Cinelit - Hybrid Approach

Dataset

```
        Dataset
       /        \
   Movie        Book
(from IMDb)   (from goodreads)
```

Movie (from IMDb)

① Cleaning

② Since we have ratings, vote-count

$$\text{Weighted avg} = \frac{(\text{vote avg} * \text{vote count}) + (\text{Mean (vote avg)} * 0.8 \text{ quant})}{\text{vote count} + 0.8 \text{ quant (vote-count)}}$$

③ Calculate score $= \text{Weighted avg} * 0.4 + \text{popularity} * 0.6$

④ hybrid-df $= $ title | bag of words

⑤ Convert to TfidVector. $= Ⓜ$

Book (from goodreads)

① Cleaning

② title | bag of words

③ convert to TfidVector $= Ⓑ$

① $X = \text{cosinesim} (M, B)$

② $Y = \text{cosinesim} (M)$

③ $Z = \text{cosinesim} (B)$

Prediction

Ⓥⓟ ← just_finished_title

```
        Prediction
          (VP)
        /      \
     Movie     Book
```

Movie

→ Find index in Ⓨ

→ Get similarity

→ Arrange in descending order

→ Recc. top 5 movies

Book

→ Find index in Ⓧ

→ Get the first book with highest similarity

→ Send that title and get index in Ⓩ

→ Get similarity within Books dataset and recc 5 books

More efficient results than directly from Ⓧ