



Proposition de sujet d'IMPACT

Extraction de périodicité dans des données séquentielles.



Contexte et enjeux :

Au sein du DataLab de Crédit Agricole SA, l'équipe Data Science est amenée à réfléchir sur de nombreux sujets de recherche et développement. L'analyse de données séquentielles en est un exemple.

Les champs d'applications de l'analyse des séries temporelles sont nombreux. On pourra citer, à titre d'illustration : les données économiques, les données transactionnelles, les cours d'instruments financiers, les logs de serveurs informatiques,...

Sujet :

Dans le cadre du projet IMPACT de l'Ecole Centrale de Lille, l'étudiant devra proposer et tester une ou plusieurs méthodes pertinentes pour déterminer et extraire la périodicité de données séquentielles volatiles et fortement bruitées. Cette méthode devra être efficace en termes de temps de calculs (volumétrie à traiter de plusieurs centaines de Go) et produire des résultats pertinents.

Dans un premier temps, les développements effectués par l'étudiant se feront sur un jeu de données fictives fourni par l'encadrant. Dans un second temps, ceux-ci viendront enrichir le savoir-faire de l'équipe Data Science.

Disposant d'une grande liberté dans la méthodologie à adopter, l'étudiant devra être force proposition et prêt à explorer et implémenter de nouvelles méthodologies de machine learning, d'optimisation, de traitement de l'information pour parvenir à apporter une solution concrète.

Ce sujet peut être adapté pour un mémoire de recherche du master 2 de mathématiques appliquées. Dans ce cas, un effort de formalisation et de généralisation du problème sera à prévoir. Le sujet pourra également être prolongé dans le cadre d'un stage de fin d'étude.

Encadrant Crédit Agricole :

Emmanuel-Lin Toulemonde

emmanuel-lin.toulemonde@credit-agricole-sa.fr

Tel : +33 (0)1 43 23 15 45