

# Chapitre 1

## Exercice 1

### 1.1 Question 1

Chaque instance est un mail qui sera déterminé comme étant un spam ou non. Il y a 4601 instances. Pour déterminer si une instance est un spam, on étudie la fréquence d'un ensemble de mots, divers calculs sur la longueur des chaînes de caractères n'étant que des lettres capitales et la fréquence de certains caractères.

### 1.2 Question 2

Reorder reorganise l'ordre d'études des attributs. Standardize va "décaler" les valeurs numériques pour que leurs moyennes soient à 0.

### 1.3 Question 3

- OneR 78.0917% Correct, 477 mails classés spams, 531 spams passent, temps 0.17 seconds
- NaiveBayes 79.2871% Correct, 865 mails classés spams, 88 spams passent, 0.12 seconds
- J48 92.9798% Correct, 156 mails classés spams, 167 spams passent, temps 0.92 seconds
- RandomForest 94.8272% Correct, 78 mails classés spams, 160 spam passent, temps 1.16 seconds
- MultilayerPerceptron 91.4366% Correct, 192 mails classés spams, 202 spams passent, temps 95.83 seconds
- SMO 90.4151% Correct, 134 mails classés spams, 307 spams passent, temps 0.63 seconds

RandomForest est relativement rapide et est le plus efficace sur les faux positifs