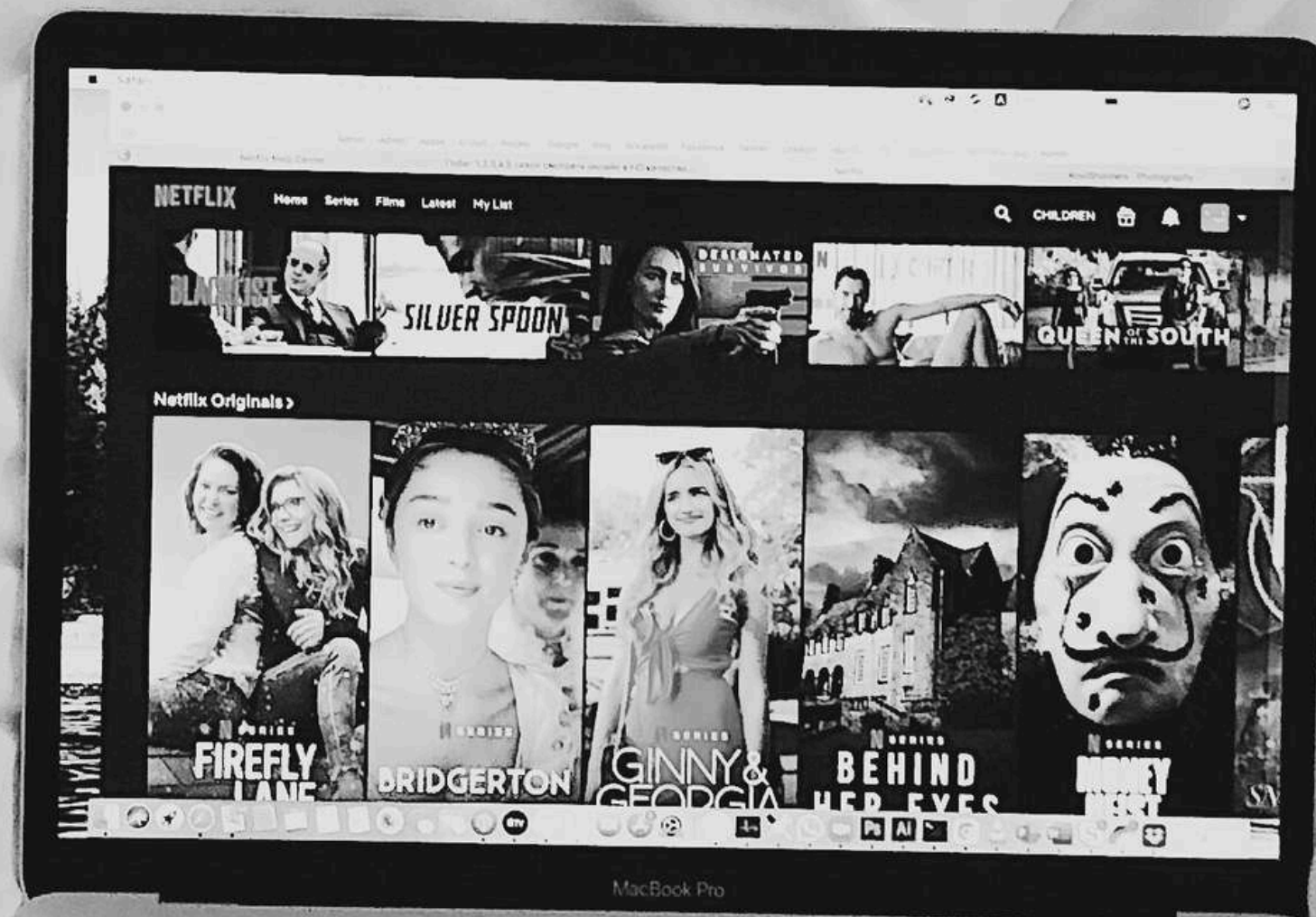


MACHINE/DEEP LEARNING

SENTIMENT ANALYSIS

LOU - VICKY - HLOIS



 **GIRLZZZ**
PRODUCTIONS



OBJECTIF

En tant que data scientist, ML engineer, vous devez travailler en équipe pour proposer un modèle de NLP qui répond aux besoins du projet. Pour cela, vous disposez de différentes sources de données pour chaque projet.

Projet: Movie Sentiment Analysis

KANBAN

Sentiment Analysis

Backlog

▼

Priority board

Team items

Roadmap

In review

My items

+ New view

Filter by keyword or by field

Discard

Save

Backlog

0 / 5

Estimate: 0

...

This item hasn't been started

+ Add item

Ready

0

Estimate: 0

...

This is ready to be picked up

+ Add item

In progress

4 / 3

Estimate: 0

...

This is actively being worked on

Draft

Présentation Canva

Draft

Déploiement Azure

Draft

Monitoring

Draft

Docker

+ Add item

In review

5 / 5

Estimate: 0

...

This item is in review

Draft

EDA

Draft

Modèle

Draft

Optimisation & amélioration

Draft

API

+ Add item

Done

This has been completed

+ Add item



DATA TRAIN

Train.tsv

colonnes : Phraseld , Sentenceld , Phrase , Sentiment

lignes : 156 060

nombre de terme avec des caractère identique qui ce répète + 2 = 7

	Phraseld	Sentenceld	Phrase	Sentiment
0	1	1	A series of escapades demonstrating the adage ...	1
1	2	1	A series of escapades demonstrating the adage ...	2
2	3	1	A series	2
3	4	1	A	2
4	5	1	series	2



DATA TEST

Test.tsv

colonnes : Phraseld , Sentenceld , Phrase

lignes : 66 292

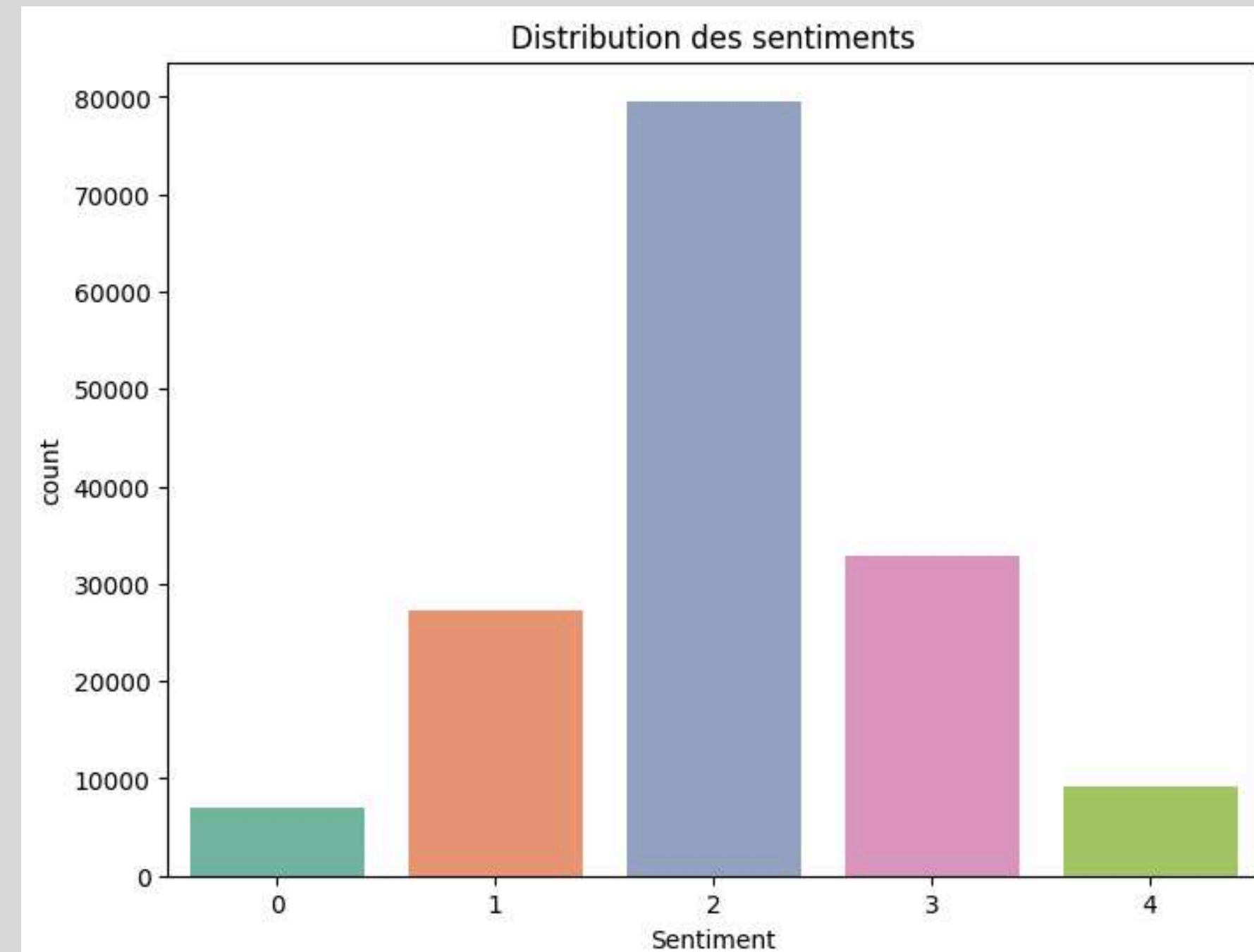
	Phraseld	Sentenceld	Phrase
0	156061	8545	An intermittently pleasing but mostly routine ...
1	156062	8545	An intermittently pleasing but mostly routine ...
2	156063	8545	An
3	156064	8545	intermittently pleasing but mostly routine effort
4	156065	8545	intermittently pleasing but mostly routine
...
66287	222348	11855	A long-winded , predictable scenario .
66288	222349	11855	A long-winded , predictable scenario
66289	222350	11855	A long-winded ,

REPARTITIONS DES SENTIMENTS

0	102.508088
1	103.569102
2	99.206549
3	108.055482
4	104.750601

SENTIMENTGROUP

Négatif	103.220625
Neutre	99.206549
Positif	106.884025



DISTRIBUTION DES SENTIMENTS :

- Les points rouges représentent des phrases associées à des sentiments négatifs, et ils sont concentrés principalement dans la zone inférieure de la polarité ($Y < 0$).
- Les points bleus symbolisent des phrases avec des sentiments positifs, majoritairement situés au-dessus de l'axe des ordonnées ($Y > 0$).
- Les points verts représentent les phrases avec des sentiments neutres, groupés autour de la polarité de 0.

RELATION ENTRE LA LONGUEUR DES PHRASES ET LA POLARITE :

- La longueur des phrases semble avoir un impact limité sur la polarité générale. Les points se concentrent de manière dense vers le centre (autour de la polarité neutre).
- Les phrases plus courtes (< 50 caractères) montrent plus de variance dans la polarité, tandis que les phrases plus longues (jusqu'à 300 caractères) tendent à avoir une polarité plus proche de zéro (neutre), avec une moindre dispersion.

DETECTION DES LANGUES

langue_detectee	
en	8314
fr	32
af	18
tl	16
da	15
cy	13
no	12
de	12
ro	11
it	11
ca	10
nl	9
et	7
pt	7
sv	6
es	5
so	5

sk	4
hu	4
id	4
sq	4
pl	3
hr	2
sw	2
cs	2
unknown	1

26 langues détectées

Suppression des langues autres que l'Anglais

MOTS LES PLUS FREQUENTS

Film	1265
Movie	1118
One	569
Like	480
Story	381
Make	363
Character	342
Time	298
Comedy	271
Good	266
Much	262





MOTS LES MOINS FREQUENTS

chortle

avuncular

herrmann

laughingly

goldie

servicable

eastvs

ganginfested

underworld

smallbudget

Nuage de mots - Sentiments très positifs



Nuage de mots - Sentiments très négatifs



Nuage de mots - Sentiments très négatifs



POLARITY

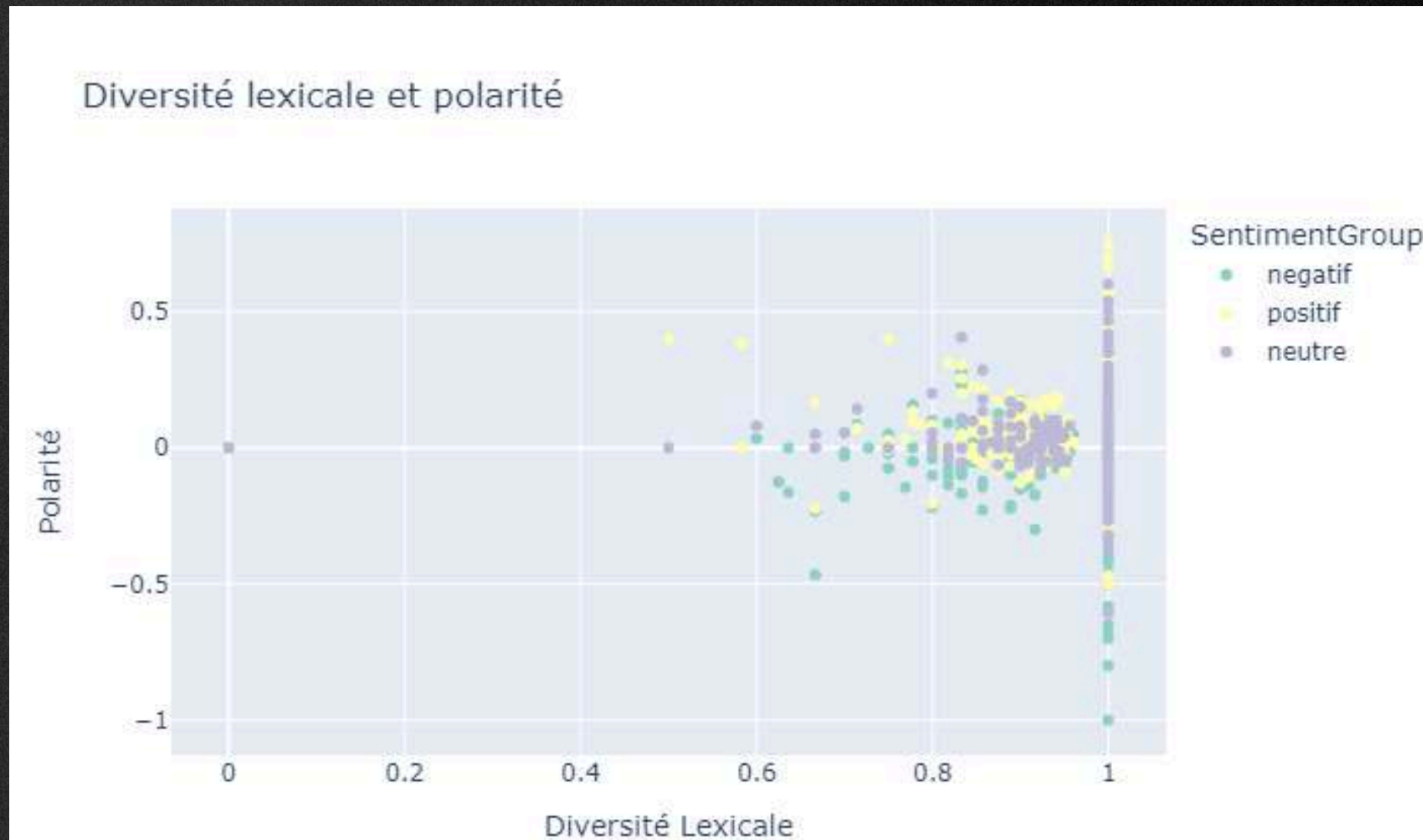
Negatif	-0.014353
Neutre	0.021938
Positif	0.064755

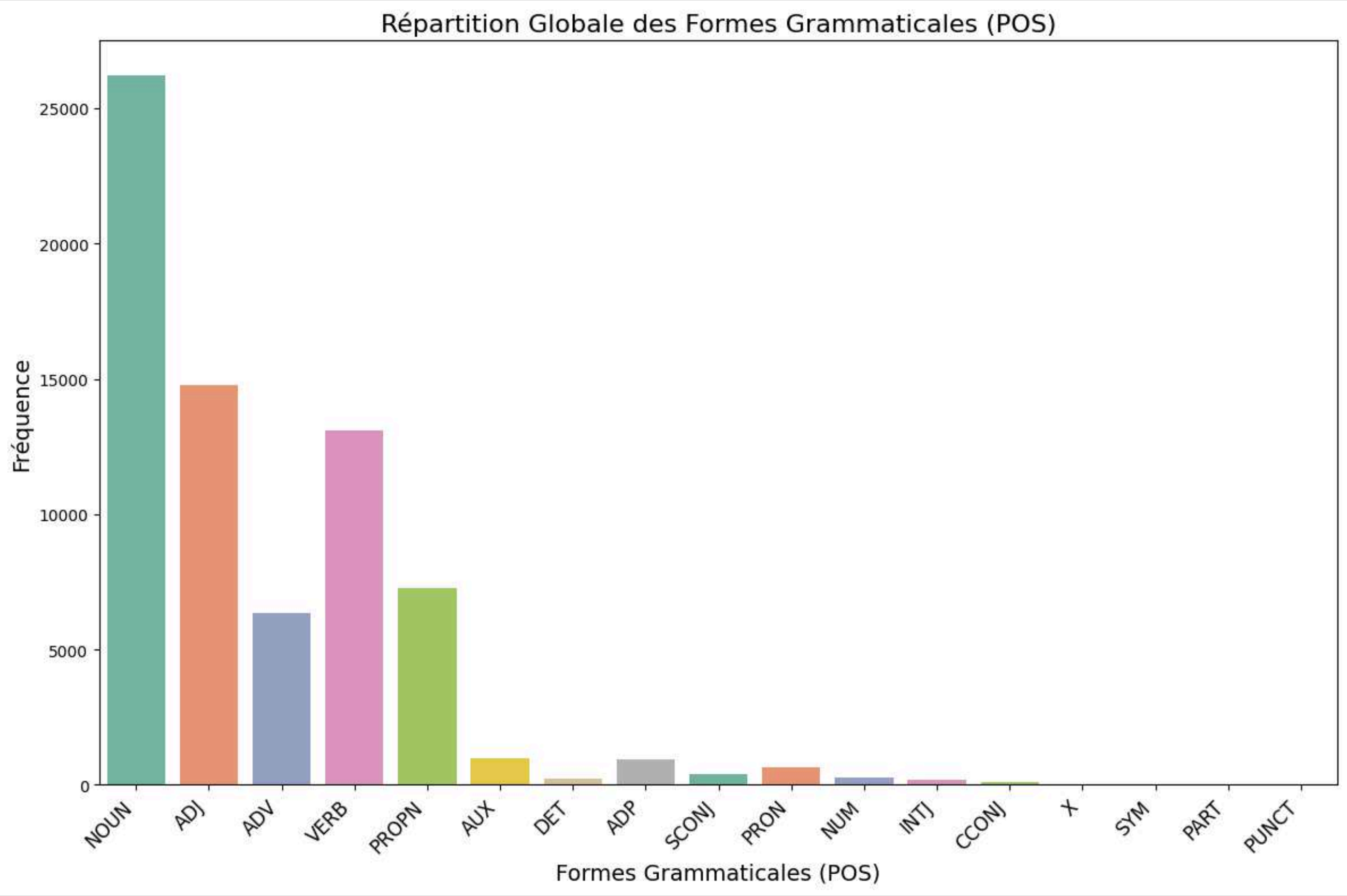
La polarité dans un texte fait référence à l'orientation émotionnelle ou au sentiment exprimé. Elle peut être positive (sentiment favorable, joyeux), négative (sentiment défavorable, triste) ou neutre (absence de sentiment fort). La polarité est souvent utilisée dans l'analyse de sentiments pour classifier les opinions ou émotions dans des textes tels que des avis, commentaires, ou critiques.



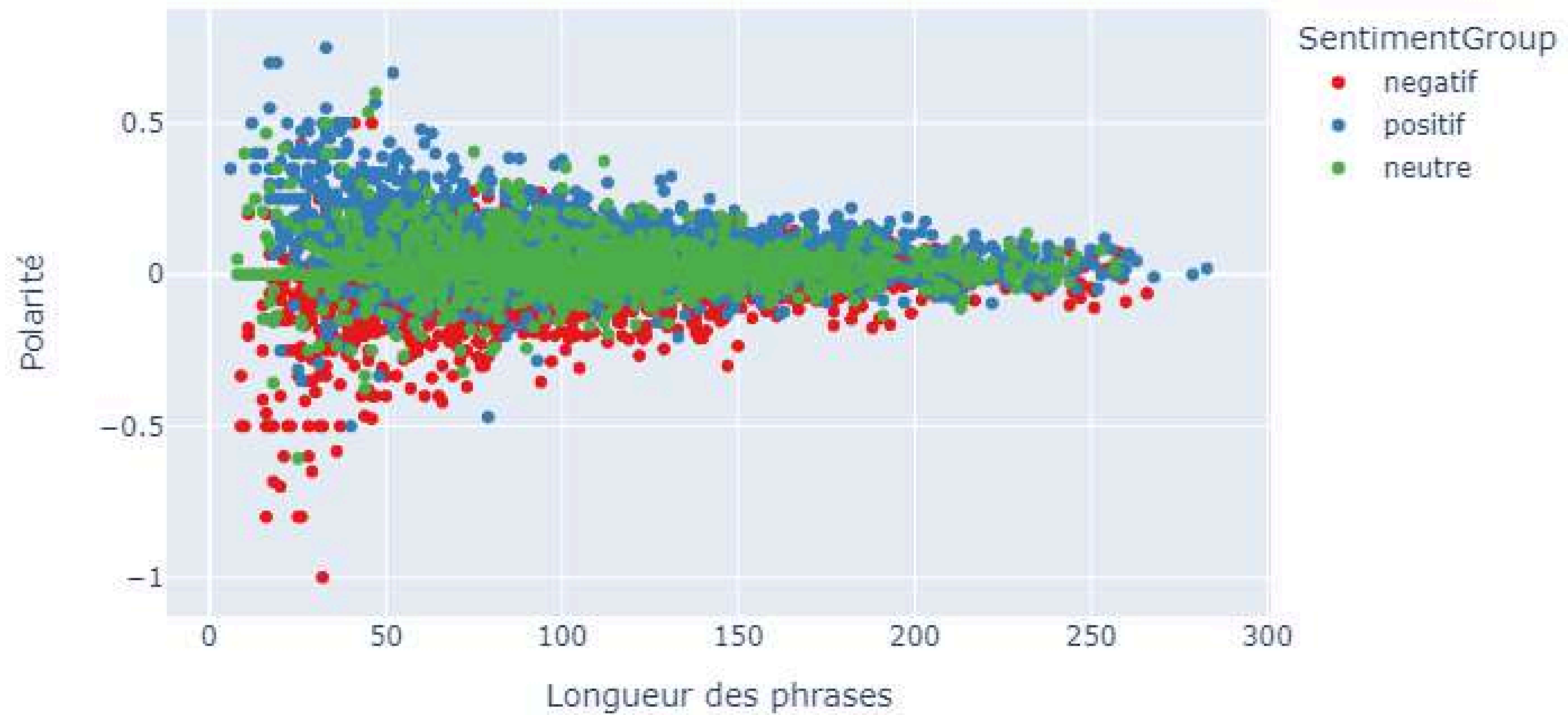
DIVERSITE LEXICALE

- La majorité des phrases ont une diversité lexicale élevée (proche de 1).
- Peu de phrases présentent une faible diversité lexicale (moins de 0.5).
- Aucune corrélation forte entre la diversité lexicale et la polarité des sentiments.
- Les phrases avec une grande diversité lexicale couvrent toute la gamme de polarités (négatif à positif).
- Les phrases avec une faible diversité montrent une polarité plutôt négative ou légèrement positive, mais ces cas sont rares.
- Les sentiments positifs, négatifs et neutres sont concentrés autour de la diversité élevée, sans schéma clair liant diversité et polarité.



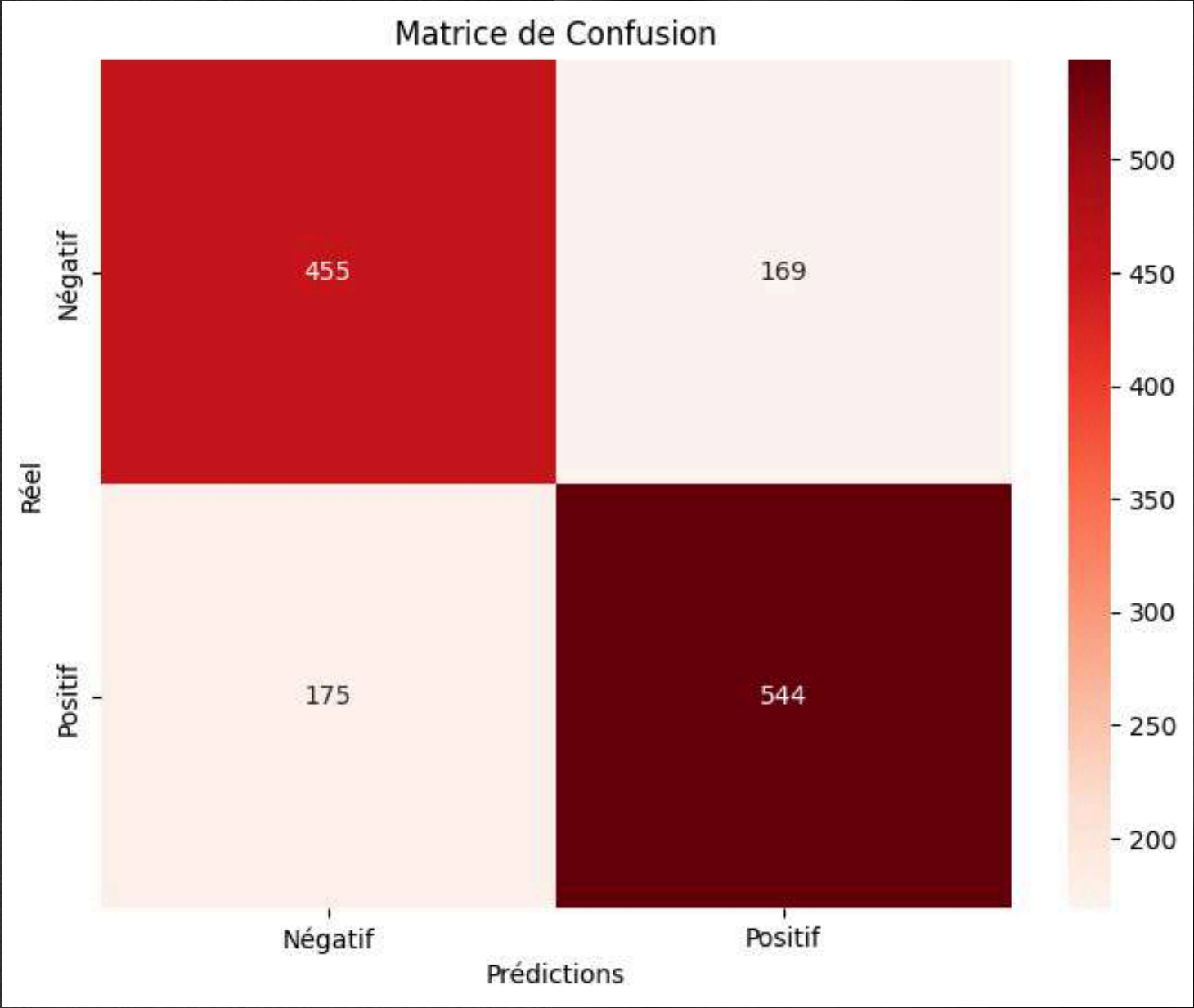


Relation entre la longueur des phrases et la polarité



BERNOULLI

	precision	recall	f1-score	support
Négatif	0.72	0.73	0.73	624
Positif	0.76	0.76	0.76	719
accuracy			0.74	1343
macro avg	0.74	0.74	0.74	1343
weighted avg	0.74	0.74	0.74	1343



MONITORING MLFLOW

Parameters (5)

<div><div></div><div>Search parameters</div></div>	
Parameter	Value
vectorizer__ngram_range	(1, 1)
vectorizer__min_df	1
vectorizer__max_df	0.5
classifier__alpha	1
vectorizer__max_features	None

Metrics (8)

<div><div></div><div>Search metrics</div></div>	
Metric	Value
recall_positif	0.7566063977746871
accuracy	0.7438570364854803
best_score_cv	0.4600490106750943
precision_negatif	0.7222222222222222
f1_score_positif	0.7597765363128491
precision_positif	0.7629733520336606
recall_negatif	0.7291666666666666
f1_score_negatif	0.7256778309409888

TEST UNITAIRE

```
===== test session starts =====
platform linux -- Python 3.10.6, pytest-8.3.3, pluggy-1.5.0
rootdir: /home/louve/code/wolfwithwanderlust/TP/Arturo/Sentiment Analysis/Sentiment-Analysis
plugins: anyio-4.4.0
collected 3 items

test_model.py ... [100%]

===== 3 passed in 0.88s =====
```


DEPLOIEMENT - CI/CD

Summary

Jobs

✓ build

✓ deploy

Run details

Usage

Workflow file

Triggered via push 1 hour ago

Status

Total duration

Artifacts

Anne68 pushed

→ f35d3e6

master

Success

18m 3s

1

master_streamlitsentiment.yml

on: push

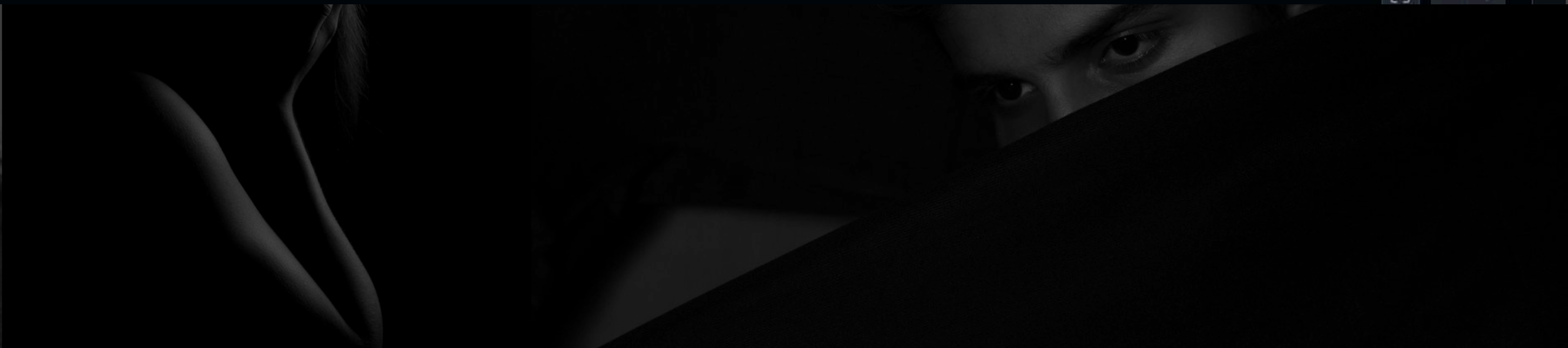
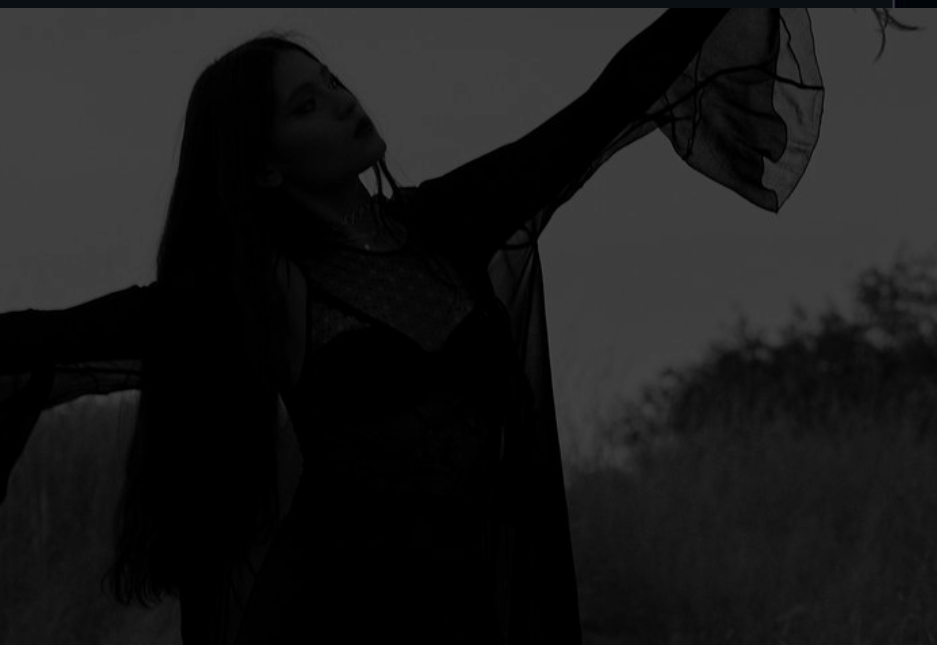
✓ build

2m 35s

✓ deploy

15m 5s

<https://streamlitsentiment-e9dpc6graxehe...>



FASTAPI

POST [/predict-sentiment/](#) Predict Sentiment



POST [/clean-csv/](#) Clean Csv



POST [/predict-from-cleaned/](#) Predict From Cleaned



POST [/predict-text/](#) Predict Text



DEMO

Fichier CSV/TSV Commentaire texte Historique

 Upload un fichier TSV 

Choisir un fichier TSV



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

Browse files

STREAMLIT