



- Home
- Explore
- Notifications
- Messages
- Bookmarks
- Lists
- Profile
- More



Identifier le leader d'opinions le plus influent sur twitter

Everyone can reply



Tweet



Siva @SivaChane

Introduction

Etudiants au sein de l'EFREI et de PSB, nous avons participé à un challenge public sur la célèbre plateforme management & Data science.

Le challenge met en perspective l'étude de la science des graph et l'analyse des données de twitter. L'objectif est d'identifier les leaders d'opinions sur le réseau social Twitter autour du sujet de l'IA.

Les enjeux de cette analyse ont un impact business direct, car cela permettra à une entité (entreprise) d'identifier les personnes qui ont les plus grandes capacités de diffusions et de médiatisation pour relayer leurs messages.

Qui sont les leaders d'opinions sur Twitter pour la thématique de l'IA ?



Lucas @LucasBillaud

Les critères utilisés

Au cours de notre analyse nous avons définis plusieurs critères d'identification :

- Nombre de Tweet crée
- Nombre de retweet effectuer
- Nombre de fois ou son tweet a été retweeter
- Nombre d'interaction
- Nom de followers
- Nombre d'abonnement
- Durée d'existence du compte

Ces différents critères nous serviront à différents niveaux dans nos analyses.



Maxime @MaximeAllakereHomo

Présentation du jeu de donnée utilisé et des statistique descriptives

Notre travail se focalise autour du #IA , cependant nous avons également récolté des données concernant des sous-catégorie de l'intelligence artificiel (ex : #deep_learning, #machine_learning) pour une meilleure granularité. Afin d'identifier ces leaders d'opinions, nous avons effectué un « scrapping » des différents Tweet contenant ces hashtags.

Le data set contient 4047 tweets (incluant des tweets et retweets), avec les libellés suivants :

- Username_screen_name : Contient le nom de la personne ayant publié le tweet ou le retweet
- Post : Contient les tweet récupérer
- Hashtags : Contient les hashtags utilisé dans le tweet
- Screen_name_original_post : Contient le nom de la personne à l'origine du tweet
- Mention_person : Contient le nom des comptes mentionner dans le tweet
- Create_date : Contient les dates de publication des tweets.

Parmi cette data nous avons 2468 retweets, 1579 tweets originaux (tweet crée par la personne posté). Nous avons identifié 932 personnes différents.

Topics to follow

- Augmented reality Technology [Follow](#)
- Internet of things Technology [Follow](#)
- Information security Technology [Follow](#)
- Cybersecurity Technology [Follow](#)
- Web development Technology [Follow](#)

Top 5 des Twittos les plus retweetés

- FlashTweet @DigitalTrans4mF [Follow](#)
- AnthonyRochand @AnthonyRochand [Follow](#)
- Fabienne Billat @fadouce [Follow](#)
- DigitalTrans4m_FR @DigitalTrans4mF [Follow](#)
- Denis Lefebvre @Lefden [Follow](#)

Statistiques des leaders d'opinions

- 15K Nombre de followers moyen
- 36K Nombre de tweets moyen

Statistiques du réseau

- Degré 2,048
- Average Path 6,33
- Densité 0,001
- Coefficient de clustering 0,105



- Home
- Explore
- Notifications
- Messages
- Bookmarks
- Lists
- Profile
- More

Tweet

Un deuxième data set à également été extraite, se focalisant sur les 50 acteurs les plus influents (personnes les plus retweeté). Voici les libellés du deuxième DataFrame

- User Name : le nom du compte
- Followers : nombre de followers
- Following : nombre de followings
- Description : description présente sur le profil
- Tags : hashtags présents dans les descriptions des profils
- Post : nombre de post réaliser par le profil

Nous remarquons qu'en moyenne ces profils ont environs 15000 followers et ont posté plus de 36000 tweets depuis la création de leur compte.



Claire @ClaireMazzucato

La méthode et les étapes de notre étude

La première partie de notre étude c'est focalisé sur la récolte et nettoyage de données, que nous avons réalisé sous python (tweepy et Pandas) et R (rtweet, tidy, readr).

Une fois que notre dataset à été rendu exploitable, nous avons réalisé un graph grâce au différents liens (tweet, retweet etc) afin de visualiser le réseau et identifier les principaux clusters, acteur.

Une analyse sur les différents éléments du graphique sera réalisée afin de mieux comprendre sa structure.

Ensuite une étude parallèle sera réalisée afin d'identifier précisément les principaux acteur (avec des chiffres).

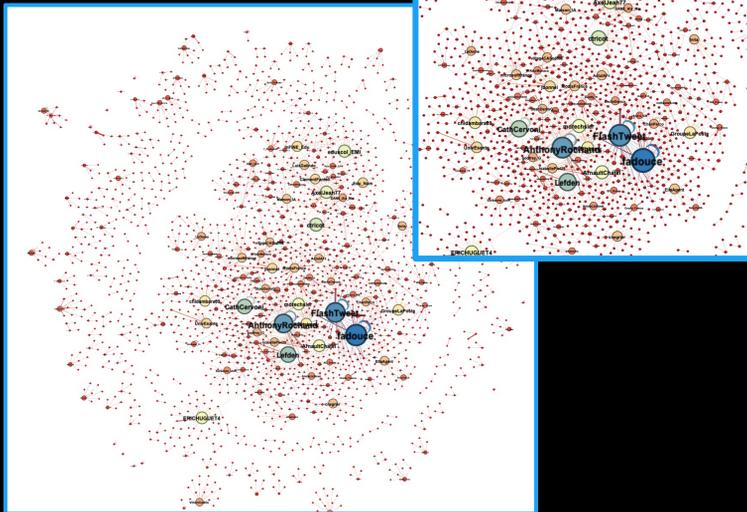
Les tweets et les hashtags seront également analysé grâce au NLP (Natural language processing) afin de dégager les principales tendances.

Nous terminerons notre étude avec une recommandation, sur les combinaisons d'hashtag à employer afin d'optimiser sa communication et essayer de devenir un leader d'opinion



Siva @SivaChane

Les résultats obtenus



De façon générale, le graphique à un degré de 2.048, ce qui signifie que nous avons en moyenne 2 liens par nœud. La distance est de 6.33, chaque personne prise au hasard est séparée de 6 individus en moyenne. Le coefficient de clustering général est de 0.105, a première vue le réseau paraît peu dense. Ensuite nous allons nous focaliser sur un échantillon (les nœuds > 20 liens) afin de ne pas biaiser notre analyse avec les multitudes de micro-clusters.

Prenons "Anthony Rochand" et "FlashTwee" deux profils à degré similaire, on observe que leur coefficient de clustering est significativement différent. En effet, le coefficient de Clustering de Anthony est dix fois plus élevé que celui de FlashTwee.

Topics to follow

- Augmented reality Technology [Follow](#)
- Internet of things Technology [Follow](#)
- Information security Technology [Follow](#)
- Cybersecurity Technology [Follow](#)
- Web development Technology [Follow](#)

[Show more](#)

Top 5 des Twittos les plus retweetés

- FlashTwee @Digitaltrans4mF [Follow](#)
- AnthonyRochand @AnthonyRochand [Follow](#)
- Fabienne Billat @fadouce [Follow](#)
- DigitalTrans4m_FR @Digitaltrans4mF [Follow](#)
- Denis Lefebvre @Lefden [Follow](#)

Statistiques des leaders d'opinions

- 15K** Nombre de followers moyen
- 36K** Nombre de tweets moyen

Statistiques du réseau

- Degré** 2,048
- Average Path** 6,33
- Densité** 0,001
- Coefficient de clustering** 0,105

