

Université du Québec à Rimouski
300, allée des Ursulines, C. P. 3300, succ. A
Rimouski (Québec) G5L 3A1, Canada.
www.uqar.ca

Lévis, 09 Mars 2021

Yacine Yaddaden
Professeur à l'UQAR

Objet : *Remarques et suggestions après évaluation du manuscrit de thèse de Florentin Thullier*

Cher Florentin,

C'est avec grand plaisir que j'ai lu ton manuscrit de thèse que j'ai trouvé très intéressant que ce soit du point de vue technique que recherche. J'aimerais saisir cette occasion pour te féliciter par rapport à la qualité de ton manuscrit que ce soit du point de vue du fond que de la forme.

Néanmoins, j'ai quelques remarques et suggestions qui permettront éventuellement d'enrichir un peu plus le contenu de ton manuscrit de thèse que je listerai ci-dessous :

Chapitre 2

- Pour la réduction de la dimensionnalité, je trouve que tu pourrais mettre un peu plus de détails surtout qu'on a *réduction de la dimensionnalité* par transformation des données et la *sélection des attributs* qui permet d'ordonner par importance les caractéristiques (*Random Forest*). Pour chacun, il y a une ribambelle de techniques que tu peux utiliser et je trouve d'ailleurs dommage que ce ne soit pas inclus dans ton pipeline.
- Tu as parlé d'un certain nombre de techniques d'apprentissage automatique, je pense qu'il serait de diviser ça en deux sous-sections (*supervisés* et *non supervisés*).
- Dans la partie Clustering ou segmentation, tu parles de classification alors qu'habituellement, terme est réservé aux méthodes d'apprentissage supervisées.
- Pour la partie Machine à Vecteurs de Support, il manquerait une référence (Cortes & Vapnik, 1995). Tu devrais également faire référence à son utilisation pour la résolution de problèmes multiclasse en utilisant les stratégies One-vs-One ou One-vs-All.
- Tu devrais également faire référence au k-NN et Random Forest même si tu en parles après, juste pour ne pas laisser le lecteur sur sa fin ou qu'il se rende compte que tu n'as pas parlé d'une méthode basique (k-NN).
- Pour la partie Réseau de Neurones, tu devrais parler ou commencer par introduire le *Perceptron*, c'est une bonne base pour comprendre les réseaux multicouches.
- Tu parles de surapprentissage, mais il y a de nombreuses méthodes pour y remédier, ce n'est pas une fatalité.
- Je sais que tu n'as pas utilisé de Deep Learning dans le cadre de ton travail de thèse, mais il vaut la peine d'en parler un peu plus surtout que beaucoup de travaux récents se basent dessus. D'ailleurs, la limite que tu as citée dans les SVM concernant son inefficacité sur de

grandes quantités de données peut être résolue avec un réseau de neurones profond. De plus, ça permet de simplifier ton pipeline, car l'extraction des caractéristiques se fait automatiquement. De plus, il s'intègre bien avec le *Cloud Computing*.

Chapitre 3

- Il y a beaucoup de travaux ciblant la reconnaissance d'activité et qui se basent sur l'utilisation de capteurs haut niveau (caméras), ça vaut la peine d'en parler un peu.
- Dans le contexte de capteurs physiologique, tu as parlé de ECG, mais généralement, sur les bracelets et montres intelligentes on parle de capteurs PPG.
- Tu as parlé de capteurs, de technologies et de protocoles de communication, mais tu as omis de parler de sécurité des informations récoltées par les capteurs. C'est une problématique réelle, tu devrais faire référence à des méthodes de chiffrement par pour encrypter les données.

Chapitre 4

- Pour les méthodes de k-NN et Random Forest, tu devrais mettre deux sous-sections afin d'avoir une meilleure organisation.
- Pour la méthode de k-NN, elle est dite paresseuse et lourde quand on a une grande quantité de données, mais il y a des méthodes permettant de structurer et d'optimiser la recherche du plus proche voisin, exemple : KD-Tree et Ball-Tree.

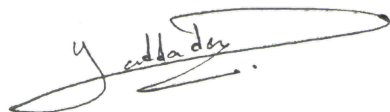
Chapitre 5

- Il y a une petite erreur dans l'URL utilisée dans la Figure 5.5.

Chapitre 6

- Le *Workbench* créé est censé s'inspirer des autres que tu as définis dans le deuxième chapitre, mais je trouve que ce que tu as présenté n'est pas trop détaillé sur un point de vue flexibilité. C'est-à-dire que je vois bien que le pipeline de ton premier article s'y implémente bien, mais si je souhaite implémenter ou tester d'autres qui n'ont pas forcément les mêmes blocs, est-ce que c'est possible et comment ça va marcher ?

Les remarques ci-dessus n'enlèvent rien à la qualité de ton travail ou du manuscrit. N'hésite pas à revenir vers moi pour plus de détails.



Yacine Yaddaden, Ph. D.
Professeur en informatique
Département de mathématique, d'informatique et de génie
Université du Québec à Rimouski, Campus de Rimouski
<https://yyaddaden.github.io>