

# Prédiction des adventices de la canne à sucre à La Réunion : une approche par apprentissage supervisé multi-labels

Frédéric Fabre Ferber<sup>\*,\*\*\*</sup>, Thomas Le Bourgeois<sup>\*\*</sup>, Marion Schwartz<sup>\*\*\*</sup>, Aude Ripoche<sup>\*\*\*</sup>, Dominique Gay<sup>\*</sup>, Sandrine Auzoux<sup>\*\*\*</sup>

<sup>\*</sup>LIM, Université de La Réunion, Sainte Clotilde, La Réunion  
frederick.fabre@hotmail.com, dominique.gay@univ-reunion.fr

<sup>\*\*</sup>CIRAD, UMR AMAP, Montpellier, France  
thomas.le\_bourgeois@cirad.fr

<sup>\*\*\*</sup>CIRAD, UPR AIDA, La Bretagne, La Réunion  
marion.schwartz@cirad.fr,aude.ripoche@cirad.fr,sandrine.auzoux@cirad.fr

## 1 Introduction

Dans les régions tropicales, la nuisibilité des mauvaises herbes est peu étudiée, alors que nulle part ailleurs elles ne causent autant de pertes de rendement. À La Réunion, les conditions climatiques durant la saison de culture (chaleur et pluies abondantes) favorisent le développement des adventices, pouvant entraîner des pertes de rendements considérables dans les parcelles de canne à sucre. Connaître cette flore adventice à l'avance permettrait aux agriculteurs de mieux gérer cet enherbement et ainsi limiter les pertes de récoltes. Il n'est pas aisé de prédire un enherbement, bien que l'on puisse caractériser la tendance de certaines espèces ou groupes d'espèces à être présents à un lieu donné en fonction de facteurs agro-écologiques (Marnotte et al., 2021). Aucune référence dans la littérature ne concerne l'utilisation d'algorithmes d'apprentissage pour prédire l'enherbement d'une parcelle. Nous présentons dans cet article, les résultats d'une étude expérimentale sur des méthodes de prédiction de l'enherbement d'une parcelle et formalisons le problème de prédiction de l'enherbement comme un problème d'apprentissage supervisé multi-labels – un label représentant la présence/absence d'une adventice sur une parcelle donnée.

## 2 Protocole expérimental

La problématique de ce travail est de prédire les différentes adventices d'une parcelle en fonction de facteur environnementaux de celle-ci. Nous pouvons ramener cette problématique à un problème d'apprentissage supervisé multi-labels où nous évaluons les différents algorithmes de l'état de l'art selon plusieurs mesures de performances, plusieurs groupes d'espèces et plusieurs pré-traitements. Pour cette évaluation nous

utilisons 2 familles d'algorithmes (Ganda et Buch, 2018) : (i) Les méthodes de transformations de problème qui transforment un problème de classification multi-labels en un problème de classification binaire : Binary Relevance (BR), Classifier Chain (CC) et Label Powerset (LP). (ii) Les algorithmes adaptés qui peuvent directement travailler dans un cadre de classification multi labels : Random Forest (RF), ML- $k$ NN et MLARAM. Les données utilisées par nos différents modèles d'apprentissage sont issues du DataVerse AMATROP (Le Bourgeois et al., 2020) du CIRAD, une plateforme open-source pour le stockage de données scientifiques. Après cette évaluation, nous comparons les prédictions des algorithmes les plus efficaces avec des profils écologiques, couramment utilisés en écologie pour analyser le lien entre une espèce et les différentes modalités d'une variable descriptive. Cette comparaison est exprimée en terme d'un pourcentage de concordance.

### 3 Résultats

L'évaluation montre que les algorithmes ML- $k$ NN et ML-ARAM apparaissent être les plus performants pour prédire la présence/absence des adventices. Malgré les scores assez hauts de ML- $k$ NN et de ML-ARAM (entre 70 et 82% de bonnes prédictions en fonction des groupes d'espèces) les résultats des prédictions d'espèces ne sont pas systématiquement en concordance avec les résultats des profils écologiques, ce qui peut laisser entendre que la quantité de données pour cette étude est peut être insuffisante ou que les facteurs environnementaux pris en compte dans les jeux de données de l'étude ne sont pas encore suffisants pour prédire correctement la flore adventice. L'intégralité de ces résultats est disponible dans le mémoire de fin d'étude de Fabre Ferber (2021).

### Références

- Fabre Ferber, F. (2021). Prédire l'enherbement par apprentissage supervisé pour optimiser le contrôle des adventices - Agritrop.
- Ganda, D. et R. Buch (2018). A survey on multi label classification. *Recent Trends in Programming Languages* 5(1), 19–23.
- Le Bourgeois, T., S. Auzoux, P. Marnotte, et B. Fayolle (2020). Amatrop : Tropical weed studies dataverse.
- Marnotte, P., L. Le Mezo, J. Martin, J.-J. Esther, D. Marion, et T. Le Bourgeois (2021). La flore des mauvaises herbes de la canne à sucre à La Réunion - Caractérisation à partir des témoins des essais d'herbicides - 2005 - 2016. Publisher : CIRAD.

### Summary

In agricultural surveys, weeds are considered as the most harmful pests. We propose a comparative study of multi-label classification algorithms to predict their weediness. The results show that the ML-ARAM and ML- $k$ NN are the most efficient with a need to further improve the predictions.