

# BI Data Mining - TP4 - Classification

Ludovic Denoyer et Laure Soulier

## 1 Iris

Nous allons travailler sur le jeu de données Iris disponible sous RapidMiner.

### 1.1 K-means

- Appliquer une 3-means (K-means en 3 classes) sur les données. Observer/Commenter les résultats. Vous devez retrouver :
  - L'attribution des individus dans les classes
  - La distribution des variables pour chaque classe (de façon graphique et numérique)
- Les clusters correspondent-ils aux labels assignés initialement aux données ? Dans quels cas y a-t-il une différence ?
- Faire varier le nombre de classe et les mesures. Interpréter/regarder les résultats.

### 1.2 Classification hiérarchique

- Réaliser une classification hiérarchique sur les données (Agglomerative clustering). Commenter les résultats. Vous devez retrouver :
  - L'attribution des individus à des notions de distance
  - Le dendrogramme
- Comment choisir le nombre de classes ?
- Faire varier les notions de distance. Interpréter/comparer les dendrogramme.