

Practiced of Recommendation System

Agenda → Real world interview:

\Rightarrow No major ~~algos~~ \rightarrow DS Maturity

- ↳ architecture
- ↳ design
- ↳ ideas

⇒ Design a Web System

- refail
- social media
- e-commerce
- Dating

 Use a ~~apriori~~ algo for rec system $\rightarrow \rightarrow \} =$
 \hookrightarrow Customer also bought

- A*

 - → Define the data you will use →
User profile / Preferences
Past order date
 - → Define your feature →
 - Model description →
 - How to test your model →
 - How will you improve →

Q) How can you use clustering for a supervised learning /
rec system ?? → data

- *) Cluster user and collectively recommend them according to their cluster instead of individual person
 - ↳ individual people don't have high amount of data.

$$\rightarrow I_1 \sim 2-3 \\ I_2 \sim 2- \\ I_3 \\ I_4 \\ I_5$$

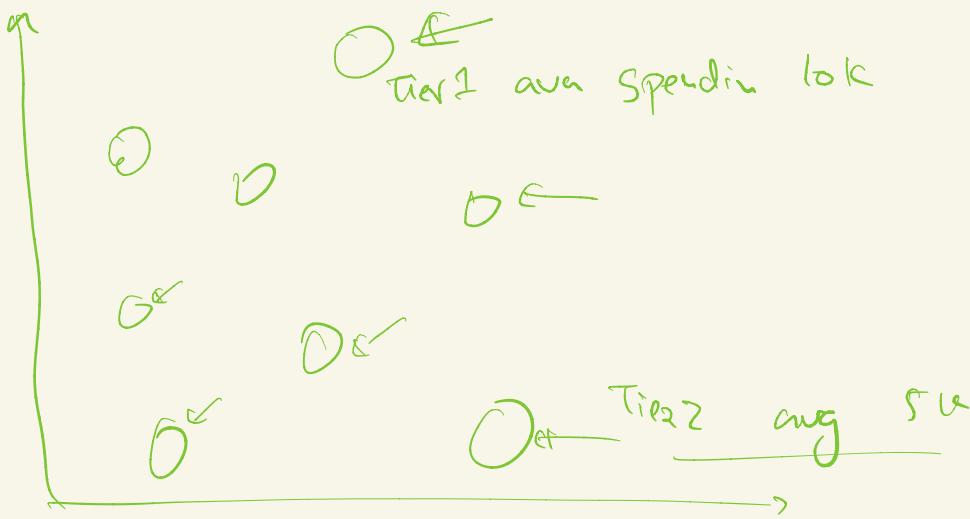
10-15
adapt

② We use clusters to define segments → Separate models for each of them

idea → • Clothes →
• watches →
• Shoes →
• → Sales pattern would be different

User segment

- Location based
- Consumer based → iphone → Cheapest
→ design cover



→ 1000 prod
 10^5 prod 10^7 users →

Q How do you measure a rec. system ?? ↗

1) Instagram →

- Time spent
- likes

⇒ Offline → No user testing, only use data available
 MAE

MAPE, coverage, diversity, accuracy.

* Online → A/B Testing

• You deploy your model → check if time
 spent, likes, ...

⇒ Finance influencer →
 • Finance article } → (%)
 → { random items }

\Rightarrow Depends on type Problem (retail
social media)

① Movies recommendation (Netflix) \rightarrow

a) Did we actually recommend things \rightarrow got liked by viewer

Online way \rightarrow Avg rating of first k recommender movies.

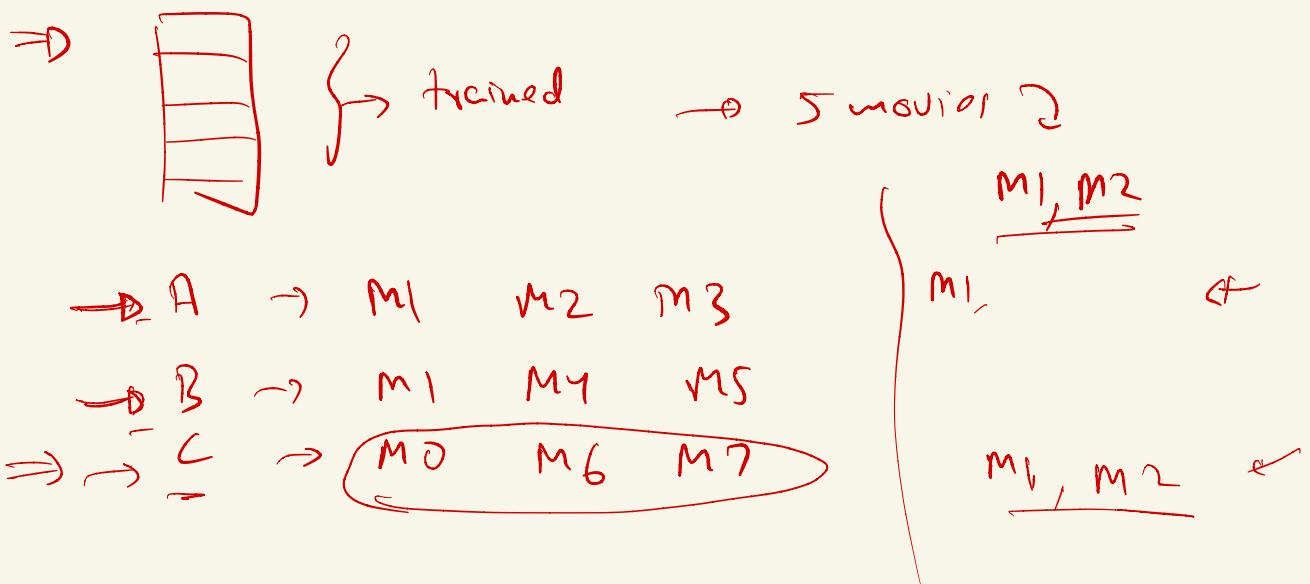
Recommend

\hookrightarrow to movies \rightarrow 2 movies

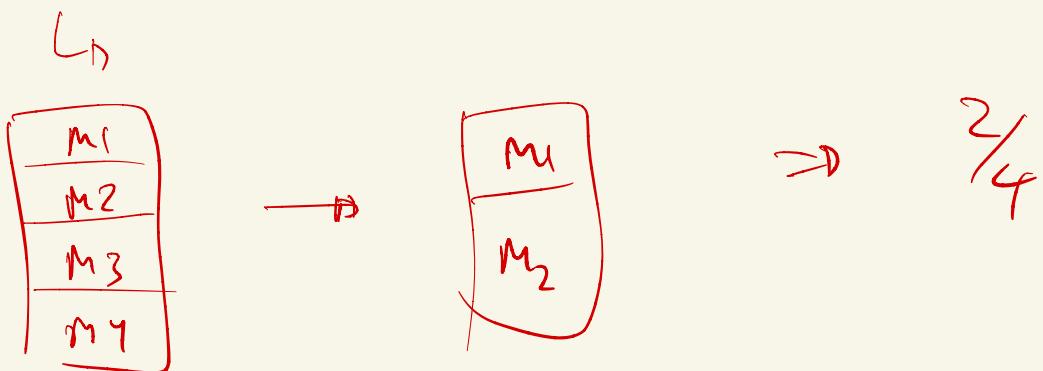
Offline way \rightarrow 1M rows \rightarrow 800k \rightarrow
 \rightarrow 200k
 \hookrightarrow 50 users who get recommended
and have also seen the movie.

\hookrightarrow 200k \rightarrow 100 \rightarrow probably recom

\rightarrow \hookrightarrow 300 \rightarrow seen \rightarrow ratings provided



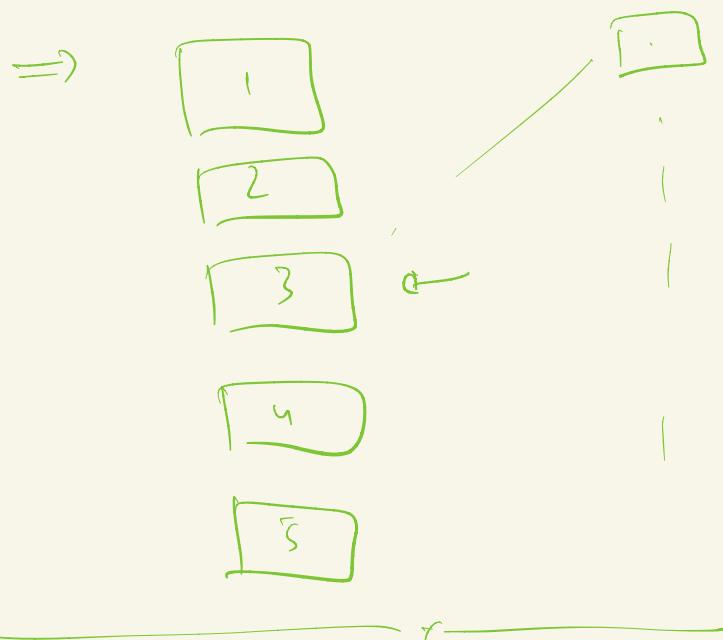
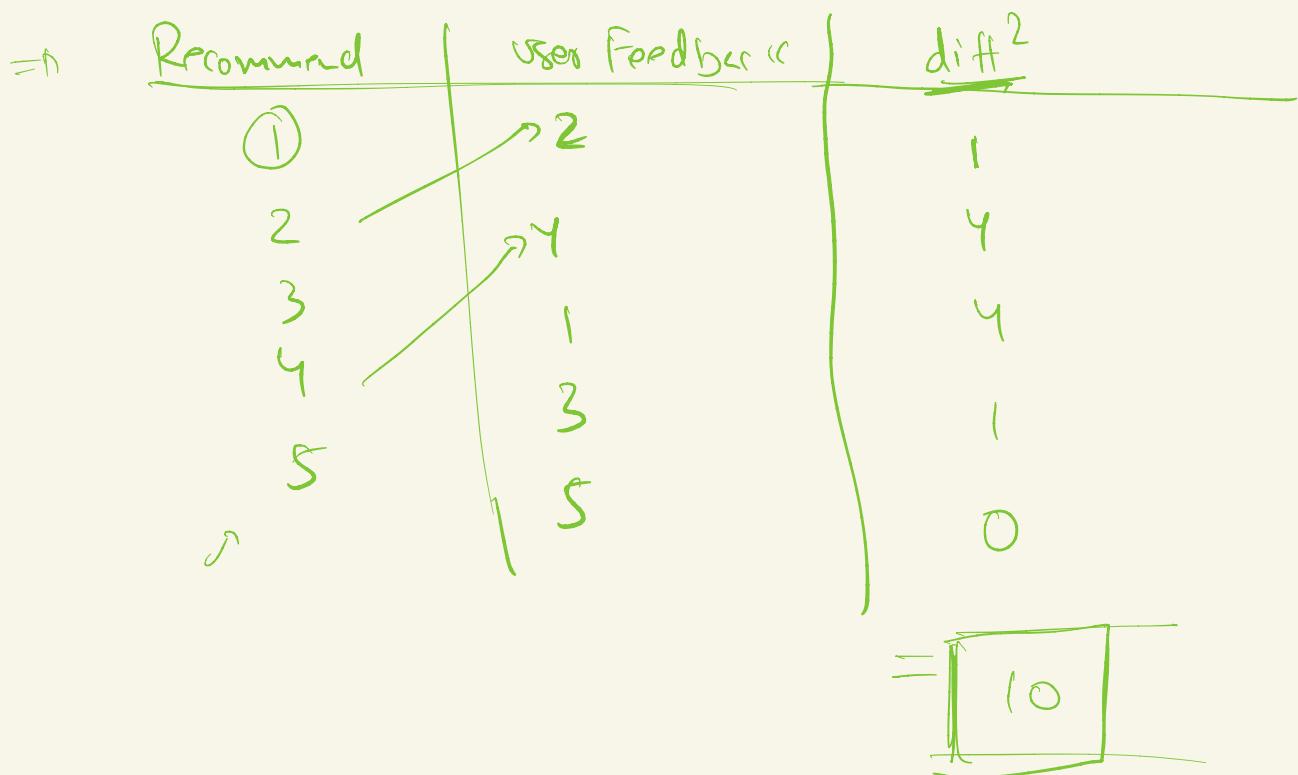
\Rightarrow K-overall \rightarrow positive outcomes out of the first K recommended values



- b) Product/E-commerce \rightarrow
- Did they click on it
 - Sales \uparrow

c) Social Media :-

Rank diff²



* Diversity → Liked a car image
 ↓
 all next 30 posts → Cars related to

\rightarrow 5 movies/songs \rightarrow
 \Rightarrow 5 products \rightarrow 1
 at least be some diversity in terms of clusters.

* Coverage \rightarrow SM

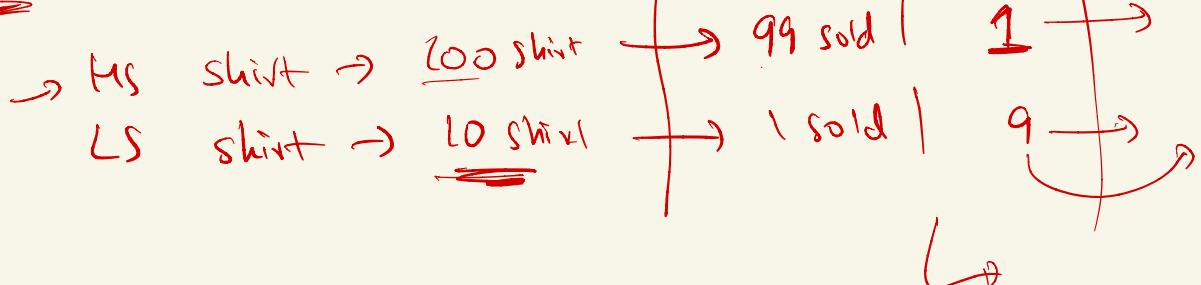
100 K products \rightarrow

$$\frac{100\text{ K}}{SM} = \frac{10}{500} \approx 2\%$$

\Rightarrow 100 products \rightarrow 98 products \rightarrow never sold.

\Rightarrow 90% Coverage \rightarrow

SOH \rightarrow Stock on Hand



Stock / Inventory goes down \rightarrow Jek up the price \uparrow

Q Other than the model / central algo → what else should you keep in mind??

→ Coverage

→ Recency → • New website → more preference
• Youtube →

→ • day since launched

→ Exploration vs Exploitation

→ To showcase new content

→ New users → something to watch

Small % of recomm → random [diversity]

⇒ Some checks for non-sense recommendation

e.g. → Dating App → • Male - Male
↳ Preference → Female

② OOS →

↳ item not available

③ Completely differ category product → Shirt with iPhone

* Loopholes to avoid ??

⇒ ① Power law → 80/20 rule

↳ Top 20% of products will attract 80% of users



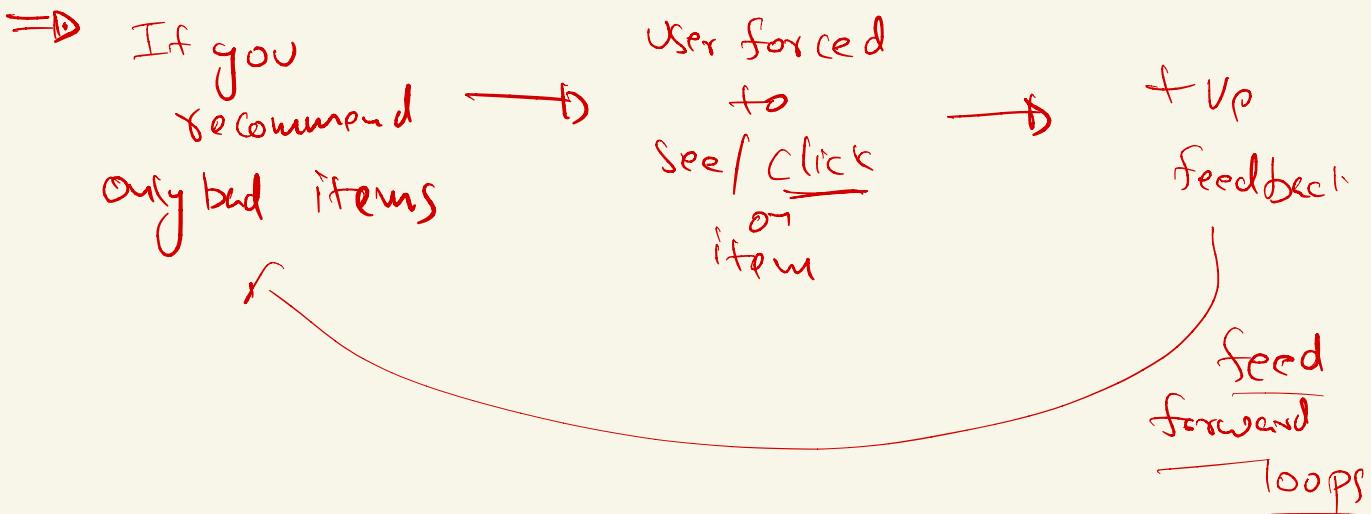
① 80% jobs → 20% candidates ← ↗

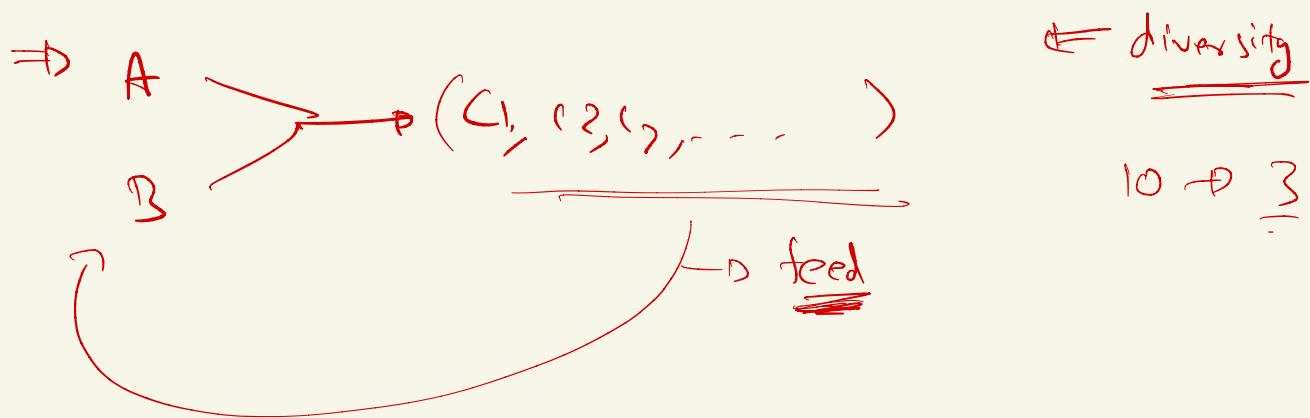
Job
↳ 80% candidates → 20% jobs ← (Job market is crowded)

② 80% inc → 20% of people

Dating

③ E-commerce → high coverage





~~Social Responsibility~~ :-

- Biased Caste / Gender / Ethnicity / Color
- Offensive / wrong Content

⇒ • Started showing ads of height Extension →

Height param

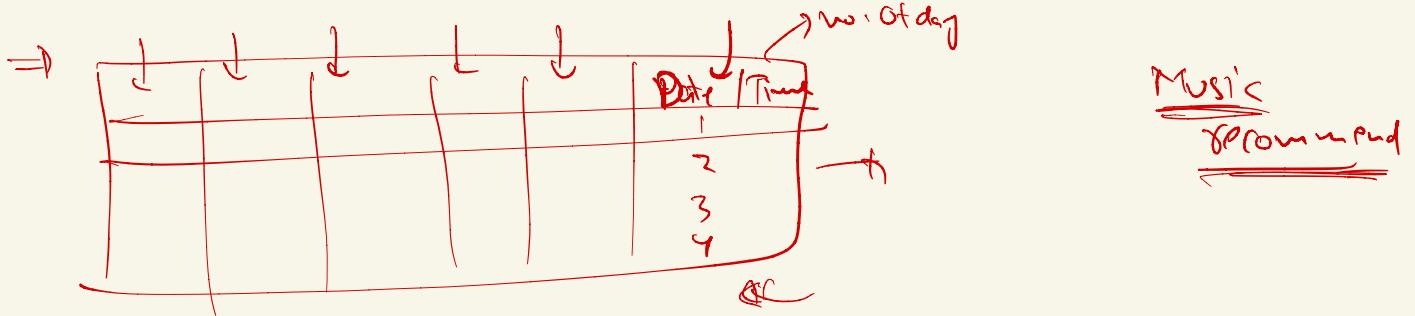
- User data →
 - Age →
 - Region
 - location

Interaction

~~Receiving bias~~ :-

↳ Just because something new → It is good.

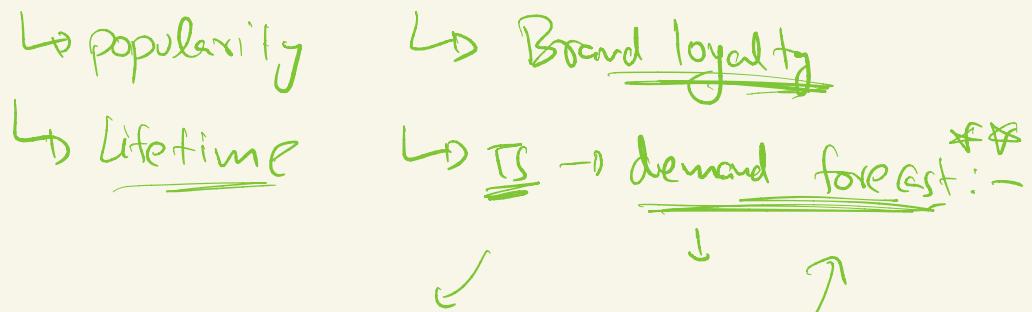
- Category / Specific / Clusters Specific →



 Can you think of features for a e-commerce product recommendation ??

- Product attributes →
 - Trendy products
 - Location
 - Recent searches / Likes
 - gender
 - Collaborative filtering feature → . user interactions ✓
 - Geographical
 - User profiling → age, gender, past purchase, likes,
 - ↓
clusters
 - Paired frequencies →
 - Conditional prob with another product → Apriori

→ Indirect Product attributes



→ Ts of overall winter clothes
→ How sales look in each month

⇒ Business Metrics :-

↳ ① Profit from product

$$\begin{array}{c} \text{↳ } \frac{A}{B} \rightarrow 4.3^* \\ \text{↳ } 4.1^* \rightarrow 1000 \text{ Rs} \end{array}$$

→ 10 Rs

- ② Inventory on hand
- ③ Discount/ promos/ marketing
- ④ Sponsored
- ⑤ New vertical →

⇒ Do this question for every industry you know

Q What is Cold Start problem & how to solve this?)

⇒ Don't have data for forming your model.

→ Solutions

↳ Random

↳ Popularity based

↳ geolocation

↳ Recency bias

↳ User preference

↳  ↳ only used till first
100 songs

↳ Reinforcement learning:

↳ Rule based Recommendation →

↳ .

•



