

Protocole d'évaluation des modèles du Livrable 3

Afin de choisir un modèle de groupement des déclarations d'incident, nous vous proposons un protocole d'évaluation. Afin de le réaliser, il faut utiliser le Notebook **visualisation.ipynb** et utiliser le fichier Excel **analyse.xlsx** pour stocker les résultats.

1ère étape : évaluation du modèle des thèmes (topic model) :

- Analyse de 30% des topics :
 - 10% les plus cohérents
 - 10% les moins cohérents
 - 10% moyennement cohérents
- Attribution d'un score de cohérence du topic allant de 0 à 2 :
 - 0: pas cohérent (topic poubelle)
 - 1: moyennement cohérent (topic à améliorer)
 - 2 : topic cohérent (topic ayant un niveau de cohérence suffisant)
- Si topic suffisamment cohérent (2), associer un titre au topic
- Commentaire d'impression générale concernant le modèle

2nde étape : évaluation du modèle des groupements (cluster model)

- Analyse de 50% des clusters de manière aléatoire
- Attribution d'un score de cohérence du cluster allant de 0 à 2 :
 - 0 : Groupement à revoir
 - 1 : Groupement devant être amélioré
 - 2 : Groupement de qualité suffisante pour le besoin exprimé
- Si cluster suffisamment cohérent (2), associer un titre au cluster
- Commentaire d'impression générale concernant le modèle et les autres visualisations :
 - Groupement suffisamment séparé ? (cf visualisation globale et matrice)
 - Distance inter-cluster faisant sens d'après la comparaison des groupes proches

3ième étape : analyse de documents

- Analyse de 20 documents choisis de manière aléatoire : les thèmes et les clusters associés font-ils sens ?
- Attribution d'un score allant de 0 à 2 :
 - 0 : Les thèmes et cluster choisis ne font pas sens
 - 1 : Certains font sens
 - 2 : Ils font tous sens