

# Business Intelligence per i Big Data

*Progetti di analisi di dati*



Data Base and Data Mining Group of Politecnico di Torino

*AA 2016-2017 - Politecnico di Torino*



# Obiettivo

- Utilizzo di una (o più di una) tecnica di data mining per analizzare un dataset reale
  - Analisi del dataset per caratterizzare la distribuzione dei dati
  - Tecniche di analisi di dati
    - *Regole di associazione*
      - *Effettuare più sessioni di analisi*
      - *Variare gli indici di qualità (e.g., supporto, confidenza, lift)*
    - *Clustering*
      - *Effettuare più sessioni di analisi con un algoritmo specifico (e.g., K-Means)*
      - *Variare i parametri di input*
      - *Valutare gli indici di qualità (e.g., SSE)*



# Regole

- Gruppi di due persone
- Ogni gruppo deve
  - Caratterizzare il dataset
  - Effettuare diverse sessioni di analisi su un dataset utilizzando il tool RapidMiner
  - Analizzare i risultati e sintetizzarli in grafici



# Regole

- Preparare una breve ma completa presentazione sulle attività svolte
  - Caratterizzazione del dataset
  - Analisi effettuata (e.g., configurazione ottimale dell'algoritmo selezionato)
  - Risultati ottenuti
- Presentare i risultati in 15 minuti
  - 5 minuti di presentazione a testa e 5 di domande



# Consulenze per il progetto

## ■ Consulenza

- Giovedì 08/06 11:30-14:30 presso il LEP
- Lunedì 12/06 11:30-13:00 presso l'Aula 11
- Giovedì 15/06 11:30-14:30 presso il LEP



# Materiale da consegnare

- Preparare il materiale
  - Processo di rapid miner e file memorizzati nel repository
  - Insieme di lucidi
  - Sorgenti dei grafici (e.g., file excel)



## Date importanti

- Consegnare i lucidi via mail a Tania Cerquitelli
  - entro il giorno **20/6/2017**
    - Le presentazioni saranno svolte il 21/6 13:00-18:00 aula da definire
  - entro il giorno **2/7/2017**
    - Le presentazioni saranno svolte il 03/7 14:00-18:00 aula da definire
  - entro il giorno **18/7/2017**
    - Le presentazioni saranno svolte il 19/7 14:00-18:00 aula da definire

- *Consegnare tutto il materiale tramite pen*



# Valutazione

- Ogni studente del gruppo sarà valutato con un punteggio in trentesimi
  - In caso di lode, viene considerato 32 (per calcolare il voto finale)
- Il voto della tesina sarà mediato con il seguente punteggio
  - Voto conseguito all'esame scritto incrementato di
    - 1/30 se lo studente ha consegnato l'esercitazione su MapReduce
    - 1/30 se lo studente ha consegnato l'esercitazione sulle tecniche di classificazione
  - La lode viene riconosciuta se il voto finale è  $\geq 31$