

Instructions sur le rendu du projet 4IF « Fouille de Données » de Janvier 2022

Enseignants : Jean-François Boulicaut, Irène Gannaz, Diana Nurbakova

Le projet est à réaliser en binôme. Un rapport est à fournir et sera noté. Il est dit dans le sujet que vous répondez à un appel d'offre mais il est clair qu'ici nous sommes dans la phase préliminaire de démonstration des possibilités (« Proof of Concept ») pour la découverte partiellement automatisée de centres d'intérêts sur une agglomération comme celle de Lyon.

Ce rapport doit comporter deux parties. La première va être un compte rendu d'expérimentations (de 10 à 15 pages sans compter les insertions de figures ou de copies d'écran). Elle doit montrer votre effort dans la mise en œuvre des méthodes demandées (sélection et préparation des données, algorithmes de classification automatique, méthodes de caractérisation des groupes calculés, etc). Les étapes du processus de découverte de connaissances doivent être discutées et vous devez montrer votre compréhension et votre rigueur (méthodologie scientifique) dans les décisions prises (données sélectionnées, choix des composants lorsque c'est non trivial, paramétrage des composants et notamment choix des principaux paramètres des algorithmes utilisés). Vous devez également discuter des principaux résultats obtenus lors de ces expérimentations (qu'avez-vous trouvé dans les données finalement traitées ? Doit-on et peut-on interpréter les groupes de points calculés avec les différents algorithmes ... en général ? ... sur un groupe calculé spécifique ? Quel est l'intérêt pratique de la classification automatique de données spatio-temporelles ? ... Dans une seconde partie, vous devez répondre de façon synthétique (5 pages au maximum) à destination des clients visés : est-il possible de réaliser de la découverte de centres d'intérêts au moyen de workflow KNIME sur des données comme des enregistrements FLICKR ? Il s'agit ici d'extrapoler les expérimentations réalisées pour répondre à cette question, quelles sont les qualités mais aussi les limites de ce que nous avons fait pour développer de nouveaux services basés sur la découverte de centres d'intérêts à partir de photos géo-localisées ? Quelles recommandations pour les équipes qui doivent poursuivre ce développement ?

Le document à rendre une semaine après votre dernière séance.