PROJECT 5

Best hyperparameters found:{'num\_topics': 2, 'alpha': 0.61, 'eta': None, 'cv': -1.7413564061349542}.

ANALYSIS

* Are the TSNE graphs correct ?
  + Tu devrais faire un wordcloud par mot unique par tag.

UNSUPERVISED

* Is it ok to inly have 2 topics or should I put more ?

TESTS

* Fais un E2E avec Flask, une question une reponse et verifie ca.

INFERRING SERVICE

* Utilise <https://www.gradio.app/> pour l-interface, c-est plus une webapp et quand tu entres question et titre il te file les tags suggeres.
* Save en db les payloads recus et tags predits, capture d-events.

SUPERVISED

- Hyperoptimize with xgboost as well (lightgb?) to compare with the jaccard score as scorer.

**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) ???

* La mise en oeuvre d’un **pipeline**, de codage de tous les steps standardisés de la préparation des données jusqu’à l’élaboration des modèles (outils de type Kedro, pipeline / recipes de MLFlow, …)
* La mise en oeuvre du **suivi de l’utilisation et de la performance du modèle en production**, analyse du “model drift” (de type ”data drift” ou “concept drift”) à l'aide outils de type evidentlyAI, Promotheus, ou Popmon.

En attendant leur mise œuvre ultérieure, vous avez prévu de vérifier la stabilité du modèle dans le temps sur 1 an, en mesurant mensuellement l’évolution des mesures et scores des questions de chacun de ces mois.

##### ****Définir et mettre en œuvre une stratégie de suivi de la performance d’un modèle en production et en assurer la maintenance afin de garantir dans le temps la production de prédictions performantes.****

CE1 Le candidat a défini une stratégie de suivi de la performance du modèle. Dans le cadre de ce projet, il s’agit dans la note technique de l’étude MLOps de :

* Présenter les concepts liés au suivi de la performance du modèle : “model drift” -”data drift”, “concept drift” et leur mise oeuvre possible sur le projet
* Lister les outils candidats (de type evidentlyAI, Promotheus, Popmon)

CE2 Le candidat a réalisé un système de stockage d’événements relatifs aux prédictions réalisées par l’API et une gestion d’alerte en cas de dégradation significative de la performance. Dans le cadre de ce projet, il s’agit dans la note technique de l’étude MLOps de :

* Présenter la conception concrète du système de suivi de la performance adapté au projet : les indicateurs et mesures à mettre en oeuvre, les types d’alerte préconisées (il n’est pas demandé de le développer)
* Présenter comment utiliser les outils envisagés pour mettre en oeuvre le système de suivi (il n’est pas demandé d’implémenter les outils)

CE3 Le candidat a analysé la stabilité du modèle dans le temps et défini des actions d’amélioration de sa performance. Dans le cadre de ce projet, il s’agit de :

* Comparer l’évolution des mesures et scores pour plusieurs séries mensuelles de questions (sur 1 an), et de stocker les résultats dans MLFlow tracking

After having trained the model get some questions, predict and calculate the scores and upload them on MlFlow.

* 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).