

DATA KONTEST

KPMG a le plaisir de vous accueillir pour cette première édition du Data Kontest.

Ce concours vous permettra d'appliquer vos compétences en Machine Learning et Intelligence Artificielle à des problématiques du monde de la finance, et de vous mesurer à d'autres équipes de data-scientists motivées et innovantes. Tout au long de la semaine, vous serez soutenus, conseillés, orientés... par nos équipes de professionnels des sujets de data appliqués à la finance. Profitez-en, leurs conseils sont à consommer sans modération.

A vos machines, prêts, partez !

Objectif

Le monde de la finance est l'un des secteurs les plus demandeurs lorsqu'il s'agit d'améliorer les processus via l'utilisation de méthodologie de machine learning et d'intelligence artificielle.

Utilisateurs avertis de modèles divers (modèles de valorisation, modèles de calcul des risques de marché...) les opérateurs (Traders, Structurers, Gérants de Fonds...) réfléchissent depuis plusieurs années déjà à l'utilisation de méthodes automatisées pour améliorer les processus et choix d'investissement. L'une des utilisations les plus connues de ce type de méthodes est le trading algorithmique qui consiste à automatiser les transactions permettant de tirer avantage des anomalies de marché (on parle d'« arbitrage »).

Votre objectif aujourd'hui est d'aborder un sujet qui fait aujourd'hui l'objet de nombreux travaux de recherche au sein des banques et des fonds d'investissement : comment prévoir l'évolution à court terme (1 à plusieurs jours) des cours de bourse et construire des stratégies d'investissement efficaces sur cette base.

Vous disposez d'un capital de 1 000 000 €, et entraînez un algorithme permettant de prévoir la hausse ou la baisse de cet indice le jour suivant. Vous testerez cet algorithme en comparant les rentabilités suivantes :

J'achète pour 1 000 000 € de titres CAC40 à $T_{Initial}$ et les revends à T_{Final}

J'achète pour 1 000 000 € de titres CAC40 lorsque mon modèle d'apprentissage prévoit une hausse, et revends ces titres lorsque le modèle prévoit une baisse. Seule une décision d'investissement peut-être prise par jour (et ce entre $T_{Initial}$ et T_{Final}).

Présentation des données de l'approche

Vos travaux s'appuieront sur des données historiques de cours de plusieurs indices boursiers et indicateurs économiques extraits des principales plateformes financières parmi lesquelles : le CAC40, le S&P500, le FTSE... etc. Une liste détaillée exhaustive de ces indices, de leurs descriptions et valeurs vous est fournie au format Excel.

Vous pourrez par ailleurs alimenter votre modèle en données externes de votre choix. L'utilisation du Sentiment Analysis sur la base de Tweet est un exemple, et nous vous engageons à être créatifs (« Think Out of The Box »).

N'oubliez pas à titre indicatif qu'un projet de modélisation inclut à minima les étapes que suivent :

- **Nettoyage des données** : il convient de bien s'assurer que les données sont consistantes, sans valeurs aberrantes ni manquantes,
- **Analyse exploratoire** : il s'agit de capter les différentes informations pertinentes que renferment vos données et de les présenter (sous formes de graphes ou de test statistiques). Vous serez amenés à faire de l'analyse bivariée ou multivariée,
- **Modélisation à l'aide du Machine Learning** : implémentation d'algorithme correspondant au cas étudié,
- **Evaluation et validation du modèle sur une base test** : calcul et présentation des métriques de performance adaptés,
- **Amélioration du modèle** : construction de nouvelles variables, utilisation d'autres algorithmes, Hyper optimisation de paramètre ou calibrage du modèle ... etc.
- Vous pourrez vous appuyer sur des librairies externes, du moment que vous êtes en mesure d'en comprendre et d'en expliquer le fonctionnement.

Planning du concours

Planning du Data Kontest		
Horaires	Matin	Après-Midi
Lundi 19/03		14h00-17h00 - Accueil des étudiants - Lancement du concours - Début des travaux
Mardi 20/03		
Mercredi 21/03		14h00-17h00 - Suivi d'avancement
Jeudi 22/03		
Vendredi 23/03		14h00-17h00 - Suivi d'avancement
Samedi 24/03		
Dimanche 25/03		
Lundi 26/03	10h00-13h30 - Soutenances	13h30-14h30 - Délibérations 14h30-15h30 - Annonces de l'équipe gagnante - Verre de clôture

Les professionnels seront à votre disposition lors des sessions de travail dans nos locaux pour répondre à vos questions et vous donner leur avis. N'hésitez donc pas à les solliciter.

Critères d'évaluation

Chaque groupe sera évalué sur la base des informations présentées dans le cadre de la soutenance. Cet exercice de présentation de la démarche et des résultats obtenus durera 25 minutes, dont 15 minutes de présentation des travaux, et 10 minutes de questions réponses.

Le jury, composé d'enseignants chercheurs spécialistes de sujets de statistiques / data-science et de spécialistes KPMG sur les sujets data et finance, s'appuiera sur les éléments suivants pour départager les équipes :

- Qualité de la présentation écrite (présentation PowerPoint) : le professionnalisme, la qualité / clarté des slides seront pris en compte ;
- Qualité de la présentation orale: le partage de la parole au sein du groupe, ainsi que la qualité / clarté des explications feront partie intégrante des critères de choix ;
- La qualité des travaux, précision des analyses et résultats obtenus, la maîtrise des techniques utilisées seront également analysés par le jury.

Bonne chance à tous !