

Owlint : Test technique Data Science

Owlint a besoin de toi ! Depuis peu, nous recevons des centaines de spams sur notre page de contact. Des propositions pour augmenter notre nombre de clients, notre visibilité, nos ventes, etc.

Ta mission, si tu l'acceptes (ce que l'on te conseille quand même dans le cadre de cet entretien), construire un classifieur de messages dans les catégories Spam/Non Spam.

Dans cette mission, tes armes seront :

- Un jeu de données que l'on te fournit;
- Tes indiscutables compétences ainsi que ton esprit analytique;
- Ton courage et ton sens de l'organisation;

Cette mission commence dès maintenant et se termine mardi 30/04 à 23h59, heure à laquelle nous prévoyons une grosse attaque de bots spammeurs sur notre infrastructure. Après cette heure, il sera trop tard. Ta chance sera passée.

Résultat *optimal* :

- Un classifieur permettant de faire la différence entre des messages de spam et des messages légitimes;
- Des métriques de performance du fonctionnement du classifieur;
- Un manuel permettant d'expliquer à la team comment entraîner, tester et ré-entraîner le classifieur;
- Le classifieur doit être utilisable par des humains ou d'autres applications et doit ainsi disposer d'une API permettant de faire des requêtes de classification afin de vérifier si les messages sont légitimes;
- Nous souhaitons une vraie traçabilité des classifications faites. Cela nous permettra d'éplucher périodiquement l'historique des classifications et de vérifier qu'aucun message légitime n'a été oublié.
- Nous souhaitons pouvoir indiquer si une classification qui a été faite est fausse afin que le modèle puisse s'améliorer automatiquement lors d'un ré-entraînement;
- Le travail rendu doit être disponible sur ton repo Git pour les autres développeurs de Owlint ;

Les conseils du jeune sage :

- Le bonheur ne se trouve pas en haut dans la montagne mais dans la manière de la graver (autrement dit, on s'intéresse moins au résultat qu'à la logique pour y arriver);
- Rome ne s'est pas faite en un jour. Il faut savoir découper et prioriser;
- On ne peut correctement apprendre si nos livres sont inexacts;
- Ce n'est pas la force mais la persévérance qui fait les grandes œuvres.

Bonne chance pour cette Dernière Ligne Droite¹ !

¹ Dépasse-toi, rejoins-nous et tu comprendras !