

Business Intelligence per Big Data

Progetti di analisi di dati



Politecnico
di Torino

AA 2023-2024 - *Politecnico di Torino*

- Attività da svolgere
 - Scegliere uno use case tra i due proposti
 - Caratterizzare il dataset
 - Esplorazione dei dati
 - Effettuare diverse sessioni di analisi sul dataset utilizzando il tool RapidMiner e/o Python e/o altri tools noti al team
 - Almeno 2/3 algoritmi
 - Analisi di sensitività dei parametri
 - Analizzare i risultati e sintetizzarli in grafici
 - Analisi comparativa

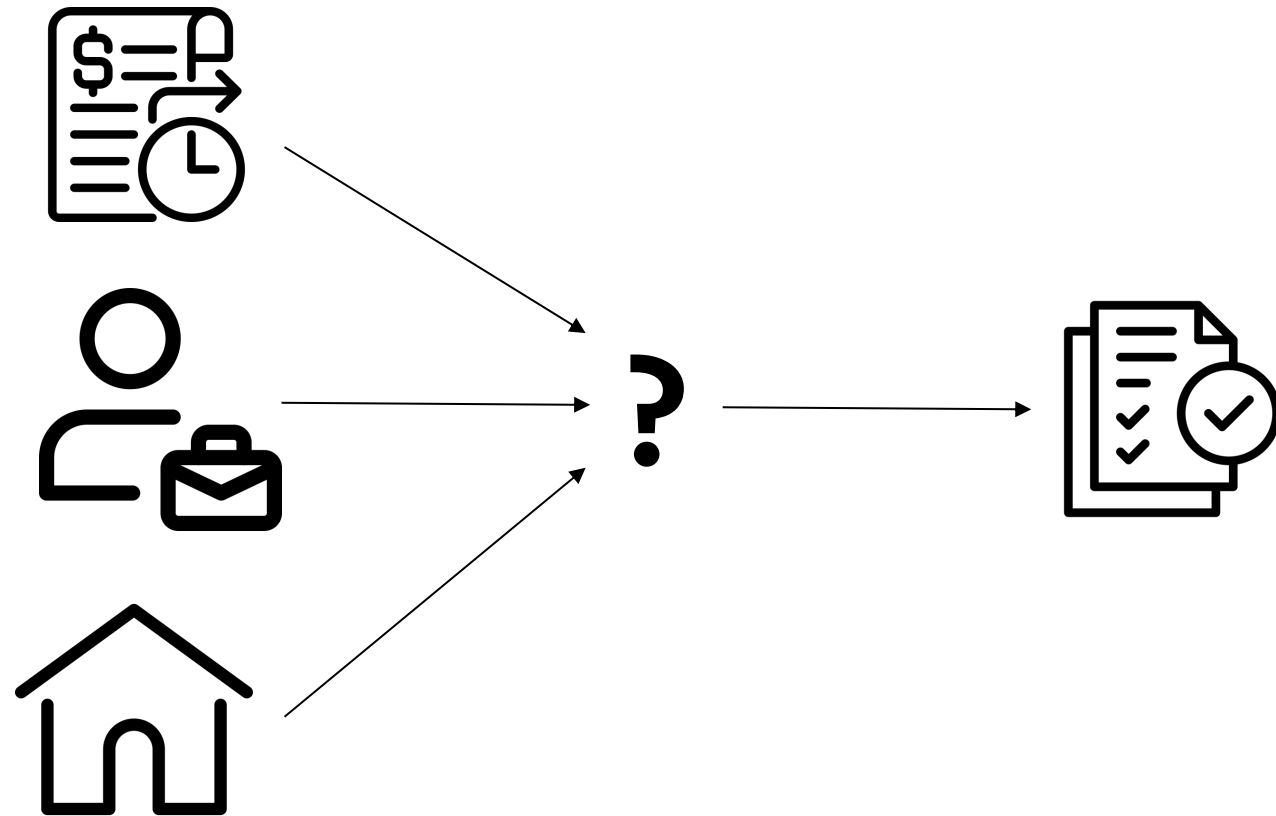
- Attività da svolgere (cont.)
 - Discutere come sfruttare la conoscenza estratta in un'applicazione di business
 - Scrivere il **report scientifico sintetico** sulle attività svolte
 - Utilizzare il template overleaf
 - Rispettare la struttura data
 - Rispettare il numero di pagine massimo (4 pagine massimo)

Formazione dei gruppi



- Formazione dei gruppi
 - Compilare il google form al link <https://forms.gle/CeQN16fze6VMj4668> entro 6 Maggio 2024
 - 1 compilazione per gruppo
- Scelta dello use case
- Informazioni dei componenti del gruppo
 - Matricola, Nome, Cognome
- Le assegnazioni degli use case ai gruppi sarà resa disponibile entro il 8 Maggio 2024

Use Case: Contratto di credito



Caratteristiche del Problema

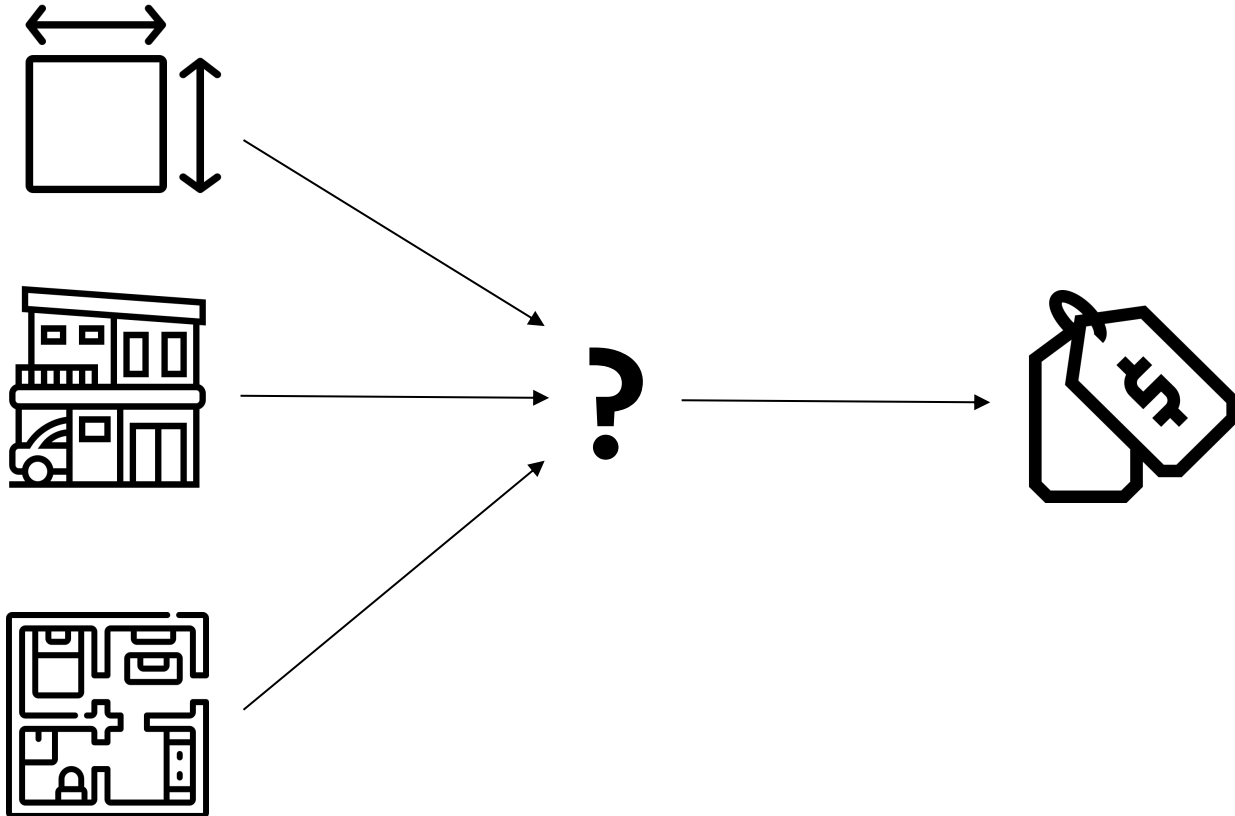


Pochi attributi (ca 20)



Classi sbilanciate

Use Case: Valore di un immobile



Caratteristiche del Problema



Classi bilanciate



Molti attributi (ca 80)

Consulenze per le attività progettuali



- Slot di laboratorio
- Mail
 - eleonora.poeta@polito.it
 - daniele.regecambrin@polito.it
 - + eliana.pastor@polito.it
- Consulenze
 - Orario e slot da concordare

Materiale da consegnare



- Il gruppo deve consegnare
 - Report scientifico sintetico
 - progetto overleaf (sorgenti latex)
 - Processo di rapid miner e file memorizzati nel repository e/o codice
 - File excel/csv con i completi dati degli esperimenti (opzionale, risultati completi e/o aggiuntivi rispetto a quanto riportato nel report)
- Il gruppo deve compilare
 - 1-2 questionari online (obbligatori)

Date di consegna



- Entro 7 giorni prima della data della prova scritta
- Il gruppo deve consegnare un'unica cartella zip con il materiale indicato in slide 'Materiale da consegnare' effettuando l'upload sul Portale della Didattica, 'Sezione Elaborati'.

- Ogni studente del gruppo sarà valutato con un punteggio in trentesimi
 - Con valutazione pari 30 e lode, viene considerato 32 (per calcolare il voto finale)
 - Completezza, Metodo, Analisi Dati, Presentazione.
- Il voto della tesina sarà mediato con il voto conseguito all'esame scritto incrementato di
 - 1/30 in caso di consegna dell'homework su Google Data studio
 - 1/30 in caso di consegna dell'homework su MapReduce/MongoDB
- La lode viene riconosciuta se il voto non arrotondato è ≥ 31