



Air Paradise

Analyse de sentiments des tweets

Tristan Hérou - MIC



Analyse de sentiments



Sommaire

Analyse des données

MLOPS

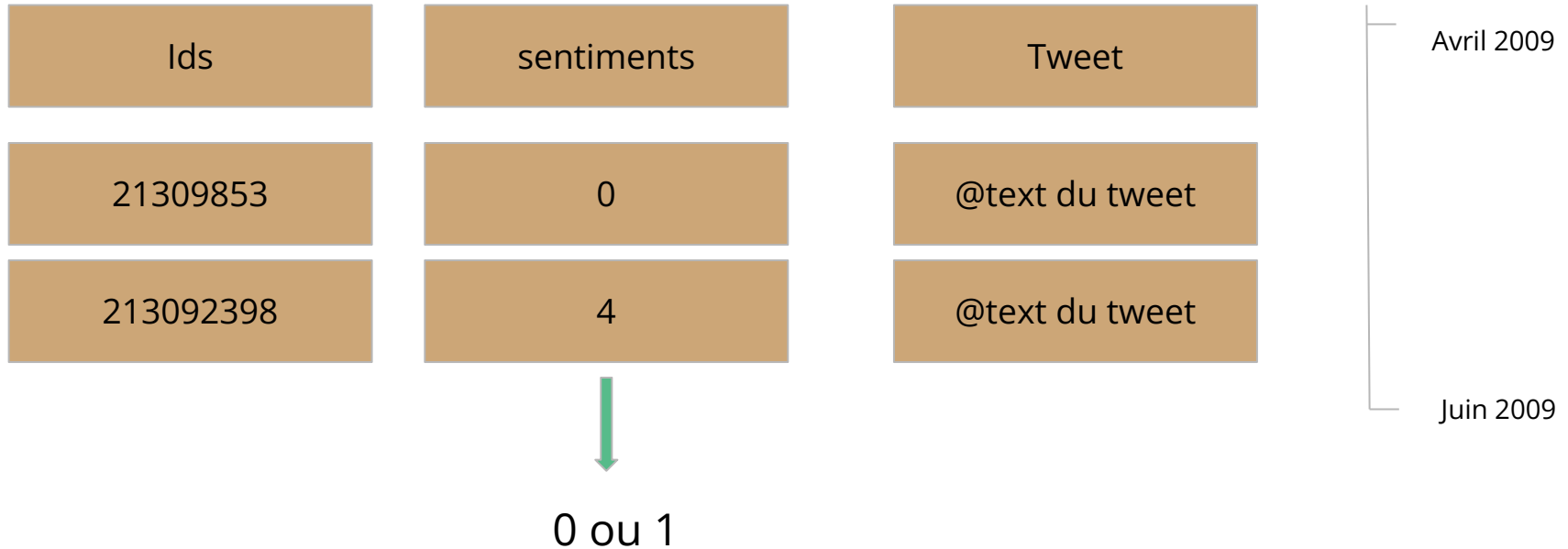
Entrainements

API

Monitoring

Tweets

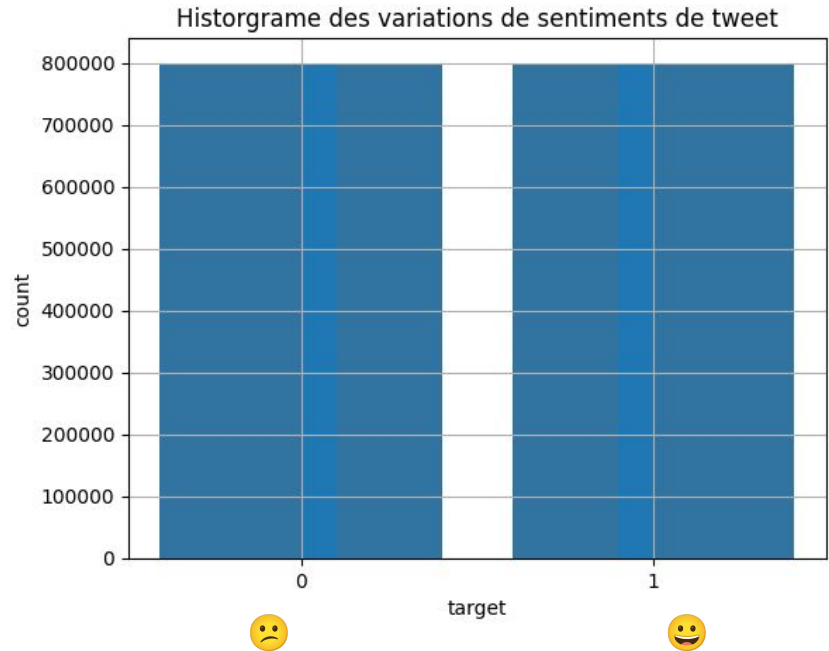
Base de données csv



Tweets

count	1.6e+06
mean	7.4e+01
std	3.6e+01
min	6.0e+00
25%	4.4e+01
50%	6.9e+01
75%	1.0e+02
max	3.7e+02

Dataset équilibré



Tweets

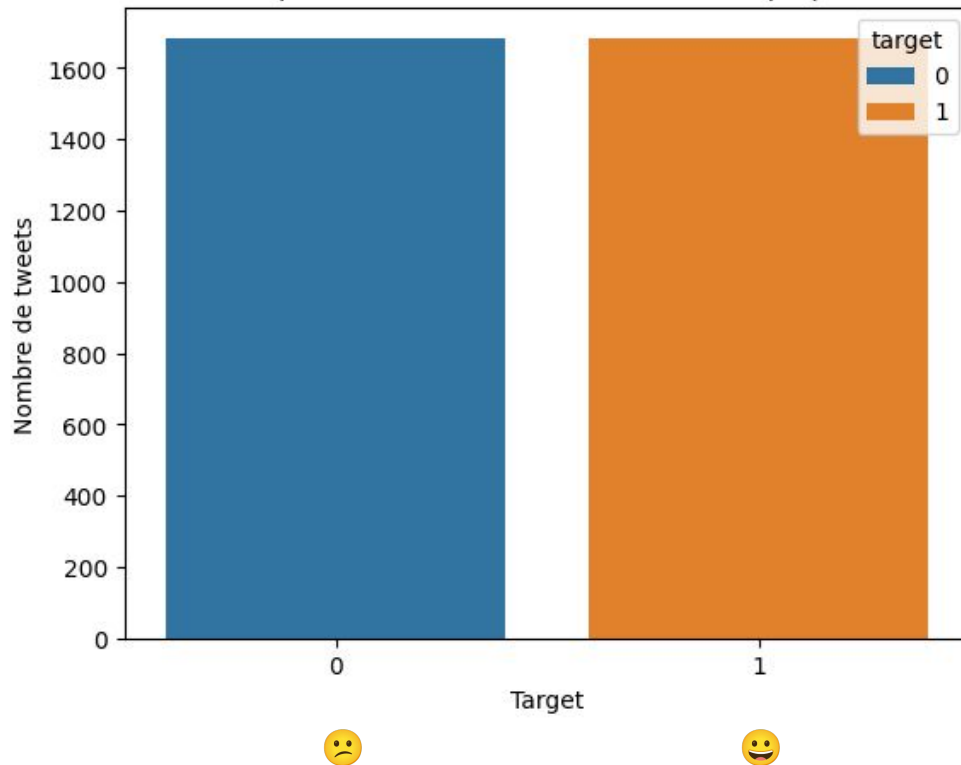
Duplications
1685



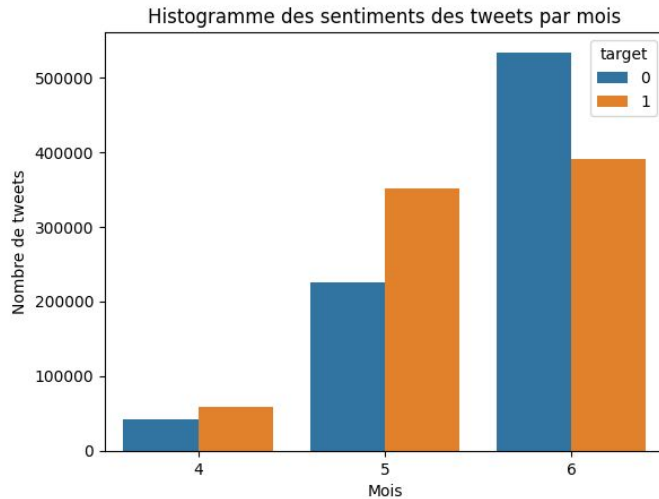
Exemple:

"Haven't tweeted nearly all day Posted
my website tonight, hopefully that goes
well Night time!"

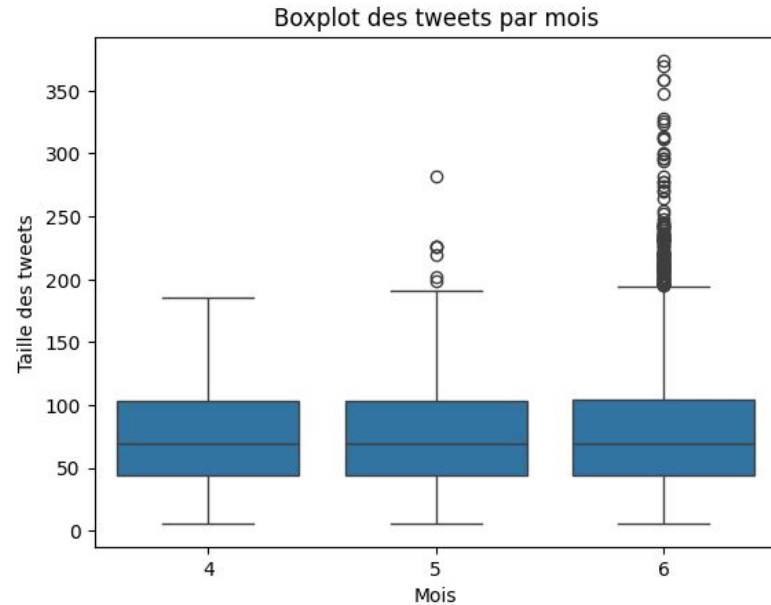
Boxplot des sentiments des tweets dupliques



Tweets

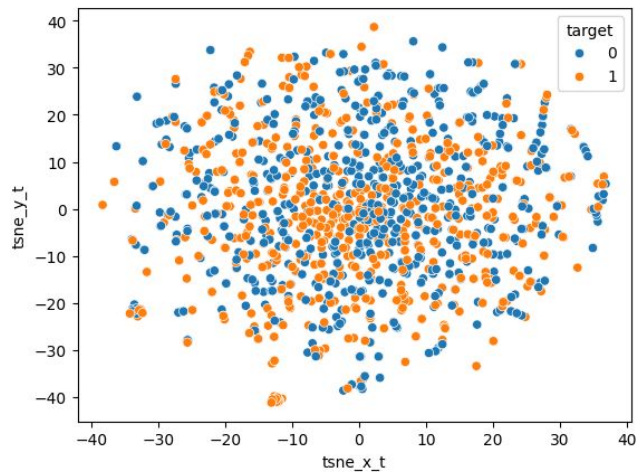


Boxplot de la longueur des tweets par mois



Boxplot de la longueur des tweets par mois

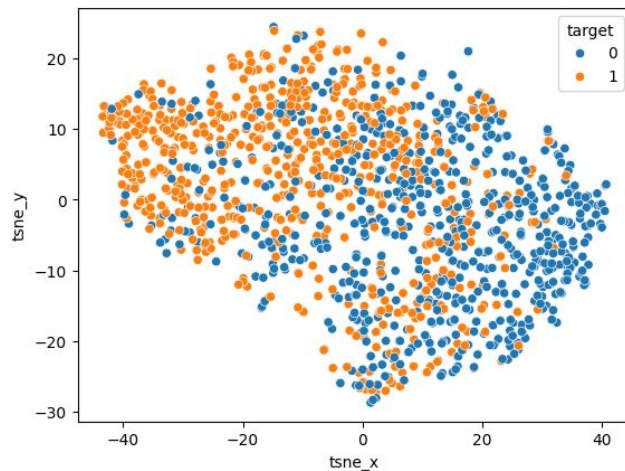
Tweets



Analyse de features d'embedding tfidf par tsne



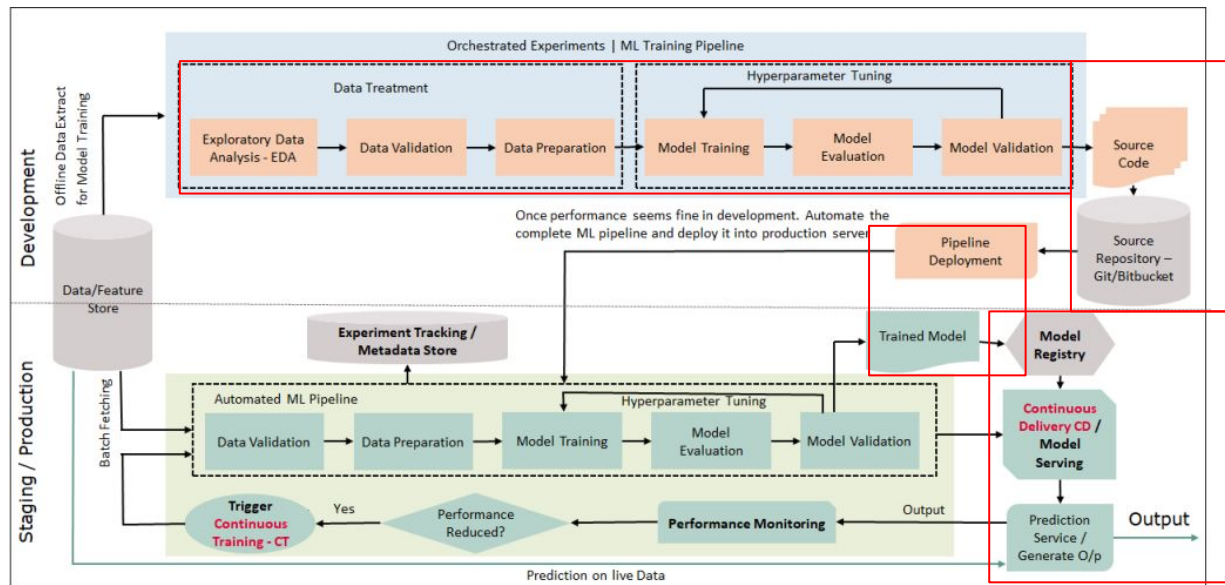
Données non séparables



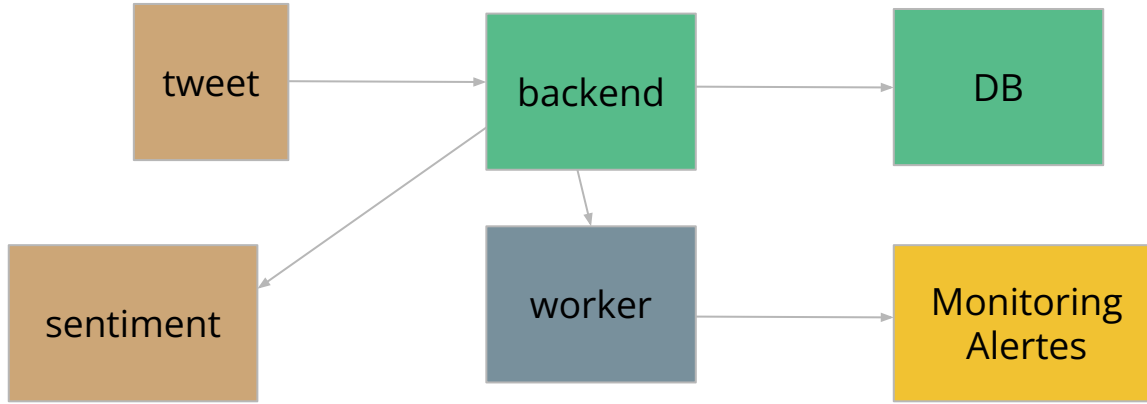
Analyse de features embedding bert par tsne

MLOPs

- CI/CD
- Monitoring
- Réentraînement



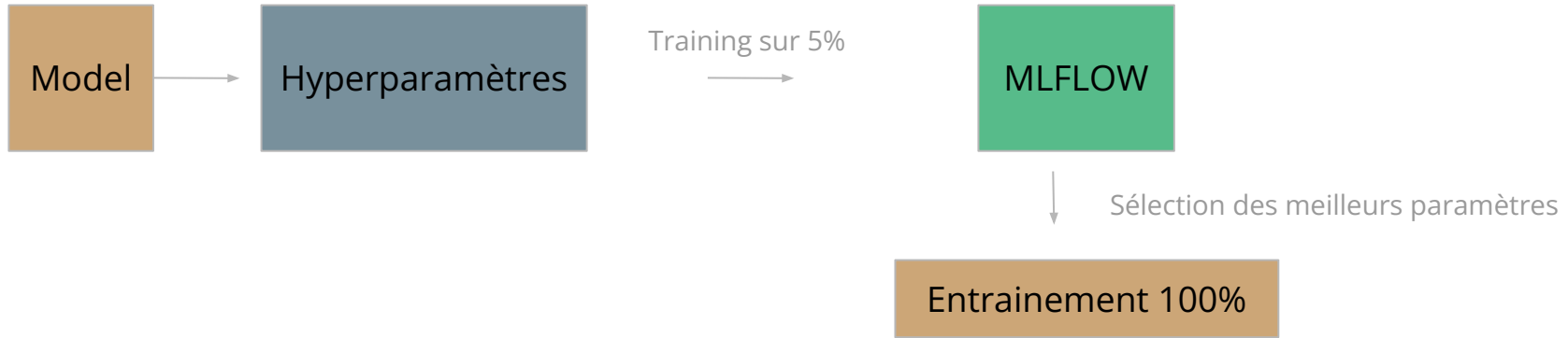
Architecture



tweet sentiments analysis

- git
- docker
- monitoring
- différents modèles

Entrainement



Modèles disponibles

DL

- Bert
- Roberta
- LSTM

ML

- RandomForest
- Roberta
- LSTM

Entrainement

Modèles disponibles

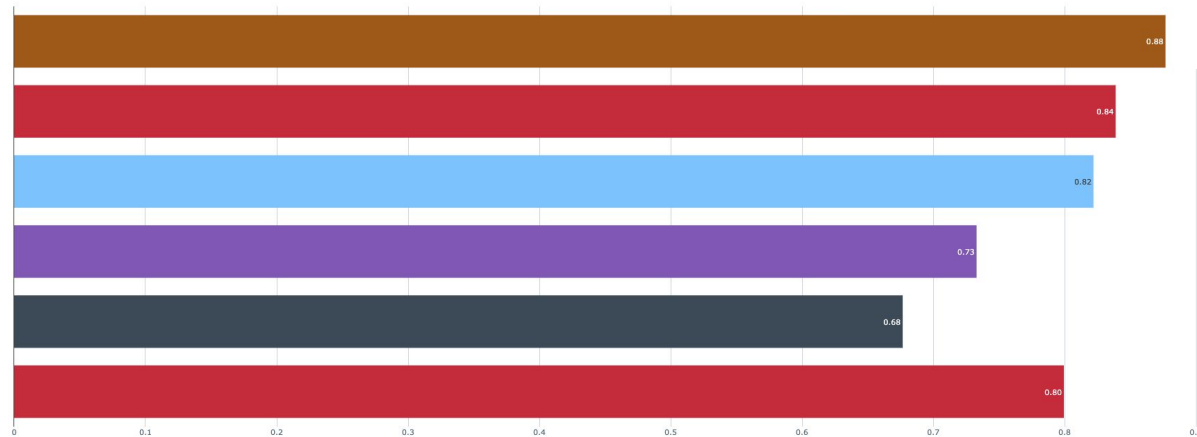
DL

- Bert
- Roberta
- LSTM

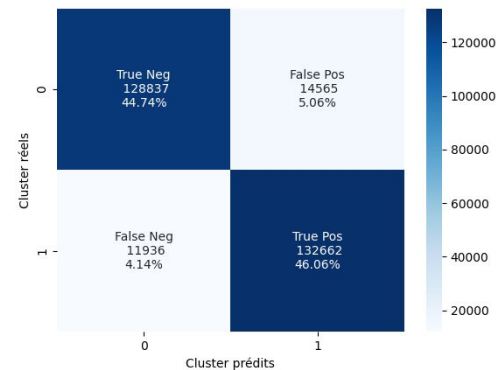
ML

- RandomForest
- Logistic Regression
- LightGBM

0_f1-score



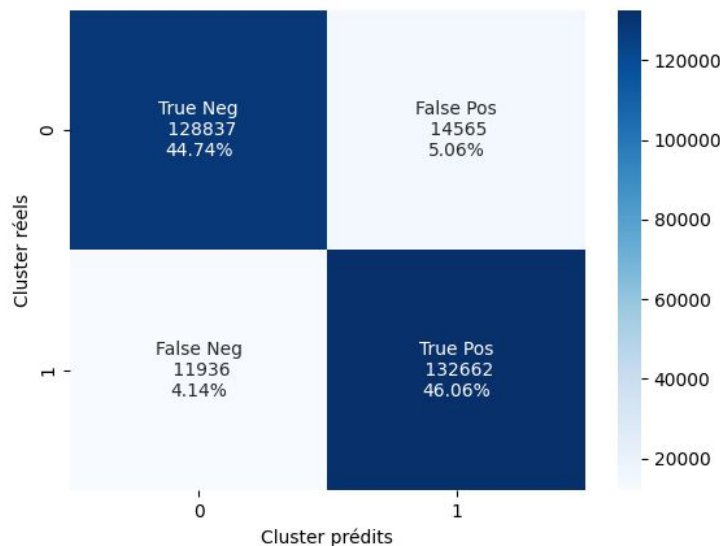
LogisticRegressionModel-bestParamOptimBayzian RandomForestModel LightGBMModel LSTMModel BertModel RobertaModel



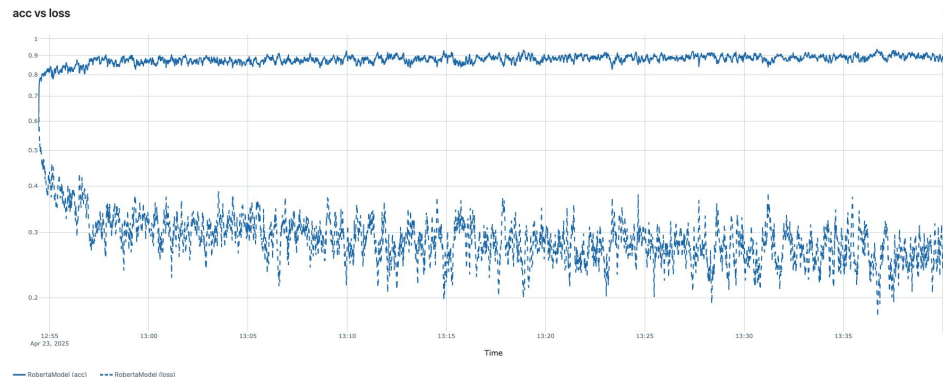
matrice de confusion Roberta

Entrainement

Roberta



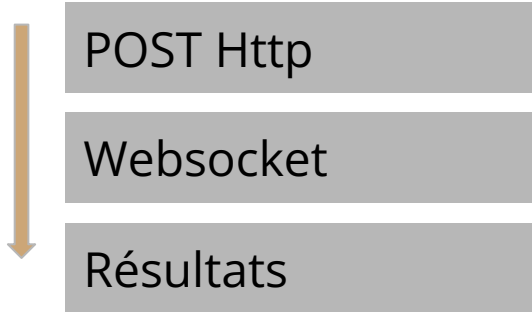
matrice de confusion



Courbes d'apprentissage loss et accuracy en fonction du temps

- 1 h d'entraînement
- 2 GPU
- Entrainement parallélisé

API



DEMO

Bienvenue Sur TweetSentimentPredict

Ajoute un tweet

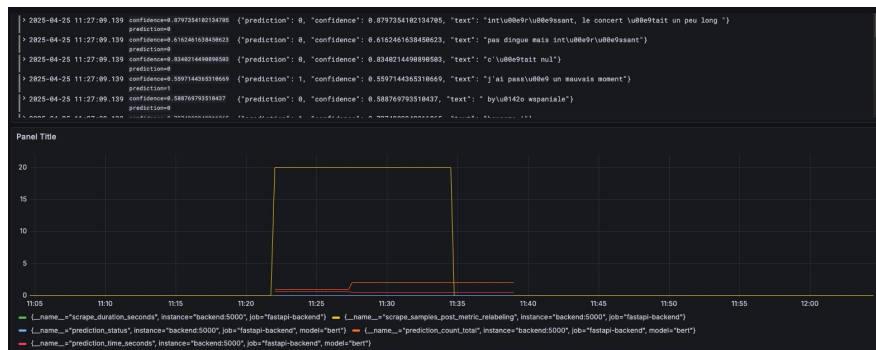
ok, cool	Supprimer
franchement, pas top	Supprimer
comment dire c'était bof	Supprimer
howsome !	Supprimer
byto wspaniale	Supprimer
j'ai passé un mauvais moment	Supprimer

Monitoring



Alerte mail

- $\text{Jobs} > 5$
- $\text{Confiance prédiction} < 0.6$



DEMO

Perspectives

Amélioration des prédictions

Fine tuning

Multi langues (Espagnole, Chinois, etc...)

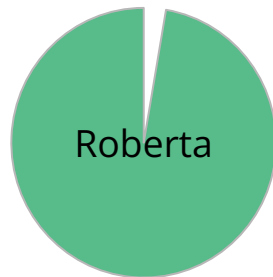
Amélioration du monitoring

Alertes

CPU bounds

Réentraînement avec nouveaux tweets

Conclusion



Coordonnées

Tristan Hérou

shift.python.software@.com

www.shift-python-software.com

