PROJECT 5

ANALYSIS

* Are the TSNE graphs correct ? How to interprate them ? Have all git/python questions a similar number of tokens ?

SUPERVISED

* I’ll use MultiOutputClassifier, Binary Relevance scheme and the Classifier Chain Scheme, should I use another one as well ? Or only one ? **You can use both self : but check what scoring is used by default.**
* Is Hamming Loss a good way to evaluate a model ? YES.
* Word2Vec(Continuous Bag of Words or The skip gram model) or Sentence2Vec or Doc2Vec or Glove ? Doc2Vec seems to also consider the context of the words. → Check les Acceptance criteria si plus sont utilises.
* Si trop de temps a train tu pourrais splitter en 40-30-30 et utiliser un set de validation.

MLOPS

* What should be unit tested here ? The words embedding functions ?
  + Test ta methode de clean et preprocessing.
  + Bert et USE tu peux comparer les shapes avant apres.

Check what matrice creuse est pour prez.