PROJECT 5

UNSUPERVISED

* Tu peux faire un tfidf puis un kmeans pour voir un wordcloud des groupes pour voir ce qui apparaît et si ça ressemble au lda topics.

SUPERVISED

- max depth of 2 is quite low but ok with 100 estimators. En utilisant USE tu dois avoir une dimension discriminante qui est la plus utilisée dans ton arbre vu qu’il est coupé qu’une fois. Le jaccard score est bien pour ce genre de prédiction.

NOTE TECHNIQUE

- Check mes notions de data/model drift et concept drift, pas sûr qu’ils soient 100% correct.

VERIFICATION

- Ca va mes metriques ? Je pense pas que je peux calculer le drift avec ma single feature, creuse un peu plus mais bon.

- J-ai assez de question par mois ? Je pense pas.

Word clouds for java will include the javascript words as well:)

##### ****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).