

TP Final d'algorithmique

Sofiane LASRI-TRIENPONT

Matrices de confusion

Attention, mes matrices ne sont pas drôles. Je ne sais pas pourquoi, pourtant j'ai ajouté pleins de tweets, mais elles ne bougent pas beaucoup. :c

Modèle positif

	Prédit haineux	Prédit non haineux
Réel haineux	5	1
Réel non haineux	0	7

Modèle négatif

	Prédit non haineux	Prédit haineux
Réel non haineux	5	1
Réel haineux	0	7

Analyse détaillée des performances du modèle

Préambule

Le modèle généré par l'api n'est pas infallible. En effet, il paraît évident qu'il ne sait pas reconnaître les biais de langage. Un message ayant un sous-entendu négatif pouvant blesser quelqu'un n'est pas toujours reconnu. Toutefois, il est important de noter que les scores de ce genre de message est toujours relativement proche de 0.

Prenons cet exemple :

« Tu es un bon à rien, fais quelque chose de tes journées ! »

Cette phrase est reconnue comme haineuse par l'api, et obtiens le score de -0,78. L'api a su reconnaître le sens de la phrase. Toutefois, ce n'est pas le cas pour le second exemple qui, attention, pourra choquer la sensibilité de plus d'un.

« Ostie de calis mon char est remplie de merde là, tabernacle si je vois ce maudit niaiseux d'oiseau je lui fais sa peau »

Tout bon québécois vous dira sans aucune hésitation qu'il s'agit là d'une phrase remplie d'injures. Pourtant, l'api lui donne le score de 0,38.

Passons désormais à l'analyse des performances (page suivante).

Modèle positif

	Précision	Rappel	F1 Score
0	1	0.83	0.91
1	0.88	1	0.93

Modèle négatif

	Précision	Rappel	F1 Score
0	1	0.83	0.91
1	0.88	1	0.93

Améliorations possibles

Entraîner le modèle régulièrement avec un plus grand jeu de donnée est de toute évidence la meilleure amélioration que l'on puisse apporter à cette api. Mais nous pouvons également envisager de corriger le jugement donné par l'api par la suite, par exemple, via une interface d'administration qui serait gérée par les modérateurs du site. On pourrait imaginer que cela aurait un impact non négligeable sur les phrases dont le sens est implicite.