PROJECT 5

ANALYSIS

* Are the TSNE graphs correct ?

SUPERVISED

* Can I use XGBoost instead of RF ?Same result less parameters to optimize.
* Should I include a confusion matrix ?
* Si trop de temps a train tu pourrais splitter en 40-30-30 et utiliser un set de validation.

TESTS

* Is what I did enough ?
* I think the tests should be about the inferring service instead, right ? What to test as I process a single string at a time, a given input/output pair ?

INFERRING SERVICE

* Should this service process one body/title at a time or a bunch ?
* 16.0gb memory seems not enough (t3a.xlarge), can USE be done client side ?

**Une note technique**présentant une étude sur les approches et outils qui permettraient de généraliser l’approche MLOps (pipeline de codage des steps d’élaboration du modèle, et suivi de la performance du modèle en production) ???

* Présentation de la problématique et du prétraitement effectué, requête SQL et requête API de récupération des questions, exploration, cleaning et feature engineering (5 min).
* Présentation de l’approche non supervisée (4 min).
* Présentation de l’approche supervisée, de la visualisation avec MLFlow UI de la synthèse des résultats, et du modèle final sélectionné (5 min).
* Présentation des principes du MLOps et du pipeline de mise en production réalisé (3 min).
* Démonstration de l’API et conclusion (3 min).