

CR1 Fake News 2019-2020

Nouvelière Benjamin / Durand Pierre / Auger Nathan / Lopez Julio

Github : https://github.com/nouveliere-benjamin/Fake-News-2019_2020

Planning du groupe Fake News :

Jour	Date	Heures
Mardi	05/11	9H-12H
Jeudi	14/11	16H-18H
Vendredi	22/11	13H30-16H30
Mardi	26/11	9H-12H
Jeudi	28/11	13H30-16H30
Jeudi	05/12	9H-12H
Mardi	10/12	13H30-16H30
Vendredi	13/12	10H-12H
Mercredi	08/01	13H30-16H30
Jeudi	16/01	13H30-16H30
Mercredi	22/01	9H-12H
Mardi	28/01	13H30-16H30
Mercredi	29/01	9H-12H
Mardi	04/02	13H30-16H30
Jeudi	13/02	13H30-16H30
Lundi	17/02	13H30-16H30
Mardi	03/03	13H30-16H30
Mardi	10/03	9H-12H
Mercredi	11/03	10H-12H
Lundi	16/03	15H30-17H30
Jeudi	19/03	9H-12H
Lundi	23/03	9H-12H
Mardi	24/03	9H-12H

Total : 65H

Langages considérés :

Langages	Spécificités
Python	<ul style="list-style-type: none">- Large panel de bibliothèques (Numpy, Panda, NLTK, SpaCy, SciKit-learn)- Intègre facilement des APIs- Performant en temps- Communauté développée et active
JVM	<ul style="list-style-type: none">- Bon panel de bibliothèques (CoreNLP, ND4J, DL4J)- Calcule sur CPU- Machine virtuelle- Communauté développée
C/C++	<ul style="list-style-type: none">- Quelques bibliothèques (CUDA, TensorFlow, Caffe)- Utilisation de pointeurs- Calcule sur GPU- Accède à des APIs de haut niveau et flexibles
JavaScript	<ul style="list-style-type: none">- Très peu de bibliothèques- Intègre mal les APIs

Nous trouvons plus intéressant de travailler sur Python, comme on peut le voir, il possède un large panel de bibliothèques, des APIs faciles à intégrer et une communauté développée et active. Nous verrons au fur et à mesure du projet si d'autres langages sont nécessaires.

Technologies :

Technologies	Spécificités
Régression linéaire	<ul style="list-style-type: none">- Facile et rapide à calculer- Réajustement manuel de la courbe
Réseau de neurones	<ul style="list-style-type: none">- Réajustement automatique de la courbe

Nous trouvons plus intéressant d'utiliser des réseaux de neurones car la majorité des projets similaires l'utilisent, ce qui pourra nous aider pour la recherche d'informations. De plus le modèle se réajuste automatiquement avec les entrées d'informations.

Axes de recherche envisagés :

Nous allons prendre le réseau social Twitter pour réaliser le projet Fake News car les messages sont limités en nombre de caractères, ce qui permettra une analyse plus efficace.

Nous allons effectuer des traitements du langage sur les tweets pour en déduire les thématiques qui nous permettront de déterminer si une new est vraie ou non.

Axe 1 :

Se baser sur une banque de compte Twitter ou des journaux et classer par nous-même si une new est fake ou non.

Axe 2 :

Trouver une corrélation entre les tweets connus faux et les tweets connus vrais sur les mots ou la structure des messages.