

# Capitolo 1

## Implementazione

### 1.1 Data Preprocessing

La fase di preprocessing dei dati consiste nel preparare i dati per l'addestramento del modello. Una volta acquisiti i dati e averne studiato le peculiarità e le caratteristiche, si procede con la pulizia dei dati.

Abbiamo usato il metodo *pivot\_table* di Pandas per creare una tabella pivot che contiene i genome-score degli utenti per ogni film che lo possiede (andando così a ridurre la cardinalità da 60k a 13.816). In seguito a ciò verrà effettuata una merge con un i generi dei film applicando una One-Hot Encoding, attraverso la funzione *get\_dummies()*. Ci siamo inoltre andati a focalizzare sui voti da parte degli utenti, raggruppandoli per film ci abbiamo calcolato la media dei voti.

### 1.2 Modeling

#### Tabular Data