

# RANDOM FOREST

## EXERCICES

### I. EXERCICE 1

1. Prédire la survie d'un passager de Titanic en utilisant l'algorithme de Random Forest.
2. Evaluer et discuter vos résultats.

### II. EXERCICE 2

A partir de mails labélisés, détecter les spams en utilisant l'algorithme de Random Forest.

Pour cela, vous disposez de fichiers contenant des mails et des spams répartis dans [deux dossiers](#) : train-mail et test-mail.

1. Ecrire une fonction qui prend en paramètre le nom d'un dossier et qui renvoie un dictionnaire avec les 3000 mots les plus utilisés dans l'ensemble des mails comme clé et le nombre d'occurrences comme valeur.  
Résultat attendu :
2. Ecrire une fonction qui prend en paramètre le nom d'un dossier et qui renvoie
  - a. Un tableau de deux dimensions (features) qui contient en colonne tous les mots les plus utilisés dans les mails et en ligne les numéros des emails. Chaque case correspond alors au nombre d'occurrence d'un mot dans un email.
  - b. Un tableau à une dimension (target) qui contient 1 s'il s'agit d'un spam et 0 dans le cas échéant.
3. Elaborer un modèle d'apprentissage de type Random Forest en utilisant les features et les targets construits avec la fonction précédente.
4. Evaluer votre modèle et discuter les résultats.