## Movie Recommendation Engine

#### Presented by

Aditya Kumar(191520001) Anil Kumar Vishwakarma(181520007)

Department of Computer Engineering and Applications

Institute of Engineering and Technology

Under the guidance of Mr. Rahul Pradhan



GLA University
Mathura-281406,INDIA
2022-2023

# **OBJECTIVE**

- To build Movie Recommendation System based on TMDB Dataset.
   We will create a content-based recommendation system using Cosine similarity.
- The idea behind creating this project was to build a model that can recommend movie on the basis of a keyword we are giving as an input.
- It will be a basic model of an actual recommendation System.

### MOTIVATION

One key reason why we need a recommender system in modern society is that people have too much options to use from due to the prevalence of Internet. In the past, people used to shop in a physical store, in which the items available are limited. For instance, the number of movies that can be placed in a Blockbuster store depends on the size of that store. By contrast, nowadays, the Internet allows people to access abundant resources online. Netflix, for example, has an enormous collection of movies. Although the amount of available information increased, a new problem arose as people had a hard time selecting the items they actually want to see. This is where the recommender system comes in.

# Hardware & SOFTWARE REQUIREMENTS

#### Hardware:-

Personal computers with min 4 GB RAM.

#### Software:-

- Jupiter Notebook / Google Collab
- Python

## Libraries used

- Pandas
- Numpy
- Sklearn
- AST

### **DataSet**

For this Python project we'll be using **TMDB** dataset. Here we are using two data sets – **Movies.csv**, **Credits.csv**.

- Movies Data has 4803 rows and 20 columns.
- Movies.csv has following features Budget, Genres, Homepage, id, keywords, original\_language, original\_title, overview, popularity, production\_companies, production\_countries, release\_date, revenue, runtime, spoken\_languages, status, tagline, title, vote\_average, vote\_count.
- Credits Data has 45 rows and 4 columns.
- Credits.csv has following features Movie\_id, title,cast, crew.

# IMPLEMENTATION

# Data Exporting (Importing libraries)

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

movies = pd.read_csv("tmdb_5000_movies.csv")
credits = pd.read_csv("tmdb_5000_credits.csv")
```

# Data Wrangling

budget	genres	homepage	id k	eywords orig	inal_language	original_title	overview	popularity	production_companies	production_countries	release_date	revenue	runtime	spoken_languages	status	tagline	title vote_average	vote_cou
237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam	http://www.avatarmovie.com/		[{"id": 1463, "name": "culture clash"}, {"id":	en	Avatar	In the 22nd century, a paraplegic Marine is di	150.437577	[{"name": "Ingenious Film Partners", "id": 289	[{"iso_3166_1": "US", "name": "United States 0	2009-12-10	2787965087	<b>1</b> 62.0	[{"iso_639_1": "en", "name": "English"}, {"iso	Released	Enter the World of Pandora.	Avatar 7.2	118
300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "	http://disney.go.com/disneypictures/pirates/		[{"id": 270, "name": 'ocean"}, 'id": 726, "na	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha	139.082615	[{"name": "Walt Disney Pictures", "id": 2}, {"	[{"iso_3166_1": "US", "name": "United States 0	2007-05-19	961000000	169.0	[{"iso_639_1": "en", "name": "English"}]	Released	At the end of the world, the adventure begins.	Pirates of the Caribbean: 6.9 At World's End	45

v Os [9] movies.shape

(4803, 20)

#### [8] credits.head()

	movie_id	title	cast	c	crew
0	19995	Avatar	[{"cast_id": 242, "character": "Jake Sully", "	[{"credit_id": "52fe48009251416c750aca23", "c	de
1	285	Pirates of the Caribbean: At World's End	[{"cast_id": 4, "character": "Captain Jack Spa	[{"credit_id": "52fe4232c3a36847f800b579", "c	de
2	206647	Spectre	[{"cast_id": 1, "character": "James Bond", "cr	[{"credit_id": "54805967c3a36829b5002c41", "c	de
3	49026	The Dark Knight Rises	[{"cast_id": 2, "character": "Bruce Wayne / Ba	[{"credit_id": "52fe4781c3a36847f81398c3", "c	de
4	49529	John Carter	[{"cast_id": 5, "character": "John Carter", "c	[{"credit_id": "52fe479ac3a36847f813eaa3", "c	de

```
[11] credits.shape
(4803, 4)
```

# Merging both data sets

```
[12] movies = movies.merge(credits,on='title')

movies.shape
(4809, 23)
```

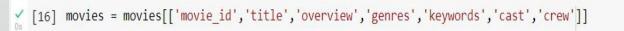
### Features Selection

#### All features which are present in our dataset are as follows:-

- Budget
- Genres
- Homepage
- id
- keywords
- original\_language
- original\_title
- overview

- popularity
- production\_countries,
- production\_companie s
- release\_date,
- revenue,
- runtime,
- spoken\_languages,
- status,

- tagline,
- title
- vote\_average,
- vote\_count
- Movie\_id,
- cast,
- crew



↑ ↓ ⊖ **目 ‡** ॄ **:** :

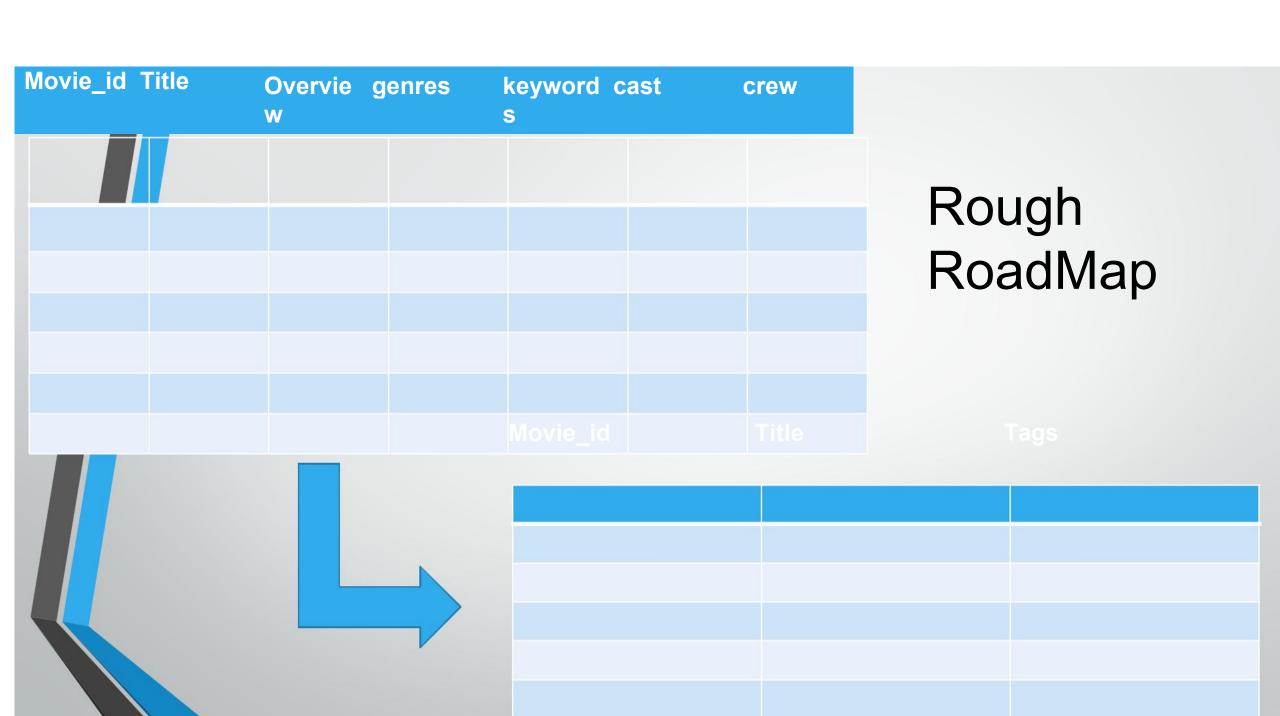
1		
0	movies.head	(

N.	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di	[{"id": 28, "name": "Action"}, {"id": 12, "nam	[{"id": 1463, "name": "culture clash"}, {"id":	[{"cast_id": 242, "character": "Jake Sully", "	[{"credit_id": "52fe48009251416c750aca23", "de
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha	[{"id": 12, "name": "Adventure"}, {"id": 14, "	[{"id": 270, "name": "ocean"}, {"id": 726, "na	[{"cast_id": 4, "character": "Captain Jack Spa	[{"credit_id": "52fe4232c3a36847f800b579", "de
2	206647	Spectre	A cryptic message from Bond's past sends him o	[{"id": 28, "name": "Action"}, {"id": 12, "nam	[{"id": 470, "name": "spy"}, {"id": 818, "name	[{"cast_id": 1, "character": "James Bond", "cr	[{"credit_id": "54805967c3a36829b5002c41", "de
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve	[{"id": 28, "name": "Action"}, {"id": 80, "nam	[{"id": 849, "name": "dc comics"}, {"id": 853,	[{"cast_id": 2, "character": "Bruce Wayne / Ba	[{"credit_id": "52fe4781c3a36847f81398c3", "de
4	49529	John Carter	John Carter is a war-weary, former military ca	[{"id": 28, "name": "Action"}, {"id": 12, "nam	[{"id": 818, "name": "based on novel"}, {"id":	[{"cast_id": 5, "character": "John Carter", "c	[{"credit_id": "52fe479ac3a36847f813eaa3", "de

# Checking if there are null values or not and deleting

```
[16] movies.isnull().sum()
       movie id
       title
       overview
       genres
        keywords
       cast
        crew
       dtype: int64

√ [17] movies.dropna(inplace=True)
[18] movies.isnull().sum()
       movie id
       title
       overview
       genres
        keywords
        cast
        crew
       dtype: int64
```



```
[20] movies['genres'][0]

'[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
```

```
[23] def convert(text):
    L = []
    for i in ast.literal_eval(text):
        L.append(i['name'])
    return L
```

Workin g on column Genres

)

crew	cast	keywords	genres	overview	title	movie_id	Ľ÷
[{"credit_id": "52fe48009251416c750aca23", "de	[{"cast_id": 242, "character": "Jake Sully", "	[{"id": 1463, "name": "culture clash"}, {"id":	[Action, Adventure, Fantasy, Science Fiction]	In the 22nd century, a paraplegic Marine is di	Avatar	19995	0
[{"credit_id": "52fe4232c3a36847f800b579", "de	[{"cast_id": 4, "character": "Captain Jack Spa	[{"id": 270, "name": "ocean"}, {"id": 726, "na	[Adventure, Fantasy, Action]	Captain Barbossa, long believed to be dead, ha	Pirates of the Caribbean: At World's End	285	1
[{"credit_id": "54805967c3a36829b5002c41", "de	[{"cast_id": 1, "character": "James Bond", "cr	[{"id": 470, "name": "spy"}, {"id": 818, "name	[Action, Adventure, Crime]	A cryptic message from Bond's past sends him o	Spectre	206647	2
[{"credit_id": "52fe4781c3a36847f81398c3", "de	[{"cast_id": 2, "character": "Bruce Wayne / Ba	[{"id": 849, "name": "dc comics"}, {"id": 853,	[Action, Crime, Drama, Thriller]	Following the death of District Attorney Harve	The Dark Knight Rises	49026	3
[{"credit_id": "52fe479ac3a36847f813eaa3", "de	[{"cast_id": 5, "character": "John Carter", "c	[{"id": 818, "name": "based on novel"}, {"id":	[Action, Adventure, Science Fiction]	John Carter is a war-weary, former military ca	John Carter	49529	4

[25] movies['keywords'] = movies['keywords'].apply(convert)
 movies.head()

Working on column Keywords

	movie_id	title	overview overview	genres	keywords	cast	crew
C	19995	Avatar	In the 22nd century, a paraplegic Marine is di	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon	[{"cast_id": 242, "character": "Jake Sully", "	[{"credit_id": "52fe48009251416c750aca23", "de
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india	[{"cast_id": 4, "character": "Captain Jack Spa	[{"credit_id": "52fe4232c3a36847f800b579", "de
2	206647	Spectre	A cryptic message from Bond's past sends him o	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi	[{"cast_id": 1, "character": "James Bond", "cr	[{"credit_id": "54805967c3a36829b5002c41", "de
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i	[{"cast_id": 2, "character": "Bruce Wayne / Ba	[{"credit_id": "52fe4781c3a36847f81398c3", "de
4	49529	John Carter	John Carter is a war-weary, former military ca	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel	[{"cast_id": 5, "character": "John Carter", "c	[{"credit_id": "52fe479ac3a36847f813eaa3", "de

```
[26] def convert3(text):
    L = []
    counter = 0
    for i in ast.literal_eval(text):
        if counter < 3:
            L.append(i['name'])
        counter+=1
    return L</pre>
```

Working on column

Cast

```
[27] movies['cast'] = movies['cast'].apply(convert3)
movies.head()
```

crew	cast	keywords	genres	overview	title	novie_id	
[{"credit_id": "52fe48009251416c750aca23", "de	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[culture clash, future, space war, space colon	[Action, Adventure, Fantasy, Science Fiction]	In the 22nd century, a paraplegic Marine is di	Avatar	19995	0
[{"credit_id": "52fe4232c3a36847f800b579", "de	[Johnny Depp, Orlando Bloom, Keira Knightley]	[ocean, drug abuse, exotic island, east india	[Adventure, Fantasy, Action]	Captain Barbossa, long believed to be dead, ha	Pirates of the Caribbean: At World's End	285	1
[{"credit_id": "54805967c3a36829b5002c41", "de	[Daniel Craig, Christoph Waltz, Léa Seydoux]	[spy, based on novel, secret agent, sequel, mi	[Action, Adventure, Crime]	A cryptic message from Bond's past sends him o	Spectre	206647	2
[{"credit_id": "52fe4781c3a36847f81398c3", "de	[Christian Bale, Michael Caine, Gary Oldman]	[dc comics, crime fighter, terrorist, secret i	[Action, Crime, Drama, Thriller]	Following the death of District Attorney Harve	The Dark Knight Rises	49026	3
[{"credit_id": "52fe479ac3a36847f813eaa3", "de	[Taylor Kitsch, Lynn Collins, Samantha Morton]	[based on novel, mars, medallion, space travel	[Action, Adventure, Science Fiction]	John Carter is a war-weary, former military ca	John Carter	49529	4

Working on column Crew

```
[31] movies['crew'] = movies['crew'].apply(fetch_director)
```

```
[33] movies['overview'] = movies['overview'].apply(lambda x:x.split())
movies.sample(5)
```

ì	movie_id	title	overview	genres	keywords	cast	crew
611	4614	The Sum of All Fears	[When, the, president, of, Russia, suddenly, d	[Thriller, Action, Drama]	[cia, terrorist, atomic bomb, cold war, nuclea	[Ben Affleck, Morgan Freeman, James Cromwell]	[Phil Alden Robinson]
4128	22913	24 7: Twenty Four Seven	[In, a, typical, English, working-class, town,	[Comedy, Drama, Romance]	[transporter, sport, friends, nottingham]	[Bob Hoskins, Danny Nussbaum, Bruce Jones]	[Shane Meadows]
4047	356216	The Rise of the Krays	[Follows, the, early, years, of, two, unknown,	[Crime]	0	[Kevin Leslie, Simon Cotton, Phil Dunster]	[Zackary Adler]
1852	8699	Anchorman: The Legend of Ron Burgundy	[It's, the, 1970s,, and, San, Diego, super-sex	[Comedy]	[journalism, sexism, ladder, panda, tv show in	[Will Ferrell, Christina Applegate, Paul Rudd]	[Adam McKay]
646	4244	The Kid	[Powerful, businessman, Russ, Duritz, is, self	[Fantasy, Comedy, Family]	[age difference, midlife crisis, suppressed pa	[Bruce Willis, Spencer Breslin, Emily Mortimer]	[Jon Turteltaub]

```
[34] def collapse(L):
        L1 = []
        for i in L:
            L1.append(i.replace(" ",""))
        return L1
```

```
[35] movies['cast'] = movies['cast'].apply(collapse)
movies['crew'] = movies['crew'].apply(collapse)
movies['genres'] = movies['genres'].apply(collapse)
movies['keywords'] = movies['keywords'].apply(collapse)
```

## Joining all columns in single column named tags

```
[39] movies['tags'] = movies['overview'] + movies['genres'] + movies['keywords'] + movies['cast'] + movies['crew']

[42] new = movies.drop(columns=['overview', 'genres', 'keywords', 'cast', 'crew'])

#new.head()

[43] new['tags'] = new['tags'].apply(lambda x: " ".join(x))

new.head()

movie_id

title

tags

Avatar In the 22nd century, a paraplegic Marine is di...
```

Captain Barbossa, long believed to be dead, ha...

Following the death of District Attorney Harve...

John Carter is a war-weary, former military ca ...

Spectre A cryptic message from Bond's past sends him o...

Pirates of the Caribbean: At World's End

The Dark Knight Rises

John Carter

206647

49026

49529

3

# Text vectorization

```
/ [37] from sklearn.feature extraction.text import CountVectorizer
        cv = CountVectorizer(max features=5000, stop words='english')
  [38] vector = cv.fit transform(new['tags']).toarray()
  [40] vector
       array([[0, 0, 0, ..., 0, 0, 0],
               [0, 0, 0, \ldots, 0, 0, 0],
               [0, 0, 0, ..., 0, 0, 0],
               [0, 0, 0, ..., 0, 0, 0],
               [0, 0, 0, ..., 0, 0, 0],
               [0, 0, 0, ..., 0, 0, 0]])
  [39] vector.shape
        (4806, 5000)
```

Using
Bag of
words
Technique

```
[42] cv.get_feature_names()
       ustronuut,
      'astronauts',
      'asylum',
      'atheist',
      'athlete',
      'atlantic',
      'atomicbomb',
      'attack',
      'attacked',
      'attacks',
      'attempt',
      'attempting',
      'attempts',
      'attending',
      'attends',
      'attention',
      'attic',
      'attitude',
      'attorney',
      'attracted',
      'attraction',
      'attractive',
      'audience',
      'audiences',
      'audition',
      'august',
      'aunt',
      'austin',
      'australia',
      'australian',
      'author',
      'authorities',
```

# Stemming

```
[43] import nltk

[44] from nltk.stem.porter import PorterStemmer
    ps = PorterStemmer()

[45] def stem(text):
        y=[]
        for i in text.split():
            y.append(ps.stem(i))
        return " ".join(y)
[48] new['tags']=new['tags'].apply(stem)
```

After stemming, again repeat vectorization steps.

# **Using Cosine similarity**

## Helper function for Recommendation

# Results

```
[59] recommend('Gandhi')

Gandhi, My Father
Guiana 1838
The Wind That Shakes the Barley
Mr. Turner
A Passage to India
```

```
[62] recommend('Titanic')

The Notebook
Under the Same Moon
Ghost Ship
The Bounty
Pirates of the Caribbean: On Stranger Tides
```