

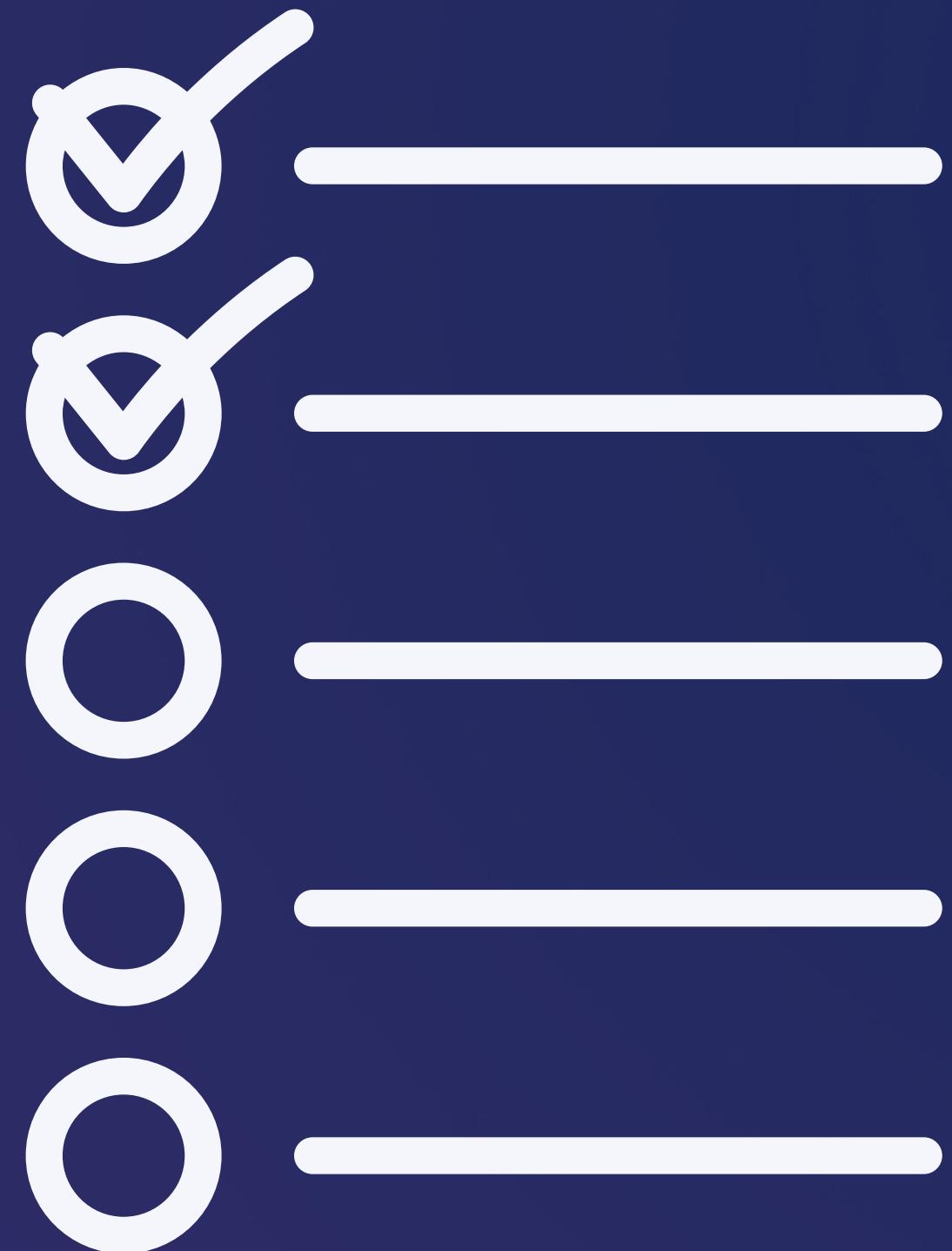
Analyse : Sentiments, Emotions et Sarcasm

Master Science des données 2023 - 2024



Plan :

1. Introduction
2. Analyse des sentiments
3. Analyse des émotions
4. Analyse du sarcasm
5. Déploiement avec streamlit
6. Conclusion





Introduction :

Ce projet se concentre sur le développement d'un système avancé d'analyse des sentiments, d'émotions et de détection du sarcasme spécifiquement adapté à la langue française.

Objectifs :

- Développer un modèle de prédiction des sentiments.
- Explorer les nuances émotionnelles.
- Détection du sarcasme.
- Développement d'une interface utilisateur conviviale.

Prédiction des sentiments :



L'analyse des sentiments, également connue sous le nom d'opinion mining, est une tâche de traitement du langage naturel (NLP) qui vise à déterminer l'attitude ou le sentiment exprimé dans un texte, qu'il s'agisse d'une phrase, d'un paragraphe, d'une critique, ou même d'un document entier.

Prédiction des sentiments :

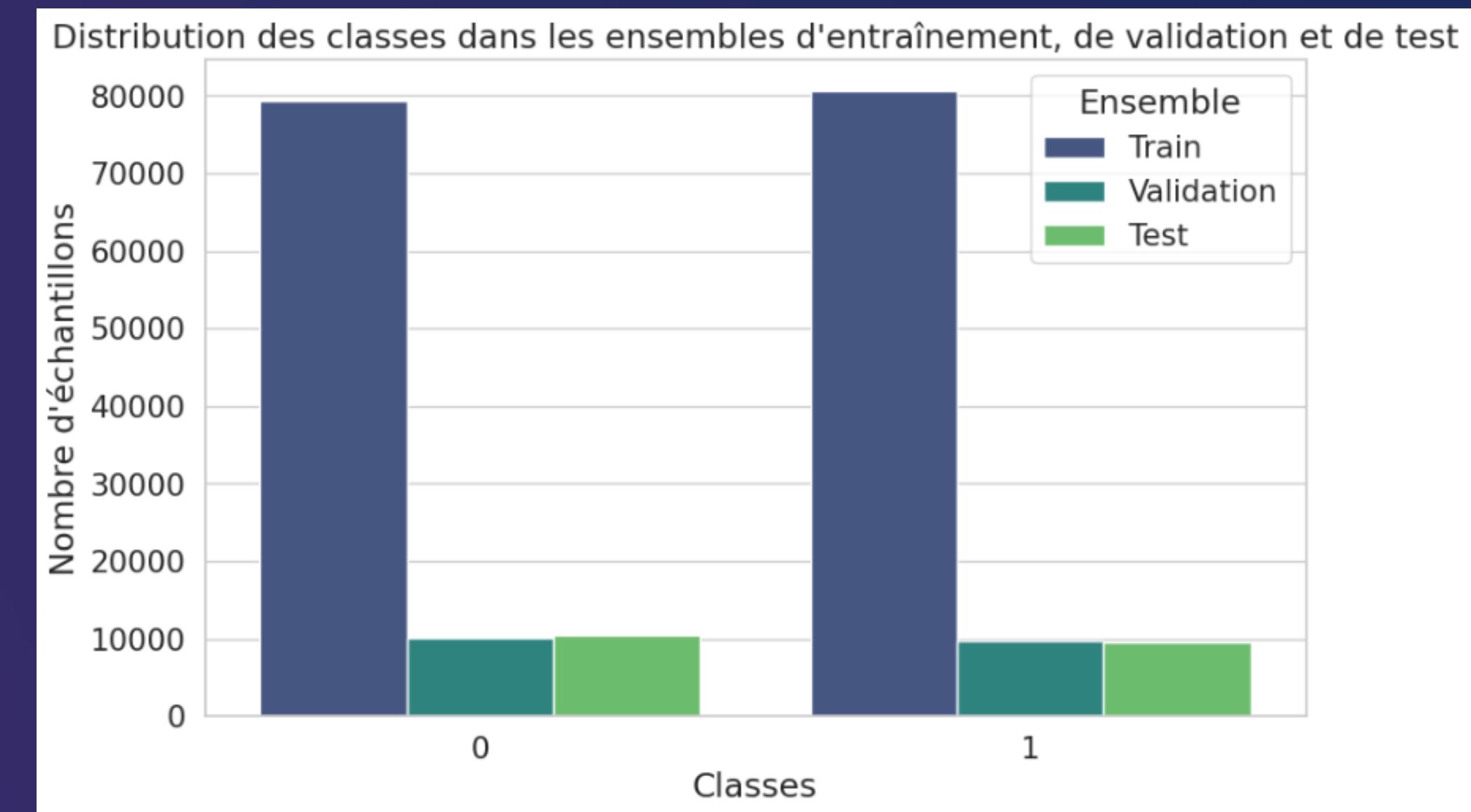
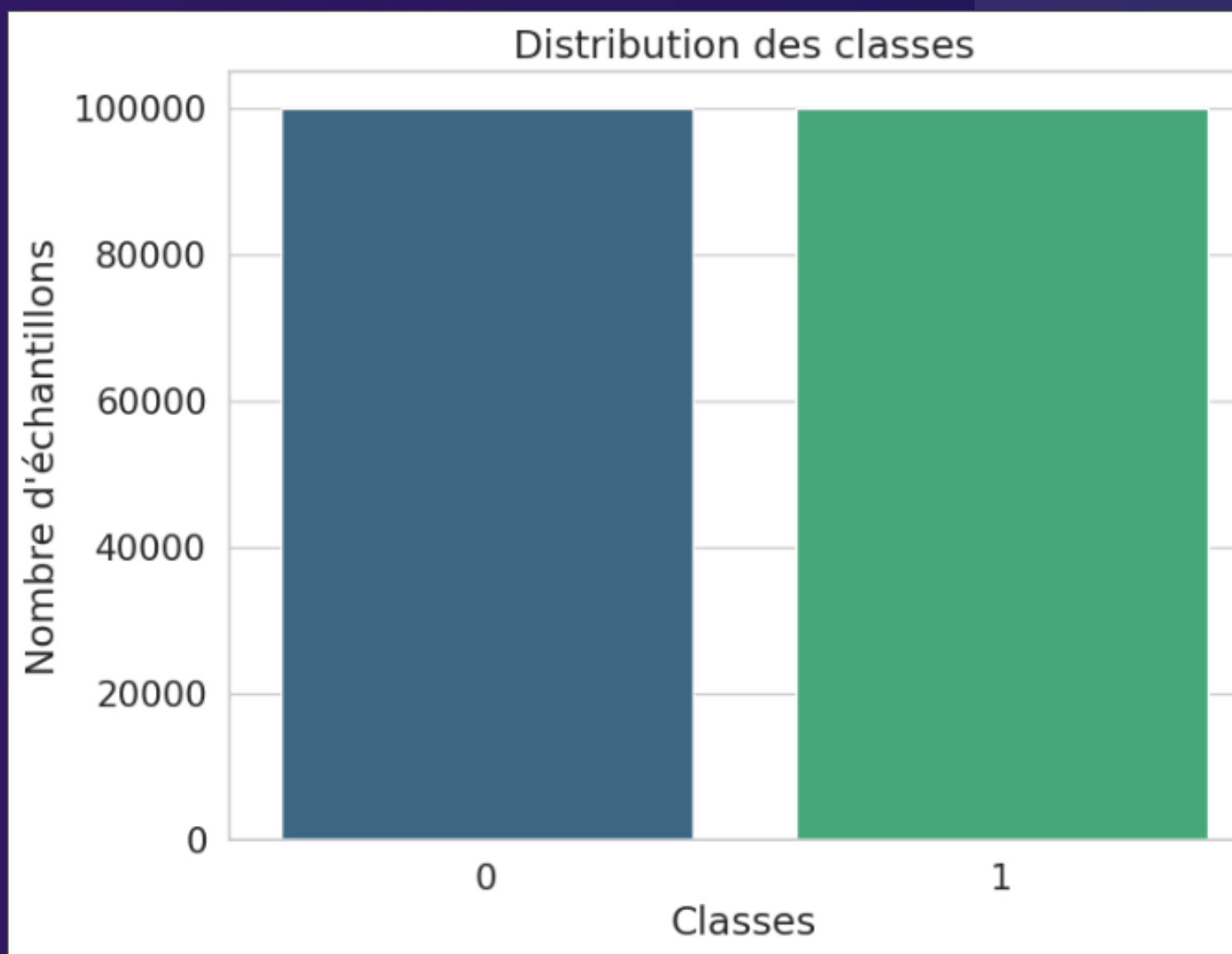
APERÇU DU DATASET “ALLOCINE DATASET” :

Mon dieu ! Quelle métaphore filée ! Je suis abasourdi ! Mais réellement, ici on a une métaphore filée de l'acceptation que sa mère puisse à nouveau avoir...	0 neg
Premier film de la saga Kozure Okami, "Le Sabre de la vengeance" est un très bon film qui mêle drame et action, et qui, en 40 ans, n'a pas pris une ride.	1 pos
L'amnésie est un thème en or pour susciter le mystère. Encore faut-il être capable de construire un scénario qui se tienne. Celui-ci est boursouflé et...	0 neg
Tout commence comme une comédie légère avant un drame soudain qui fait basculer le film dans un...suspense incroyable mâtiné d'une étude...	1 pos
un excellent film qui merite ses quatre étoiles. tellement bon que les ricains en font déjà une suite, qu'il s'empresseront de pourrir...	1 pos

- **Source** : Les membres de la communauté “Allociné.fr” ont généré ce dataset (critiques de films).
- **Période** : Les critiques de films couvrent la période de 2006 à 2020.
- **Taille** : Entraînement (160k), validation (20k), et test (20k).
- **Équilibre des Classes** : Chaque ensemble contient un nombre équivalent de critiques positives et négatives.
- **Langue** : Les critiques sont rédigées en français
- **Format** : Fichiers .jsonl et d'un Fichier .pickle.

Prédiction des sentiments :

APERÇU DU DATASET “ALLOCINE DATASET” :



Prédiction des sentiments :

MODÈLE PRÉdictives :

LOGISTIC REGRESSION AVEC TF-IDF

- TF-IDF pour représenter les documents texte.
- La régression logistique.
- Paramètres par défaut pour le TF-IDF et la régression logistique.

LOGISTIC REGRESSION AVEC TF-IDF (OPTIMISÉ)

- TF-IDF avec des paramètres personnalisés (bigrammes, limite de fréquence maximale).
- Régression logistique avec des hyperparamètres ajustés (regularisation, type de regularisation).
- Optimisé pour une meilleure performance.

RÉSEAU DE NEURONES AVEC EMBEDDING

- ANN simple avec une couche d'embedding.
- L'embedding apprend des représentations denses des mots.

Prédiction des sentiments :

MODÈLE PRÉdictives :

LOGISTIC REGRESSION
AVEC TF-IDF

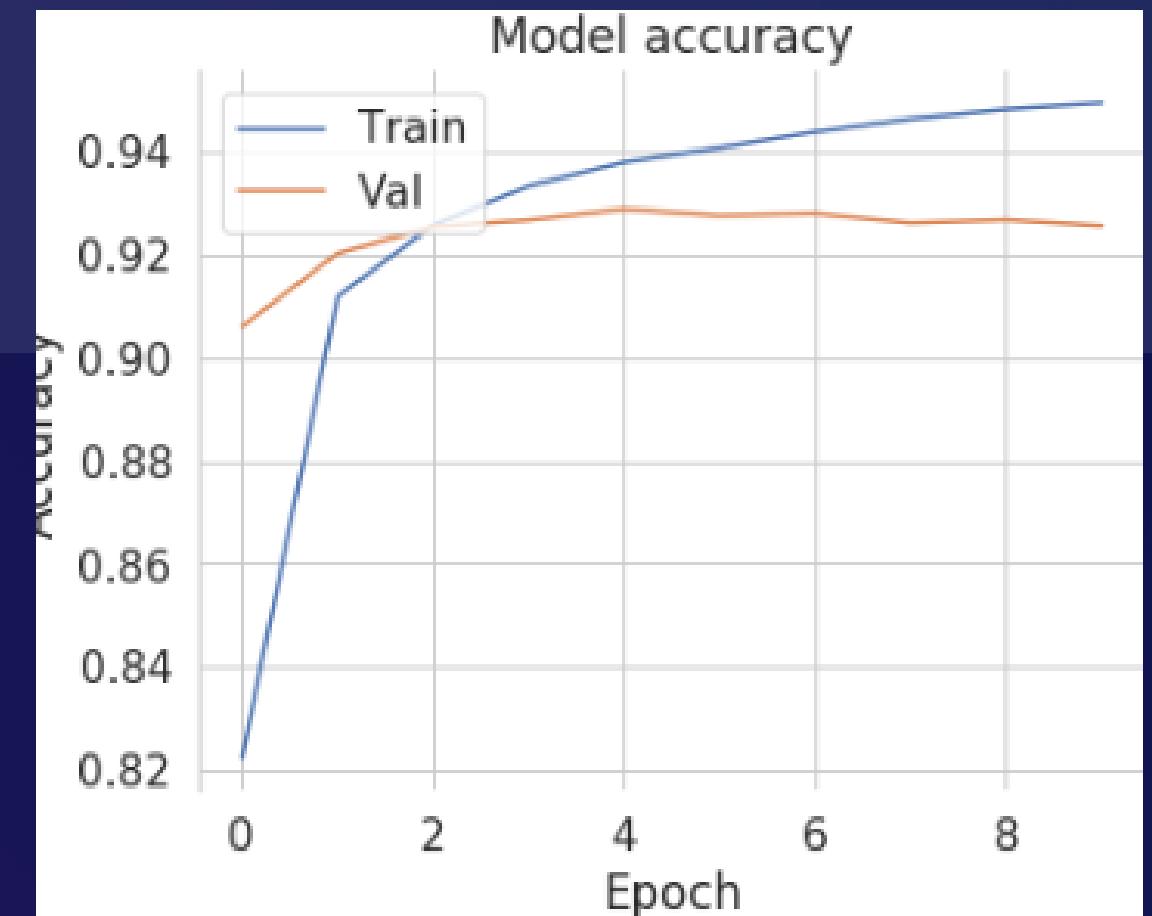
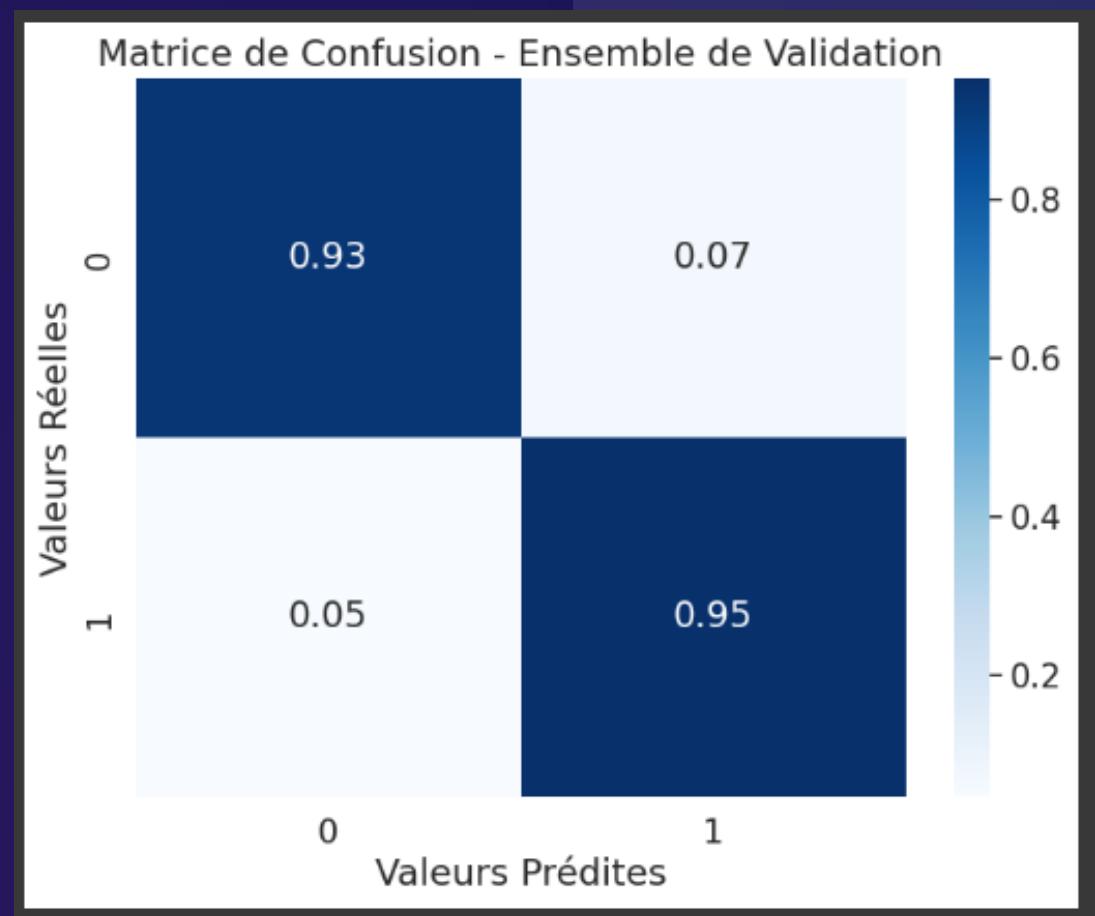
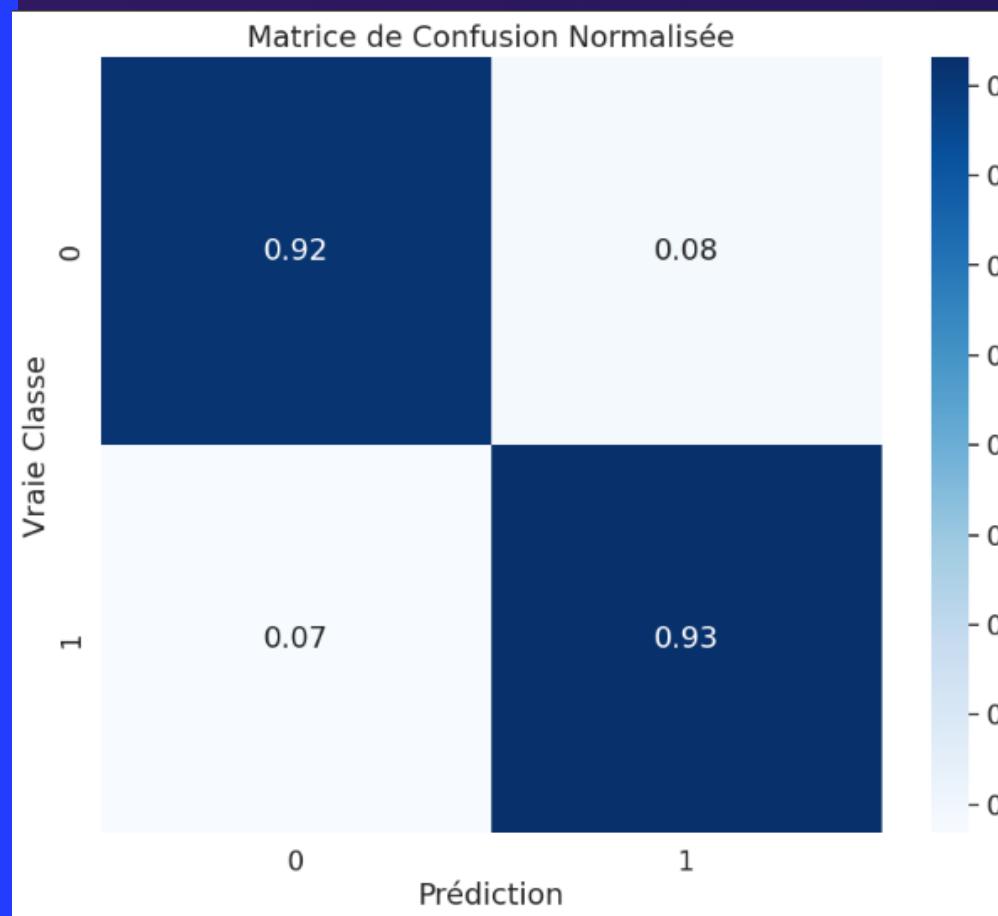
LOGISTIC REGRESSION
AVEC TF-IDF (OPTIMISÉ)

RÉSEAU DE NEURONES
AVEC EMBEDDING

Accuracy : 0.92

Accuracy : 0.94

Accuracy : 0.93



Prédiction des sentiments :

CHOIX DU MODÈLE :

Le Modèle “Logistic Regression avec TF-IDF ”est simple et rapide, le Modèle “Logistic Regression avec TF-IDF (Optimisé)” est une version optimisée avec des paramètres ajustés, et le Modèle “ Réseau de Neurones avec Embedding” utilise un réseau de neurones pour apprendre des représentations des mots mais il est plus lent . Le choix tombe sur le deuxième qui le plus optimal pour ce jeu de donnée .



Analyse des emotions



La tâche de l'analyse des émotions dans NLP revêt une importance cruciale, car il permet de comprendre les tonalités et les nuances émotionnelles présentes dans le langage humain.

Il existe plusieurs types d'émotions fréquemment observées chez les individus. Voici quelques exemples courants :



Joie



Peur



Tristesse



Surprise



Colère



Amour



Dataset Overview



File Edit View

je ne me suis pas senti humilié ; tristesse
je peux passer d'un sentiment de désespoir à un sentiment d'espoir juste en étant avec quelqu'un qui se soucie de moi ; joie
qui est éveillé ; tristesse
je prends une minute pour poster, je me sens avide, je me sens mal ; colère
Je me sens toujours nostalgique de la cheminée, je saurai qu'elle est toujours sur la propriété ; amour
je me sens grincheux ; colère
Je me sentais un peu accablé ces derniers temps, je ne savais pas pourquoi ; tristesse
J'ai pris ou milligrammes ou fois la quantité recommandée et je me suis endormi beaucoup plus vite mais je me sens aussi très drôle ; surprise
je me sens aussi confus à propos de la vie en tant qu'adolescent ou aussi blasé qu'un homme d'un an ; peur
Je travaille chez Petronas depuis des années. Je pense que Petronas a bien performé et a réalisé dénormes bénéfices ; joie
je me sens romantique aussi ; amour
j'ai l'impression que je dois donner un sens à la souffrance que je vois ; tristesse
je pense que courir est une expérience divine et que je peux m'attendre à vivre une sorte de rencontre spirituelle ; joie
Je pense que c'est la période de l'année la plus facile à ressentir de l'insatisfaction ; colère
je me sens faible en énergie, j'ai juste soif ; tristesse
J'ai une immense sympathie avec le point général, mais en tant que proto-écrivain potentiel essayant de trouver temps d'écrire dans les coins de la vie et sans aucun signe d'agent et encore moins de contrat d'édition, cela semble un peu précieux ; joie

- Source de données : Notre Dataset provient de Haggling face. Initialement, il était au format Excel et en anglais. Ensuite, nous l'avons converti au format CSV.
- L'étape suivante consiste à traduire notre dataset. Nous avons utilisé l'API de traduction de Microsoft pour traduire nos données en français
- Customer review tweeter



MICROSOFT TRANSLATOR TEXT API

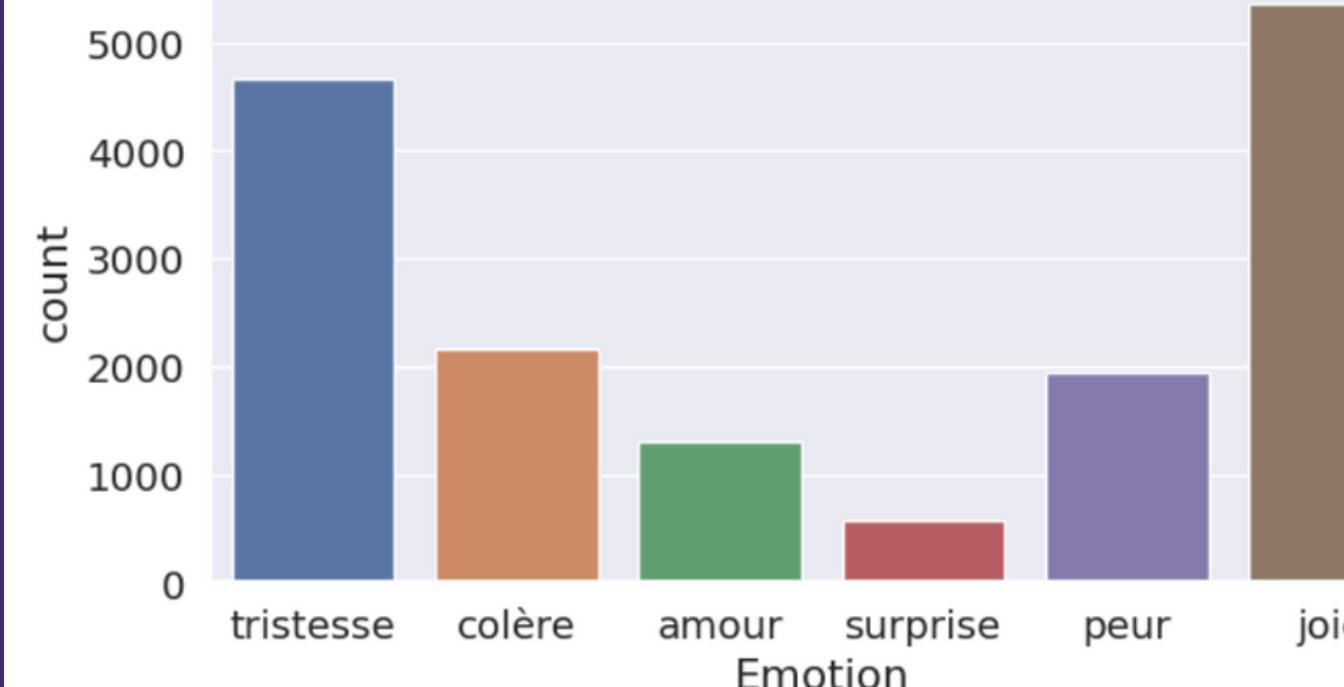
Inscription et accès à l'API :

- Créez un compte Microsoft Azure :
<https://portal.azure.com/>
- Créez une ressource Translator. Une fois connecté à Azure
- Obtenir les clés d'API pour utiliser dans la traduction en utilisant python

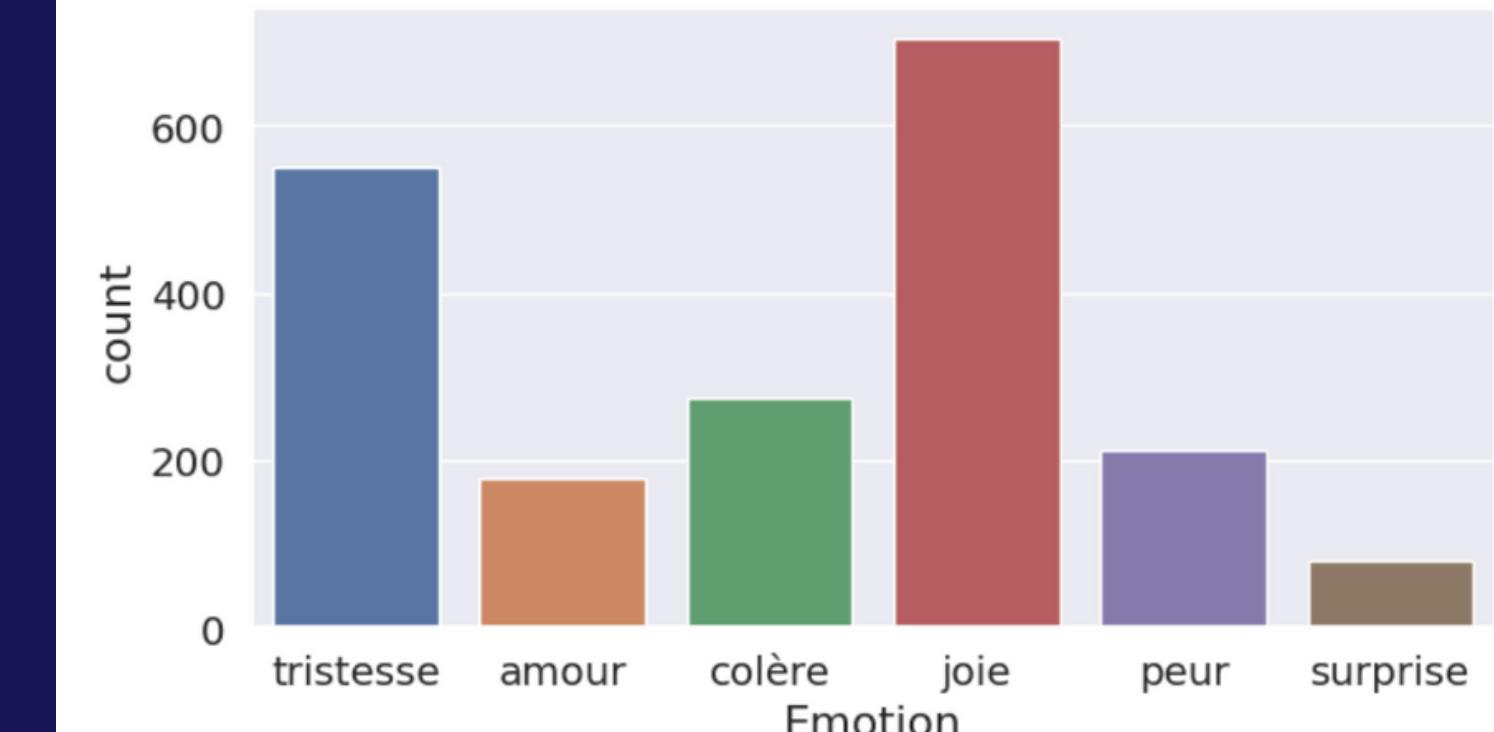
Data visualisation



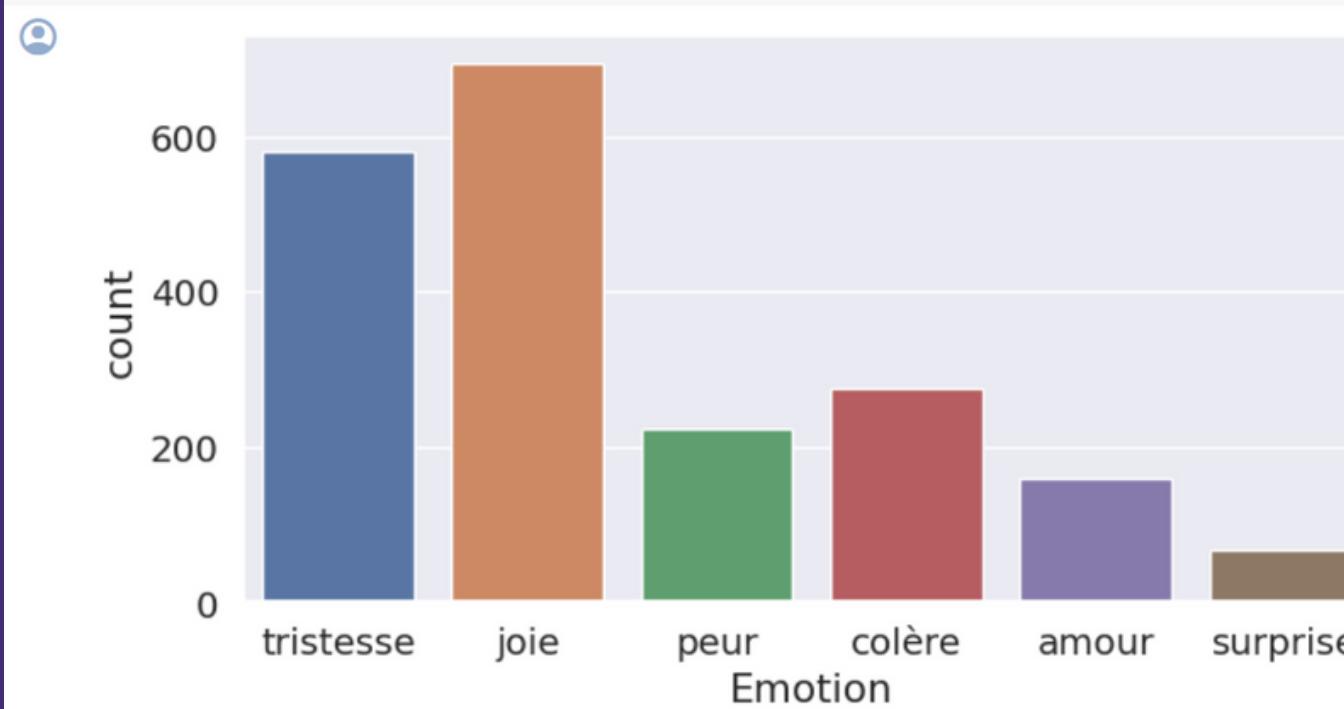
```
plt.figure(figsize=(8,4))
sns.countplot(x='Emotion', data=df_train);
```



```
plt.figure(figsize=(8,4))
sns.countplot(x='Emotion', data=df_val);
```



```
▶ plt.figure(figsize=(8,4))
sns.countplot(x='Emotion', data=df_test);
```



```
✓ 0s ▶ #print the shape of the data set
print(df_train.shape)
```

```
(15993, 2)
```

```
▶ #print the shape of the data set
print(df_val.shape)
```

```
(1999, 2)
```

```
[ ] #print the shape of the data set
print(df_test.shape)
```

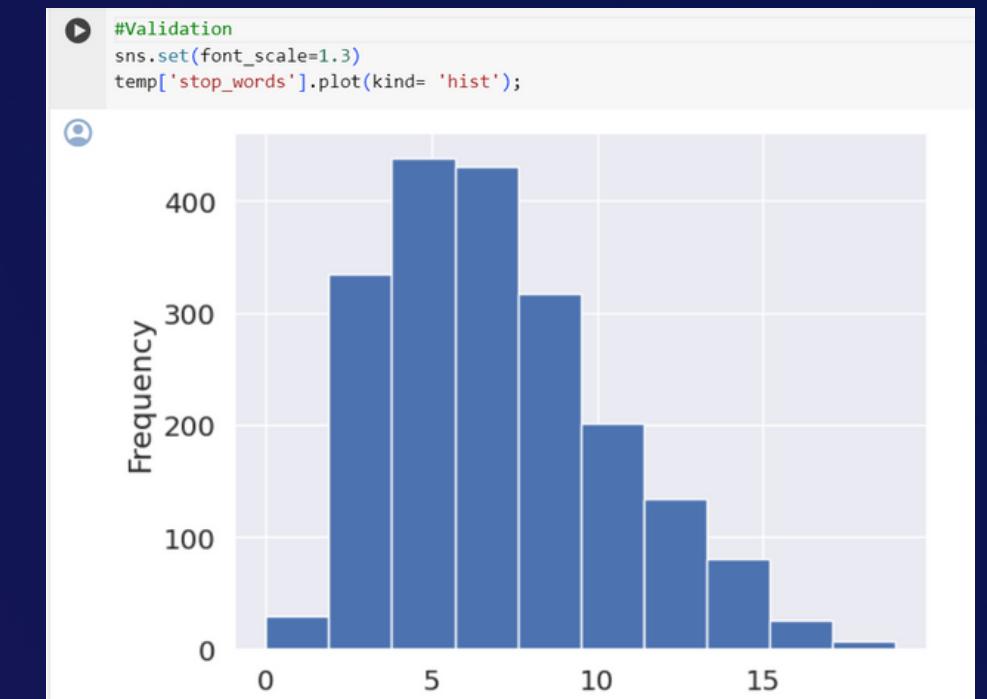
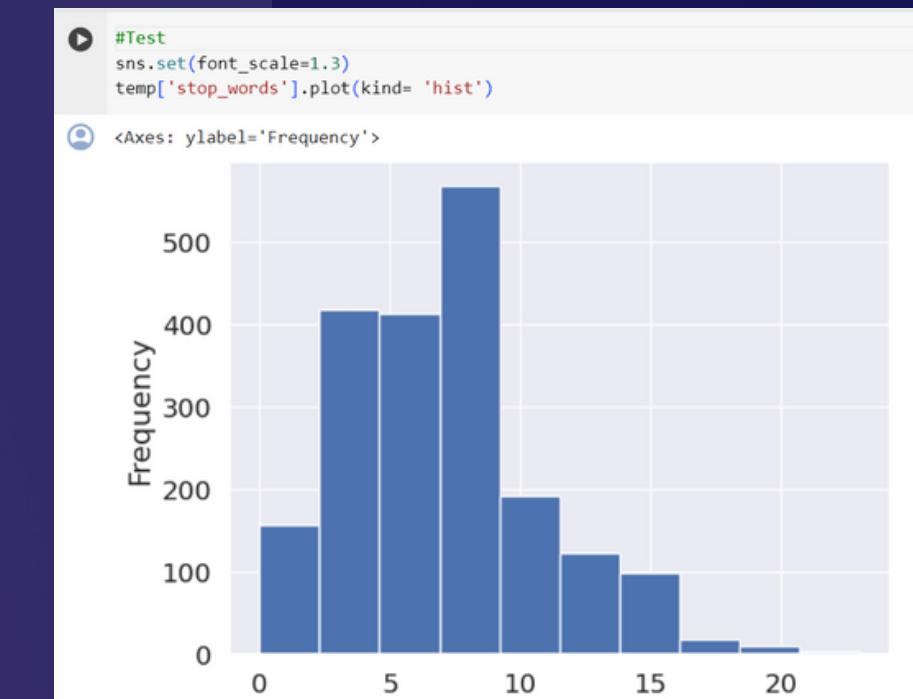
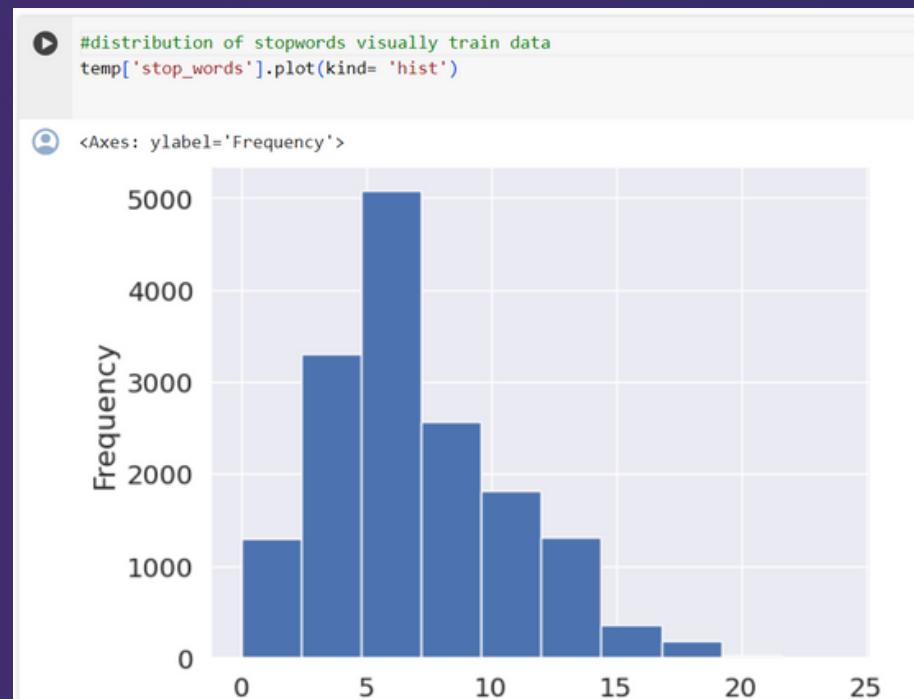
```
(1999, 2)
```



Data preprocessing



- Vérifiez si les données contiennent des valeurs nulles et des doublons
- Vérifiez également s'il y a des lignes ayant le même texte mais avec des émotions différentes
- Eliminer les stopwords



- Comparez les lignes dataset test , train, val
- Cleaning data : lemmatization , remove_stop_words , Removing_numbers, lower_case, Removing_punctuations, Removing_urls, remove_small_sentences,



Modeling



LOGISTIC REGRESSION



DECISION TREE

LSTM

SUPPORT VECTOR
MACHINE

RANDOM FOREST

Logistic Regression

▼ les classes ne sont pas équilibré

```
##Classification Report
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
amour	0.63	0.30	0.40	159
colère	0.84	0.67	0.74	275
joie	0.74	0.93	0.83	694
peur	0.80	0.59	0.68	224
surprise	0.88	0.45	0.60	66
tristesse	0.78	0.86	0.82	580
accuracy			0.77	1998
macro avg	0.78	0.63	0.68	1998
weighted avg	0.77	0.77	0.76	1998

Decision Tree

```
##Classification Report
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
amour	0.47	0.47	0.47	159
colère	0.73	0.71	0.72	275
joie	0.74	0.71	0.73	694
peur	0.67	0.67	0.67	224
surprise	0.57	0.58	0.57	66
tristesse	0.74	0.78	0.76	580
accuracy			0.70	1998
macro avg	0.65	0.65	0.65	1998
weighted avg	0.70	0.70	0.70	1998

Random Forest

```
##Classification Report
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
amour	0.73	0.40	0.52	159
colère	0.84	0.71	0.77	275
joie	0.75	0.92	0.83	694
peur	0.77	0.64	0.70	224
surprise	0.79	0.56	0.65	66
tristesse	0.81	0.82	0.81	580
accuracy			0.78	1998
macro avg	0.78	0.68	0.71	1998
weighted avg	0.78	0.78	0.77	1998

Support Vector Machine

```
##Classification Report
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
amour	0.62	0.29	0.39	159
colère	0.84	0.64	0.73	275
joie	0.74	0.93	0.82	694
peur	0.75	0.62	0.68	224
surprise	0.88	0.45	0.60	66
tristesse	0.79	0.84	0.82	580
accuracy			0.76	1998
macro avg	0.77	0.63	0.67	1998
weighted avg	0.76	0.76	0.75	1998

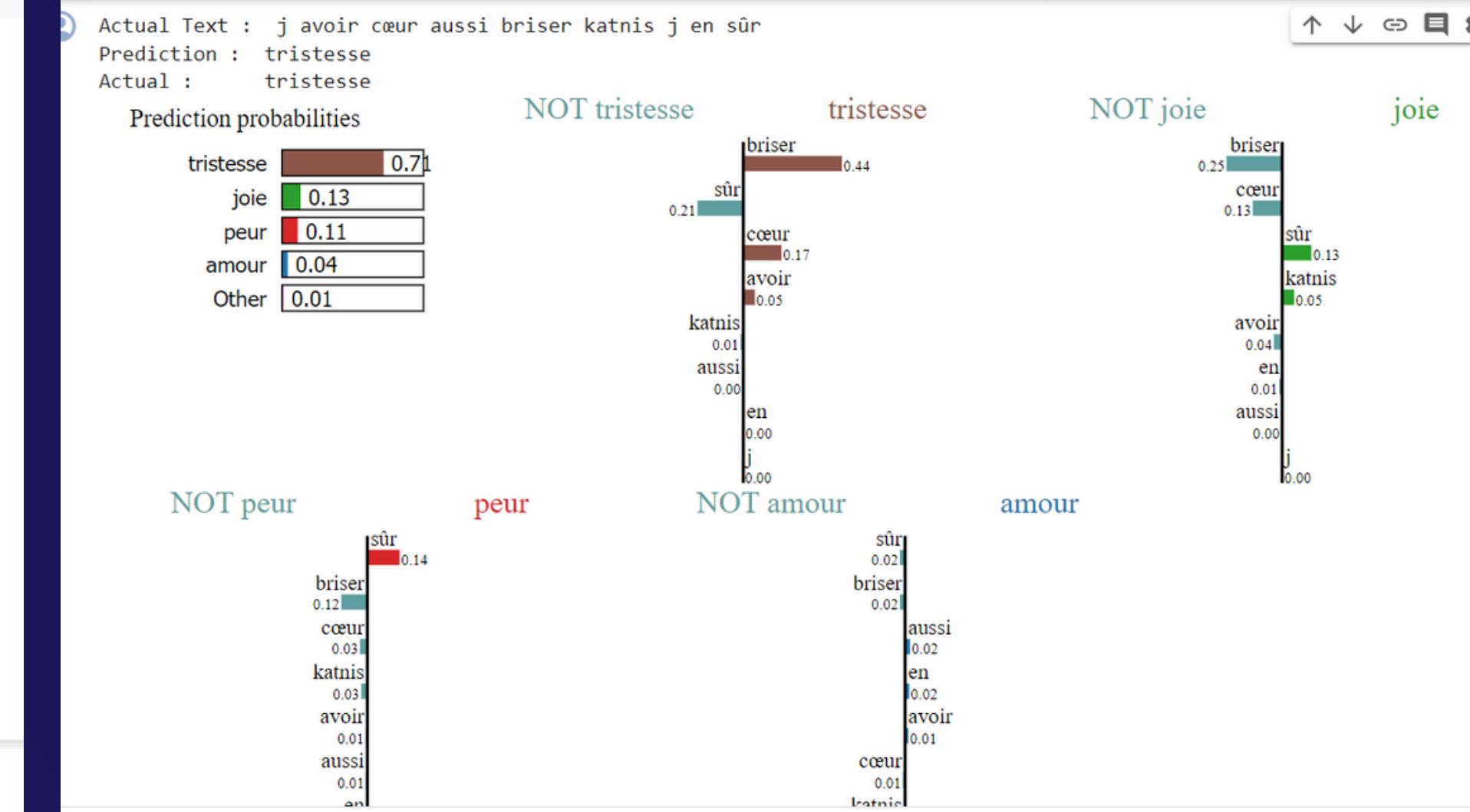
Comparaison et résultat



COMPARAISON

LIME : RF

	Model	Accuracy
0	Random Forest	0.78
1	Logistic Regression	0.77
2	Support Vector Machine	0.76
3	Decision Tree	0.70

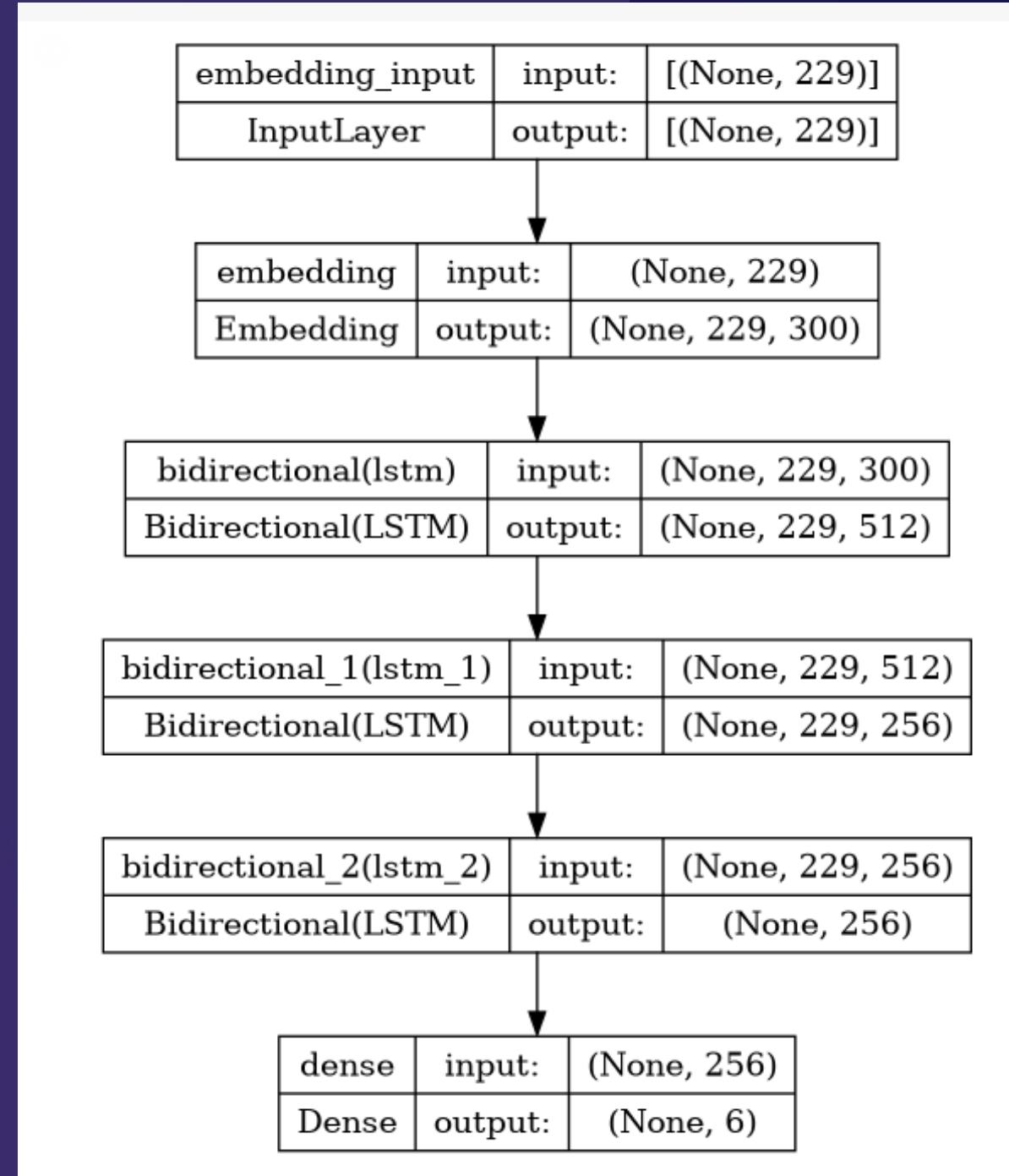


LSTM

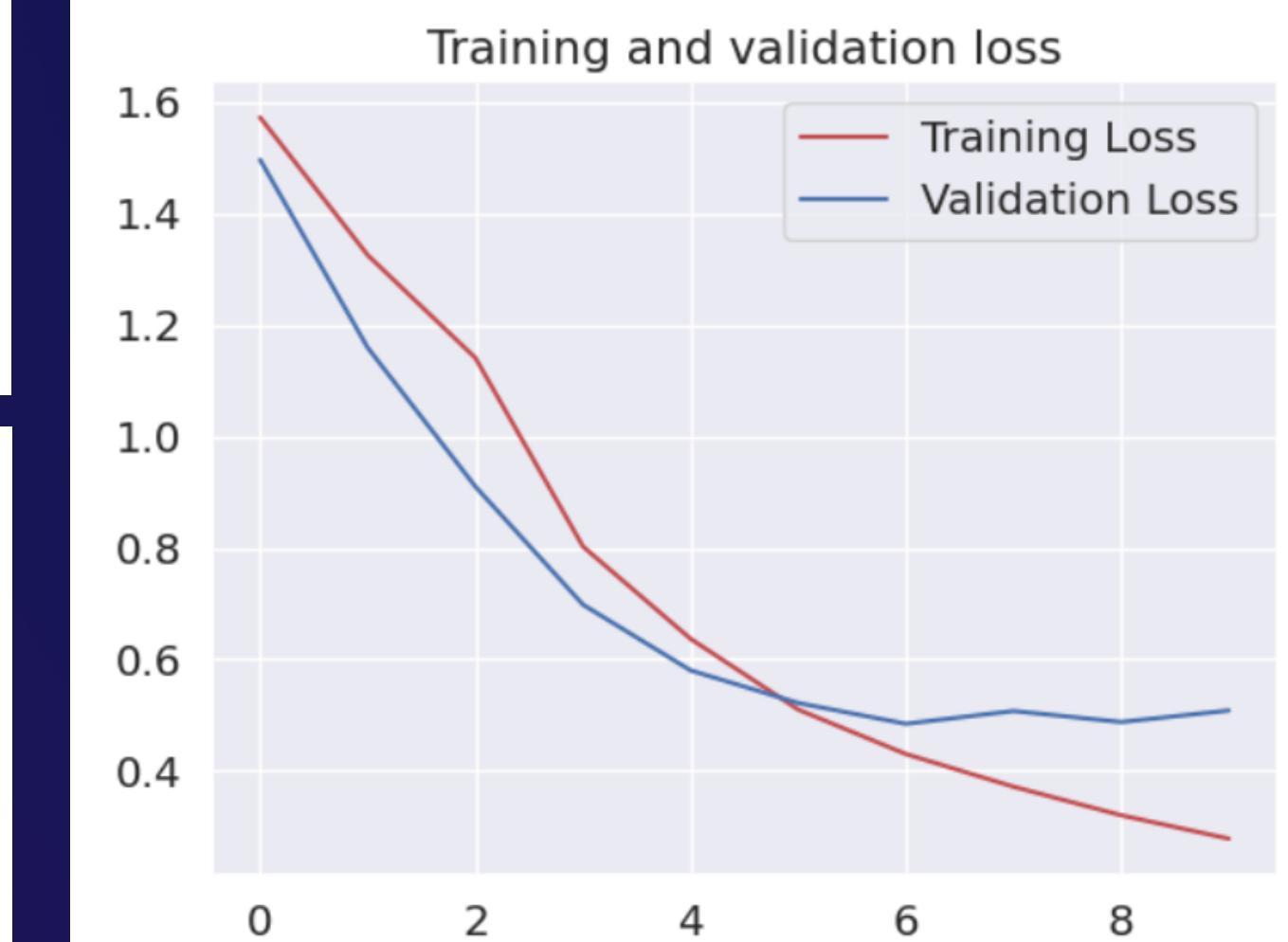
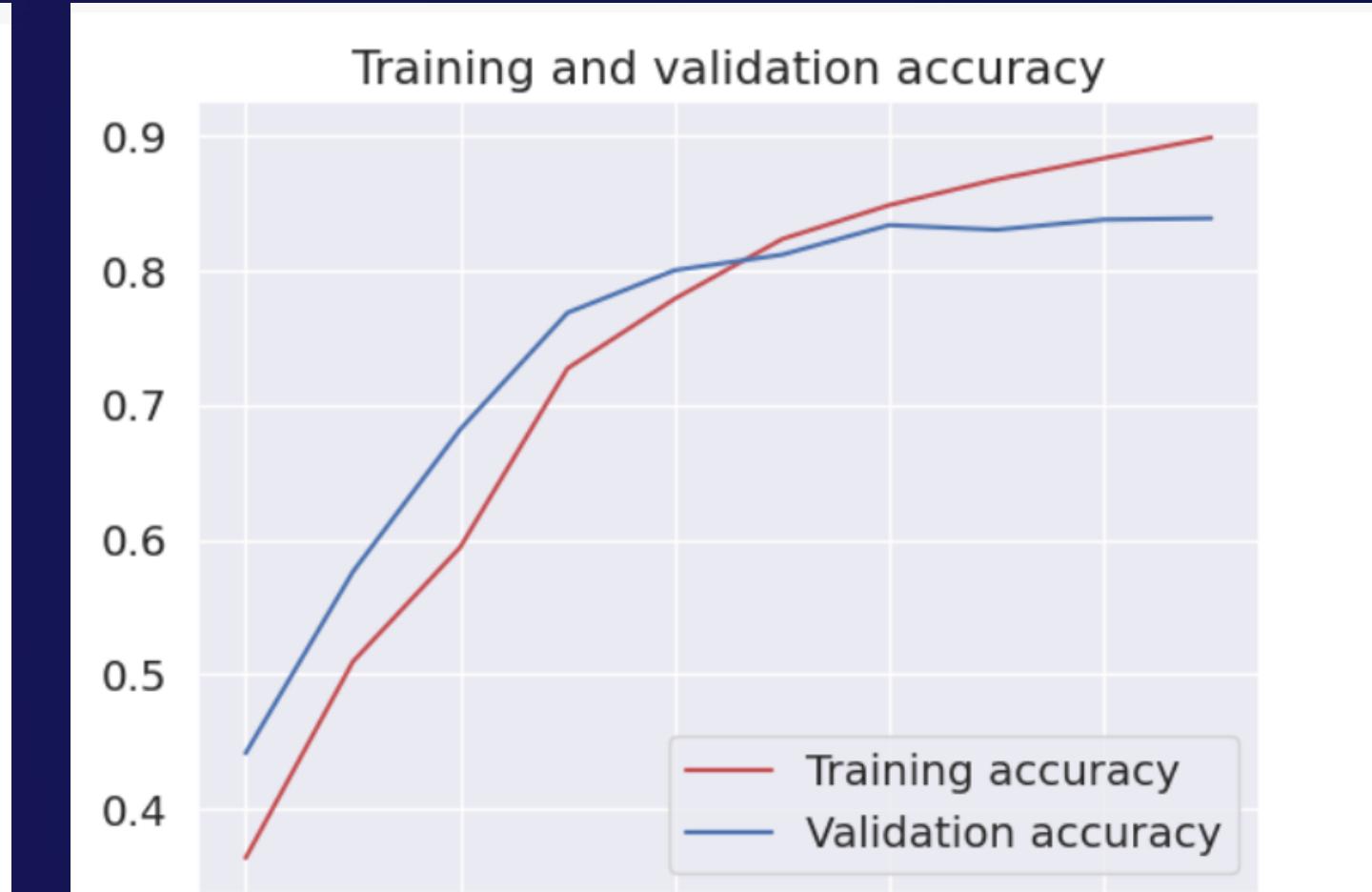


WordEmbedding

- Fair



```
63/63 [=====] - 50s 758ms/step
      precision    recall   f1-score   support
      0       0.63     0.71     0.67     159
      1       0.81     0.75     0.78     275
      2       0.85     0.88     0.87     694
      3       0.82     0.73     0.77     224
      4       0.65     0.80     0.72      66
      5       0.88     0.85     0.86     580
accuracy                           0.82     1998
macro avg       0.77     0.79     0.78     1998
weighted avg    0.82     0.82     0.82     1998
```



Sarcasm

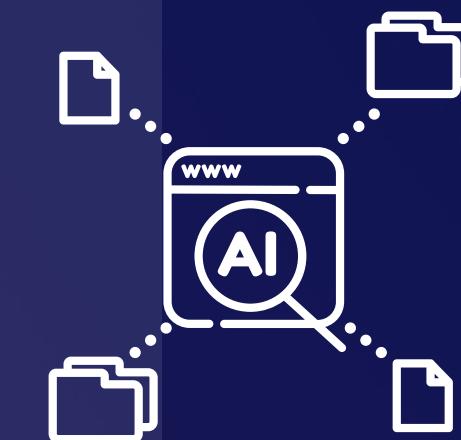


L'analyse du sarcasme se présente comme une tâche significative en NLP, nécessitant une compréhension subtile des nuances linguistiques pour interpréter et identifier le sarcasme dans les textes, contribuant ainsi à une meilleure compréhension contextuelle des discours.

Dataset Overview

```
File Edit View
mon héritage blanc ;0
5 façons de déclarer vos impôts avec moins de stress ;0
Richard Branson fait un don pour lutter contre le réchauffement climatique presque autant que le coût des voyages en montgolfière ratés ;1
le gouvernement fantôme devient trop important pour se réunir dans la salle de conférence Marriott b ;1
beaucoup de parents connaissent ce scénario ;0
cette lesbienne est considérée comme un père dans l'Indiana et c'est un incroyable en plus ;0
Amanda Peet a dit à sa fille que le sexe était un câlin spécial ;0
ce qu'il faut savoir sur les traitements actuels contre Ebola ;0
Chris Christie suggère qu'Hillary Clinton était responsable de l'enlèvement de centaines d'écolières par Boko Haram ;0
Ford développe un nouveau SUV fonctionnant uniquement à l'essence ;1
Le PDG d'Uber, Travis Kalanick, démissionne du conseil consultatif économique de Trump ;0
le garçon de la zone entre dans la phase de saut et de toucher le haut des portes ;1
l'homme de la région effectue la plupart de ses déplacements en civière ;1
ne laisser aucune personne handicapée de côté ;0
lin manuel miranda aimeraient vous rappeler de ranger votre téléphone ;0
60 journalistes tués en 2014 alors que les attaques contre la presse internationale se multiplient ;0
```

commentary website Reddit



kaggle



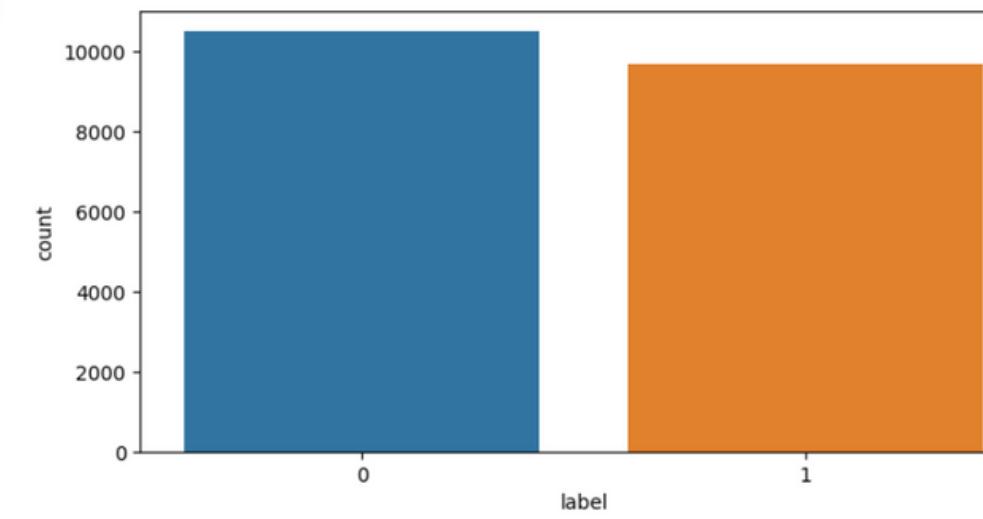
Scraping

Data visualisation

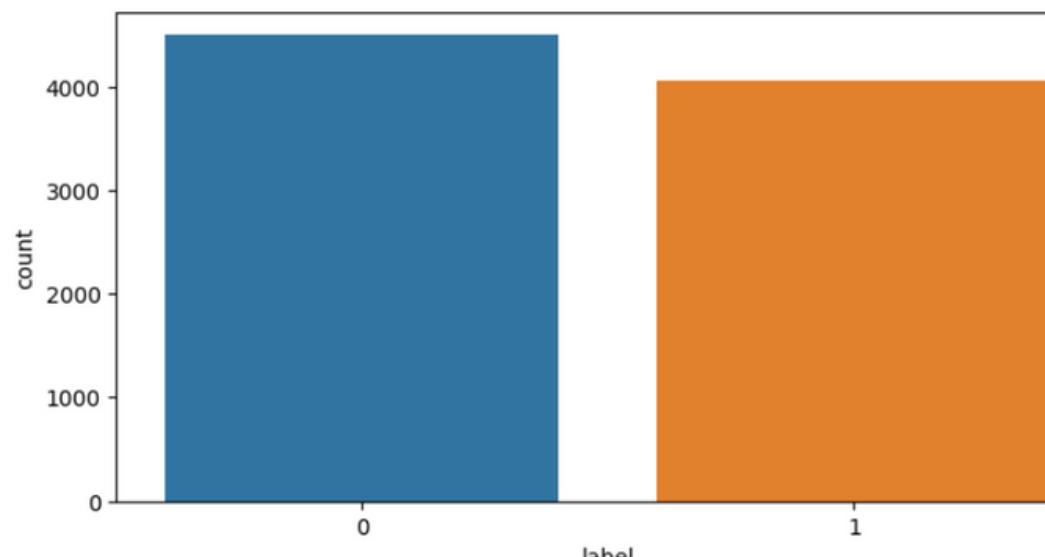
```
▶ print(df_train.columns)
print(df_train.count()) ## small dataset
```

```
⠼ Index(['Text', 'label'], dtype='object')
Text      20032
label     20032
dtype: int64
```

```
⠼ import matplotlib.pyplot as plt
plt.figure(figsize=(8,4))
sns.countplot(x='label', data=df_train);
```

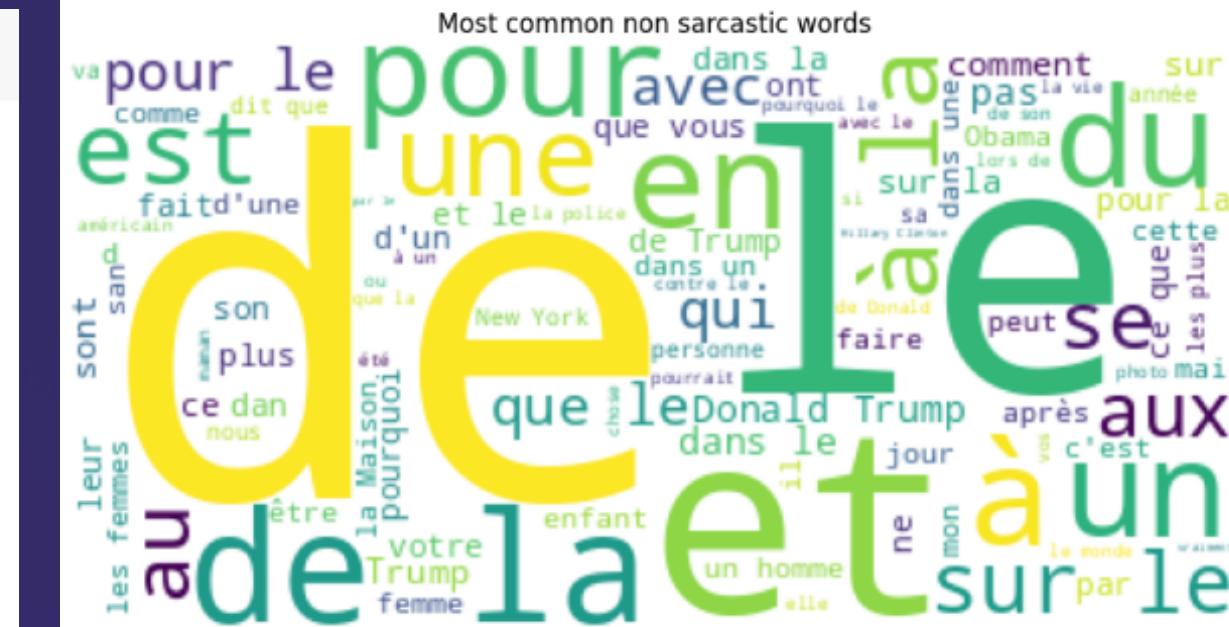


```
⠼ import matplotlib.pyplot as plt
plt.figure(figsize=(8,4))
sns.countplot(x='label', data=df_test);
```



```
⠼ ⚡ print(df_test.columns)
print(df_test.count()) ## small dataset
```

```
⠼ Index(['Text', 'label'], dtype='object')
Text      8561
label     8561
dtype: int64
```



Modeling

BERT

BERT est un modèle de langage pré-entraîné basé sur des réseaux de neurones transformer, développé par Google. Il est conçu pour comprendre le contexte et la signification des mots dans un texte en utilisant une approche bidirectionnelle

DISTILBERT

```
[ ] bert = TFAutoModel.from_pretrained('bert-base-cased')

model.safetensors:  0% | 0.00/436M [00:00<?, ?B/s]
Some weights of the PyTorch model were not used when initializing the TF 2.0 model TFBertModel: ['cls.predictions.bias', 'cls.predictions.transform.dense']
- This IS expected if you are initializing TFBertModel from a PyTorch model trained on another task or with another architecture (e.g. initializing a TFBertModel from a PyTorch RobertaModel)
- This IS NOT expected if you are initializing TFBertModel from a PyTorch model that you expect to be exactly identical (e.g. initializing a TFBertForSequenceClassification from a PyTorch BertForSequenceClassification)
All the weights of TFBertModel were initialized from the PyTorch model.
If your task is similar to the task the model of the checkpoint was trained on, you can already use TFBertModel for predictions without further training
```

```
| model = TFDistilBertModel.from_pretrained('distilbert-base-cased')

tokenizer_config.json:  0% | 0.00/29.0 [00:00<?, ?B/s]
vocab.txt:  0% | 0.00/213k [00:00<?, ?B/s]
```

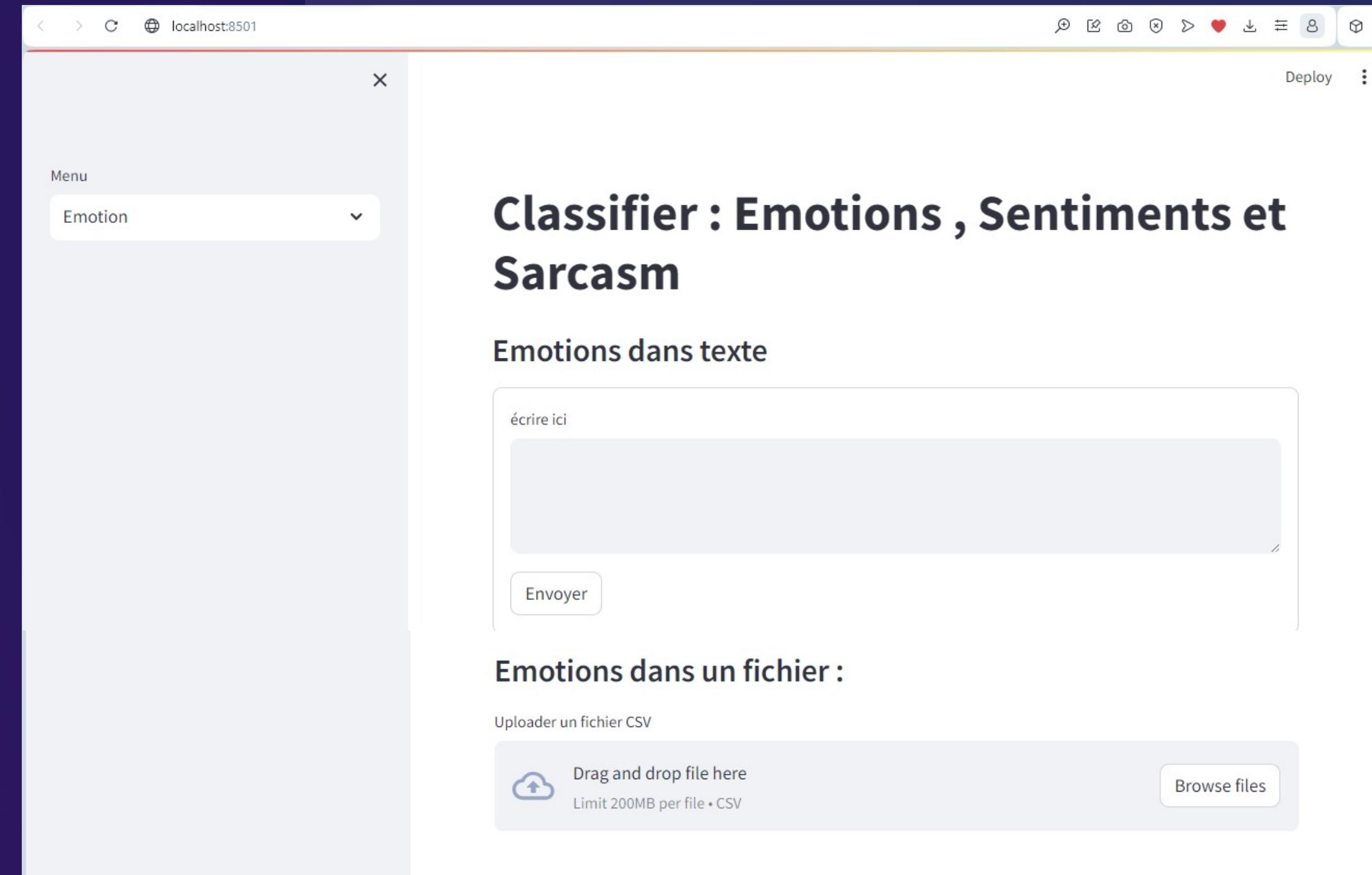
DistilBERT est une version réduite et plus rapide du modèle de langage BERT. Il a été créé par Hugging Face en utilisant des techniques de distillation, visant à conserver les performances de BERT tout en réduisant sa complexité.

Choix du modèle DistilBert

- **Meilleure performance en termes d'accuracy** : DistilBERT affiche une précision (accuracy) de 0,93, comparée à celle de BERT qui est de 0,78.
- **Emprunte mémoire plus légère** : DistilBERT présente un avantage en termes de taille mémoire avec seulement 238 MB, tandis que BERT requiert 448 MB.
- **Temps de déploiement plus rapide** : En plus de sa taille réduite, DistilBERT bénéficie d'une vitesse de déploiement plus rapide par rapport à BERT.

Déploiement :

DÉPLOIEMENT AVEC STREAMLIT :



Déploiement :

DÉPLOIEMENT DE PRÉDICTION DES ÉMOTIONS :

Emotions dans texte

écrire ici

joyeux anniversaire

Envoyer

Text original

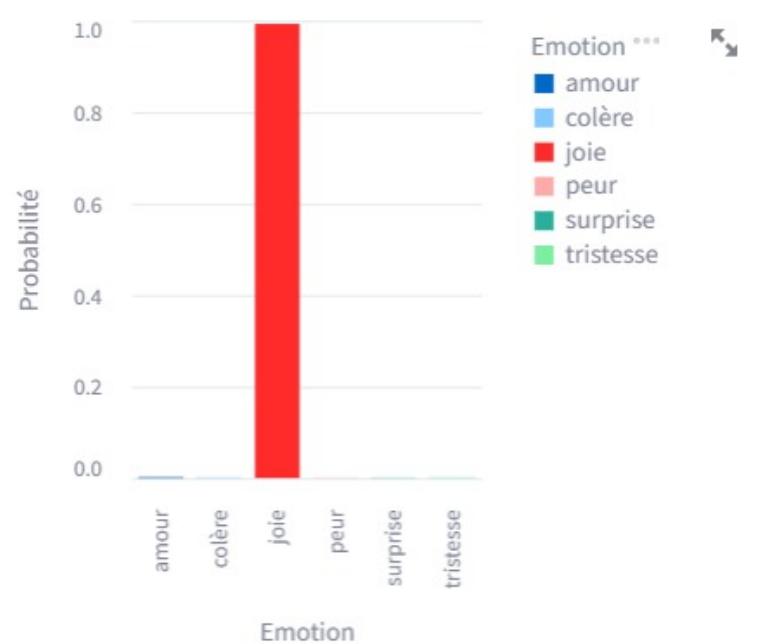
joyeux anniversaire

Prediction

joie: 😊

Confidence: 99.30%

Probabilité de prédition



Emotions dans texte

écrire ici

c'est avec une grande tristesse j'ai reçu cette nouvelle

Envoyer

Text original

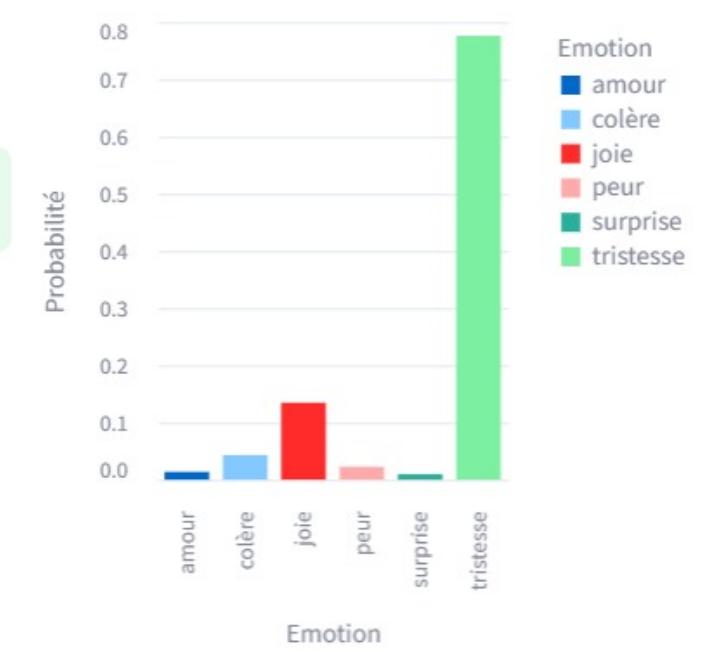
c'est avec une grande tristesse j'ai reçu cette nouvelle

Prediction

tristesse: 😢

Confidence: 77.63%

Probabilité de prédition



Déploiement :

DÉPLOIEMENT DE PRÉDICTION DES ÉMOTIONS :

Emotions dans texte

écrire ici

j'ai été surpris par son soudain changement d'avis

Envoyer

Text original

j'ai été surpris par son soudain changement d'avis

Prediction

surprise: 😲

Confidence: 97.35%

Probabilité de prédition

A bar chart titled "Probabilité de prédition" showing the probability of six emotions for the input text. The x-axis is labeled "Emotion" and lists "amour", "colère", "joie", "peur", "surprise", and "tristesse". The y-axis is labeled "Probabilité" and ranges from 0.0 to 1.0. A single teal bar for "surprise" reaches a height of 1.0, indicating it is the most probable emotion.

Emotion	Probabilité
amour	0.00
colère	0.00
joie	0.00
peur	0.02
surprise	1.00
tristesse	0.00

Emotion

- amour
- colère
- joie
- peur
- surprise
- tristesse

Emotions dans texte

écrire ici

la peur a saisi mon cœur face à cette situation angoissante

Envoyer

Text original

la peur a saisi mon cœur face à cette situation angoissante

Prediction

peur: 😱😱

Confidence: 99.62%

Probabilité de prédition

A bar chart titled "Probabilité de prédition" showing the probability of six emotions for the input text. The x-axis is labeled "Emotion" and lists "amour", "colère", "joie", "peur", "surprise", and "tristesse". The y-axis is labeled "Probabilité" and ranges from 0.0 to 1.0. A single red bar for "peur" reaches a height of 1.0, indicating it is the most probable emotion.

Emotion	Probabilité
amour	0.00
colère	0.00
joie	0.00
peur	1.00
surprise	0.00
tristesse	0.00

Emotion ***

- amour
- colère
- joie
- peur
- surprise
- tristesse

Déploiement :

DÉPLOIEMENT DE PRÉDICTION DES ÉMOTIONS :

Emotions dans texte

écrire ici

chaque geste décrit son amour inconditionnel

Envoyer

Text original

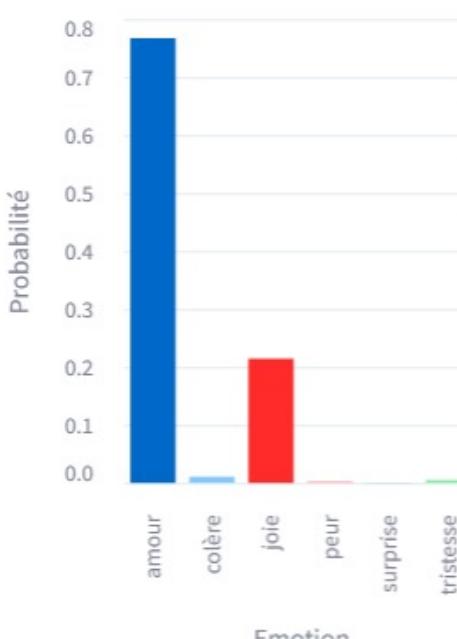
chaque geste décrit son amour inconditionnel

Prediction

amour: 😊

Confidence: 76.74%

Probabilité de prédiction



Emotion
■ amour
■ colère
■ joie
■ peur
■ surprise
■ tristesse

Emotions dans texte

écrire ici

le rouge de la colère teinte ces joues

Envoyer

Text original

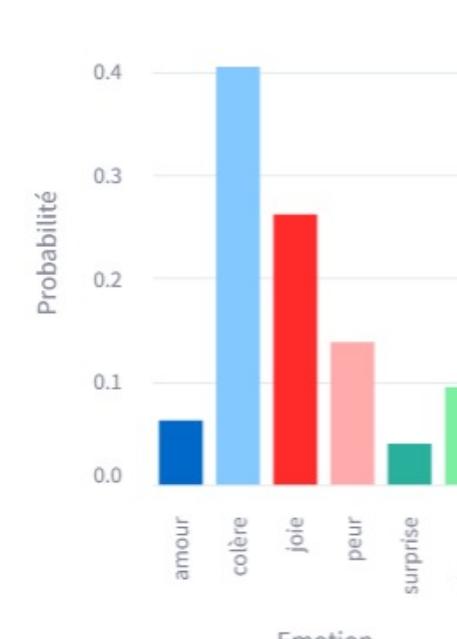
le rouge de la colère teinte ces joues

Prediction

colère: 😡

Confidence: 40.42%

Probabilité de prédiction



Emotion ***
■ amour
■ colère
■ joie
■ peur
■ surprise
■ tristesse

Déploiement :

DÉPLOIEMENT DE PRÉDICTION DES ÉMOTIONS :

	A	B	C	D	E	F
1	Texte					
2	Je me sens tellement heureux quand je suis avec mes amis.					
3	La nouvelle sur l'accident est vraiment triste et dâchirante.					
4	Ce film est hilarant ; il me fait toujours rire.					
5	Je suis anxieux ... propos de l'examen ... venir					
6	La fâte surprise d'anniversaire m'a rempli de joie et de gratitude.					
7	Recevoir un compliment renforce ma confiance et ma joie.					

Emotions dans un fichier

Uploader un fichier CSV

Drag and drop file here
Limit 200MB per file • CSV

TÉLÉCHARGÉ

predictions (1).csv predictions.csv wheat-seeds.csv

< >

Afficher tous les fichiers ...

Browse files

EMOTIONS dans un fichier :

Uploader un fichier CSV

Drag and drop file here
Limit 200MB per file • CSV

émotions.csv 355.0B x

Browse files

Contenu du fichier CSV

	Texte	Prediction	Confidence
0	Je me sens tellement heureux quand je suis avec mes amis.	joie: 😊	99.70%
1	La nouvelle sur l'accident est vraiment triste et dâchirante.	tristesse: 😢	93.56%
2	Ce film est hilarant ; il me fait toujours rire.	joie: 😊	95.76%
3	Je suis anxieux ... propos de l'examen ... venir	peur: 😰	99.72%
4	La fâte surprise d'anniversaire m'a rempli de joie et de gratitude.	surprise: 😲	87.75%
5	Recevoir un compliment renforce ma confiance et ma joie.	joie: 😊	99.35%

Télécharger les prédictions CSV

	A	B	C	D	E	F	G
1	Texte,Prediction,Confidence						
2	Je me sens tellement heureux quand je suis avec mes amis.,joie,0.99701595						
3	La nouvelle sur l'accident est vraiment triste et dâchirante.,tristesse,0.9356431						
4	Ce film est il me fait toujours rire.,joie,0.9575976						
5	Je suis anxieux ... propos de l'examen ... venir ,peur,0.9972283						
6	La fâte surprise d'anniversaire m'a rempli de joie et de gratitude.,surprise,0.87753177						
7	Recevoir un compliment renforce ma confiance et ma joie.,joie,0.99348104						

Déploiement :

DÉPLOIEMENT DE PRÉDICTION DES SENTIMENTS :

Sentiments dans texte

écrire ici

c'est le meilleur film du monde

Envoyer

Text original

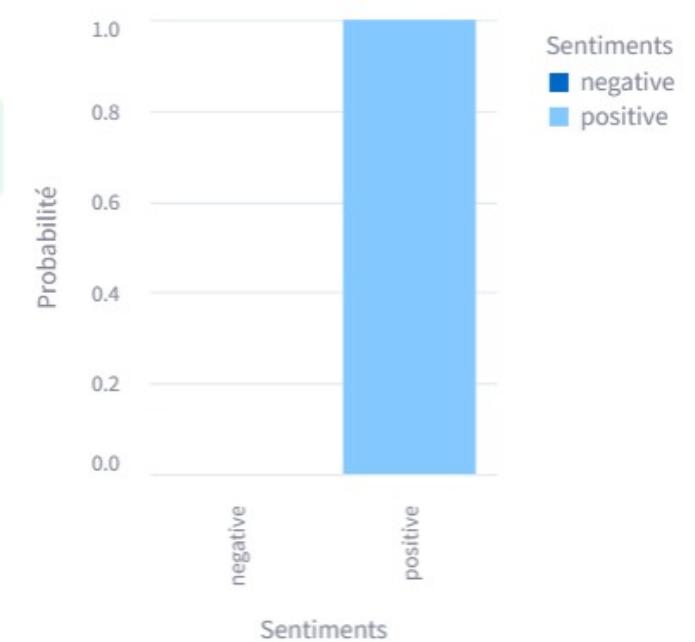
c'est le meilleur film du monde

Prediction

positive: +

Confidence: 100.00%

Probabilité de prédition



Sentiments dans texte

écrire ici

mauvaise chanson

Envoyer

Text original

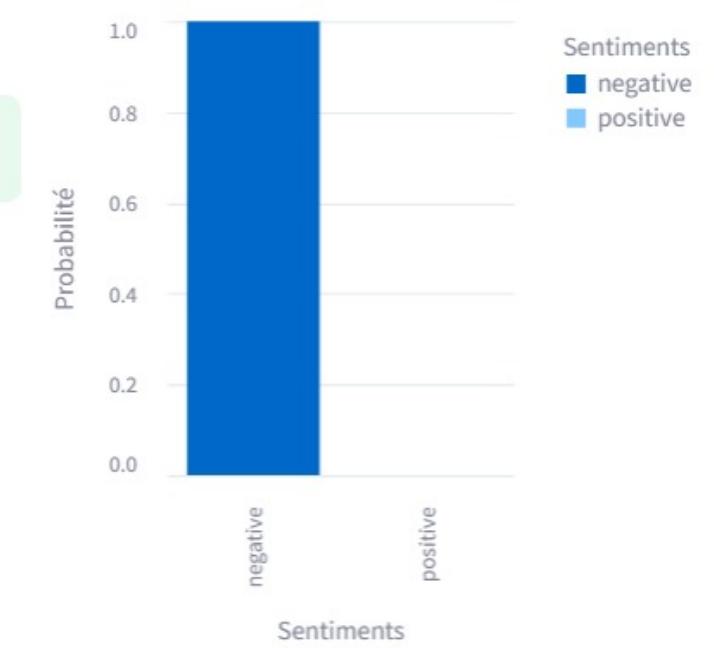
mauvaise chanson

Prediction

negative: -

Confidence: 100.00%

Probabilité de prédition



Déploiement :

DÉPLOIEMENT DE PRÉDICTION DES SENTIMENTS :

	A	B	C	D	
1	Texte				
2	J'adore ce film il est amusant				
3	J'aime cette musique elle est géniale				
4	C'est un mauvais jeu				
5	Le contrôle est difficile				

SENTIMENT

TÉLÉCHARGÉ

predictions (2).csv predictions (1).csv predictions.csv

< >

Sentiments dans un fichier

Uploader un fichier CSV

Drag and drop file here
Limit 200MB per file • CSV

Afficher tous les fichiers ...

Browse files

Sentiments dans un fichier :

Uploader un fichier CSV

Drag and drop file here
Limit 200MB per file • CSV

sentiments.csv 127.0B x

Browse files

Contenu du fichier CSV

	Texte	Prediction
0	J'adore ce film il est amusant	positive: +
1	J'aime cette musique elle est géniale	positive: +
2	C'est un mauvais jeu	negative: —
3	Le contrôle est difficile	negative: —

Télécharger les prédictions CSV

	A	B	C	D	E
1	Texte, Prediction				
2	J'adore ce film il est amusant, positive				
3	J'aime cette musique elle est géniale, positive				
4	C'est un mauvais jeu, negative				
5	Le contrôle est difficile, negative				

Déploiement :

DÉPLOIEMENT DE PRÉDICTION DES SARCASM :

Sarcasm dans texte

écrire ici

bonjour , tout va bien ?

Envoyer

Text original

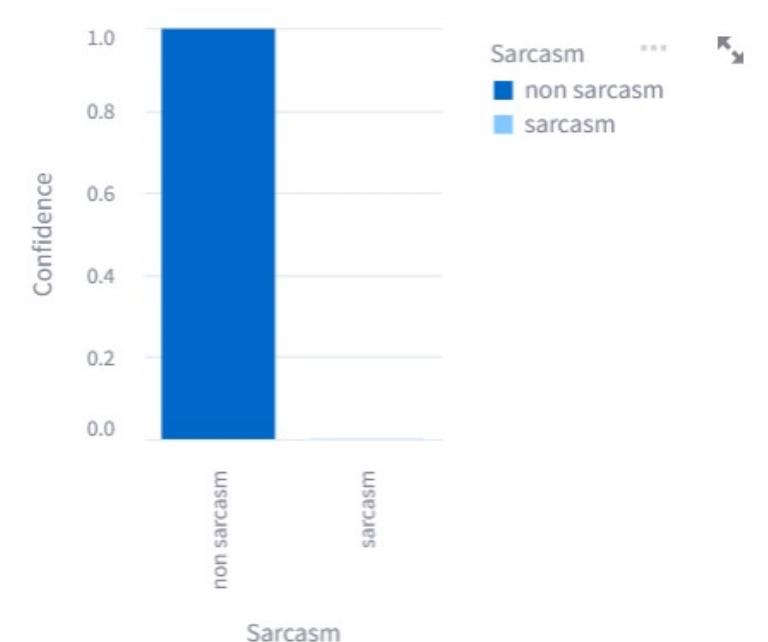
bonjour , tout va bien ?

Prediction

non sarcasm: ✗

Confidence: 99.86%

Probabilité de prédiction



Sarcasm dans texte

écrire ici

félicitations , tu viens de résoudre tous les problèmes du monde

Envoyer

Text original

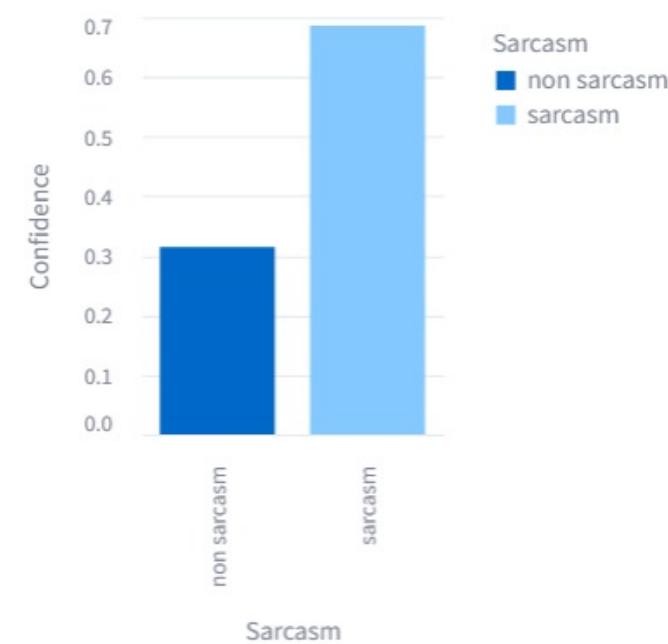
félicitations , tu viens de résoudre tous les problèmes du monde

Prediction

sarcasm: ✓

Confidence: 68.58%

Probabilité de prédiction



Déploiement :

DÉPLOIEMENT DE PRÉDICTION DES SARCASM :

Texte								
L'homme adulte aime tromper les enfants								
Bonjour								
un responsable de la Maison Blanche aurait déclaré que les tirs de masse constituaient un répit face au chaos								

Contenu du fichier CSV

	Texte	Prediction	Confidence
0	L'homme adulte aime tromper les enfants	non sarcasm: 	79.36%
1	Bonjour	non sarcasm: 	99.94%
2	un responsable de la Maison Blanche aurait déclaré que les tirs de masse constituaient un répit face au chaos	sarcasm:	98.70%

Conclusion