

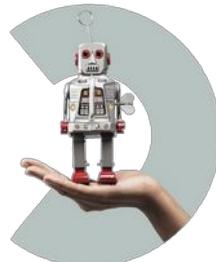


Réalisez une analyse de sentiments grâce au Deep Learning

Lucie Belleguic

AI Engineer

OPENCLASSROOMS





Mission : contexte et objectif



Modélisation et résultats



Mise en production et suivi

Mission : contexte et objectif

- ▶ Prédire le sentiment associé à un tweet



- ▶ Mettre en place une démarche MLOps

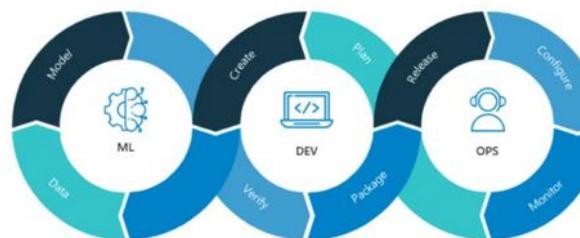
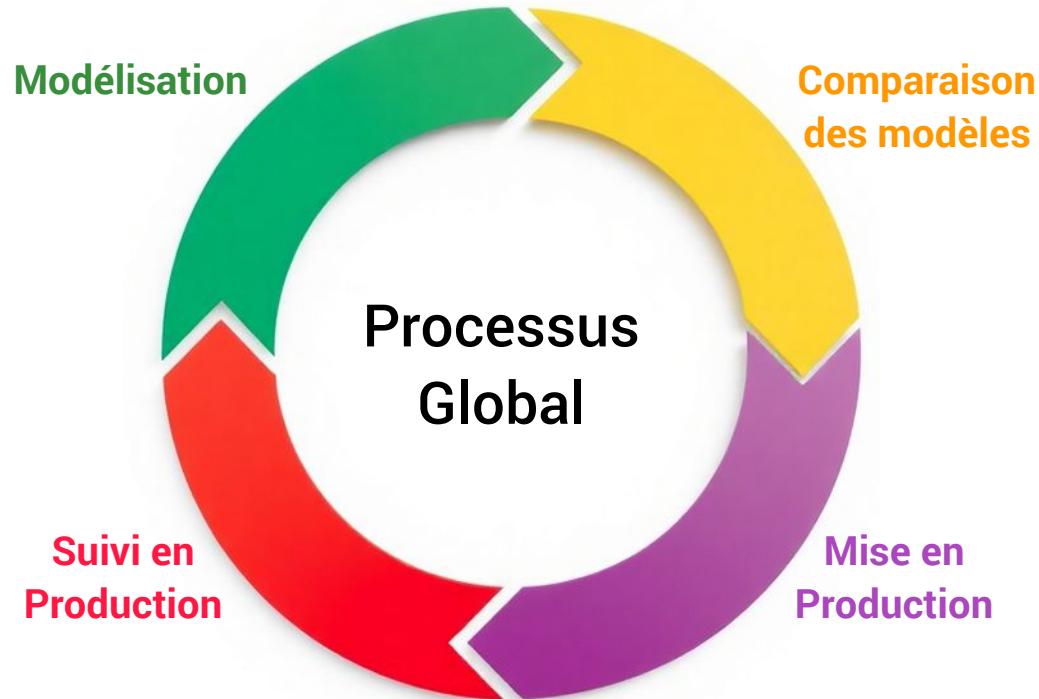


Illustration © Nutria

Mission : contexte et objectif

▶ Approche méthodologique

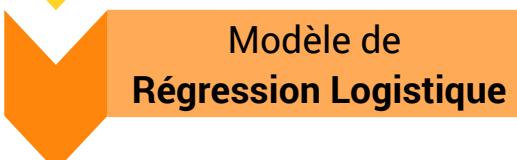
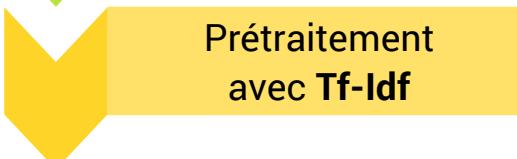
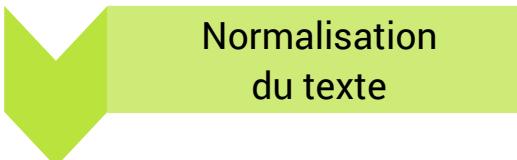


Modélisation et résultats

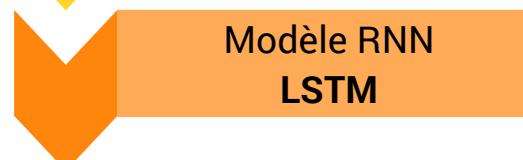
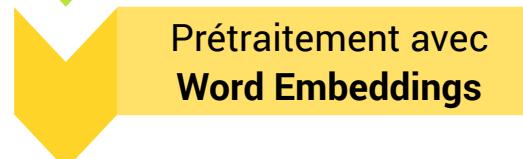
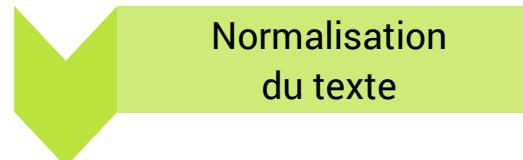
► 3 approches pour

Prédire le sentiment associé à un tweet

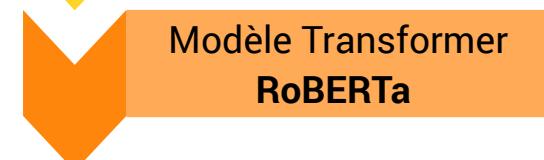
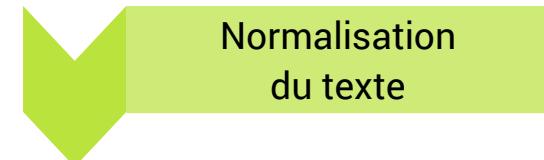
Modèle sur-mesure
Simple



Modèle sur-mesure
Avancé



Modèle avancé
RoBERTa



Modélisation et résultats

▶ Normalisation du texte

Modèle RoBERTA

- Conversion des smileys en mot-clés

Modèle
sur-mesure
Avancé

- Suppression des URL et mentions @nom
- Texte en minuscule

Modèle
sur-mesure
Simple

- Lemmatisation
- Suppression des stop words
- Suppression de la ponctuation



~~@justinstevens~~
~~https://www.adresse.com~~

Product → product

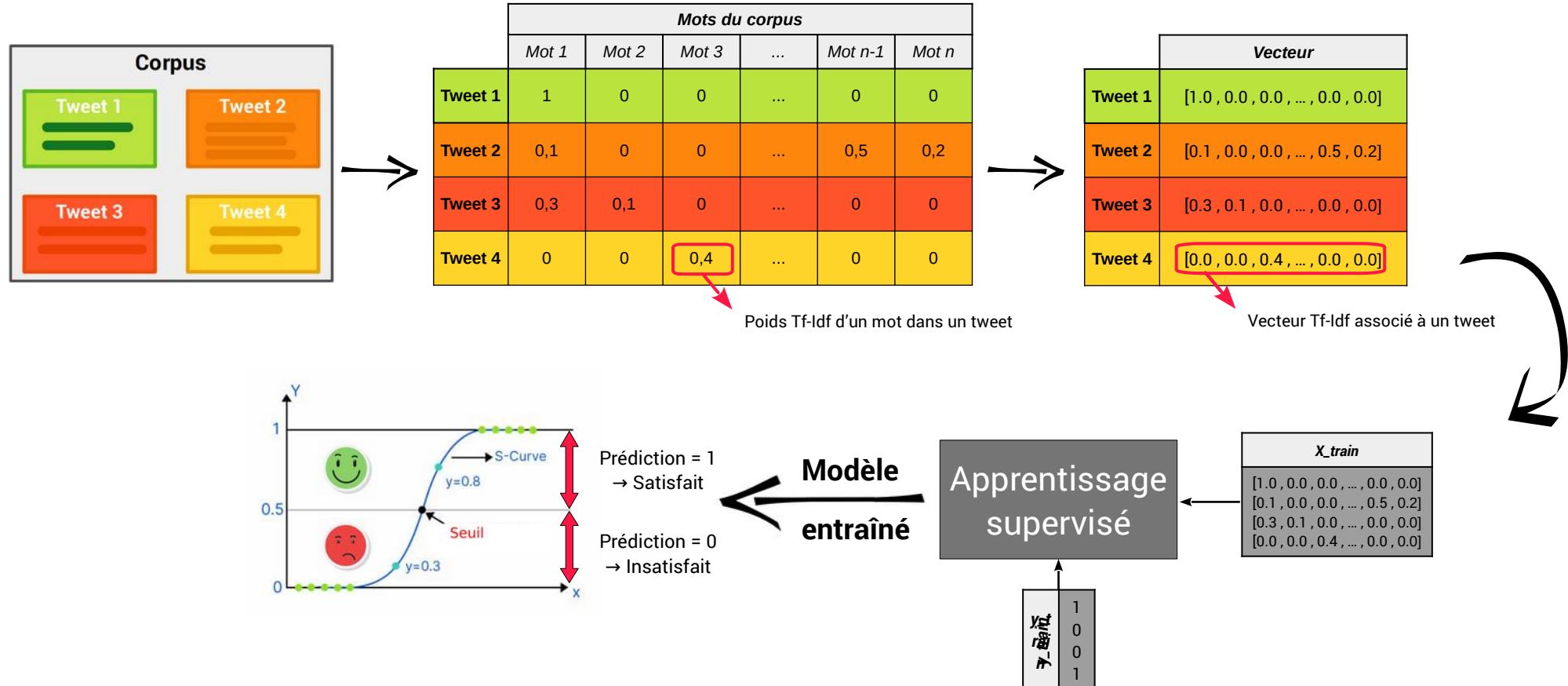
running
ran → run

~~and / the / is~~

~~... ? # ; :~~

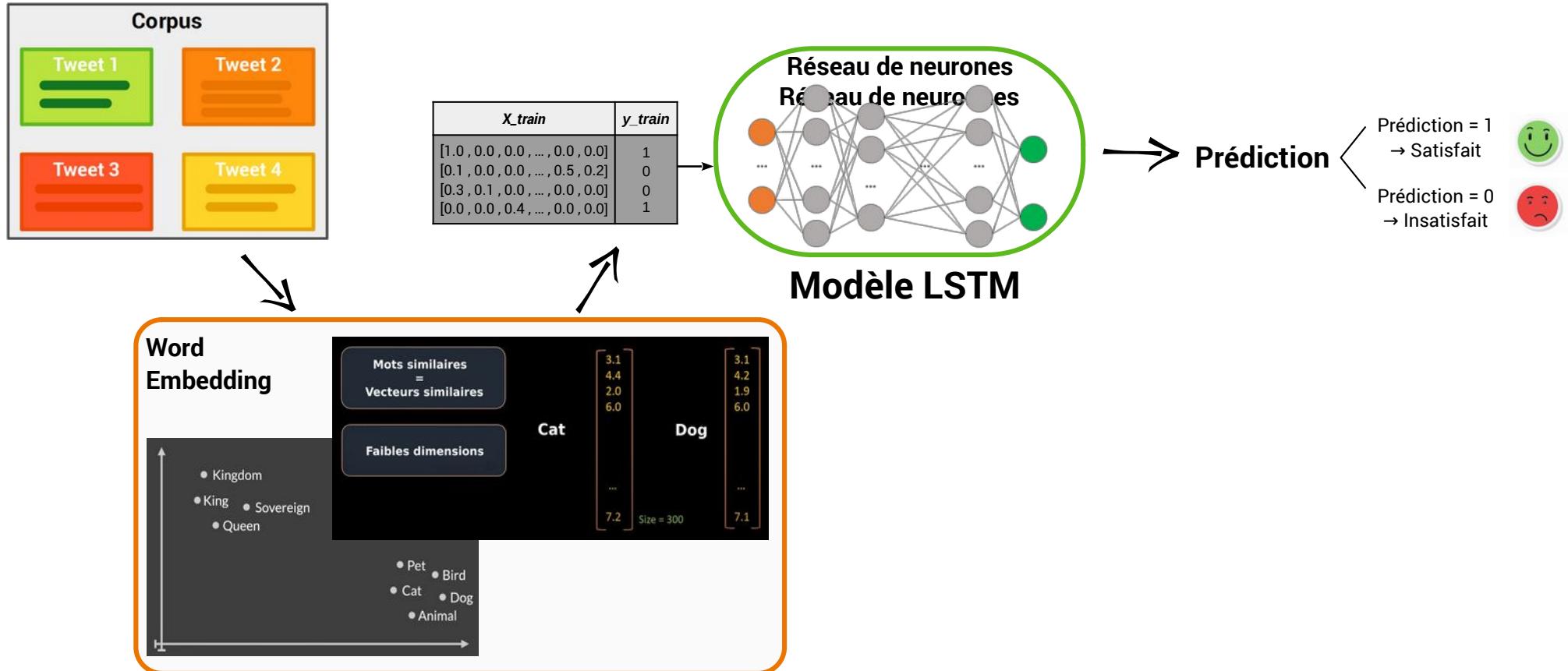
Modélisation et résultats

▶ Modèle sur-mesure simple



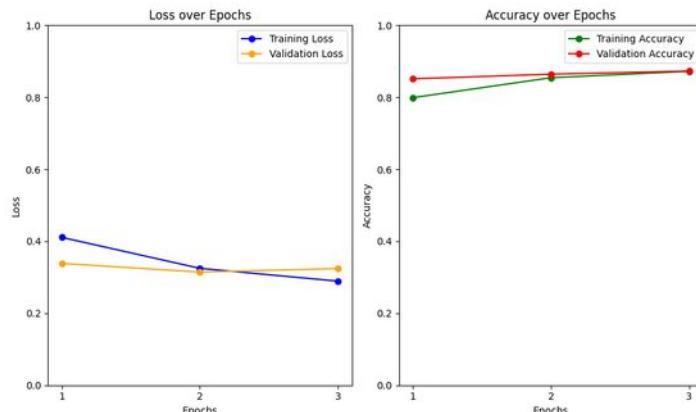
Modélisation et résultats

▶ Modèle sur-mesure avancé



Modélisation et résultats

▶ Modèle avancé RoBERTa



Tokenisation

Texte brut
"RoBERTa is a transformer-based model."

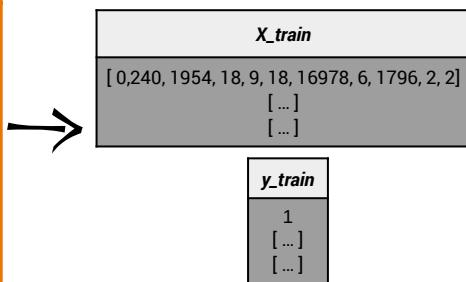
Tokens :

[ro', 'bert', 'a', 'is', 'a', 'transformer', '-', 'based', 'model', '']

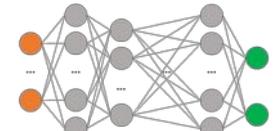
Identifiants uniques correspondant au vocabulaire du modèle

Input IDs : [0,240, 1954, 18, 9, 18, 16978, 6, 1796, 2, 2]

Attention Mask : [1, 1, 1, 1, 1, 1, 1, 1, 1, 1]



 HUGGING FACE
Transformer RoBERTa
Pré-entraîné



Prédiction = 1
→ Satisfait



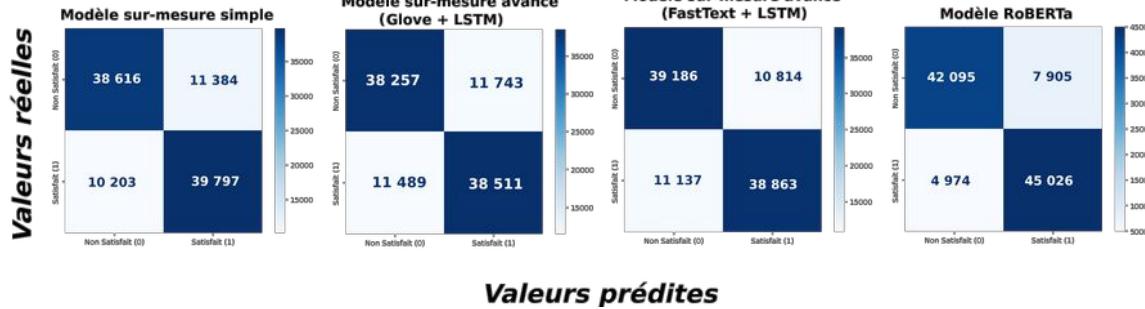
Prédiction = 0
→ Insatisfait

Prédiction

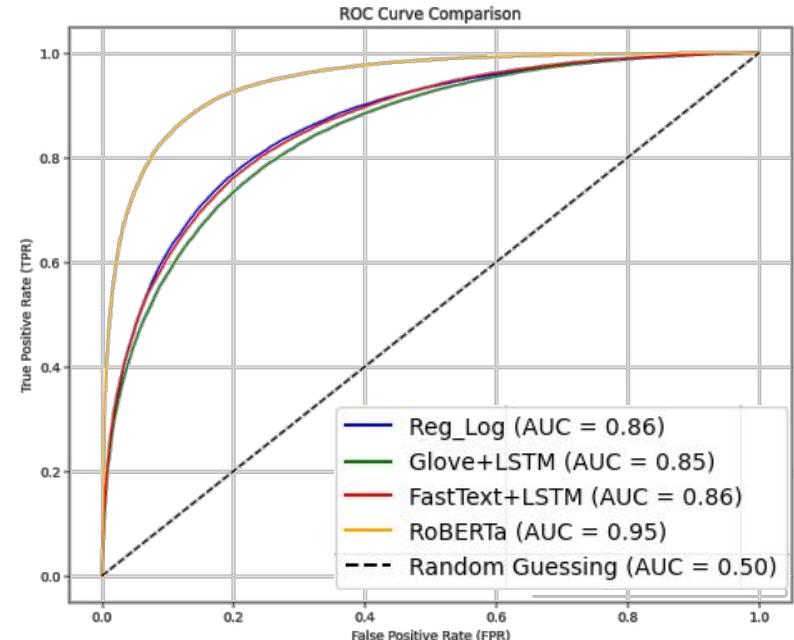


Modélisation et résultats

Résultats comparés des modèles



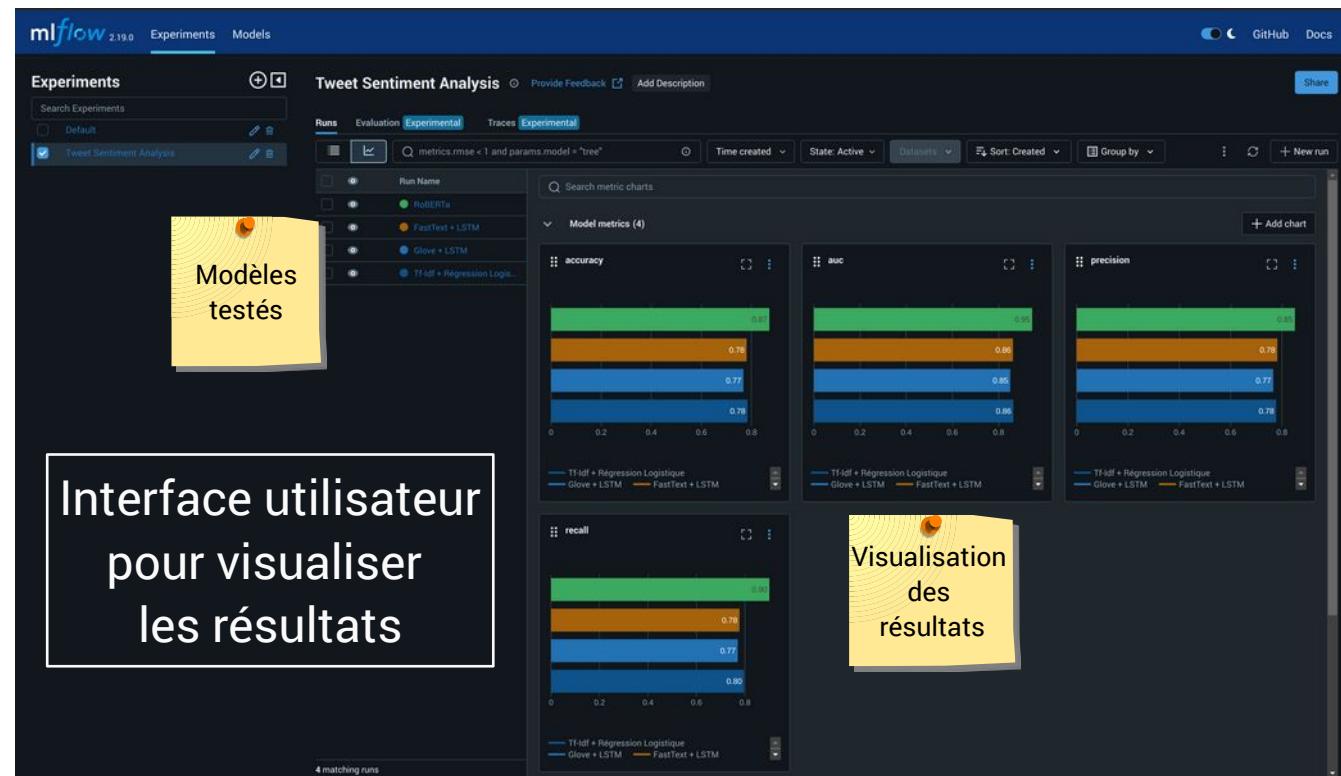
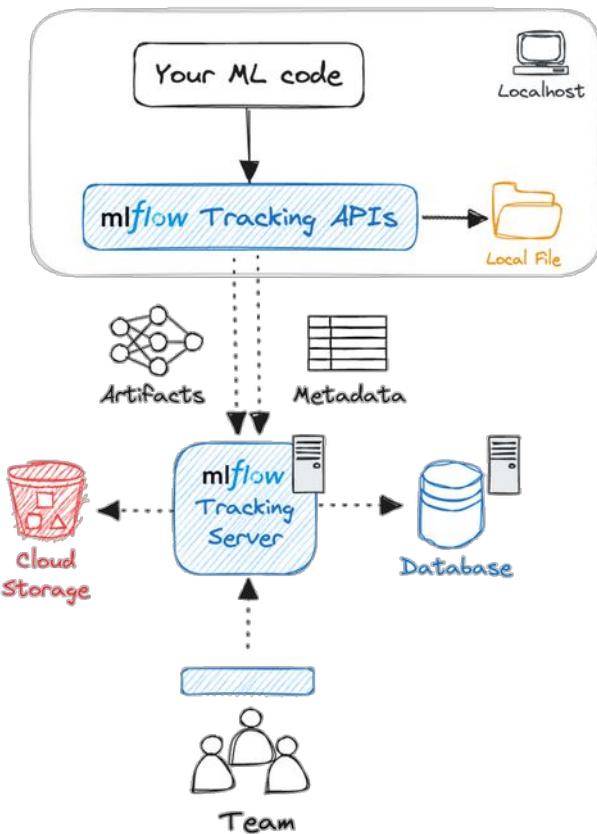
Modèle	Avantages	Inconvénients
Sur-mesure simple	Simple, rapide	Moins contextuel
Sur-mesure avancé (word embeddings Glove)	Embeddings, contextuel	Lent, moins performant
Sur-mesure avancé (word embeddings FastText)	Performant, contextuel	Lent, complexe
Avancé RoBERTa	Performances, précis	Très lent, coûteux



Modèle choisi pour
le déploiement

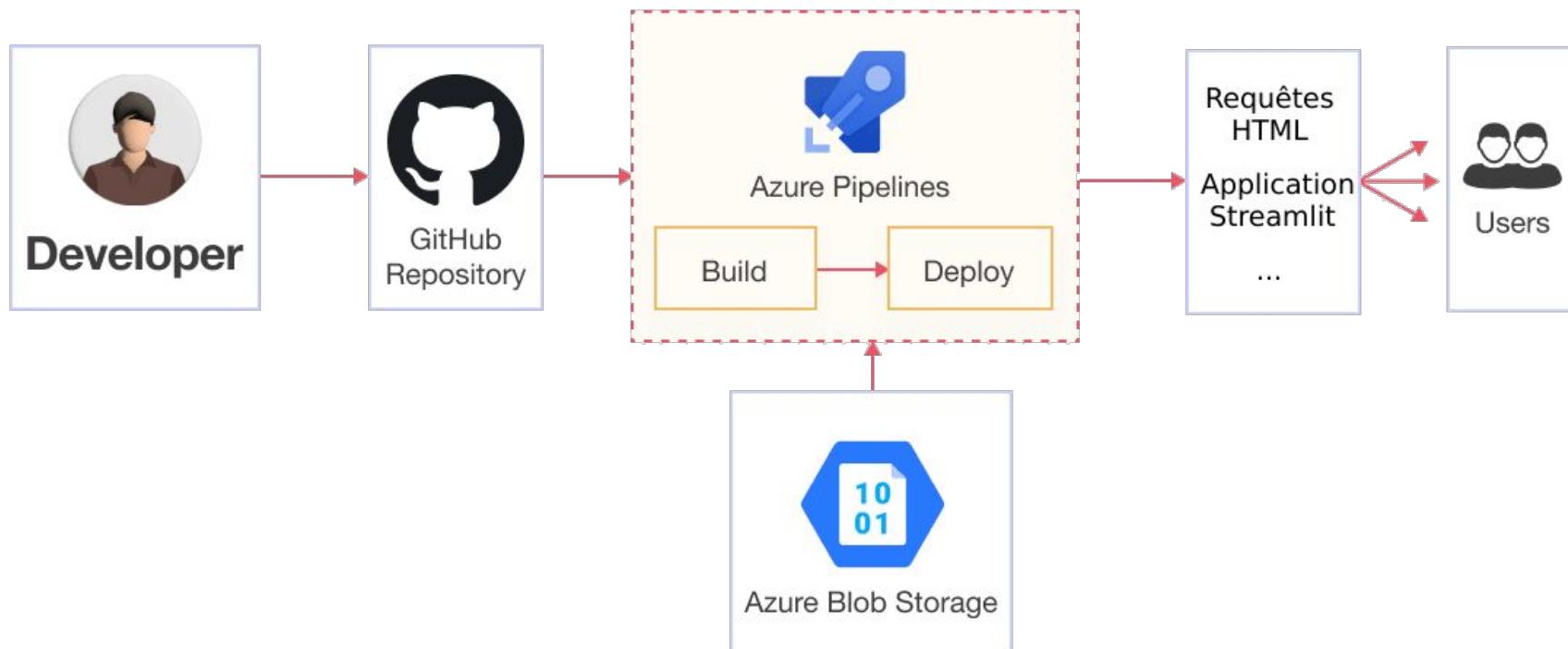
Modélisation et résultats

▶ Expérimentation avec MLflow



Mise en production et suivi

▶ Pipeline de déploiement continu

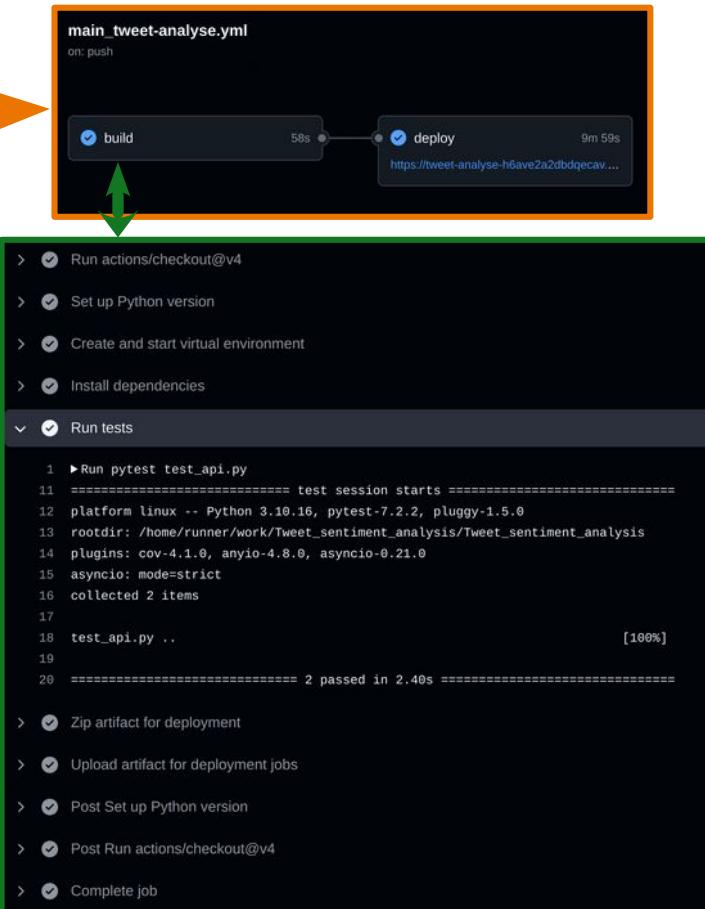


Mise en production et suivi

▶ Pipeline de déploiement continu

The screenshot shows a GitHub repository named "Tweet_sentiment_analysis" under the owner "La-Luciole". The "Actions" tab is highlighted with an orange box and arrow. The repository has 1 branch and 0 tags. The main branch has 18 commits. The commit history includes:

- Modification README by La-Luciole (2 days ago)
- Update main_tweet-analyse.yml by .github/workflows (last week)
- Ajout fichiers de modélisations et interface de test de l'API by modelisation (2 days ago)
- Réduction de la taille des embeddings FastText by .gitignore (last week)
- Modification README by README.md (2 days ago)
- Ajout fichiers de modélisations et interface de test de l'API by interface_test_API.py (2 days ago)
- Gestion des erreurs by main.py (last week)
- Gestion des erreurs by preprocess.py (last week)
- Ajout fichiers de modélisations et interface de test de l'API by requirements.txt (2 days ago)
- Modification des tests unitaires by test_api.py (5 days ago)



Mise en production et suivi

▶ Démonstration du moteur d'inférence

```
(testAPI_env) la-luciole@AER016:~/test_API$ streamlit run test_API.py  
You can now view your Streamlit app in your browser.  
Local URL: http://localhost:8501  
Network URL: http://192.168.1.34:8501
```



The screenshot shows a Streamlit application window titled "Test de l'API tweet-sentiment". At the top, it says "localhost:8501". On the right side, there are "Deploy" and three-dot menu icons. The main content area has a dark background with white text. It displays the instruction "Veuillez saisir un tweet en anglais pour obtenir une prédiction :" above a large, empty text input field.

Mise en production et suivi

▶ Suivi des performances du modèle

Test de l'API tweet-sentiment

Veuillez saisir un tweet en anglais pour obtenir une prédiction :

I don't know if I should go to the event tonight or stay home. I guess I'll decide later.

Sentiment prédit : satisfait (Probabilité : 0.52)

Tweet saisi :

I don't know if I should go to the event tonight or stay home. I guess I'll decide later.

Validez-vous la réponse ?

- Oui
 Non

Valider

Votre réponse a été enregistrée comme NON.

Nouvelle prédiction



Application
Insights

Microsoft Azure

Accueil > test_API

test_API | Journaux
Application Insights

Réponse invalide + Essayer le nouveau Log Analytics Commentaires

test_API Sélectionner une étendue Exécuter Intervalle de temps : Dernière heure Enregistrer Partager Nouvelle règle d'alerte Exporter

```
1 traces
2 | where message contains "Utilisateur a validé la réponse comme NON"
3 | summarize count() by bin(timestamp, 1s), message
4 | order by timestamp desc
5 | project timestamp, message
6
```

Résultats Graphique

timestamp [UTC]	message
> 20/01/2025 14:36:44.000	Utilisateur a validé la réponse comme NON. Tweet : I was thinking about trying out a new restaurant this weekend, but I'm not sure if it's worth it., prédiction : insatisfait, probabilité : 0.48
> 20/01/2025 14:36:15.000	Utilisateur a validé la réponse comme NON. Tweet : It's been raining all day. I guess it's good for the plants, though., prédiction : insatisfait, probabilité : 0.29
> 20/01/2025 14:34:20.000	Utilisateur a validé la réponse comme NON. Tweet : I don't know if I should go to the event tonight or stay home. I guess I'll decide later., prédiction : satisfait, probabilité : 0.52

Schema

Mise en production et suivi



Suivi des performances du modèle



Application
Insights

The screenshot shows the Azure Application Insights dashboard. At the top, it displays the total number of alerts: 1 Critique, 0 Erreur, 0 Avertissement, 1 Information, and 0 Commentaires. Below this is a table with columns: Nom (sorted by name), Gravité (severity), Ressource affectée (affected resource), Condition d'alerte (alert condition), Réponse de l'utilisateur (user response), and Temps d'incendie (incendie time). One alert is listed: 'Alerte si au moins 3 erreurs en 5min' for the 'test_api' resource, triggered by 'Déclenché' (triggered) and marked as 'Nouveau' (new) at '20/01/2025 15:36'.

Mail d'alerte envoyé
si l'utilisateur identifie
3 prédictions erronées
en moins de 5 minutes !

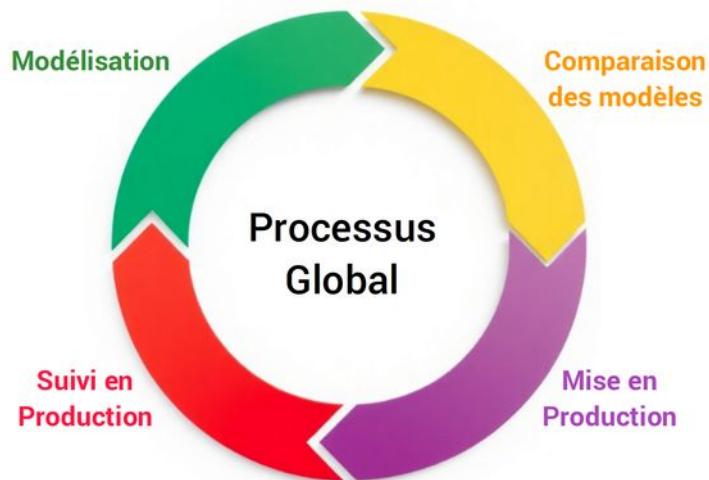


The screenshot shows an email from Microsoft Azure to the user. The subject line is 'Fired:Sev3 Azure Monitor Alert Alerte si au moins 3 erreurs en 5min on test_api (microsoft.insights/components) at 1/20/2025 2:36:17 PM'. The email contains two buttons: 'View the alert in Azure Monitor >' and 'Investigate >'. Below the buttons is a 'Summary' section with the following details:

Alert name	Alerte si au moins 3 erreurs en 5min
Severity	Sev3
Monitor condition	Fired
Affected resource	test_api
Resource type	microsoft.insights/components
Resource group	p7_analyse_de_sentiments
Description	Une alerte est envoyée par mail si au moins 3 prédictions sont identifiées comme erronées en moins de 5 min.
Monitoring service	Platform
Signal type	Metric
Fired time	January 20, 2025 14:36 UTC
Alert ID	700e2169-9c95-45bf-9dc8-1f4f791bf000
Alert rule ID	https://portal.azure.com/#resource/subscriptions/c84c520c-008d-4702-bbc1-401f140d37d2/resourceGroups/p7_Analyse_de_sentiments/providers/microsoft.insights/metricAlerts/Alerte si au moins 3 erreurs en 5min

Mise en production et suivi

► Synthèse et amélioration



- Un modèle performant
- Une automatisation complète
- Possibilité d'extension vers d'autres langues, d'autres plateformes