

Compte rendu de réunion

Le 15/09/2021 à 12h00-13h00 à l'université Paul Valéry

Objet : Première réunion sur le sujet de TER de M1.

Ordre du jour :

- Présentations de notre groupe et des encadrants du projet.
- Présentation du sujet de TER.

Légende du tableau ci-dessous : P = présent, A = absent, E = excusé, NC = non concerné par la réunion

Liste des participants au projet			
Nom	Email	Fonction	Présent
Sandra BRINGAY	sandra.bringay@univ-montp3.fr	Encadrante Pédagogique	P
Florian LOMBARDO	florian.lombardo@univ-montp3.fr	Autre encadrant	P
Laura SENECAILLE	laura.senecaille@etu.univ-montp3.fr	Etudiante	P
Anamé ROUMY	aname.roumy@etu.univ-montp3.fr	Etudiante	P
Lisa BETEILLE	lisa.beteille@etu.univ-montp3.fr	Etudiante	P
Matéo CALSACY	mateo.calsacy@etu.univ-montp3.fr	Etudiant	P
Jean CHABANOL	jean.chabanol@etu.univ-montp3.fr	Etudiant	P
Célia TEYSSIER	celia.teyssier@gmail.com	Etudiant	P

Sommaire

Présentation du contexte.....	2
Présentation du projet	2
Actions à entreprendre	3

Présentation du contexte

- ➔ Le suicide est une cause de mortalité importante chez les jeunes. Pour les 16-25 ans, c'est la deuxième cause de mortalité après les accidents de la route. Pour les 25-39 ans, c'est la première cause de mortalité. Or, il y a peu de médiatisation et de démarches publiques comparée à celles qu'on peut rencontrer dans le cas des accidents de la route par exemple.
- ➔ Pour des raisons médicales et morales, il est compliqué de recueillir des témoignages de personnes ayant fait des tentatives de suicides, le risque de récurrence au bout de 3 mois est très élevé. On utilise donc des corpus numériques, les données sont publiques puisqu'elles proviennent de réseaux sociaux ou de Wikipédia.
- ➔ Les deux dernières années, les groupes chargés de ce sujet de TER ont eu des difficultés à recueillir les données depuis Twitter, notamment à cause des modalités de téléchargement. On va donc garder les anciens corpus pour la suite de ce projet. Pour Wikipédia, le code pour récupérer les données fonctionne bien.
- ➔ **Conclusion du projet l'année dernière** : Leur algorithme prédisait si un corpus de texte venant de Reddit faisait référence au suicide. De plus, leurs variables pour représenter les textes n'étaient pas forcément les plus pertinentes.

Présentation du projet

- ➔ Les données à utiliser sont celles de Reddit, Wikipédia et Twitter.
- ➔ Les données provenant de Twitter seront fournies et celle de Wikipédia pourront être extraites à l'aide d'un algorithme donné.
- ➔ Les commanditaires souhaitent un algorithme qui identifie un comportement suicidaire sur toutes les données (plus général que l'année dernière : Twitter, Wikipédia et Reddit). L'algorithme pourra avoir une performance différente en fonction de l'origine des corpus données en entrée puisque les « sous-algorithmes » pourront être différents selon elles. Dans un premier temps, on utilisera des corpus qui font références à un suicide et non juste à un mal-être.
- ➔ Le résultat doit être présenté sous forme de site internet pour la présentation, avec des graphiques et des explications et analyses.
- ➔ Aucune contrainte sur l'utilisation de R ou de python pour réaliser ce projet.
- ➔ Réutiliser le travail de l'année dernière pour des corpus venant de Twitter et Wikipédia.

Finalité de ce projet : Les commanditaires souhaitent détecter les tweets d'une personne à risque suicidaire et être notifié pour ensuite faire de la prévention.

Lorsque l'algorithme détectera un comportement à risque, une notification serait envoyée et amènerait à un suivi médical ou à une mise en contact avec des groupes de survivants suicidaires pour apporter un soutien.

Dans la première partie de l'année, expliciter les données et faire des modèles de prédiction « classiques » (ex : forêts aléatoires). Au deuxième semestre : passer au deep learning.

Actions à entreprendre

Pour la prochaine réunion le **lundi 11 octobre à 13h30**, finir le répertoire GitHub et envoyer le lien aux commanditaires (déjà fait). Ensuite, récupérer les données et les décrire (savoir combien il y en a, faire 2-3 graphiques, les mettre sur panda..). Enfin, réussir à décrire avec nos mots l'objectif et comment ça va se passer.