

Projet data - alternative au stage de fin de formation



Le but de ce projet sera de mettre en pratique les notions abordées tout au long de la formation que vous venez de suivre avec un compte-rendu hebdomadaire de votre avancée et un rendu final qui devra être hébergé publiquement sur Github.

Votre mission sera de participer à la compétition Kaggle : <https://www.kaggle.com/c/g-research-crypto-forecasting>

Dans cette compétition, vous utiliserez votre expertise en machine learning pour prévoir les rendements à court terme de 14 cryptomonnaies populaires.

G-Research est le leader européen de la recherche quantitative en finance. Cet organisme a récolté un ensemble de données de millions de lignes de données de marché à haute fréquence remontant à 2018 que vous pouvez utiliser pour construire votre modèle.

Une fois la date limite de soumission passée, votre score final sera calculé sur les 3 mois suivants à l'aide des données cryptographiques en direct au fur et à mesure de leur collecte.

L'activité simultanée de milliers de traders garantit que la plupart des signaux seront transitoires, l'alpha persistant sera exceptionnellement difficile à trouver et le danger de sur-apprentissage sera considérable.

De plus, depuis 2018, l'intérêt pour le marché de la cryptographie a explosé, de sorte que la volatilité et la structure de corrélation des données sont susceptibles d'être fortement non stationnaires, rendant la prédiction de série temporelle presque impossible !

Déroulement du projet

Vous commencerez par prendre connaissance du sujet et des données à votre disposition. Vous devez vous inscrire obligatoirement à la compétition et accepter les règles pour pouvoir télécharger les données.

1. **Découverte du sujet** : Vous commencerez par une exploration naïve des données, vous devrez rendre au terme de votre première semaine une analyse exploratoire commentée du dataset à votre disposition. Vous mettrez ce notebook dans un repository Github et enverrez le lien en fin de semaine.
2. **Préparation des données** : Dans cette étape, vous devrez recenser les différentes techniques utilisées en analyse de trading, identifier et créer les indicateurs pouvant vous être utiles pour la suite. Vous créerez un notebook dédié uniquement à la préparation de données que vous hébergerez également sur Github.
3. **Modélisation** : Lors de cette étape, il va vous falloir opter pour le meilleur système prédictif possible. Pour cela, vous disposez d'une très grande base de connaissance, il vous faudra opter pour quelques modèles à confronter et opter pour le meilleur et l'implémenter dans un nouveau notebook qui sera encore une fois à mettre dans votre github.
4. **Évaluation** : Vous devrez soumettre vos prédictions sur Kaggle afin d'être positionné dans le leaderboard de la compétition, si vous êtes dans le top, il y a un très gros prix à gagner ! Vous continuerez l'étape d'optimisation de votre modèle jusqu'à la fin de la compétition, le 2 février 2022.
5. **Déploiement** : Pour finir, il vous restera 1 mois pour créer une application afin d'exploiter au mieux votre modèle. Vous devrez réaliser une application du type que voulez : dashboard avec un affichage temporel des évolutions, une application suivant les cours boursiers et interrogeant votre modèle prédictif, etc.

En cas de problème

Si jamais vous bloquer durant le projet, n'hésitez pas à demander à la communauté de kaggle ou stackoverflow en premier, des pratiques extrêmement courante dans la vie professionnelle. Si jamais votre question reste malgré tout sans réponse, vous pouvez bien sûr demander à votre formateur. Attention il ne répondra qu'avec un lien de votre question sur une des deux plateformes citées précédemment !

Quelques conseils

Soyez méthodiques et allez-y pas à pas.

Ayez confiance en vos capacités, vous avez tout ce qu'il faut pour réussir !