pmi (mot1, mot2) = log [probabilité (nombre de fois que les deux mots apparaissent ensemble dans un document) / probabilité (mot1) \* probabilité (mot2)].

L'information ponctuelle mutuelle est en fait une mesure de la corrélation entre les deux mots ou phrases.

K-means explication : <https://mrmint.fr/algorithme-k-means>

Lda : <http://wwwabi.snv.jussieu.fr/erocha/webthese/ADL.html>

Svd : <https://math.stackexchange.com/questions/2138119/interpretation-of-svd-for-text-mining-topic-analysis>

Tfidf oké

Nmf: oké extraction de topic voir le code

Sentiment : pyFeel est une bibliothèque python, elle facilitera votre analyse des émotions en français. pyFeel calcule sept émotions: la positivité, la joie, la peur, la tristesse, la colère, la surprise, le dégoût avec un sac de mots méthode.

Calcule l'émotion d'une chaîne avec une méthode de type 'sac de mot'.

    object.dict pour voir le sac de mot.

Calcule l’émotion en faisant la moyenne des vecteurs d’émotion de chaque mot la chaîne d'entrée et le sac de mot.

Entrée: (auto) Sortie: dict

Suppervisés : violence

Merge des dataset : le tfidf, svd, la nmf et les « emotions »

Labels correspondent aux annotations avec nom sequence et violent

On utilise un modèle de régression logistique pour modéliser les variables violences

Verbose pour avoir des infos lors de la periode de training

w 1 {\displaystyle w\_{1}}