

Rdv Samy (15/09)

Controverse : sujets qui vont attirer opinions diff ex : vaccins (débat)

Son but : quantifier ça.

Présentation des données qu'il a pas eu le temps de traiter

Fichier json qui contient les infos de requête. Bdd sur nosql en local (hashtag utilisés, id tweet, etc...). Pour chaque tweet récupéré, création fichier json.

Idées données : Evolution du nombre de posts par jour, étude de corrélation : très retweeté très liké ?, Partitionning d'un graphe, trouver les influenceurs d'une controverse

Nuage de mots avec les hashtag de tweets les plus utilisés

Il faut des évènements particuliers, par exemple dans le sport décisions arbitrales, puis comment cibler la requête afin de réussir à trouver l'info.

Via Python (connexion à l'API)

Se baser sur le format de données de Samy ; les hashtags sélectionnés vont vraiment trier les tweets à récupérer, et donc éviter le hors sujet ; possibilité paramètre à donner pour indiquer dans quelle langue récup les tweets.

Algo force Atlas : Outil Gephi, logiciel qui a directement inclu l'algo de force (d'autres outils existants, car celui-ci pas très stable)

Les shap values

Conseil : commencer avec jeux de données artificiels pour commencer avec des petits jeux