

## Obiettivo del progetto: sviluppo di un Recommendation System

Il progetto può essere svolto singolarmente o da gruppi di massimo 2 persone per gli appelli di Gennaio e Febbraio 2022.

*Progetti singoli:* devono essere svolti tutti i passi di analisi del “progetto base” più a scelta 1 analisi aggiuntiva

*Progetti di gruppo:* devono essere svolti tutti i passi di analisi del “progetto base”, e almeno 4 analisi aggiuntive.

### **PROGETTO BASE:**

Dati per l'analisi: Movie Lens 100k (<https://grouplens.org/datasets/movielens/100k/>). 100,000 ratings relativi a 1000 utenti di 1700 film.

#### **Passi di analisi base:**

1. Analisi Esplorativa (statistiche descrittive, analisi correlazione, ...)
2. Filling della matrice di rating (user-based) attraverso l'algoritmo di classificazione K-NN con cosine similarity.
3. Calcolo delle metriche di performance dell'algoritmo K-NN (MSE e RMSE).
4. Segmentazione degli utenti in base alle preferenze: algoritmo di clustering K-MEANS con cosine similarity. Visualizzazione dei cluster attraverso metodi visti a lezione.

Librerie suggerite: Surprise per il filling della matrice.

#### **Analisi aggiuntive:**

1. Individuazione del set di item da raccomandare ad ogni utente. Formulazione del problema multi-obiettivo considerando accuratezza, novelty e coverage. Sua soluzione attraverso l'algoritmo NSGA-II (Non-dominated Sorting Genetic Algorithm).
2. Analisi delle metriche di performance dell'algoritmo evolutivo (NSGA-II): Hypervolume, C-metric.
3. Filling della matrice di rating attraverso l'algoritmo K-NN usando la correlazione di Pearson oltre alla cosine similarity e confrontare i risultati ottenuti in termini di MSE e RMSE.
4. Filling della matrice di rating attraverso Matrix Factorization in aggiunta a K-NN e confrontare i risultati ottenuti in termini di MSE e RMSE.
5. Utilizzo di nuovi dataset con dimensioni maggiori. Di seguito alcuni esempi:
  - Movie Lens 1M (<https://grouplens.org/datasets/movielens/1m/>).
  - Altri esempi al seguente link (<https://grouplens.org/datasets/movielens/>).Ciascuno studente è libero di scegliere altri dataset previo confronto con il docente.
6. Strutturare i passi 1 -4 (passi di analisi progetto base) lavorando sugli items giungendo a una valutazione comparativa e critica dei risultati ottenuti.
7. Individuazione del set di item da raccomandare ad ogni utente attraverso l'algoritmo MOEA/D della libreria Pymoo. Confronto dei risultati ottenuti attraverso l'analisi delle metriche di performance (punto 2).

Librerie suggerite: Pymoo per la formulazione e risoluzione del problema multi-obiettivo.

### **Analisi “Extra strong”:**

1. Rappresentazione della matrice di rating tramite un grafo. I nodi del grafo rappresentano gli utenti che sono collegati se la loro cosine similarity è superiore ad una certa soglia.  
Librerie suggerite: pandas e networkx

## ORGANIZZAZIONE DEI RISULTATI, REPORT FINALE E PRESENTAZIONE

Ogni studente/gruppo di studenti deve produrre

- un report strutturato come segue:
  - un breve riassunto esecutivo con i principali risultati
  - un'introduzione sul problema
  - diverse sezioni che riassumono i risultati dell'analisi (raggruppati per argomento)
  - Conclusioni e interpretazione dei risultati
- una presentazione per la discussione d'esame

Ogni studente/gruppo di studenti deve inviare a Ilaria Giordani ([ilaria.giordani@unimib.it](mailto:ilaria.giordani@unimib.it)) seguendo le scadenze su Moodle il codice Python e il report prodotto.