

# Data Mining e Scoperta di Conoscenza

## Progetto 14

### **Realizzazione di un task per l'analisi di web clickstreams e transazioni di acquisto**

Si realizzi un task per l'analisi di alcune problematiche legate ai web clickstreams e transazioni di acquisto fatte sul sito *Gazelle.com* che vende prodotti on-line.

I dati forniti sono raccolti direttamente al negozio essenzialmente senza compiere nessun passo di pulizia degli stessi.

In base a quanto detto, si chiede di:

1. di preprocessare opportunamente i dati grezzi forniti;
2. di implementare modelli di mining capaci di rispondere alle seguenti domande:
  - a. Sulla base di un insieme di pagine web visitate da un utente, l'utente stesso visiterà una nuova pagina sul sito oppure deciderà di uscire?
  - b. Dato un set di pagine web visitate da un utente, quale marca di prodotti il visitatore deciderà di visitare nel resto della sessione?

I datasets completi, oltre ad un file contenente una spiegazione dettagliata degli attributi ivi contenuti, vengono forniti contestualmente al progetto.

### **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