

SARDAIGNE Théo

Le mercredi 17 avril 2019

SART24069802

Making-off

**Récolte des avis laissés par les utilisateurs de
Twitter sur différents groupes de médias au
Québec et en France**

Pour Jean-Hugues Roy

L'étude de l'opinion majoritaire du public vis-à-vis de certains médias est un projet fastidieux dans lequel nous nous sommes lancés et qui a rencontré beaucoup de limites. Premièrement, les résultats obtenus en analysant 5 000 tweets mentionnant plusieurs médias de France ou du Québec ne nous ont pas permis de tirer de conclusions quant aux divergences d'opinions entre les différents types de médias (écrit, radio, télé). Pour les pays, on observe une légère différence entre la France et le Québec. Chez ce dernier, les avis ont l'air un peu plus favorables. Mais il n'y a pas de véritable tendance et j'avoue être un peu déçu du résultat même si nous avons pas mal travaillé dessus.

Créer un programme qui récolte les tweets et qui les consigne dans un fichier csv avec les valeurs de notre choix était une tâche assez difficile, qui aurait été dure à réaliser sans les tutoriels disponibles et surtout l'assistance des étudiants de polytechnique. C'est d'ailleurs grâce à eux que l'on a pu avoir la connaissance d'un analyseur de sentiment rendant ainsi notre projet plus concret. Et bien qu'il fallait avoir beaucoup de compétences pour le mener à bien, notre projet est vraiment intéressant d'un point de vue journalistique.

En 2019, la population passe une grande partie de son temps à s'informer sur les réseaux sociaux. Ce sont des supports internet qui vont leur donner la possibilité de réagir de multiples manières et le point le plus important : d'ajouter des commentaires. On est donc en mesure de se dire que l'analyse des avis laissés par le public sur les réseaux sociaux pourrait permettre de comprendre ce qu'il pense et, par exemple, grandement intéresser les médias. Il n'y avait pas vraiment de moyen auparavant de sonder son public et cette apparition des réseaux sociaux dans le monde de l'information entrouvre énormément de possibilités.

Nous avons donc commencé avec de grands espoirs ce projet, car il est clair qu'en se servant du programmation pour essayer d'analyser l'avis du public, on avait de grandes chances de découvrir des tendances et faits intéressants.

La récolte des tweets émis par les utilisateurs ainsi que l'utilisation du logiciel vadersentiment nous permettrait de classer ces avis en commentaires négatifs ou positifs.

Malheureusement, ce projet présente quelques limites qui nécessiterait un peu plus de moyens pour le rendre concret.

L'analyseur de sentiment utilisé n'est pas forcément fiable car nous avons déjà relevé de multiples erreurs (classe des commentaires positifs comme négatifs et inversement), ce qui nous donne une marge d'erreur inconnue et potentiellement grande. Il en existe des payants qui se servent de l'intelligence artificielle pour réduire cette marge d'erreur, notamment avec le "machine learning", et qui seraient bien plus pertinents pour avoir un résultat viable.

Deuxièmement, malgré les nombreux tweets analysés, il faudrait en récupérer beaucoup plus et sur une plus longue période pour réussir à dégager une tendance fiable.

En effet, sur une semaine, les sujets de l'actualité influencent assurément les commentaires faits sur Twitter et ainsi les résultats. Par exemple, une semaine peut être bien plus compliquée en France qu'au Québec et inversement.

Et malgré tous ces paramètres, si on arrive à les régler, il nous restera une variable bien plus compliquée à prendre en compte et à traiter, la subjectivité des réseaux sociaux. Selon Rokni Siavash, professeur à l'UQAM en histoire des communications, mais aussi pas mal de spécialistes, il est affirmé que ce que l'utilisateur retrouve

dans son fil d'actualité relève du contenu personnalisé, adapté à ses préférences. Ainsi, en analysant les commentaires des utilisateurs de Twitter, le projet peut paraître intéressant et viable puisqu'une grande majorité de la population y est active et donne son avis de manière récurrente, mais le résultat sera finalement biaisé car personne n'a accès à la même information et de la même manière tout le monde a accès au contenu qu'il souhaite. Ce contenu personnalisé mériterait une thèse car en se penchant sur ce problème à la suite de ce projet de moissonnage de données, on se rend compte que si à l'avenir la totalité de la population lit les actualités sur Internet, le principe d'objectivité et d'information pour tous deviendra totalement nul et chaque utilisateur d'internet ne lira que le reflet de ce qu'il veut entendre. Pour conclure, j'imagine que les résultats très équilibrés que l'on a obtenu suite à notre moissonnage sont peut être liés à cette notion de contenu personnalisé. Néanmoins, il serait nécessaire de procéder à un nouveau moissonnage tout en solutionnant les limites de celui que l'on a effectué afin d'en avoir le coeur net.

Liens :

<https://openclassrooms.com/fr/courses/235344-apprenez-a-programmer-en-python>

<https://python.developpez.com/>

<https://developer.twitter.com/en/docs.html>

Tutoriel de Jean-hugues Roy sur la programmation d'un script pour récupérer des tweets

Les liens jupyter et python sont disponibles dans le carnet.

