

# Prédiction de défauts dans les arbres du parc végétal Grenoblois et préconisations pour les futures plantations

Yelen Per\*, Kevin Dalleau\*\*, Malika Smail-Tabbone\*\*\*

\*LORIA UMR 7503, CNRS, yelen.per@loria.fr

\*\*LORIA UMR 7503, Université de Lorraine, kevin.dalleau@loria.fr

\*\*\*LORIA UMR 7503, Université de Lorraine, malika.smail@loria.fr

**Résumé.** Nous décrivons dans cet article notre réponse au défi EGC 2017. Une analyse exploratoire des données a tout d'abord permis de comprendre les distributions des différentes variables et de détecter de fortes corrélations. Nous avons défini deux variables supplémentaires à partir des variables du jeu de données. Plusieurs algorithmes de classification supervisée ont été expérimentés pour répondre à la tâche numéro 1 du défi. Les performances ont été évaluées par validation croisée. Cela nous a permis de sélectionner les meilleurs classifieurs uni-label et multi-label. Autant sur la tâche uni-label que multi-label, le meilleur classifieur dépasse les références d'environ 2%. Nous avons également exploré la tâche numéro 2 du défi. D'une part, des règles d'association ont été recherchées. D'autre part, le jeu de données a été enrichi avec des connaissances telles que des données climatiques (pluviométrie, température, vent) ou des données taxonomiques dans le domaine de la botanique (famille, ordre, super-ordre). En outre, des données géographiques et cartographiques sont exploitées dans un outil de visualisation d'une partie des données sur les arbres.

## 1 Introduction

Les deux tâches du défi vert de Grenoble ont été abordées. La première tâche de prédiction, visant à prédire si les arbres présentent des défauts ou non, est un problème de classification supervisée. Dans un premier temps, les données ont été analysées afin de s'assurer d'un corpus d'apprentissage le plus exploitable possible. Dans un second temps, quelques algorithmes de classification sélectionnés ont été testés et évalués sur le jeu de données. La seconde tâche est quant à elle axée sur une meilleure connaissance de l'état ainsi que de l'évolution du "parc végétal" de Grenoble. Les contributions proposées sont au nombre de trois :

- une prise en compte de données climatiques ou de données sur la classification botanique des arbres ;
- une recherche de règles d'association ;
- un outil de visualisation du parc arboricole Grenoblois.

Le logiciel libre WEKA, offrant une implémentation des principaux algorithmes de classification, a été utilisé dans ce travail (Witten et Frank (2005)). Des programmes complémentaires ont été écrits principalement pour l'ingénierie des données.