



Université Paul Valéry - Montpellier 3

UFR 6 - Éducation & Sciences pour les LLASHS

Master MIASHS

Prédiction des comportements suicidaires

RENDU 5 - OUTILS ET PROGRAMMES À UTILISER



Équipe :

Lisa BÉTEILLE
Matéo CALSACY
Jean CHABANOL
Anamé ROUMY
Laura SÉNÉCAILLE
Célia TEYSSIER

2021/2022

Encadrants :

Sandra BRINGAY
Pierre LAFAYE de MICHEAUX
Florian LOMBARDO

INTRODUCTION

Pour la réalisation de ce projet il a fallu choisir les bons outils et programmes à utiliser pour ne pas nous ralentir et pour nous faciliter les tâches.

OUTILS COLLABORATIFS

Nous avons réfléchi à la mise en place d'outils collaboratifs pour nous permettre d'optimiser l'efficacité de notre travail et ainsi d'accroître notre productivité.



Pour l'organisation des rédactions de comptes rendus, de résumés d'articles et des agendas minutes nous utilisons Google Drive. Cet outil nous permet d'optimiser notre travail et d'être plus efficace en suivant en temps réel ce que chacun écrit pour donner un avis ou modifier des choses directement sur la même page.

De même pour le Google Colab, qui permet de partager nos codes de classification entre nous après les avoir terminés sur Jupyter Notebook. Cela nous permet de voir ce que nous obtenons en faisant fonctionner ces codes avant de les push sur Github. De plus, nous avons relié directement les données du Github au Google Colab pour nous faciliter le travail.



Le Github nous sert à partager les données mises à notre disposition (tweets, reddit, wikipédia, articles ...) ainsi que les codes de classification plus rapidement, et par conséquent pouvoir avoir tous les mêmes données pour travailler ou pour voir ce que les autres ont ajouté.

Nous utilisons aussi Github pour créer une “to do list” avec les éléments à faire (“to do”), en cours (“in progress”) et terminés (“done”). Cela nous permet de rester organiser, d'avoir une vue d'ensemble de tout ce que nous devons faire et d'attribuer des tâches à chaque membre pour être plus efficace.

Notre travail est notamment optimisé grâce à Discord et Whatsapp. Ces outils nous permettent de rester en contact pendant nos périodes d'alternance ou les weekends pour être plus productifs. Nous utilisons Discord pour entretenir des réunions régulièrement en appels vocaux en utilisant des partages d'écran, et Whatsapp pour des messages réguliers plus instantanés. Discord nous sert également pour entrer en contact avec Mathéo Daly et Florian Lombardo plus rapidement lorsque nous avons des soucis dans notre projet (notamment notre problème d'algorithme Wikipédia)



LANGAGE DE PROGRAMMATION ET OUTILS D'ANALYSE DES DONNÉES

Nous utilisons plusieurs langages de programmation pour mener à bien notre projet.



Il est important d'utiliser RStudio pour le nettoyage utile à l'exploration des données. Nous avons privilégié RStudio pour le nettoyage car Célia a plus l'habitude de travailler avec cet outil, nous allons donc être plus efficaces dessus. De plus, il possède des packages très utiles pour gérer rapidement des gros volumes de données.

Iramuteq nous sert pour réaliser l'exploration des données. Ce logiciel basé sur RStudio a été créé par un groupe de chercheurs. Il est très utile pour visualiser et analyser des textes, c'est pour cela que nous avons choisi de travailler dessus.



Pour finir, Python nous est très utile pour faire la classification. Nous avons pu créer des modèles de prédictions comme des SVM, Random forest et Kmeans. Nous avons choisi Python pour réaliser la classification car c'était ce qui était le plus adapté et préconisé par nos commanditaires.

Une approche reproductible signifie une approche qui peut être reproduite ou copiée. L'approche que l'on a utilisé pour notre projet peut être qualifiée de reproductible car nous utilisons des notebooks pour écrire nos codes, ce qui permet de changer l'appel des fichiers à utiliser pour pouvoir faire marcher tout le code sur des données de types différents. (1)

Il est important d'adopter une approche permettant de garder une trace de toutes les modifications apportées ainsi que de stocker et rendre plus accessible les données.

En utilisant le GitHub, cela nous permet de garder une trace des modifications et des échanges de répertoires car nous avons tous le même code à partir du moment où nous le récupérons sur le Git. Par conséquent, rien ne peut être perdu.

Grâce au Google Drive, nous pouvons également garder une trace de toutes nos modifications car nous avons accès en temps réel aux changements des fichiers établies par tous les membres du groupe. De plus, ces changements sont sauvegardés instantanément, donc nous ne perdons rien non plus.

De même pour l'utilisation de Google Colab qui nous permet de garder toutes les traces de modifications de codes que nous avons réalisées.

DIFFICULTÉS RENCONTRÉES AVEC CES OUTILS

Tous ces outils sont compatibles entre les différents OS, sauf pour Github où nous avons eu des soucis avec les ordinateurs Macbook qui créaient des dossiers DS_Store dans les fichiers. Nous obtenions donc des erreurs au moment du push et du pull. Nous avons pu néanmoins régler ce souci en utilisant un gitignore.

CONCLUSION

Ces outils et programmes nous permettent d'optimiser notre travail et de nous rendre plus efficaces et plus productifs. Les connaissances de certains membres de l'équipe sur certains programmes sont également bénéfiques pour notre productivité et notre organisation.

BIBLIOGRAPHIE

- (1) Reproductible : <https://www.cnrtl.fr/definition/reproductible>