

# Automating biomedical data science through tree-based pipeline optimization

Benjamin GUIGON

January 2021

## 1 Lien

Article écrit par Olfa Lamti

<https://github.com/OlfaLmt/PSBX/blob/main/Th%C3%A8ses/Automating%20biomedical%20data%20science%20based%20pipeline%20optimization.pdf>

## 2 TPOT

Le travail de Data scientist est depuis quelques années la clé pour faire parler toutes les données que l'on génère. Les chercheurs et mathématiciens ont développé des modèles mathématiques et les ont interprété dans plusieurs langages (Python, R etc...) pour faire gagner du temps à ces Data scientist. Pourquoi "gagner du temps" ? Parce que la tâche la plus chronophage des Data scientist est le pré-processing. Le Data scientist doit choisir quelles données traiter, comment les traiter, avec quoi les traiter et optimiser ce traitement.

TPOT est un outil d'aide qui donne des idées sur la manière de résoudre un problème d'apprentissage particulier en explorant des configurations de pipeline qui ne sont pas évidentes et que l'homme pourrait ne pas trouver. Une fois cette méthode exploratoire terminée, il fixe les paramètres les plus pertinents.

## 3 Explication

La difficulté réside dans le fait que l'ensemble de tous les comportements possibles compte tenu de toutes les entrées possibles devient rapidement trop complexe à décrire, c'est de l'explosion combinatoire.

Il faut bien voir que l'autoML est une science grandissante qui à encore beaucoup à apprendre. C'est un outil d'une puissance redoutable qu'il faut absolument connaître dans le milieu de la Data science. Ce qui est intéressant c'est de voir

l'autoML comme un rival et qu'il faut s'appuyer dessus pour trouver les segments les plus intéressants. Il ne faut pas le voir (en tous cas pas tout de suite) comme un substitut au Data scientist.

## 4 Avis

L'article est vraiment très bien expliqué, c'est très clair et on comprend très facilement l'intérêt de l'autoML via TPOT. Un script python sur un exemple simple aurait fait une différence très intéressante. Mais il faut être curieux et faire les tests de son côté. Cet article donne un avant goût à une technologie qui serait très enrichissante d'apprendre avant d'arriver sur le milieu du travail parce que sa connaissance peut être un vrai plus.