NOUVEAU : **mise à disposition de Mise à disposition de données supplémentaires**

En complément du premier jeu de données diffusé an Avril 2016 et comportant les fichiers X_geoloc_egc_t1.csv, X_tree_egc_t1.csv et Y_tree_egc_t1.csv**, un second jeu de données** X_geoloc_egc_t2.csv, X_tree_egc_t2.csv et Y_tree_egc_t2.csv **est à présent disponible sur le site** https://egc2017.imag.fr/defi. Il s'agit de données supplémentaires ayant le même format que le jeu 1.

Un troisième jeu, utilisé pour évaluer les soumissions sur la première tâche (Prédiction de défaut), sera diffusé fin août. Il comportera uniquement les fichiers X_geoloc_egc_t3.csv et X_tree_egc_t3.csv. Les participants à cette tâche de prédiction devront renvoyer en plus de leur article, un fichier de résultats contenant leur prédiction et respectant le format des fichiers Y_tree_egc_t1.csv et Y_tree_egc_t2.csv. La version finale de la soumission devra donc prendre la forme d'une archive zip ou tgz contenant le fichier résultat (fichier csv) et les sources de l'article (fichiers tex, images, styles RNTI éventuellement) et un fichier pdf de contrôle.

Défi 2017 : Un défi vert pour Grenoble !

Pour cette seconde édition du défi EGC, Big Datext (http://www.big-datext.com), entreprise Grenobloise spécialisée dans l'analyse prédictive, et la mairie de Grenoble se sont toutes deux impliquées dans la mise en place et la diffusion de la base de données du challenge. En phase avec la politique Open Data de la Ville, visant à diffuser les données publiques de la métropole, Big Datext et les services de la Ville ont souhaité axer le défi sur les données relatives aux espaces verts.

Données :
Les données concernent des arbres situés dans la ville de Grenoble et entretenus par les services municipaux. Pour chaque arbre, on dispose de variables décrivant son type, son stade de développement, sa localisation et son environnement, son état et les traitements préconisés.

Dans un premier temps, trois fichiers de données comportant 10251 enregistrements sont mis à disposition des participants.

Dans chaque fichier, chaque enregistrement concerne un arbre et les enregistrements sont classés dans le même ordre.

- Le premier fichier (X_tree_egc_t1.csv), contient les 27 variables caractérisant l'arbre et décrites dans le classeur EGC

- le second fichier (Y_tree_egc_t1.csv) contient les variables à prédire

- le troisième fichier contient la géolocalisation de l'arbre.

Ces fichiers de données au format CSV, ainsi que le descriptif des variables (EGC_description_variables_14042016.xls, classeurs EGC et Prédiction) sont disponibles sur le site : http://egc2017.imag.fr

Objectifs :
Le but de ce défi est double.

Défi 1 :
Il consiste en une tâche de prédiction visant à déterminer, à partir des données disponibles, si

l'arbre a ou non un défaut (Variable Default or not**)** et dans l'affirmative lequel (Variables Collet, Houppier, Racine, Tronc), sachant qu'un arbre peut présenter plusieurs défauts.

Pour cette première tâche, des informations complémentaires concernant les modalités de restitution des résultats seront fournies ultérieurement aux participants, en particulier un nouveau jeu de données X_tree_egc_eval.csv contenant uniquement les 27 variables descriptives à partir duquel les participants devront produire et renvoyer un fichier ayant le même format que Y_tree_egc_t1.csv et contenant leur prédiction pour chaque enregistrement de X_tree_egc_eval.csv.

L'évaluation de la tâche de classement supervisé unilabel (prédiction de la variable Default or not) sera réalisée à l'aide des critères d'exactitude, précision et rappel. Celle de la tâche de classement supervisé multilabel à l'aide de critères d'exactitude et, de précision et rappel micro et macro (Yang et Liu, 1999).

Pour information, sur la tâche de prédiction unilabel (Variable Default or not) une baseline permet d'obtenir 86% pour l'exactitude, 82% de précision et 72% de rappel tandis que sur la tâche multilabel (Variables Collet, Houppier, Racine, Tronc) les taux sont respectivement de 70% et 47 % pour la précision et le rappel micro et de 64% et 37 % en macro.

Défi 2 :

La seconde tâche, plus ouverte, vise à appliquer des techniques d'extraction et de gestion de connaissances afin de mieux connaitre l'état du « parc végétal » de Grenoble, de mieux comprendre son évolution et de fournir des préconisations pour faciliter son entretien. Pour cette seconde tâche, les participants peuvent s'ils le souhaitent avoir recours à des données externes.

Les participants peuvent traiter au choix l'une de ces tâches ou les deux et, un retour sur la qualité des données (complétude, redondance, etc) dans un contexte open data sera apprécié.

Soumission :

Pour répondre au défi, vous devez rassembler vos résultats sur ces données dans un article au format long soumis à la conférence EGC'2017 avec la mention "Défi EGC 2017" dans le titre. Le format à utiliser est la dernière version du style LaTeX RNTI :
http://www.editions-rnti.fr/files/RNTI-X-Y2.1.zip
Les modalités de soumission et d'acceptation sont les mêmes que pour les autres articles EGC, notamment l'anonymat des soumissions.

De plus les participants au premier défi devront renvoyer un fichier de résultats au même format que Y_tree_egc_t1.csv, contenant leur prédiction pour un jeu d'évaluation qui sera fourni fin août. La version finale de la soumission devra donc prendre la forme d'une archive zip ou tgz contenant le fichier résultat (fichier csv) et les sources de l'article (fichiers tex, images, styles RNTI éventuellement) et un fichier pdf de contrôle.

Présentation :

Les papiers acceptés seront présentés lors de la conférence à Grenoble en janvier 2017, très certainement dans une session spéciale « Défi EGC ».

Attribution du prix du défi EGC 2017 : 1500 euros

Un jury se réunira pour attribuer le prix du défi EGC 2017, dans le même esprit que pour les autres

prix EGC. Les critères d'attribution seront en particulier la pertinence et la qualité de l'approche méthodologique ainsi que l'originalité et l'intérêt des résultats obtenus.

Calendrier : Les dates de soumission et de notification seront les mêmes que pour la conférence EGC 2017

Contact : Vous retrouverez tous les éléments du Défi-EGC sur la page dédiée du site de l'association EGC. http://www.egc.asso.fr/

Si vous avez d'autres questions, merci de contacter Christine Largeron (largeron@univ-st-etienne.fr) en indiquant clairement « Défi EGC 2017» dans le sujet de votre mail.