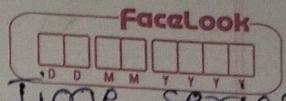


## unit-5

# recommendation systems and time series



Ques:- Explain recommendation systems with its types.

→ Recommendation system:-

- A recommendation system is a tool that predicts and suggests items to users based on various data, such as preferences, behaviors and past interactions.
- These systems are commonly used in platforms like e-commerce websites, streaming services, social media and news aggregators.

TYPES OF Recommendation systems:-

1] Collaborative Filtering :-

- Collaborative filtering is based on the assumption that users who have agreed in the past will agree in the future.

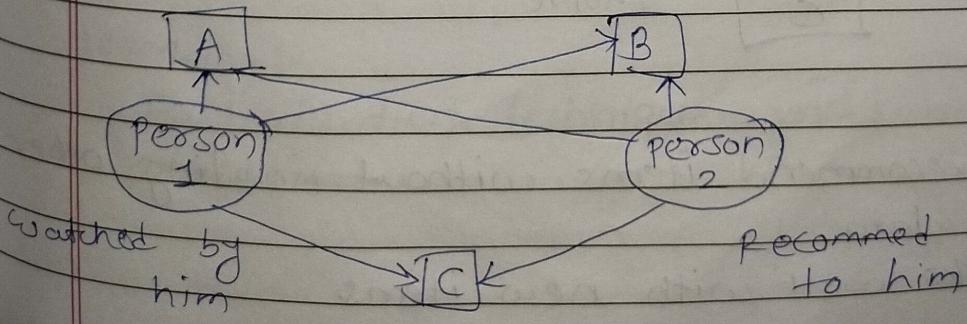
- It uses the preferences or behaviors of users to make recommendations.

- Collaborative filtering algorithm recommends the item based on the prepreference information from many users.

- There are two types of collaborative filtering :-

1) User-based collaborative filtering

2) Item-based collaborative filtering.



Pros :-

- 1) simple and effective
- 2) works well when there's a large amount of user data.

Cons :-

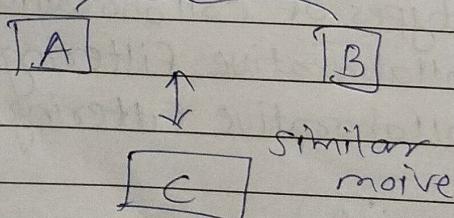
- 1) struggles with new users or new items (Cold-start Problem)
- 2) may suffer from scalability issues as the user base grows.

## 2) Content-Based Filtering :-

- Content-Based Filtering recommends items based on the characteristics or features of items that the user has shown interest in.
- In these recommendation systems uses attributes or features of an item to recommend other items similar to user preferences.
- These approaches based on similarity of item and user feature.

- Given information :-

- 1) user characteristics
- 2) item characteristics



Pros :-

- 1) can recommend items without needing other user's data
- 2) works well with new items.

cons :-

- 1) requires detailed information about items
- 2) might lead to over-specialization.

3) Hybrid systems :-

→ Hybrid recommendation systems combine different recommendation techniques such as collaborative filtering and content-based filtering to improve the overall performance and mitigate the limitations of each individual method.

Pros :-

- 1) more accurate recommendations by leveraging multiple techniques.
- 2) reduces biases and cold-start problem

cons :-

- 1) more complex to build and maintain
- 2) can be computationally expensive.

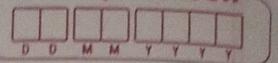
ii) Matrix Factorization :-

→ Matrix Factorization used for collaborative filtering it aims to decompose a large user item set.

Example :- rating matrix that converts into lower-two dimensional matrix

Working :-

Factorization technique works on statistical analysis.



Factorization method:-

$$R \approx U \times V^T$$

$U$  = Features of user

$V$  = Features of items

Algorithm :-

Alternating Least square (ALS)

Application :-

- 1) Netflix
- 2) Amazon

Ques - Explain Topic modeling popularity based Recommender engine with example.



- Topic Modeling is a technique used in natural language processing to discover abstract topics within a collection of documents.

- When combined with popularity-based recommendation engines, it creates a system that suggests content or products by focusing on widely liked topics and items.

How it works :-

1) Data Collection :- Gather textual data or descriptions of items.

2) Topic modeling :- Use algorithm like Latent Dirichlet Allocation (LDA) or Non-Negative matrix factorization (NMF) to extract topics from the data. Each item is associated with probabilities

of belonging to different topics.

③ popularity scoring :- determine the popularity of items, usually based on:-

- user ratings
- Number of Purchases or interactions
- views, likes or shares

④ combining Insights :-

- Identify the most popular items within each topic.
- Recommend these items to users interested in similar topics.

Example :- News Article Recommender

① Scenario :-

- A platform has news articles on various topics
- users have preferences for specific topics and tend to read articles with high popularity.

② Step-by-step Process :-

③ Topic Modeling :-

- Extract topics like "Politics", "Sports", "Technology", etc.

④ Popularity Scoring :-

- Rank article within each topic by the number of views or likes.

⑤ Recommendation :-

- A user interested in "Technology" receives the most-viewed or liked articles in the "Technology" topic.

⑥ Algorithm output :- If a user likes "Technology" the system recommends :-

- Article A : "Top 10 AI Innovations in 2024" (20,000 views)
- Article B : "Quantum Computing Breakthroughs" (18,000 views)

Advantages :-

- 1) simplicity
- 2) scalability
- 3) Topic sensitivity

Limitations :-

- 1) cold-start problem
- 2) popularity bias
- 3) lack of personalization

Ques - Explain content based recommendation engines with example.



- Content-based recommendation engines suggest items to users based on the content or attributes of the items and the user's preferences.
- These systems analyze item features and recommend items that are similar to what the user has previously liked or interacted with.

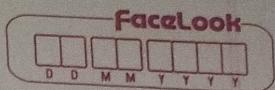
How it works :-

1) Feature Extraction :-

- Each item in the database is described by a set of features.
- The user's preferences are determined from their history or explicitly provided data.

2) Similarity calculation :-

- The system uses techniques like cosine similarity, Euclidean distance or dot product



to find items with features similar to what the user likes.

### ③ Recommendation :-

- Items with the highest similarity scores are recommended to the user.

Example :- Movie Recommendation system.

Scenario :-

- A user has watched and liked the movie "Inception".
- Features of "Inception" :-
  - Genre : sci-fi, Thriller
  - Director : Christopher Nolan
  - Lead Actor : Leonardo DiCaprio

Steps :-

#### ① Feature Extraction :-

- Each movie in the database is represented by its features.

#### ② User Profile :-

The system creates a user profile based on the liked movie, focusing on its features :- sci-fi, Thriller, Christopher Nolan, Leonardo DiCaprio

#### ③ Similarity Calculation :-

- compare the user's profile with other movies in the database using similarity metrics.
- Example :-

- Interstellar has high similarity due to shared genre and director.
- Titanic has moderate similarity due to the lead actor.

#### 4) Recommendations:-

• Recommend Interstellar and then Titanic

#### Advantages :-

- 1) Highly Personalized
- 2) Transparent recommendations.

#### Challenges :-

- 1) Limited to known features (cold-start problem)
- 2) Requires detailed and structured data about items.

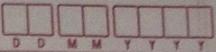
#### Applications :-

- 1) E-commerce
- 2) Music streaming
- 3) Education platforms

Ques - Explain classification based recommendation engine with example.

⇒ A classification-based recommendation engine Predicts whether a user will like or engage with an item by framing the recommendation problem as a classification task.

- This approach uses machine learning models to classify items into categories such as



"interested" or "not interested" based on user-item interactions and features.

How it works :-

① Data Preparation :-

- collect data on users, items and interactions
- Lable interactions

② Feature Engineering :-

- combine user, item and interaction context features

③ Model Training :-

- Train a classification model on labeled data.

④ Prediction :-

- Predict the probability of a user liking an item

⑤ Recommendation :-

- Recommend items with the highest predicted probabilities.

Example :- E-commerce Product Recommendation.

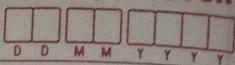
Scenario :-

- An online store wants to recommend products to users.

Steps :-

① Data :-

- collect user data.



## ② Feature Engineering :-

- combine user and product features into a single feature vector.

## ③ Model Training :-

- use a classification algorithm to predict the "Purchased" label.

## ④ Prediction :-

- For a new product, predict whether user will purchase it.

## ⑤ Recommendation :-

- If the predicted probability is high, recommend product to user.

### • Advantages :-

① Flexible

② suitable for predicting binary or multi-class outcomes.

### • challenges :-

① Require labeled data

② Performance depends heavily on feature engineering.

## Application :-

① streaming services

② E-commerce

③ Educational platform

Examine correlation in recommender engine.

- Correlation in recommender system plays a vital role in understanding relationship between users, items and their preferences.
- Correlation measures the degree to which two variables are related.
- It helps identify patterns, such as :-
- similar users who rate items in a comparable way.
- similar items that tend to receive similar ratings from users.
- For recommender systems, correlation can be used to :-
- Find similar users :- collaborative filtering
- Find similar items :- Item-based Filtering
- Improve prediction accuracy.

Types of correlation measures :-

a) Pearson correlation coefficient :-

- Measures the linear relationship between two variables.
- commonly used in user-based collaborative filtering.
- Formula :-

$$\rho = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

where,

- $x_i, y_i$  are ratings for two users on the same items.
- $\bar{x}, \bar{y}$  are the mean ratings of the respective users.

### b) cosine similarity :-

- measures the cosine of the angle between two vectors.
- common in item-based Filtering and implicit feedback systems.
- Formula :-

$$\text{Similarity} = \frac{\sum x_i \cdot y_i}{\sqrt{\sum x_i^2} \sqrt{\sum y_i^2}}$$

### c) Jaccard similarity :-

- Measures the overlap between two sets.
- often used with binary feedback.

### d) Spearman's Rank correlation :-

- A non-parametric version of Pearson correlation
- useful when ratings are ordinal or rank based.

### Applications :-

- ① User-Based collaborative Filtering
- ② Item-Based collaborative Filtering
- ③ Hybrid models

### Challenges :-

- ① Data sparsity
- ② cold start Problem
- ③ Scalability.

## Explain Time series Forecasting.

Time series Forecasting is a technique used to predict future values based on previously observed data points.

This method involves analyzing patterns, trends and seasonality in data that are indexed over time, such as daily temperatures, monthly sales or yearly stock prices.

### key components :-

#### 1) Trend :-

-The long-term direction in the data.

-Example:- Rising housing prices over years.

#### 2) Seasonality :-

-Repeating patterns or cycles in the data at regular intervals.

-Example:- Increased ice cream sales in summer.

#### 3) Cyclic Patterns :-

-Fluctuations in data with no fixed periodicity, often linked to economic or business cycles.

-Example:- Stock market booms and busts.

#### 4) Noise :-

-Random variations that cannot be explained by trend, seasonality or cyclic behavior.

-Example:- sudden spikes due to unforeseen events.

DDMMYY

- Advantages :-

- 1) Data-Driven Decision Making
- 2) Detects Trends and Seasonality
- 3) Improved Accuracy over Time
- 4) Automation Potential.

- Disadvantages :-

- 1) Data Dependency
- 2) Limited for Non-stationary Data
- 3) Sensitivity to Noise
- 4) Resource-Intensive
- 5) Short-Term Focus.

- Applications :-

- 1) Business
- 2) Finance
- 3) Weather
- 4) Healthcare
- 5) Energy.

Ques-

Differentiate between Collaborative, Content-based and Matrix Factorization.



Feature	Collaborative Filtering	Content-based	Matrix Factorization
Definition	Recommends items based on user-item interaction patterns	Recommends items similar to those a user liked before	Decomposes the user-item interaction matrix to uncover latent features
Dependency	Requires user-item interaction data	Requires detailed information about items.	Requires a user-item interaction matrix.

Type	User-User or Item-Item similarity-based	User-Item similarity based on attributes	Model-based Collaborative Filtering
old start issue	struggles with new users or items	struggles with new users but handles new items better	similar to collaborative filtering
Scalability	can become computationally expensive for large datasets.	relatively scalable if item metadata is concise	Efficient for large datasets with proper optimization
Personalization	Provides recommendations based on collective behavior.	Focuses on individual User Preferences	Balances collective behavior and latent preferences.
Example Technique	k-NN, clustering algorithms	TF-IDF, cosine similarity.	SVD, ALS (Alternating Least squares).
Application	Amazon	Netflix	Spotify