Algorithms ⟶ Preference , Behaviour

↑

Netflix ⟶ Likes / dislikes

Romantic

↓

Breakup

Romantic

Horror

1) Media Streaming ⟶ Netflix , Prime , YT , Spotify

2) Ecommerce ⟶ Product Recommendations

3) Social media ⟶ Instagram

A ..... ⟶ A^{α}

B .....

C .....

specific ⟶ Content

latest , Similar ∝

User engagement

User engagement

1) Popularity Based → Top rated, Top 25 Hindi movies

Hindi → Avg user rating

Sort → Descending Top K

Artist → Bhai

Advantage → 1) Easy to implement.

2) language → Cold start Problem

User history, likes / dislikes

3) Highly scalable → 10 lakh.
10 crore

Disadvantages → ① No personalized recommendation

② Bias

③ lack of diversity ← Hindi

② Content Based Recommendation:-

Data Sucnie → Content Based

Advantages → 1) Similar content, metadata of my current content.

SRK → SAK

2) More personalized as compared to Popularity.

3) Your user habits.

Disadvantages → ① Over specialization.

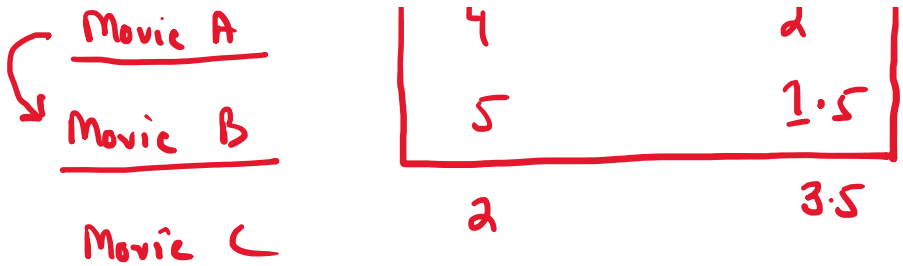
Kapil → User history.

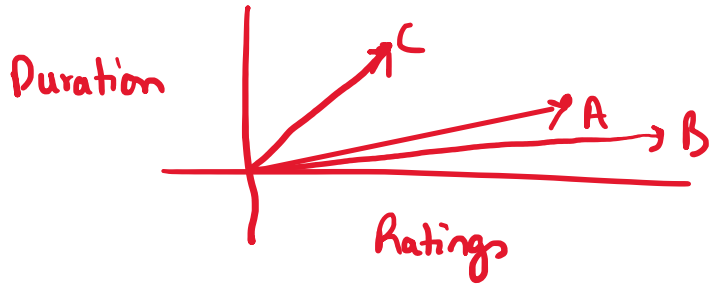
② No diversity

③ Cold start Problem.

Each row → Vector

	Rating	Duration.
Movie A	4	2





3) Collaborative Filtering

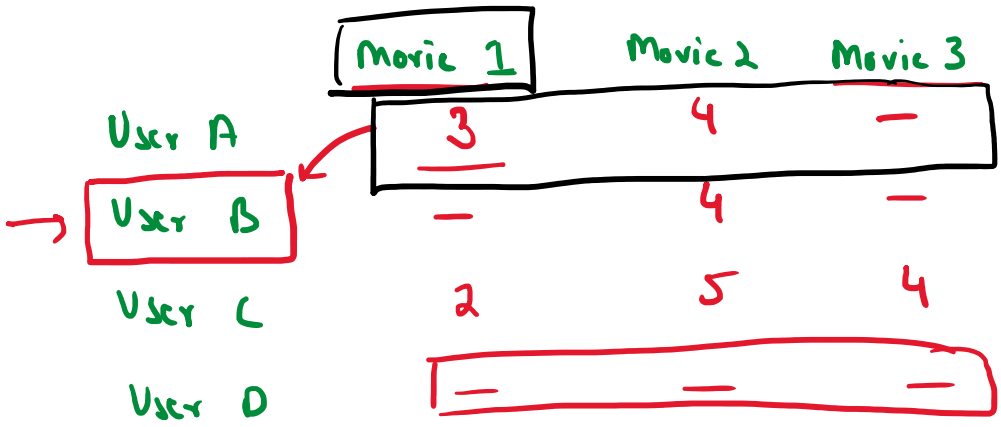
User, Item interaction

1) User Based CF.

2) Item Based CF.

1) User Based CF

5 Users, 3 Movies



User E 3 3.5 2

User B : User A = 0.8 ✓

User B : User C = 0.7

User B : User D = 0.1

User B : User E = 0.4

User B

Steps → 1) Calculate similarity of users.

2) Select the most similar user.

3) Recommend content that User B was not exposed to.

Item Based CF

Items similarity

	User A	User B	User C
Items A	✓	✓	✓
Items B	✓	✓	—
Items C	—	—	—
Items D	—	—	✓

Item A : User A, User B, User C

Item B : User A, User B

Item D : User C

No. of users >> No. of items → Use Item Based CF

Rows - Items Cols - Users

Netflix

No of items >> No. of users → User Based CF

Rows → Users Cols - Items

Instagram → Reels

Movie A, Movie B ← Movie E

Advantages → 1) Diversity

2) Scalable

3) Does not rely on metadata.

Disadvantages 1) Cold start Problem

New item, New User

2) Computationally expensive

3) Large datasets → Item User interaction

Huge

Why Hybrid systems are better

28 December 2024 15:13

Hybrid Recommendation → Hybrid → Netflix, Amazon, YT

Disadvantages

1) Popularity → Personalization

2) Content Based → Over specialization

3) Collaborative → Dynamic, New content

Hybrid

Collaborative + Content Based

↓ ↓

Variety + Personalization

More personalized and more dynamic / variety in recommendation. Hybrid

Hybrid → Weighted system

$$w_{cb} + w_{cf} = 1$$

60% CB and 40% Collaborative

$$0.6 \times CB + 0.4\, CF = \text{Similarity scores}$$

Euclidean distance, Manhattan, cosine distance

Cosine similarity

1) Curse of dimensionality

2) **-1 to 1.** $-1 = 180°$

0 = No similarity = $90°$

$1 = 0°$

Spotify Recommender System

28 December 2024 15:14

Spotify → Music streaming platform

Data Scientist

User engagement

User retention ↓

Popularity based → Personalized → YT music

Variety

Hybrid Recommendation → User engagement

User retention

1) Songs dataset → Information for all the songs on my platform.

Attributes, Metadata. → Content Based

2) User - Item interaction → User → Song → Playcount

Song name — Username — Playcount ↑

↓

Collaborative

Hybrid Recommendation System.

Content Based + Collaborative

Project Goal

28 December 2024 15:14

User engagement ↑

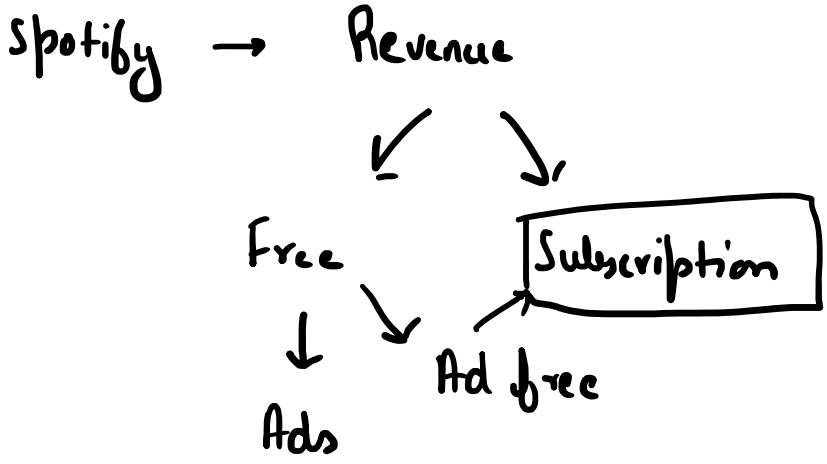
User Retention ↑

⟶ Personalized, Variety of Recommendations

Hybrid recommender system.

Improving Business Metrics

28 December 2024 15:14



1) User engagement → Music consume

↓

Subscription → monthly
→ 3
→ 9
→ Annual

Ad Based ↑

Subscription ↑

2) CTR (Click through rate) ↑ User engagement

↓

10 Recommendations → 1 Click →

3) User conversion → User engagement

↓

Ads → Plan Free → Subscription

4) lower the churn rate → Renews the subscription

Revenue ↗ Repeated customers.

How to achieve the goal

28 December 2024 15:14

Streamlit App → Song → 50,000

1) Content Based → 10 similar songs. ← Validate

2) Collaborative system → Item User interaction

Similar songs → Item based CF

10 similar songs ← Validate

3) Combine → Weighted Approach

Hybrid Recommender ← Validate

2 major challenges:-

1) Size of dataset → 97 lakh rows → 600 MB

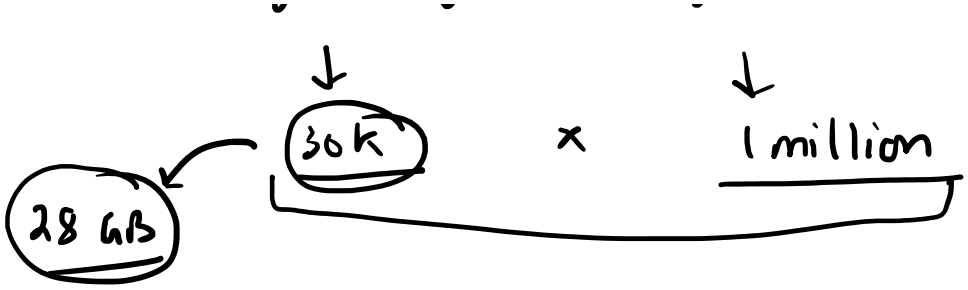
↓

Transform → Item user matrix

↙

Unique songs X unique users.

↓ ↓



Chunking → Break chunks

Song A → 1 million users → Total. Playcount

2) Weights → Collaborative ↑

→ Content ↑