

# Data Mining e Scoperta di Conoscenza

## Progetto 2

### *Implementazione di una rete neurale per il riconoscimento facciale*

1. Si generi un dataset contenente  $n > 300$  immagini:
  - a.  $N_A$  immagini di un soggetto A
  - b.  $N_B$  immagini di un soggetto B
  - c.  $N_O$  immagini di altri soggetti.

Nel generare il dataset, si deve garantire che  $N_A + N_B = 10\% N_O$ . Inoltre, le immagini dovrebbero ritrarre un soggetto specifico in varie pose tutte sullo stesso sfondo (sorridente, normale, arrabbiato, di fronte, sul lato destro, sul lato sinistro, ecc.). Le immagini devono essere del formato Jpeg.
2. Si estenda Weka con la classe `weka.core.ImageInstance.java`, che implementa la lettura di un insieme di immagini Jpeg, le trasforma in un formato matriciale di  $32 \times 32$  pixels (utilizzando ricampionamento bicubico), e lo trasforma in un dataset utilizzabile in Weka.
3. Si apprenda una rete neurale che implementa il riconoscimento facciale. In pratica, la rete prende in input un'immagine e, dopo averla trasformata nel formato desiderato, la classifica come soggetto A oppure soggetto B oppure soggetto differente da A e B.
4. Si studi la performance della rete con particolare riferimento alle curve di ROC:
  - a. Come variano le performance della rete al variare dei parametri che influenzano la topologia della rete?
  - b. Come varia la velocità di apprendimento in funzione del numero di immagini usate come training?
5. Si studi l'influenza del preprocessing sulle performances della rete. In particolare:
  - a. Si studi l'effetto dell'analisi delle componenti principali sull'errore
  - b. Si determinino delle trasformazioni dei dati (ad esempio binarizzazione, normalizzazione, ecc.) che permettano un miglioramento generale delle performances sulla rete.
6. Si implementi una strategia boosting/stacking che risolva il problema dello sbilanciamento.

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

1. Scrivi un rapporto di circa 10 pagine in cui
  - a. Descrivi analiticamente l'algoritmo che hai utilizzato e la tecnica che hai usato per la valutazione.
  - 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. Prepara delle slides Powerpoint (non più di 10 slides) in cui riassumi gli esiti del progetto