



Prédire l'enherbement par apprentissage supervisé pour optimiser le contrôle des adventices

Frédéric Fabre Ferber

Référent Entreprise :
Sandrine **Auzoux**

Référent Université :
Remy **Courdier**

- 1 Introduction
- 2 Démarche expérimentale
- 3 Résultats
- 4 Conclusion
- 5 MOOC suivi
- 6 Questions

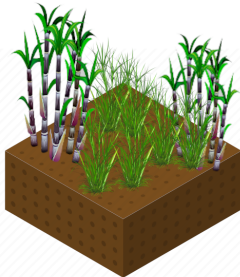
Weeds in tropical regions

- A wide variety of species
- Most harmful pests
- Favored by heat and abundant rainfall
- Result in significant yield losses

Consequences of Weeds for Sugarcane plots in Reunion Island



Maintained Sugarcane plot



Sugarcane plot invaded by weeds

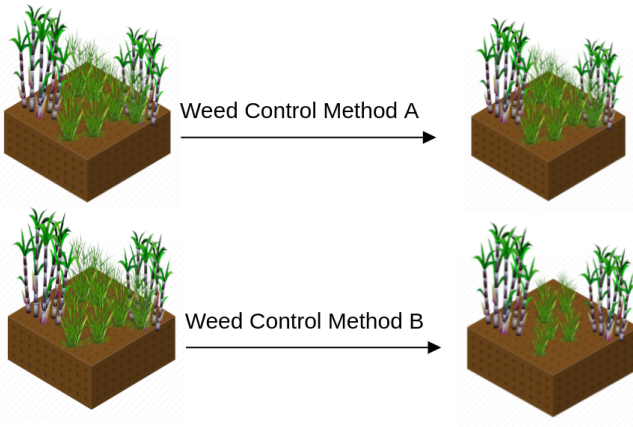
- Yield loss of 400 to 500 kg of cane per hectar per day
- Difficulty in controlling the weed flora when the plot is invaded

Exemple of invasion by weeds



Location : (Les colimaçons, Reunion Island)

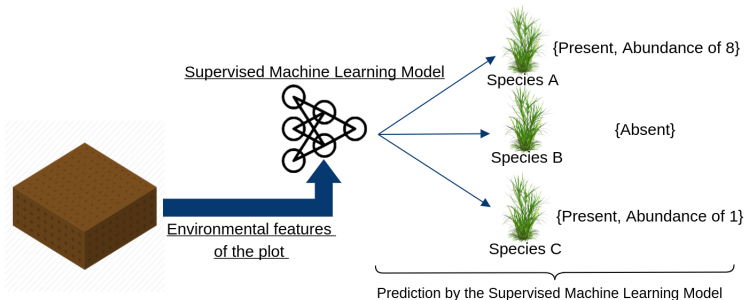
Weed Control Methods



Problem : Often, a Weed Control method works for certain species, **not all**.

Solution : Knowing in advance which weeds will be encountered.

Supervised Machine Learning for weed control



Method : Predicting weed presence and abundance using supervised machine learning

- Predict the presence : **Multi-labels classification**
- Predict the abundance : **Multi-outputs regression**

Research Question

What are the most efficient multi-labels classification algorithms to predict the presence/absence of weeds according to environmental factors - and what are the most efficient multi-outputs regression algorithms to predict the abundance of the predicted present species ?

1 Introduction

2 Démarche expérimentale

Sélection et transformation des données

Apprentissage machine supervisé - Cadre multi-labels

Test des différents algorithmes d'apprentissage machine

3 Résultats

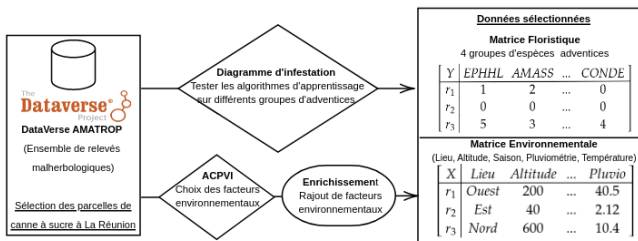
4 Conclusion

5 MOOC suivi

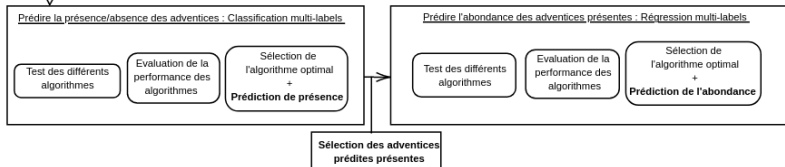
6 Questions

Schéma synthétique de la démarche expérimentale mise en oeuvre dans l'étude

Etape 1 : Sélection et transformation des données



Etape 2 : Test des différents algorithmes d'apprentissage machine



1 Introduction

2 Démarche expérimentale

Sélection et transformation des données

Apprentissage machine supervisé - Cadre multi-labels

Test des différents algorithmes d'apprentissage machine

3 Résultats

4 Conclusion

5 MOOC suivi

6 Questions

Jeu de données de l'étude



- Dataverse AMATROP du Cirad
- 30 études malherbologiques dans 11 pays
- 9180 relevés malherbologiques

Sélection des études pour la canne à sucre à La Réunion

Représentation d'une étude malherbologique

X	Lieu	Altitude	...	Pluvio
r_1	Ouest	200	...	40.5
r_2	Est	40	...	2.12
r_3	Nord	600	...	10.4

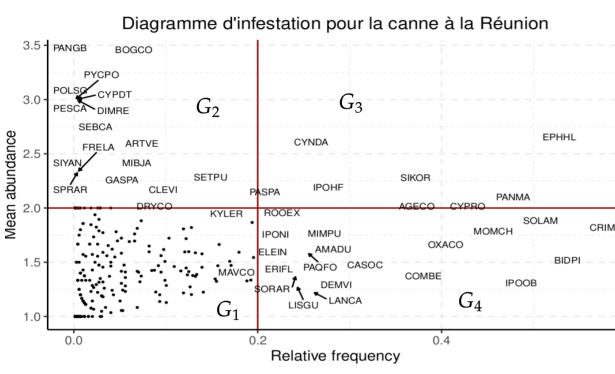
Y	EPHHL	AMASS	...	CONDE
r_1	1	2	...	0
r_2	0	0	...	0
r_3	5	3	...	4

Étude malherbologique composée d'une matrice environnementale X et d'une matrice floristique Y .

Sélection des facteurs environnementaux pertinents

- **ACPVI sur les deux matrices** → Lieu, Altitude, Saison
- **Enrichissement des facteurs environnementaux** → Mois, Année, Longitude, Latitude, Température, Pluviométrie

Sélection des adventices



- Espèces Globales : G_2 , G_3 et G_4
- Espèces Complètes : Espèces globales + espèces pertinentes de G_1
- Espèces Majeures-locales : G_2
- Espèces Générales : G_4

1 Introduction

2 Démarche expérimentale

Sélection et transformation des données

Apprentissage machine supervisé - Cadre multi-labels

Test des différents algorithmes d'apprentissage machine

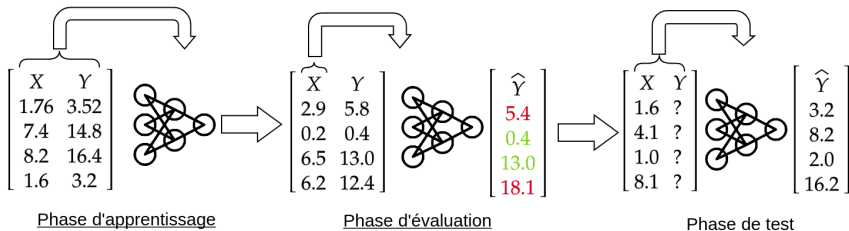
3 Résultats

4 Conclusion

5 MOOC suivi

6 Questions

Principe de l'apprentissage machine



- Données de bonne qualité
- Quantité de données suffisante

Apprentissage supervisé Multi-labels

Plusieurs variables à prédire

$$\left[\begin{array}{c|cccc} Y & EPHHL & AMASS & \dots & CONDE \\ r_1 & 1 & 2 & \dots & 0 \\ r_2 & 0 & 0 & \dots & 0 \\ r_3 & 5 & 3 & \dots & 4 \end{array} \right]$$

	$ Y = 1$	$ Y > 1$
$Card(Y) = 2$	Classification Binaire	Classification multi-labels
$Card(Y) > 2$	Classification Multi-classe	Classification multi-sorties
Variable continue	Régression	Régression multi-labels

Pré-traitements sur la matrice environnementale

Pas de pré-traitements

Standardisation

Polynomial Features Construction (PFC)

PFC + Gaussian Random Projection

1 Introduction

2 Démarche expérimentale

Sélection et transformation des données

Apprentissage machine supervisé - Cadre multi-labels

Test des différents algorithmes d'apprentissage machine

3 Résultats

4 Conclusion

5 MOOC suivi

6 Questions

Algorithmes de classification multi-labels

① Prédire la présence/absence des adventices

- **Méthode de transformation :**
 - Binary Relevance
 - Classifier Chain
 - Label Powerset
- **Algorithmes adaptés :**
 - Multi-labels KNN (MLkNN)
 - Random Forest
 - Multi-labels Hierarchical ARAM Neural Network (MLARAM)
- **Mesures de performance :** Hamming-Loss, F1, LRAP, Ranking-Loss

Algorithmes de régression multi-labels

② Prédire l'abondance des adventices présentes

- **Méthode de transformation :**

- Binary Relevance
- Regressor Chain

- **Algorithmes adaptés :**

- KNN Regressor
- Multi-target Regression Tree
- Random Forest Regressor
- Régression Linéaire
- Régression Ridge
- Régression LASSO

- **Mesures de performance :** MSE, RMSE, R^2

- 1 Introduction
- 2 Démarche expérimentale
- 3 Résultats**
- 4 Conclusion
- 5 MOOC suivi
- 6 Questions

Résultats pour la présence/absence des adventices

Performance des algorithmes pour les espèces globales (Jeu de données enrichi)

Algo	1-HM				F1				LRAP				1-RL			
	NP	SD	PCF	RM	NP	SD	PCF	RM	NP	SD	PCF	RM	NP	SD	PCF	RM
RF	0,85	0,85	0,85	0,84	0,62	0,62	0,62	0,58	0,15	0,15	0,15	0,15	0,1	0,1	0,1	0,12
LP	0,78	0,78	0,79	0,78	0,1	0,08	0,09	0,16	0,53	0,53	0,53	0,53	0,8	0,79	0,79	0,79
BR	0,85	0,85	0,85	0,84	0,62	0,62	0,63	0,59	0,15	0,15	0,15	0,15	0,11	0,11	0,11	0,13
CC	0,84	0,84	0,84	0,83	0,58	0,58	0,59	0,53	0,15	0,15	0,15	0,15	0,1	0,1	0,1	0,12
KNN	0,84	0,84	0,84	0,84	0,61	0,6	0,61	0,58	0,16	0,16	0,16	0,16	0,1	0,1	0,1	0,11
MLkNN	0,84	0,84	0,84	0,82	0,57	0,56	0,56	0,5	0,64	0,63	0,64	0,6	0,86	0,86	0,86	0,84
MLARAM	0,81	0,8	0,78	0,76	0,4	0,42	0,35	0,46	0,62	0,6	0,52	0,54	0,7	0,71	0,75	0,8

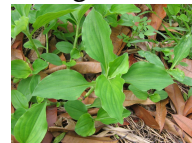
Performance des algorithmes pour les espèces majeures locales (Jeu de données enrichi)

Algo	1-HM				F1				LRAP				1-RL			
	NP	SD	PCF	RP	NP	SD	PCF	RP	NP	SD	PCF	RP	NP	SD	PCF	RP
RF	0,97	0,97	0,97	0,96	0,17	0,16	0,16	0,08	0,51	0,51	0,51	0,51	0,5	0,5	0,5	0,52
LP	0,97	0,96	0,96	0,96	0,15	0,15	0,15	0,11	0,88	0,87	0,87	0,85	0,97	0,97	0,96	0,96
BR	0,97	0,96	0,96	0,96	0,22	0,2	0,2	0,19	0,51	0,51	0,51	0,51	0,5	0,5	0,5	0,5
CC	0,96	0,96	0,96	0,96	0,21	0,21	0,21	0,21	0,51	0,51	0,51	0,51	0,5	0,5	0,5	0,52
KNN	0,96	0,97	0,96	0,96	0,18	0,21	0,2	0,19	0,51	0,51	0,51	0,51	0,49	0,49	0,49	0,49
MLkNN	0,96	0,96	0,96	0,96	0,09	0,15	0,16	0,11	0,78	0,8	0,81	0,79	0,94	0,94	0,94	0,94
MLARAM	0,92	0,92	0,88	0,8	0,24	0,18	0,09	0,16	0,76	0,67	0,58	0,69	0,8	0,67	0,57	0,86

Comparaison des profils écologiques de l'espèce *Commelina benghalensis* et des prédictions de l'algorithme MLARAM

Comparaison de la prédiction et le profil écologique de COMBE		
Modalités	Prédiction présence/absence	Profil écologique
Chaude-Sud-basse	✓	✓
Chaude-Sud-haute	✓	✓
Chaude-Est-basse	✓	✓
Chaude-Est-haute	×	×
Chaude-Nord-basse	✓	✓
Chaude-Nord-haute	✓	?
Chaude-Ouest-basse	✓	✓
Chaude-Ouest-haute	✓	✓
Froide-Sud-basse	×	✓
Froide-Sud-haute	×	✓
Froide-Est-basse	×	✓
Froide-Est-haute	×	✓
Froide-Nord-basse	×	✓
Froide-Nord-haute	×	×
Froide-Ouest-basse	×	✓
Froide-Ouest-haute	×	✓
f-Froide-Sud-basse	✓	✓
f-Froide-Sud-haute	×	✓
f-Froide-Est-basse	✓	✓
f-Froide-Est-haute	×	×
f-Froide-Nord-basse	×	✓
f-Froide-Nord-haute	✓	?
f-Froide-Ouest-basse	✓	✓
f-Froide-Ouest-haute	✓	✓
f-Chaude-Sud-basse	×	✓
f-Chaude-Sud-haute	×	✓
f-Chaude-Est-basse	✓	✓
f-Chaude-Est-haute	×	×
f-Chaude-Nord-basse	✓	✓
f-Chaude-Nord-haute	✓	?
f-Chaude-Ouest-basse	×	×
f-Chaude-Ouest-haute	×	✓

*Commelina
benghalensis*



Comparaison des profils écologiques pour la modalité "Chaude-Ouest-Basse" et des prédictions de l'algorithme MLkNN

Comparaison entre la prédiction de MLkNN et le profil écologique pour la modalité "Chaude-Ouest-Basse" pour le groupe des espèces Globales

✓ : Présente ; X : Absente

Espèces prédites présente	Profil écologiques
COMBE	✓
ROOEX	✓
SORAR	✓
AMADU	✓
AGECO	✓
BIDPI	X
MAVCO	✓
OXACO	✓
PAQFO	✓
CRIMI	✓
CYPRO	✓
CYNDA	✓
SIKOR	✓
EPHHL	✓
POROL	✓

Résultats pour la présence/absence des adventices

- MLARAM et MLkNN plus performants en moyenne (70 à 82% de performance)
- Le profil écologique ne correspond pas systématiquement aux prédictions de MLARAM et MLkNN
- Certaines mesures de performances se contredisent entre elles
- Le groupe des espèces majeures locales biaise les mesures de performance
- Le jeu de données avec les facteurs environnementaux enrichis augmente la performance

Résultats pour l'abondance des adventices

- Dépendent des résultats pour la présence/absence des adventices
- Présentés à la fin du stage, en cours de calcul

- 1 Introduction
- 2 Démarche expérimentale
- 3 Résultats
- 4 Conclusion**
- 5 MOOC suivi
- 6 Questions

Conclusion

Quels sont les algorithmes de classification multi-labels les plus efficaces pour prédire la présence/absence des adventices en fonction des facteurs environnementaux et quels sont les algorithmes de régression multi-labels les plus efficaces pour prédire l'abondance des espèces préalablement prédites présentes ?

- Les algorithmes MLARAM et MLkNN sont les plus efficaces pour prédire la présence/absence des adventices

Questionnements

- Pourquoi certaines mesures de performance se contredisent entres elles ?
- D'où viennent les contradictions entre les profils écologiques et les prédictions des algorithmes ?
- La faible quantité de données biaise t-elle la performance des algorithmes ?
- Les facteurs environnementaux du jeu de données sont-ils assez explicite pour expliquer la flore adventice ?
- Quelle est la sensibilité de ces algorithmes pour différents types et tailles de jeux de données ?

Apports du stage

- Personnels
 - Réponse à la question de recherche
 - Engouement pour le monde de la recherche
 - Regard critique et scientifique
- Techniques
 - Consolidation des connaissances en apprentissage machine
 - Nouvelles connaissances dans des domaines plus spécifiques de l'apprentissage machine
 - Nouvelles connaissances en malherbologie et en botanique
- Professionnels
 - Rencontres intéressantes dans un cadre pluri-disciplinaire
 - Immersion dans un organisme de recherche

1 Introduction

2 Démarche expérimentale

3 Résultats

4 Conclusion

5 MOOC suivi

6 Questions

Extraction et Gestion des Connaissances 2021 : Techniques d'apprentissage

- Événement annuel réunissant des chercheurs et praticiens de disciplines relevant des sciences des données et des connaissances
- Formation sur deux jours : 25 et 26 Janvier 2021
- Thèmes abordés :
 - Apprentissage supervisé avec Python
 - Text Mining
 - Techniques de Deep Learning avec Keras et Pytorch

- 1 Introduction
- 2 Démarche expérimentale
- 3 Résultats
- 4 Conclusion
- 5 MOOC suivi
- 6 Questions**

Réponses aux questions



Merci de votre attention !

Avez-vous des questions ?

