

# BioFin, un algorithme génétique pour l'optimisation des modèles de machine learning

## Edouard GOFFINET

Travail de bachelor 2021  
Informatique - Développement Logiciel et Multimédia  
Professeur: Emmanuel DE SALIS  
Expert: Jérôme MORET

### Description

Ce travail est développé en collaboration avec l'équipe de recherche de la HE-Arc dans le cadre d'un projet de recherche mandaté par l'entreprise Perréard Partners Investment.

L'objectif principal du travail est la réalisation d'une bibliothèque Python permettant l'optimisation des hyperparamètres de différents modèles d'apprentissage automatique.

La bibliothèque développée sera comparée aux autres méthodes populaires utilisées pour l'optimisation d'hyperparamètres, notamment GridSearchCV et RandomizedSearchCV de la bibliothèque scikit-learn.

Une version expérimentale prenant en charge Keras a également été développée et est fonctionnelle.

Le module développé, `genetic_search`, est disponible en deux versions : la première utilisant le framework DEAP pour la gestion de l'algorithme génétique et la deuxième utilisant un module développé durant le projet pour remplacer DEAP.

### Déroulement

Le projet s'est déroulé de la manière suivante :

- Approfondissement algorithmie génétique et scikit-learn
- Réalisation de GeneticSearchCV avec DEAP
- Réalisation d'un module d'algorithmie génétique "genetic\_algorithm"
- Développement de la version 2 de GeneticSearchCV avec le module "genetic\_algorithm"
- Adaption de GeneticSearchCV pour Keras
- Evaluation durant tout le projet de la méthode développée
- Analyse et rédaction du rapport

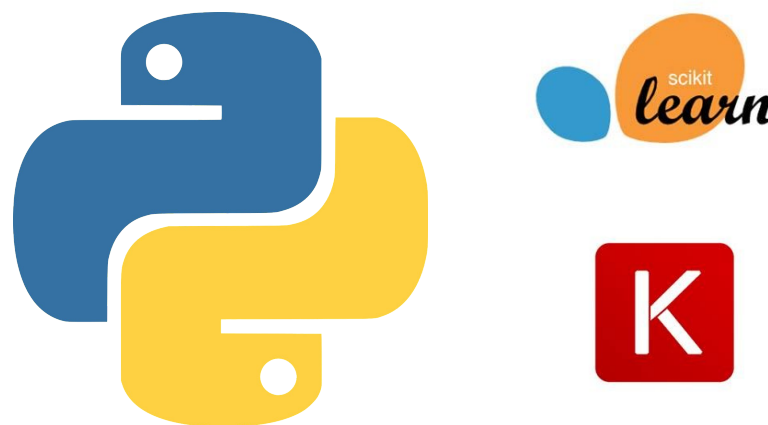
### Résultats

La bibliothèque est terminée et fonctionnelle en deux versions, ce qui permet une grande personnalisation de l'algorithme génétique afin de s'adapter à différents modèles et jeu de données.

Il ressort des multiples évaluations menées durant le travail de Bachelor que GeneticSearchCV performe aussi bien que ses concurrents.

En effet, GeneticSearchCV offre plus de stabilité que les autres méthodes rapides, comme la recherche aléatoire. Elle ne garantit pas la solution optimale comme la recherche par force brute, mais consomme bien moins de ressources. Elle est plus rapide et va entraîner beaucoup moins de modèles dans la plupart des situations.

Finalement, la recherche d'hyperparamètres à l'aide d'algorithme génétique est une méthode qui fonctionne et il ne serait pas étonnant de la voir se généraliser dans le futur.



*Tout le projet a été réalisé en Python et les frameworks scikit-learn et Keras ont été utilisés.*

### Perspectives

- Compatibilité avec Keras
- Dispersion auprès de la communauté scientifique
- Utilisation de la bibliothèque avec le jeu de données BioFin
- La compatibilité avec Keras a été explorée et une version expérimentale est disponible.