

# SISTEM DE RECOMANDARE FILME

---

Omer Tarik-IIhan (E-Guvernare)

Preda Diana (E-Guvernare)



01

# Tema

Sisteme de Recomandare pentru Filme



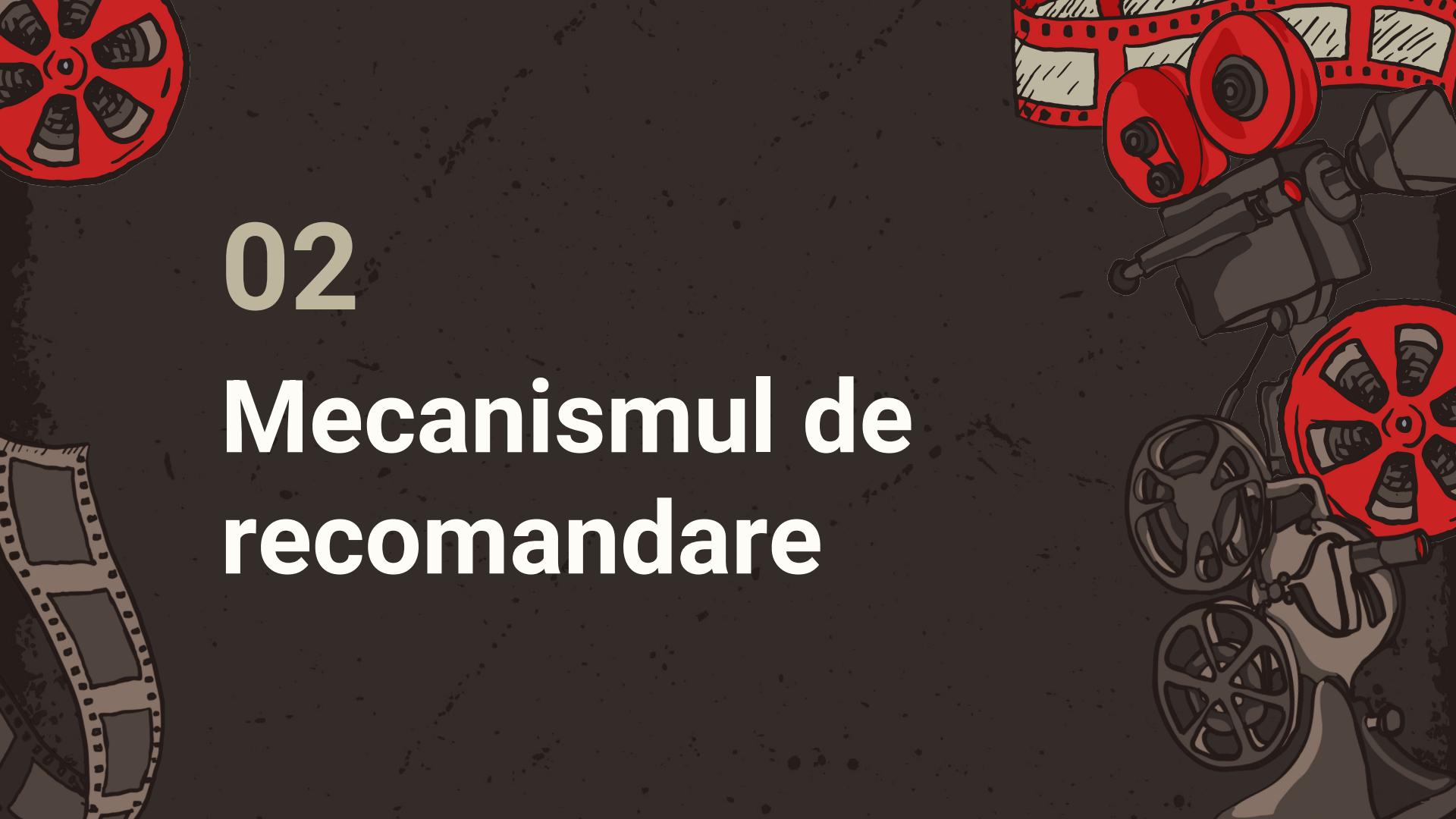
**Scop:** Dezvoltarea unui sistem intelligent care să recomande filme utilizatorilor pe baza preferințelor lor, a istoricului de vizionare și a similarităților cu alți utilizatori.

### **Problema identificată:**

Într-un context cu mii de filme disponibile pe platforme precum Netflix, HBO Max sau Amazon Prime, utilizatorii petrec mult timp alegând un film potrivit. Un sistem de recomandare eficient reduce acest timp, oferind sugestii personalizate și relevante.

### **Obiectiv:**

- Oferirea unor recomandări de filme personalizate.
- Îmbunătățirea experienței utilizatorului prin învățarea automată a preferințelor sale.
- Posibilitatea integrării cu API-uri externe pentru actualizarea datelor despre filme.



02

# Mecanismul de recomandare

# Abordare Hibridă

Vom utiliza o abordare **hibridă**, care combină **două logici** bazate pe datele noastre:

## 1. Filtrarea Colaborativă:

- **Cum folosim datele:** Sistemul analizează **istoricul de ratinguri** (din ratings.csv) pentru a găsi tipare și similarități între utilizatori (user-based) sau între filme (item-based).
- **Exemplu:** Dacă utilizatorul A și B au dat note similare la aceeași filme, sistemul presupune că au gusturi similare și le recomandă reciproc filmele apreciate.

## 2. Filtrarea Bazată pe Conținut:

- **Cum folosim datele:** Sistemul analizează **metadatele filmelor**, cum ar fi gen, regizor, actori și cuvinte cheie.
- **Exemplu:** Dacă un utilizator a apreciat un film SF regizat de "Christopher Nolan", sistemul va căuta și recomanda alte filme cu atribute similare.

03

# Dataset si API

# Dataset: The Movies Dataset (Kaggle)

## Link Dataset: [The Movies Dataset](#)

**Descriere:** Este un set de date cuprinzător, care include metadate pentru peste **45.000 de filme** (din **Full MovieLens dataset**), **26 de milioane de ratinguri** de la **270.000 de utilizatori**, plus cuvinte cheie și distribuție.

### Cum îl folosim:

- **movies\_metadata.csv & keywords.csv:** Le vom folosi pentru a încărca catalogul de iteme (filme) în Recombee. Fiecare film va avea atrbute precum **genres, keywords, overview, director**.
- **ratings.csv:** Îl vom folosi pentru a încărca istoricul de interacțiuni în Recombee. Fiecare rând devine o "interacțiune" de tip "rating".



API: Recombee

Link API: [Recombee](#)



Recombee

**Descriere:** Recombee este o platformă de recomandare ca serviciu. Vom încărca datele (filmele și ratingurile) în baza de date Recombee, iar API-ul va calcula modelele hibride.

- **Avantaj:** Recombee gestionează automat scalarea, antrenarea modelelor și problema de cold start mult mai eficient. Sistemul nostru devine un integrator care trimitе date și cere recomandări.

# Exemple de date

Am extras câteva secvențe din fișierele folosite din dataset-ul ales. Cateva date au fost șterse pentru a putea fi reprezentate în prezentare.

## Exemplu de date (movies\_metadata.csv):

<b>budget</b>	<b>genres</b>	<b>id</b>	<b>imdb_id</b>	<b>original_title</b>	<b>release_date</b>	<b>revenue</b>	<b>runtime</b>	<b>title</b>	<b>vote_average</b>	<b>vote_count</b>
30000000	[{"id": 16, "name": "Animation"}]	862	tt0114709	Toy Story	1995-10-30	373554033	81	Toy Story	7.7	5415
65000000	[{"id": 12, "name": "Adventure"}]	8844	tt0113497	Jumanji	1995-12-15	262797249	104	Jumanji	6.9	2413
0	[{"id": 10749, "name": "Romance"}]	15602	tt0113228	Grumpier Old Men	1995-12-22	0	101	Grumpier Old Men	6.5	92
16000000	[{"id": 35, "name": "Comedy"}]	31357	tt0114885	Waiting to Exhale	1995-12-22	81452156	127	Waiting to Exhale	6.1	34

## Exemplu de date (keywords.csv):

<b>id</b>	<b>keywords</b>
862	[{"id": 931, "name": "jealousy"}, {"id": 4290, "name": "toy"}, {"id": 5202, "name": "boy"}]
8844	[{"id": 10090, "name": "board game"}, {"id": 10941, "name": "disappearance"}]
15602	[{"id": 1495, "name": "fishing"}, {"id": 12392, "name": "best friend"}]
31357	[{"id": 818, "name": "based on novel"}, {"id": 10131, "name": "interracial relationship"}]

## Exemplu de date (ratings.csv):

userId	movieId	rating	timestamp
1	110	1	1425941529
1	147	4.5	1425942435
1	858	5	1425941523
1	1221	5	1425941546

04

# Problema de Cold Start

## 1. User Cold Start (Utilizator Nou)

- **Scenariul:** Un utilizator nou (user-xyz) își creează cont.
- **Provocarea:** Nu avem niciun istoric de vizionare sau rating pentru el.
- **Impact:** Filtrarea Colaborativă eșuează. Nu putem găsi "utilizatori similari" (vecini) pentru că nu știm ce preferințe are.
- **Soluția:** Folosim **filtrarea de conținut**, cerând la înregistrare preferințele utilizatorului.

## 2. Item Cold Start (Film Nou)

- **Scenariu:** Un film nou (ex: *Dune: Part Three*) este adăugat în catalog.
- **Provocare:** Nu există încă ratinguri de la utilizatori.
- **Impact:** Filtrarea colaborativă nu poate recomanda filmul, deoarece nu apare în istoricul nimăului.
- **Soluție:** Se aplică filtrarea bazată pe conținut — sistemul potrivește atributele filmului (gen, regizor etc.) cu profilurile utilizatorilor existenți.