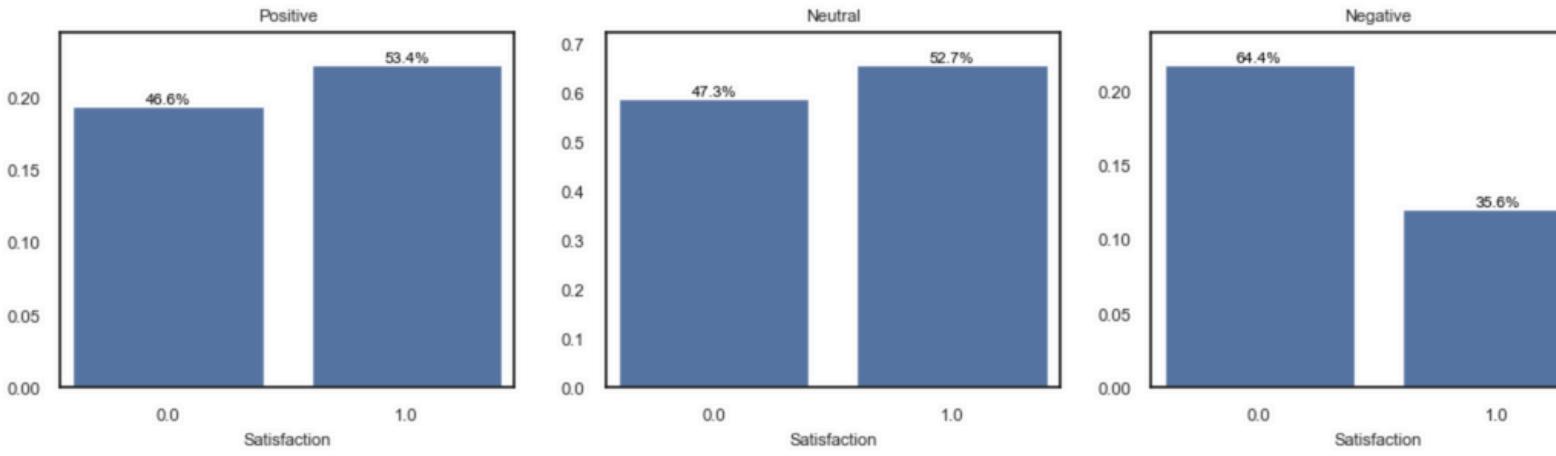
Sentiment Analysis - Target variable: 'Satisfaction' (Dataset: Feedbacks_RoBERTa)



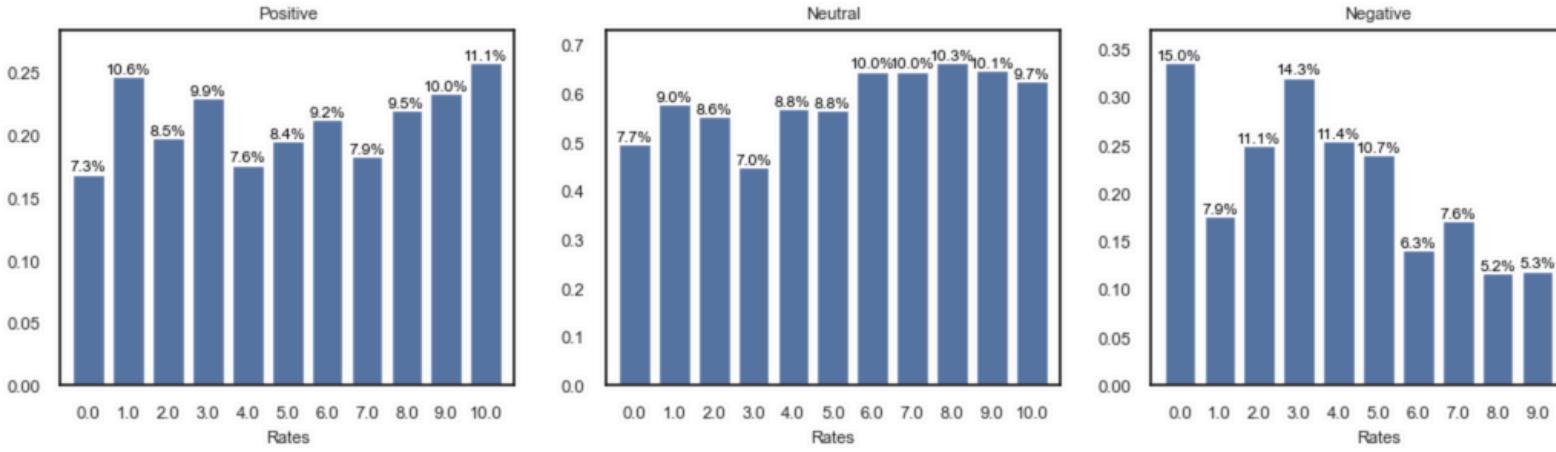
Sentiment Analysis - Target variable: 'Rates' (Dataset: Feedbacks_RoBERTa)



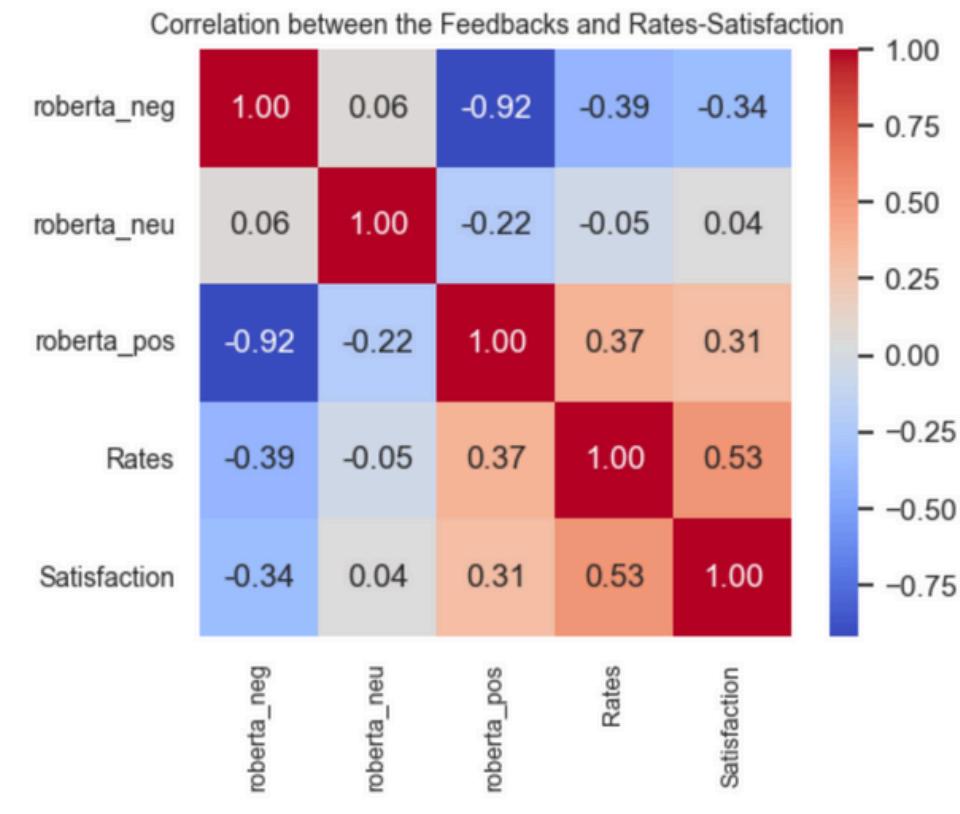
Sentiment Analysis - Target variable: 'Satisfaction' (Dataset: Suggestions_RoBERTa)



Sentiment Analysis - Target variable: 'Rates' (Dataset: Suggestions_RoBERTa)



2. Analyse de sentiments : RoBERTa avec le modèle "cardiffnlp/twitter-roberta-base-sentiment"



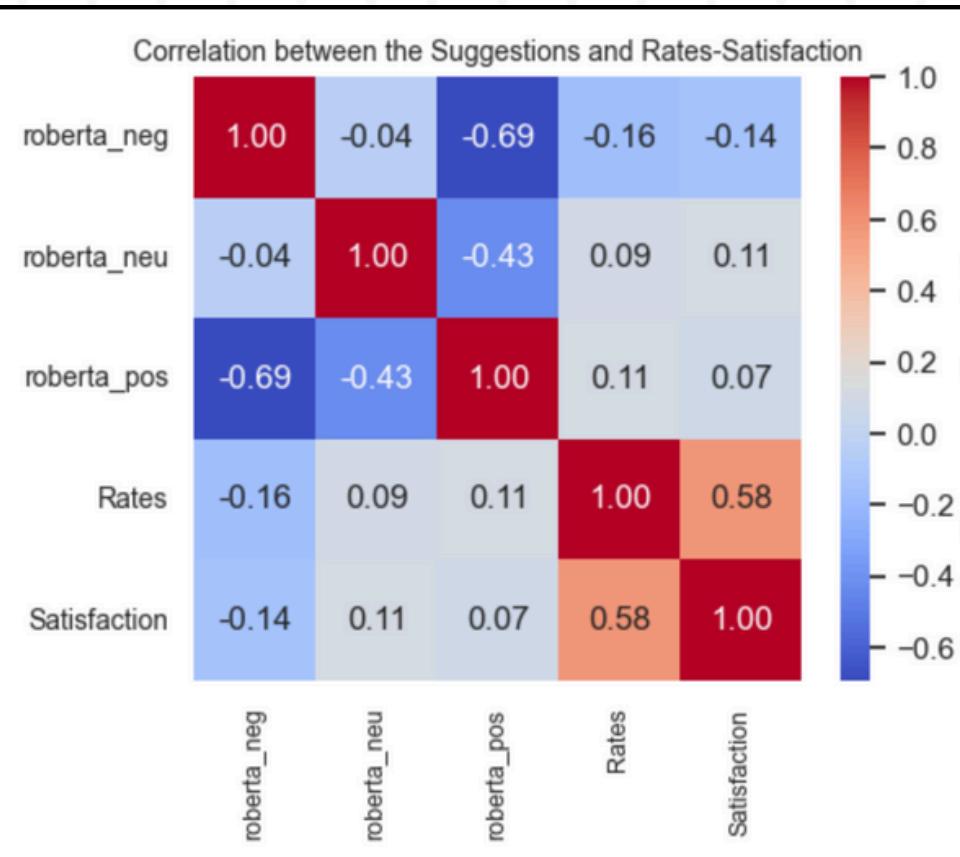
DataFrame : df_feedback

DataFrame : df_suggestion

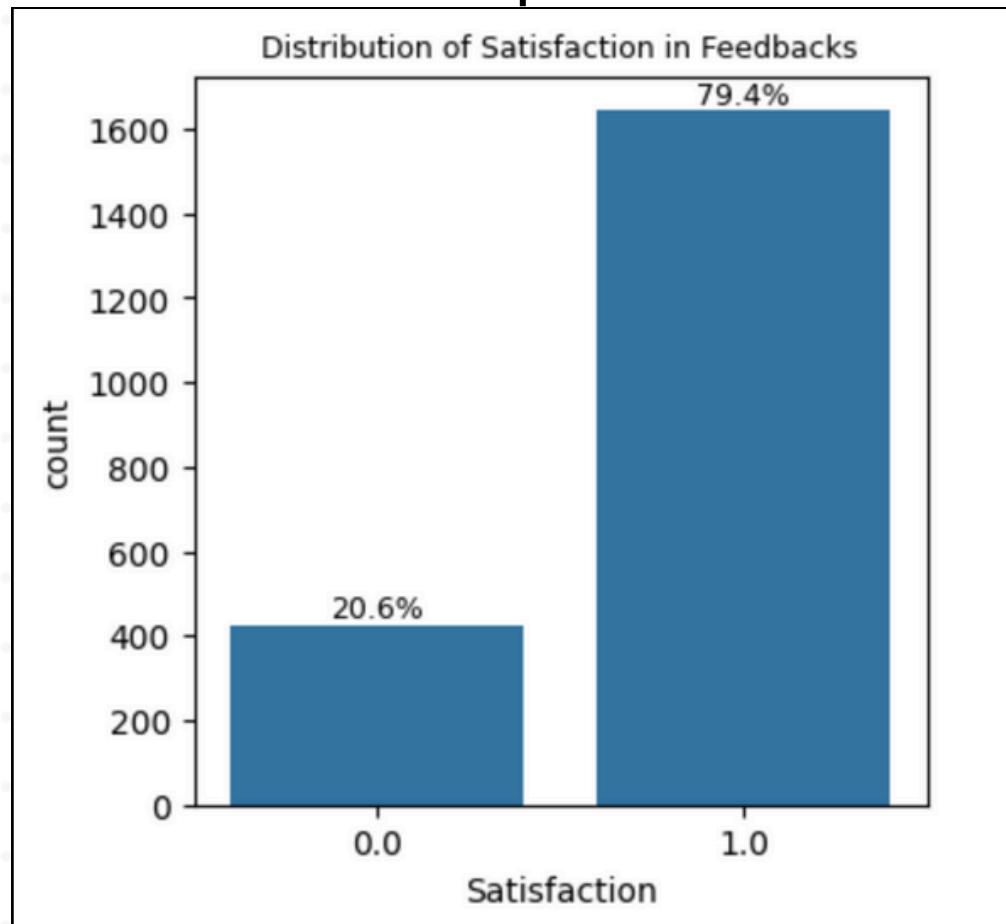
Function : Scoring_roberta

Function : plot_sentiment

Function : heatmap

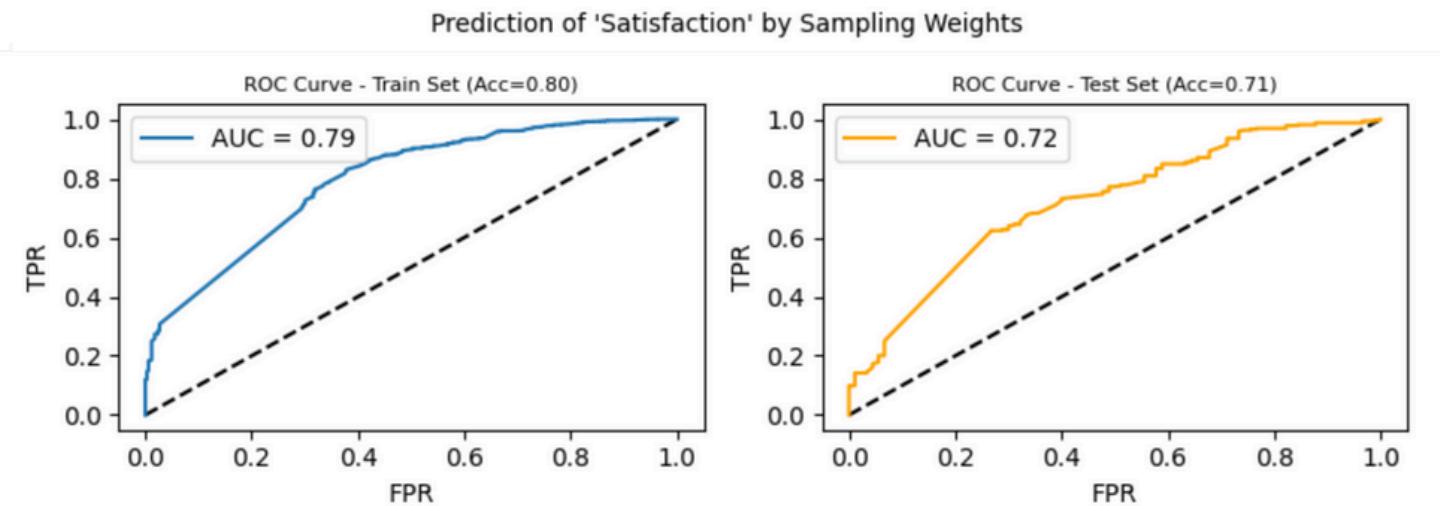


3.1. Prediction de Satisfaction à partir de Feedbacks

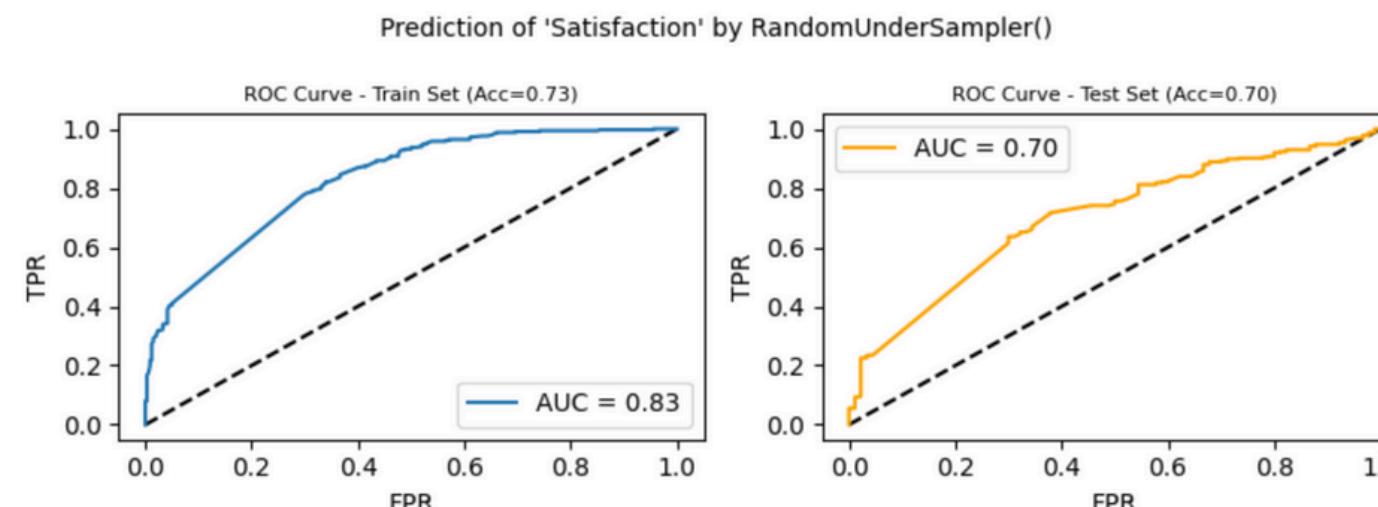


TF-IDF

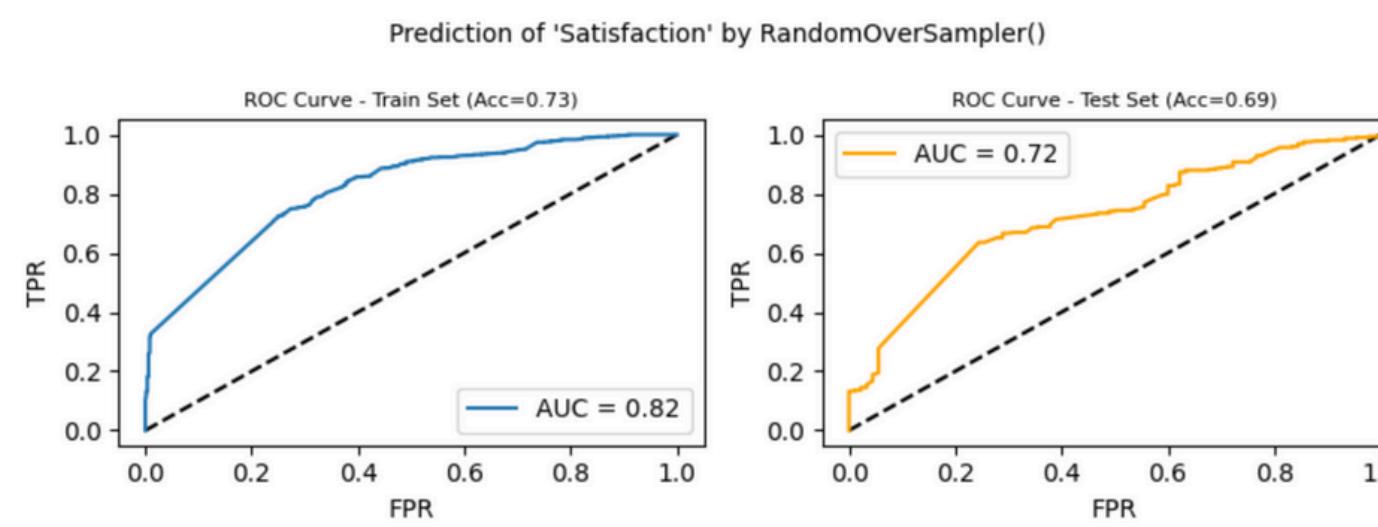
XGBoost



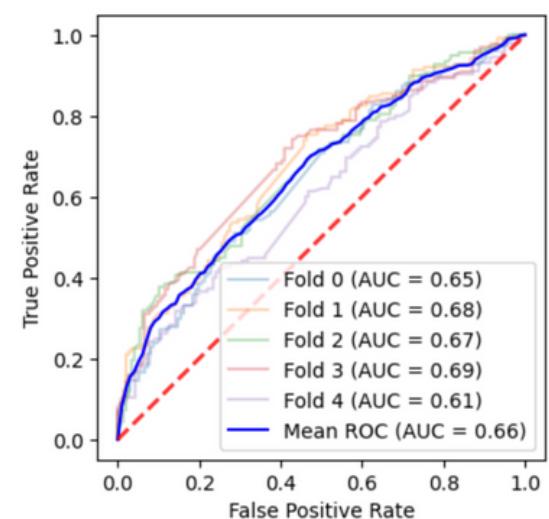
Methode : Sampling
Weights



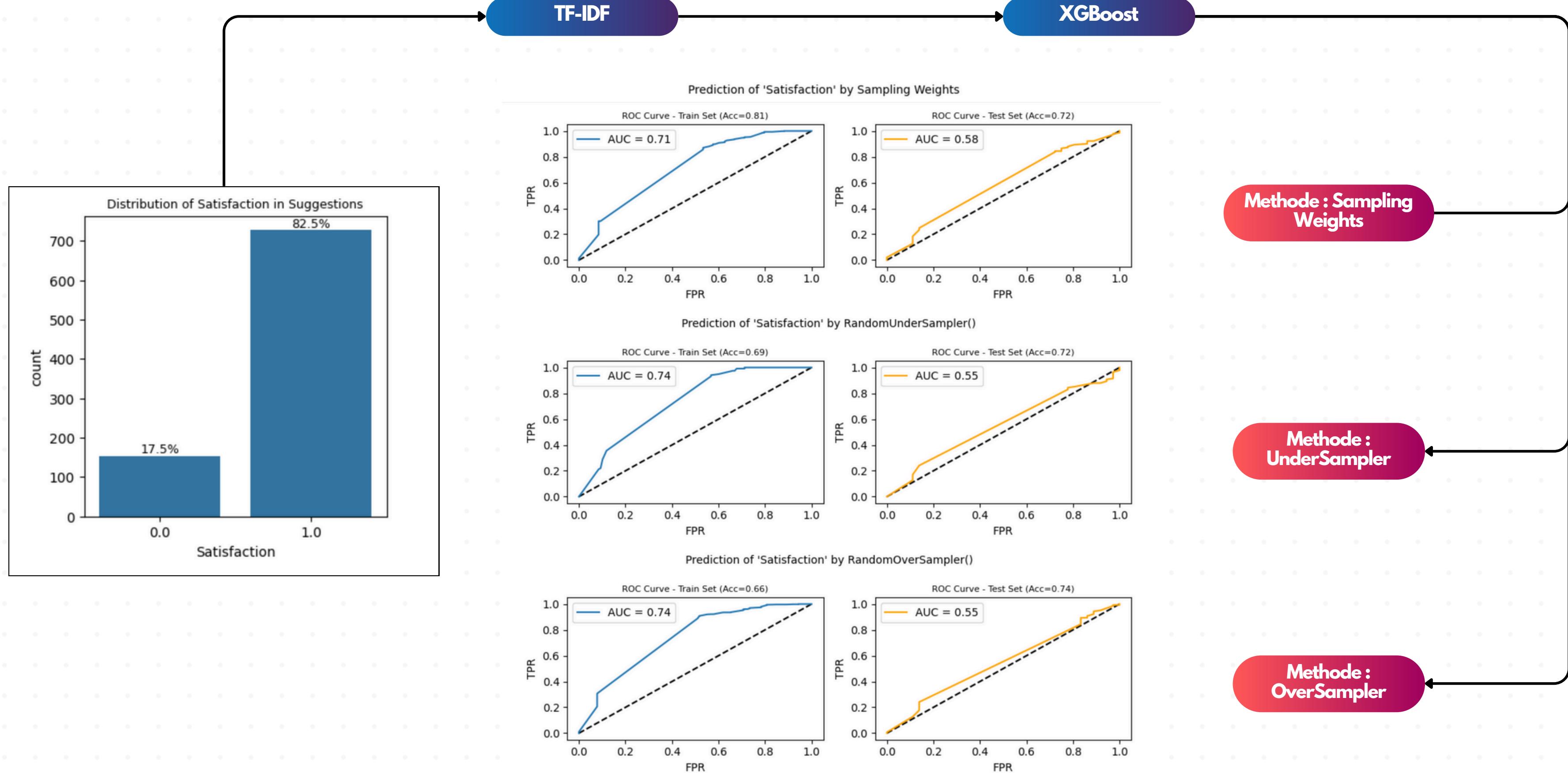
Methode :
UnderSampler



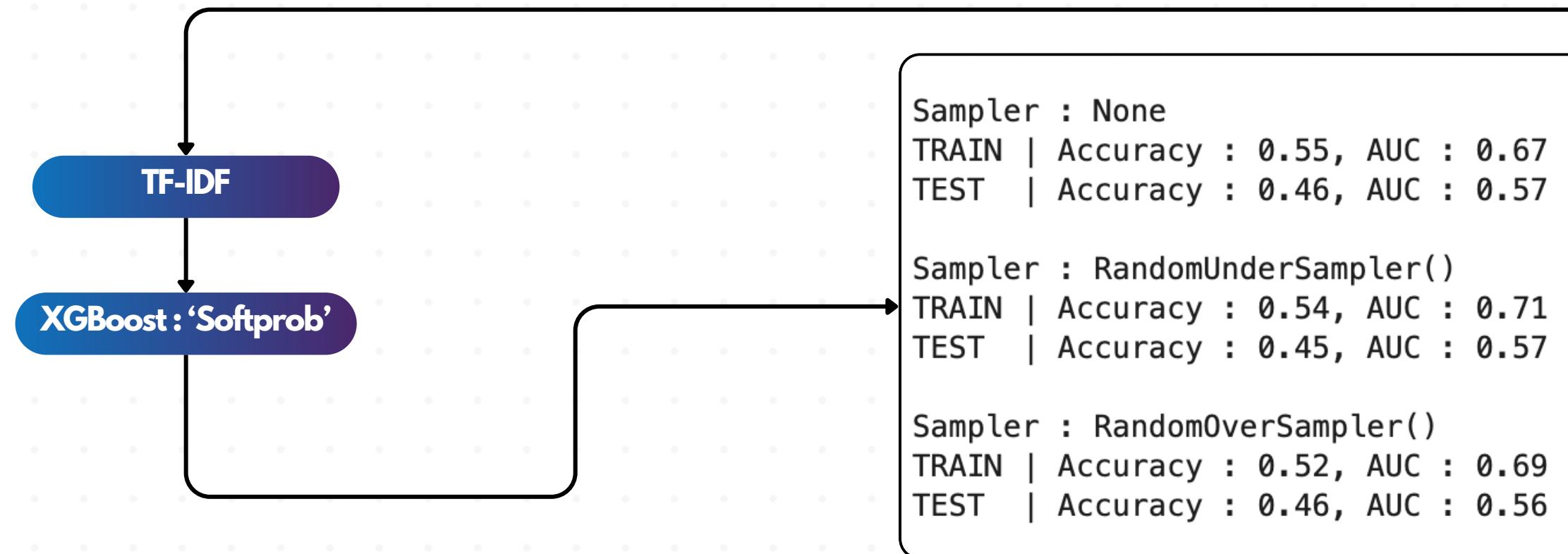
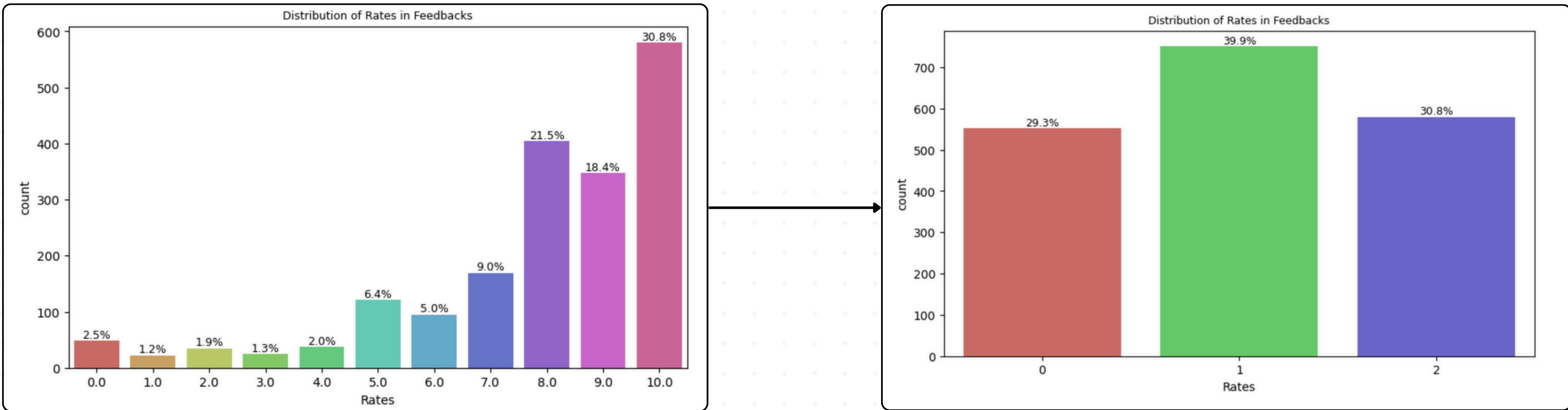
Methode :
OverSampler



3.2. Prediction de Satisfaction à partir de Suggestions



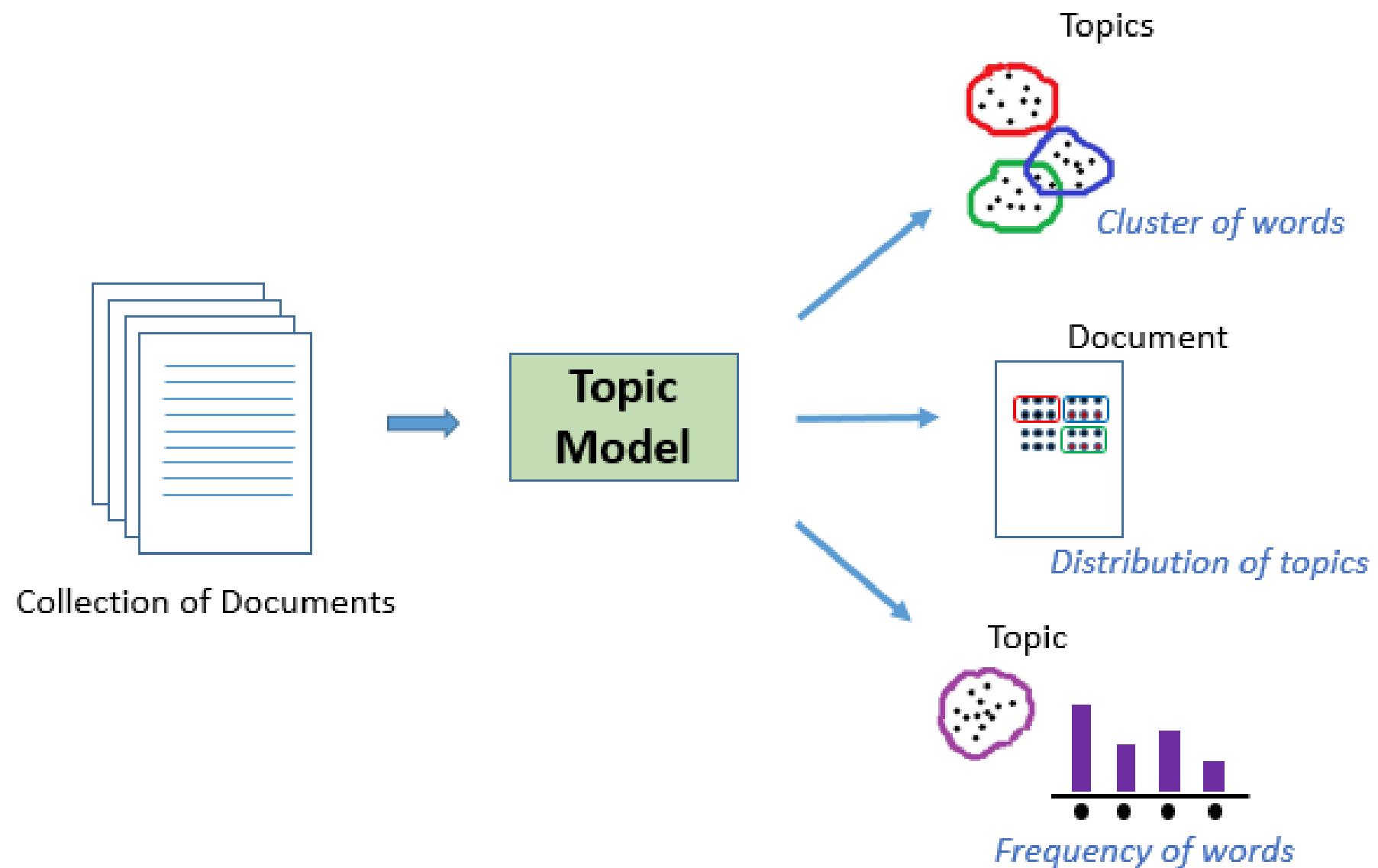
4. Prediction des notes à partir de Feedbacks



5. Topic Modeling

Objectifs :

- Regrouper les feedbacks par thèmes
- Déterminer quels thèmes se retrouvent dans les mauvaises expériences



Principe :

- Considérer uniquement les feedbacks associés à une note < 7
=> 383 observation
- Pour chaque feedback, on obtient son/ses bigrams
- Ne garder que les bigrams présents à au moins 5 reprises
=> 63 commentaires ont au moins un bigram associé
- LDA sur les bigrams obtenus afin de les regrouper par cluster

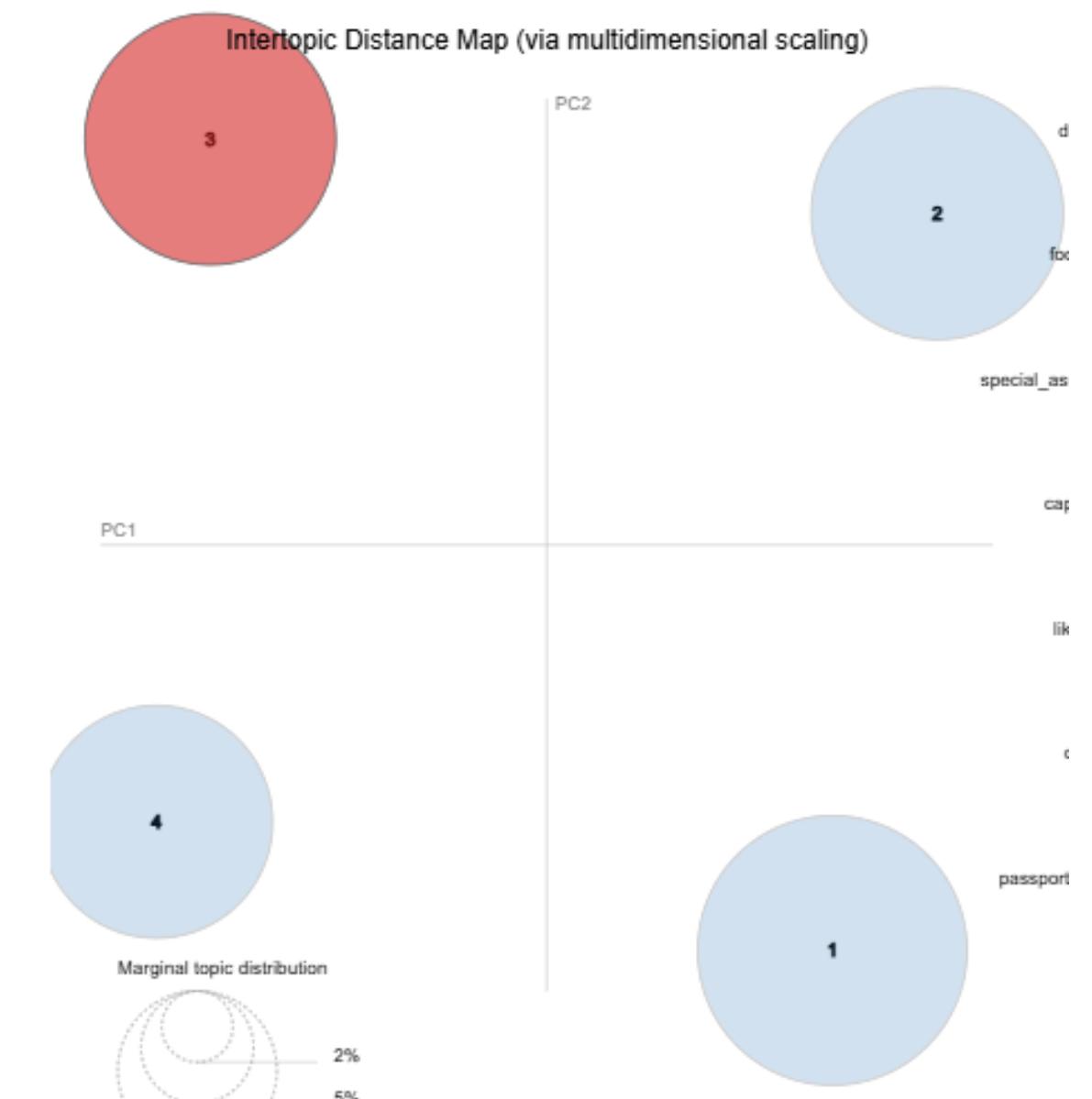
Feedbacks_clean	ngrams
passport control for arrivals is disorganised and lack coordination there should be a family lane	[passport_control]
keep the duty free open for last flights and other cafe and basic shops	[duty_free]

On trouve 4 clusters :

- food_drink & duty_free
- special_assitance (& capital_city)
- passport_control
- like_cattle & car_park

Ensuite, on reprend le dataframe et on associe à chaque feedback son cluster

On va pouvoir analyser les axes d'amélioration de l'aéroport par thématique.



Reproches récurrents des client de l'aéroport

Basés sur les feebacks associés à des mauvaises notes cluterisés par thème

Food/drink/duty free

- Choix de nourriture restreint
- Prix excessifs
- Ouvrir les duty free jusqu'aux derniers vols
- Proposer un autre chemin d'accès aux portes d'embarquement pour éviter de devoir traverser tous les duty free
- les clients se sentent agressés par toutes les pubs, effet centre commercial

Carpark

- manque de signalisation
- manque de places
- prix excessifs, même pour déposer des gens il faut payer
- parking assez éloigné de l'aéroport, il faudrait proposer un service de navettes
- système de reconnaissance de plaques pas au point pour les réservations prépayées
- les bornes devraient accepter le paiement par téléphone

Special assistance

- le service d'assistance manque de personnel
- pas assez d'ascenseurs/escalators

Passeport control

- file d'attente trop longue, mal organisée
- contrôle automatique pas au point
- créer une file pour les familles

Duty Free

- **Comportement du staff** : Une employée de la zone duty-free a été décrite comme impolie et peu professionnelle.
- **Parcours imposé** : Les passagers déplorent d'être obligés de traverser la zone duty-free, qualifiant cette expérience de "misérable" et de "honteuse".
- **Surcharge commerciale** : La quantité excessive de publicités et de boutiques donne l'impression d'un centre commercial plutôt que d'un aéroport.
- **Problèmes de travaux** : Les chantiers en cours créent des embouteillages, particulièrement après la zone duty-free.

Car Park, Car Parking

- **Couts extrêmement élevés** : Les passagers décrivent les frais de stationnement comme "absurdes", surtout pour les zones de dépôt minutes, les qualifiant de "pure extorsion".
- **Places limitées** : Le parking longue durée manque de places, obligeant les conducteurs à se garer sur les trottoirs ou à marcher longtemps. Aucune place garantie malgré une réservation préalable.
- **Systèmes de paiement obsolètes** : Les distributeurs de tickets n'acceptent pas les paiements mobiles, causant des retards et de la frustration.
- **Mauvaise organisation et accessibilité** : Une disposition confuse, de longues marches jusqu'aux terminaux et aucune option de dépôt gratuit.

Luggage Trolley (chariots de bagages)

- **Frais injustes** : Les passagers s'opposent fermement au paiement des chariots à bagages, qualifiant cette pratique de "pathétique", "humiliante" et "scandaleuse".
- **Image négative** : Ces frais nuisent à la réputation de l'Écosse auprès des visiteurs et ternissent l'image de l'aéroport d'Édimbourg en tant que hub d'une capitale.

Passeport control / Queues

- **Files d'attente longues et chaotiques** : Les portiques électroniques tombent souvent en panne, obligeant les passagers à passer par des contrôles manuels après des retards inutiles.
- **Aménagement mal conçu** : L'arrivée implique des escaliers, des couloirs étroits et des passages sales. Le personnel des frontières reste parfois inactif alors que les files d'attente débordent.
- **Gestion inefficace** : Aucune flexibilité pour contourner les portiques électroniques défectueux ; les contrôles manuels devraient être disponibles immédiatement. Les voyageurs d'affaires et les personnes âgées peinent à rester debout si longtemps.

Résumé des commentaires par profil

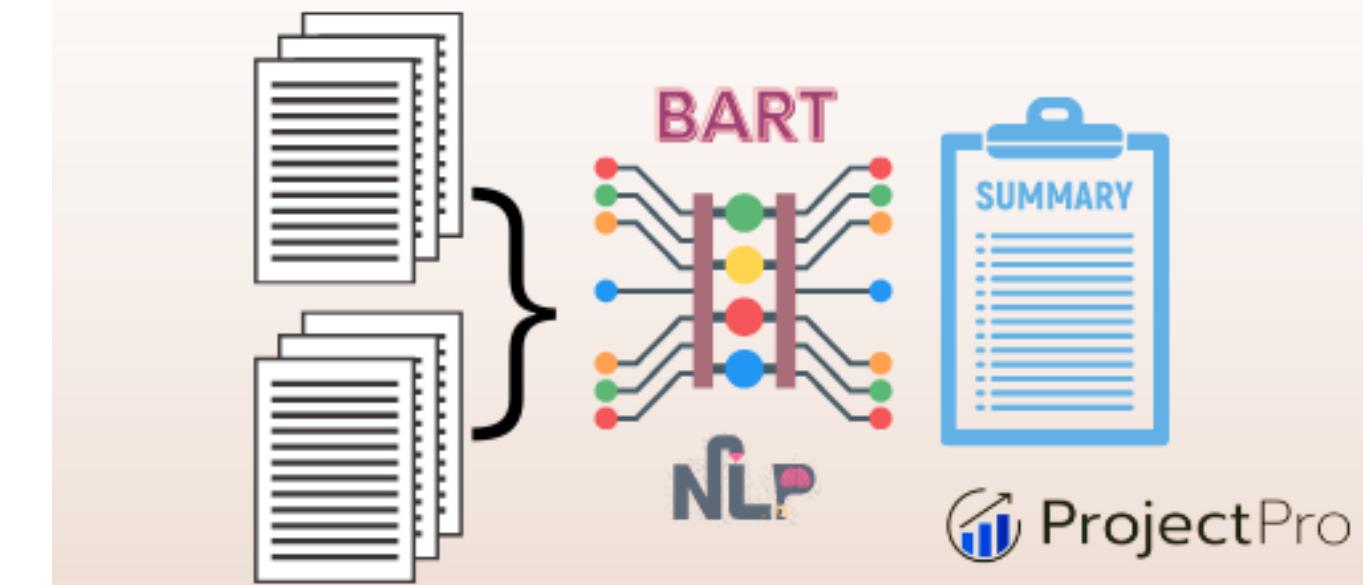
Basés sur tous les feedbacks [Colonne: Do you have any other feedback or suggestions?]

“PRE PROCESS”

- Suppression des observations vides
- Création d'une variable *sentiment* pour trier les individus selon leurs notes

Modèle BART

TRANSFORMERS-BART MODEL



Création d'une interface à l'aide de Streamlit

- Créer une interface utilisateur simple
- Rendre dynamique les résumés des feedback

Limites

- Si un même profil est sélectionné le Feedback sera le même
- Le modèle BART prend 1024 token maximum, donc limite la précision des feedback

Merci pour votre attention



UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE