

# Business Intelligence per Big Data

*Progetti di analisi di dati*



Data Base and Data Mining Group of Politecnico di Torino

AA 2021-2022 - *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
    - Analisi esplorativa 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 uno o più algoritmi (e.g., K-Means, DBSCAN)*
      - *Variare i parametri di input (analisi di sensitività)*
      - *Valutare i diversi esergli indici di qualità (e.g., SSE)*



# Regole

- Gruppi di due persone
  - Registrarsi sul google sheet
- Ogni gruppo deve
  - Caratterizzare il dataset
  - Effettuare diverse sessioni di analisi su un dataset utilizzando il tool RapidMiner e/o altri tools noti al gruppo di studenti
  - Analizzare i risultati e sintetizzarli in grafici
  - Discutere come sfruttare la conoscenza estratta in un'applicazione di business



# Regole

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



# Consulenze per il progetto

## ■ Consulenza

- Alcuni slot da 1,5h
- Da definire negli slot delle prossime lezioni e/o fuori orario se necessario



# 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 precedente la prova scritta
    - Le presentazioni saranno svolte a partire dal giorno successivo alla prova scritta
    - Primo e secondo appello (Giugno-Luglio)
- *Consegnare tutto il materiale tramite link ad una cartella condivisa*



# 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 sulle tecniche di classificazione
    - 1/30 se lo studente ha consegnato l'esercitazione su MapReduce/MongoDB
  - La lode viene riconosciuta se il voto finale è  $\geq 31$