

# Data Mining e Scoperta di Conoscenza

## Progetto 12

### *Implementazione dell'algoritmo di clustering EM*

1. Si estenda Weka con una nuova classe `weka.core.sequence.java` che consenta la rappresentazione ottimale dei dati di input sotto forme di sequenze.
2. Si estenda Weka con una nuova classe `weka.clusterers.EM.java` che implementi l'algoritmo di clustering *EM* i cui dettagli sono riportati nel seguente articolo "A General Probabilistic Framework for Clustering Individuals and Objects " fornito insieme al progetto.
3. Si applichi l'algoritmo implementato:
  - a. sul dataset UCI
  - b. sui dati forniti contestualmente al progetto valutando come varia l'accuratezza predittiva
    - a) rispetto agli altri algoritmi di clustering forniti in Weka.
    - b) al variare dei parametri dell'algoritmo.

### **NOTE PER L'ESECUZIONE DEL PROGETTO**

1. Scrivi un rapporto di circa 10 pagine in cui
  - a. Descrivi analiticamente l'algoritmo che hai implementato.
  - b. Commenti le parti essenziali del codice Java che hai scritto, e metti in un'appendice l'intero codice
  - c. commenti e illustri graficamente e quantitativamente gli esperimenti effettuati.
2. Prepare delle slides Powerpoint (non più di 10 slides) in cui riassumi gli esiti del progetto