



Proposition de sujet d'IMPACT

Contexte:

Dataiku est une startup Française basée à Paris et fondée en 2013 qui édite un logiciel collaboratif de traitement de la donnée appelé Dataiku DSS (Data-Science Studio). Ce logiciel permet de réaliser des projets data de A à Z, depuis la connexion native aux données sources jusqu'à la mise en production (en passant par les étapes de nettoyage, enrichissement, visualisation et machine learning).

Aujourd'hui Dataiku compte une centaine d'employés répartis dans ses trois bureaux de Paris, New-York et Londres, et vient de réaliser une levée de fond (série B) de 28M d'euros.

Ce projet d'IMPACT sera réalisé sous la supervision de l'équipe de Data-Scientist basée à Paris et visera à enrichir le logiciel avec des plugins d'analyse automatique de séries temporelles.

Projet:

L'idée de ce projet est de pouvoir intégrer des outils d'analyse/prédiction de séries temporelles utilisables en « click-bouton » au sein du logiciel (via la réalisation de plugin).

On souhaiterait idéalement pouvoir lancer en « un click » l'entraînement de N modèles de Machine Learning pour prédire une série temporelle multidimensionnelle et ceci grâce à la création automatique de variables via des packages Python/R comme TsFresh ou Prophet.

Concrètement, ce projet impliquera de l'étudiant qu'il soit capable de :

- explorer la littérature sur le sujet
- tester plusieurs solutions et package disponibles
- coder un plugin en python pour intégrer la solution retenue au sein du logiciel.

Contact : Pierre PFENNIG
pfennig@dataiku.com