PROJECT REPORT

(Project Term Jan-May 2023)

Movie Recommender System

Submitted by

Abimanyu S

12109470

Course Code: INT216

Under the guidance of

Mr. Ved Prakash Chaubey: 63892

School of Computer Science and Engineering,

LOVELY PROFESSIONAL UNIVERSITY PHAGWARA, PUNJAB.

<u>DECLARATION</u>

I **Abimanyu S 12109470**, hereby declare that the project work entitled **Movie Recommender System** is an authentic record of my own work carried out as requirements of Capstone Project for the award of B.Tech degree in CSE Data Science (AI & ML) from Lovely Professional University, Phagwara, under the guidance of our INT216(Python Project) staff Mr. Ved Prakash Chaubey(63892). All the information furnished in this capstone project report is based on my own intensive work and is genuine.

Abimanyu S,

12109470.

May 5, 2023

Signature of Staff   **Signature of Student**

<u>ACKNOWLEDGEMENT</u>

<u>TABLE OF CONTENT</u>

1. Abstract:

The Movie Recommender System is an artificial intelligence-based system that suggests movies to users based on their viewing history and preferences. In this project, we have developed a movie recommender system using natural language processing techniques and machine learning algorithms with Python and its libraries such as Pandas, NumPy, and Scikit-learn. We have used a dataset of movies that includes information about the title, cast, crew, keywords, genres, and overview. We have pre-processed this data, extracted relevant features, and created a movie tags vector that represents each movie. We have then used cosine similarity to find the similarity between the tags vector of the given movie and all the other movies in the dataset. The user interface of the system is built using the Tkinter library in Python. Finally, we have recommended the top 5 movies that are most similar to the given movie.

2. Introduction:

Movies are a form of entertainment enjoyed by millions of people worldwide. However, choosing a movie to watch can be a daunting task due to the vast number of available options. In recent years, movie recommender systems have become popular because they help users find movies that match their interests. The primary goal of a movie recommender system is to make personalized movie recommendations to users based on their preferences.

This project aims to build a movie recommender system using Python and its libraries. The system analyzes information about movies and users' preferences to make personalized recommendations. The system's user interface is built using the Tkinter library in Python, making it easy for users to interact with the system.

3. Objective of the Project:

The objective of this project is to create a movie recommendation system that can suggest movies to users based on their preferences. The system will use a dataset of movies and their attributes such as genre, cast, crew, keywords, and plot summary to generate a set of tags for each movie. These tags will be used to calculate the similarity between movies and recommend movies that are similar to the ones that the user has liked in the past.

The project aims to develop a functional and user-friendly recommendation system that can provide accurate and relevant movie recommendations to users.

Additionally, the project aims to explore various techniques for text processing, natural language processing, and machine learning to build an efficient and effective recommendation system.

Summarization of the objective

- Analyze information about movies and users' preferences.
- Make personalized movie recommendations based on users' interests.
- Provide a user-friendly interface for users to interact with the system.

4. Description of the Project:

The source code for the movie recommender system is written in Python. We have used several libraries such as numpy, pandas, scikit-learn, and tkinter. The system uses a dataset containing information about movies and users, preprocesses the data, and then uses the cosine similarity technique to make personalized movie recommendations. The code is divided into several parts:

- Data pre-processing: We have used pandas to read the movies dataset and pre-process the data.

- Feature extraction: We have used natural language processing techniques to extract relevant features from the data such as genres, keywords, cast, crew, and overview. We have then created a movie tags vector that represents each movie.

- Similarity calculation: We have used cosine similarity to find the similarity between the tags vector of the given movie and all the other movies in the dataset.

- User interface: We have used tkinter to create a simple user interface that takes a movie name as input and recommends the top 5 movies that are most similar to the given movie.

5. Source Code:

The source code for the Movie Recommender System project is as follows:

```
In [5]: pd.set_option('display.max_columns',None)
        pd.set_option('display.max_rows',None)
```

```
In [6]: credits_df
```

Out[6]:

	movie_id	title	cast	crew
0	19995	Avatar	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_id": "52fe4232c3a36847f800b579", "de...
2	206647	Spectre	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_id": "54805967c3a36829b5002c41", "de...
3	49026	The Dark Knight Rises	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"credit_id": "52fe4781c3a36847f81398c3", "de...
4	49529	John Carter	[{"cast_id": 5, "character": "John Carter", "c...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...
5	559	Spider-Man 3	[{"cast_id": 30, "character": "Peter Parker / ...	[{"credit_id": "52fe4252c3a36847f80151a5", "de...
6	38757	Tangled	[{"cast_id": 34, "character": "Flynn Rider (vo...	[{"credit_id": "52fe46db9251416c91062101", "de...
7	99861	Avengers: Age of Ultron	[{"cast_id": 76, "character": "Tony Stark / Ir...	[{"credit_id": "55d5f7d4c3a3683e7e0016eb", "de...
8	767	Harry Potter and the Half-Blood Prince	[{"cast_id": 3, "character": "Harry Potter", "...	[{"credit_id": "52fe4273c3a36847f801fab1", "de...
9	209112	Batman v Superman: Dawn of Justice	[{"cast_id": 18, "character": "Bruce Wayne / B...	[{"credit_id": "553bf23692514135c8002886", "de...
10	1452	Superman Returns	[{"cast_id": 3, "character": "Superman / Clark...	[{"credit_id": "553bef6a9251416874003c8f", "de...

```
In [7]: movies_df
```

Out[7]:

	budget	genres	homepage	id	keywords	original_language	original_title	
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	en	Avatar	In the a par...
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	en	Pirates of the Caribbean: At World's End	Cap long...
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name...	en	Spectre	A cr fro...
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	http://www.thedarkknightrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853,...	en	The Dark Knight Rises	Follo of D...

```
In [8]: credits_df.head()
```

Out[8]:

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

```
In [9]: movies_df.tail()
```

Out[9]:

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity
4798	220000	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	NaN	9367	[{"id": 5616, "name": "united states\u2013mexi...	es	El Mariachi	El Mariachi just wants to play his guitar and ...	14.269792
4799	9000	[{"id": 35, "name": "Comedy"}, {"id": 10749, "...	NaN	72766	[]	en	Newlyweds	A newlywed couple's honeymoon is upended by th...	0.642552
4800	0	[{"id": 35, "name": "Comedy"}, {"id": 18, "nam...	http://www.hallmarkchannel.com/signedsealeddel...	231617	[{"id": 248, "name": "date"}, {"id": 699, "nam...	en	Signed, Sealed, Delivered	"Signed, Sealed, Delivered" introduces a dedic...	1.444476
4801	0	[]	http://shanghaicalling.com/	126186	[]	en	Shanghai Calling	When ambitious New York attorney Sam is sent t...	0.857008
4802	0	[{"id": 99, "name": "Documentary"}]	NaN	25975	[{"id": 1523, "name": "obsession"}, {"id": 224...	en	My Date with Drew	Ever since the second grade when he first saw ...	1.929883

```
In [10]: movies_df = movies_df.merge(credits_df, on = 'title')
```

```
In [11]: movies_df.shape
```

Out[11]: (4808, 23)

```
In [12]: movies_df.head()
```
Out[12]:

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	production_comp
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	[{"name": "Inge Film Partners
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	[{"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 [Pictures", "id": 2
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name...	en	Spectre	A cryptic message from Bond's past sends him o...	107.376788	[{"name": "Col Pictures", "i {"
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	http://www.thedarkknightrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853,...	en	The Dark Knight Rises	Following the death of District Attorney Harve...	112.312950	[{"name": "Lege Pictures", "id": 92:
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://movies.disney.com/john-carter	49529	[{"id": 818, "name": "based on novel"}, {"id":...	en	John Carter	John Carter is a war-weary, former military ca...	43.926995	[{"name": "Walt [Pictures", "i

```
In [13]: movies_df.info()
```
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4808 entries, 0 to 4807
Data columns (total 23 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   budget                4808 non-null   int64
 1   genres                4808 non-null   object
 2   homepage              1713 non-null   object
 3   id                    4808 non-null   int64
 4   keywords              4808 non-null   object
 5   original_language     4808 non-null   object
 6   original_title        4808 non-null   object
 7   overview              4805 non-null   object
 8   popularity            4808 non-null   float64
 9   production_companies  4808 non-null   object
 10  production_countries  4808 non-null   object
 11  release_date          4807 non-null   object
 12  revenue               4808 non-null   int64
 13  runtime               4806 non-null   float64
 14  spoken_languages      4808 non-null   object
 15  status                4808 non-null   object
 16  tagline               3964 non-null   object
 17  title                 4808 non-null   object
 18  vote_average          4808 non-null   float64
 19  vote_count            4808 non-null   int64
 20  movie_id              4808 non-null   int64
 21  cast                  4808 non-null   object
 22  crew                  4808 non-null   object
dtypes: float64(3), int64(5), object(15)
memory usage: 901.5+ KB
```

```
In [14]: movies_df = movies_df[['movie_id','title','overview','genres','keywords','cast','crew']]
```

```
In [15]: movies_df.head()
```
Out[15]:

	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, "...	[{"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...

```
In [16]: movies_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4808 entries, 0 to 4807
Data columns (total 7 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   movie_id  4808 non-null   int64
 1   title     4808 non-null   object
 2   overview  4805 non-null   object
 3   genres    4808 non-null   object
 4   keywords  4808 non-null   object
 5   cast      4808 non-null   object
 6   crew      4808 non-null   object
dtypes: int64(1), object(6)
memory usage: 300.5+ KB
```

```
In [17]: movies_df.isnull().sum()
```

```
Out[17]: movie_id    0
         title       0
         overview    3
         genres      0
         keywords    0
         cast        0
         crew        0
         dtype: int64
```

```
In [18]: movies_df.dropna(inplace=True)
```

```
In [19]: movies_df.isnull().sum()
```

```
Out[19]: movie_id    0
         title       0
         overview    0
         genres      0
         keywords    0
         cast        0
         crew        0
         dtype: int64
```

```
In [20]: movies_df.duplicated().sum()
```

```
Out[20]: 0
```

```
In [21]: movies_df.iloc[0].genres
```

```
Out[21]: '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
```

```
In [22]: import ast
```

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

```
In [24]: movies_df['genres'] = movies_df['genres'].apply(convert)
         movies_df['keywords'] = movies_df['keywords'].apply(convert)
         movies_df.head()
```

Out[24]:

	movie_id	title	overview	genres	keywords	cast	crew
0	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...

```
In [25]: def convert2(obj):
             L = []
             counter = 0
             for i in ast.literal_eval(obj):
                 if counter !=3:
                     L.append(i['name'])
                     counter += 1
                 else:
                     break
             return L
```

```
In [26]: movies_df['cast'] = movies_df['cast'].apply(convert2)
         movies_df.head()
```

Out[26]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington]	[{"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 ...	[Johnny Depp]	[{"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...	[Daniel Craig]	[{"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...	[Christian Bale]	[{"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...	[Taylor Kitsch]	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

```
In [27]: def fetch_director(obj):
             L=[]
             for i in ast.literal_eval(obj):
                 if i['job']=='Director':
                     L.append(i['name'])
             return L
```

```
In [28]: movies_df['crew']=movies_df['crew'].apply(fetch_director)
```

```
In [29]: movies_df.head()
```

Out[29]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington]	[James Cameron]
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 ...	[Johnny Depp]	[Gore Verbinski]
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[Daniel Craig]	[Sam Mendes]
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...	[Christian Bale]	[Christopher Nolan]
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[Taylor Kitsch]	[Andrew Stanton]

```
In [30]: movies_df['overview'][0]
```

Out[30]: 'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes torn between following orders and protecting an alien civilization.'

```
In [31]: movies_df['overview']=movies_df['overview'].apply(lambda x:x.split())
```

```
In [32]: movies_df
```

Out[32]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington]	[James Cameron]
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[Johnny Depp]	[Gore Verbinski]
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[Daniel Craig]	[Sam Mendes]
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...	[Christian Bale]	[Christopher Nolan]
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[Taylor Kitsch]	[Andrew Stanton]
5	559	Spider-Man 3	[The, seemingly, invincible, Spider-Man, goes,...	[Fantasy, Action, Adventure]	[dual identity, amnesia, sandstorm, love of on...	[Tobey Maguire]	[Sam Raimi]
6	38757	Tangled	[When, the, kingdom's, most, wanted-and, most ...	[Animation, Family]	[hostage, magic, horse, fairy tale, musical, p...	[Zachary Levi]	[Byron Howard, Nathan Greno]

```
In [33]: movies_df['genres']=movies_df['genres'].apply(lambda x:[i.replace(" ","") for i in x])
         movies_df['keywords']=movies_df['keywords'].apply(lambda x:[i.replace(" ","") for i in x])

         movies_df['crew']=movies_df['crew'].apply(lambda x:[i.replace(" ","") for i in x])
```

```
In [33]: movies_df['genres']=movies_df['genres'].apply(lambda x:[i.replace(" ","") for i in x])
         movies_df['keywords']=movies_df['keywords'].apply(lambda x:[i.replace(" ","") for i in x])

         movies_df['crew']=movies_df['crew'].apply(lambda x:[i.replace(" ","") for i in x])
```

```
In [34]: movies_df
```

Out[34]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacewar, spacecolony, ...	[Sam Worthington]	[JamesCameron]
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...	[Adventure, Fantasy, Action]	[ocean, drugabuse, exoticisland, eastindiatrad...	[Johnny Depp]	[GoreVerbinski]
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...	[Daniel Craig]	[SamMendes]
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...	[Christian Bale]	[ChristopherNolan]
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p...	[Taylor Kitsch]	[AndrewStanton]
			[The, seemingly				

```
In [35]: movies_df['tags']=movies_df['overview']+movies_df['genres']+movies_df['keywords']+movies_df['cast']+movies_df['crew']
```

```
In [36]: movies_df
```

Out[36]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacewar, spacecolony, ...	[Sam Worthington]	[JamesCameron]
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...	[Adventure, Fantasy, Action]	[ocean, drugabuse, exoticisland, eastindiatrad...	[Johnny Depp]	[GoreVerbinski]
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...	[Daniel Craig]	[SamMendes]
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...	[Christian Bale]	[ChristopherNolan]
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p...	[Taylor Kitsch]	[AndrewStanton]

```
In [37]: new_df = movies_df[['movie_id','title','tags']]
```

```
In [38]: new_df
```

27	44833	Battleship	[When, mankind, beams, a, radio, signal, into,...
28	135397	Jurassic World	[Twenty-two, years, after, the, events, of, Ju...
29	37724	Skyfall	[When, Bond's, latest, assignment, goes, grave...
30	558	Spider-Man 2	[Peter, Parker, is, going, through, a, major, ...
31	68721	Iron Man 3	[When, Tony, Stark's, world, is, torn, apart, ...
32	12155	Alice in Wonderland	[Alice,, an, unpretentious, and, individual, 1...
33	36668	X-Men: The Last Stand	[When, a, cure, is, found, to, treat, mutation...
34	62211	Monsters University	[A, look, at, the, relationship, between, Mike...
35	8373	Transformers: Revenge of the Fallen	[Sam, Witwicky, leaves, the, Autobots, behind,...
36	91314	Transformers: Age of Extinction	[As, humanity, picks, up, the, pieces,, follow...
37	68728	Oz: The Great and Powerful	[Oscar, Diggs,, a, small-time, circus, illusio...
38	102382	The Amazing Spider-Man 2	[For, Peter, Parker,, life, is, busy., Between...
39	20526	TRON: Legacy	[Sam, Flynn,, the, tech-savvy, and, daring, so...

```
In [39]: new_df['tags'] = new_df['tags'].fillna('').apply(lambda x: ' '.join([tag for tag in x if isinstance(tag, str)]))
```

```
C:\Users\sabim\AppData\Local\Temp\ipykernel_11412\4276168096.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ve
rsus-a-copy
  new_df['tags'] = new_df['tags'].fillna('').apply(lambda x: ' '.join([tag for tag in x if isinstance(tag, str)]))
```

```
In [40]: new_df['tags']=new_df['tags'].apply(lambda X:X.lower())
```

```
C:\Users\sabim\AppData\Local\Temp\ipykernel_11412\4064007771.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ve
rsus-a-copy
  new_df['tags']=new_df['tags'].apply(lambda X:X.lower())
```

```
In [41]: from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(max_features=5000,stop_words='english')
```

```
In [42]: cv.fit_transform(new_df['tags']).toarray().shape
```

Out[42]: (4805, 5000)

```
In [43]: vectors = cv.fit_transform(new_df['tags']).toarray()
```

```
In [44]: vectors[0]
```

Out[44]: array([0, 0, 0, ..., 0, 0, 0], dtype=int64)

```
In [45]: len(cv.get_feature_names())
```

```
C:\Users\sabim\anaconda3\lib\site-packages\sklearn\utils\deprecation.py:87: FutureWarning: Function get_feature_names is deprec
ated; get_feature_names is deprecated in 1.0 and will be removed in 1.2. Please use get_feature_names_out instead.
  warnings.warn(msg, category=FutureWarning)
```

Out[45]: 5000

14

```python
In [46]: import nltk
```

```python
In [47]: from nltk import PorterStemmer
         ps = PorterStemmer()
```

```python
In [48]: def stem(text):
             y=[]
             for i in text.split():
                 y.append(ps.stem(i))
             return" ".join(y)
```

```python
In [49]: new_df['tags']=new_df['tags'].apply(stem)
```

```
C:\Users\sabim\AppData\Local\Temp\ipykernel_11412\3514595201.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ve
rsus-a-copy
  new_df['tags']=new_df['tags'].apply(stem)
```

```python
In [50]: from sklearn.metrics.pairwise import cosine_similarity
```

```python
In [51]: cosine_similarity(vectors)
```

```
Out[51]: array([[1.        , 0.08858079, 0.05905386, ..., 0.02423931, 0.08219949,
                 0.        ],
                [0.08858079, 1.        , 0.06451613, ..., 0.02648136, 0.        ,
                 0.        ],
                [0.05905386, 0.06451613, 1.        , ..., 0.02648136, 0.02993422,
                 0.        ],
                ...,
                [0.02423931, 0.02648136, 0.02648136, ..., 1.        , 0.07372098,
                 0.04129201],
                [0.08219949, 0.        , 0.02993422, ..., 0.07372098, 1.        ,
                 0.046676  ],
                [0.        , 0.        , 0.        , ..., 0.04129201, 0.046676  ,
                 1.        ]])
```

```python
In [52]: cosine_similarity(vectors).shape
```

```
Out[52]: (4805, 4805)
```

```python
In [53]: similarity=cosine_similarity(vectors)
```

```python
In [54]: similarity[0]
```

```
Out[54]: array([1.        , 0.08858079, 0.05905386, ..., 0.02423931, 0.08219949,
                0.        ])
```

```python
In [55]: similarity[0].shape
```

```
Out[55]: (4805,)
```

```python
In [56]: sorted(list(enumerate(similarity[0])), reverse=True, key=lambda x:x[1])[1:6]
```

```
Out[56]: [(539, 0.25756371316446414),
          (507, 0.24840131369742968),
          (260, 0.24012009007506566),
          (1194, 0.23995690956687135),
          (1216, 0.23995690956687135)]
```

```python
In [61]: import tkinter as tk
         from tkinter import messagebox

         def recommend():
             movie = entry.get()
             if movie not in new_df['title'].values:
                 messagebox.showerror("Error", "Movie not found. Please try again.")
             else:
                 movie_index = new_df[new_df['title'] == movie].index[0]
                 distances = similarity[movie_index]
                 movies_list = sorted(list(enumerate(distances)), reverse=True, key=lambda x: x[1])[1:6]
                 recommendations = "Top 5 recommendations for you:\n"
                 for i in movies_list:
                     recommendations += new_df.iloc[i[0]].title + "\n"
                 messagebox.showinfo("Recommendations", recommendations)

         root = tk.Tk()
         root.title("Movie Recommender")
         root.geometry("400x200")

         label = tk.Label(root, text="Enter a movie name:")
         label.pack()

         entry = tk.Entry(root)
         entry.pack()

         button = tk.Button(root, text="Recommend", command=recommend)
         button.pack()

         root.mainloop()
```
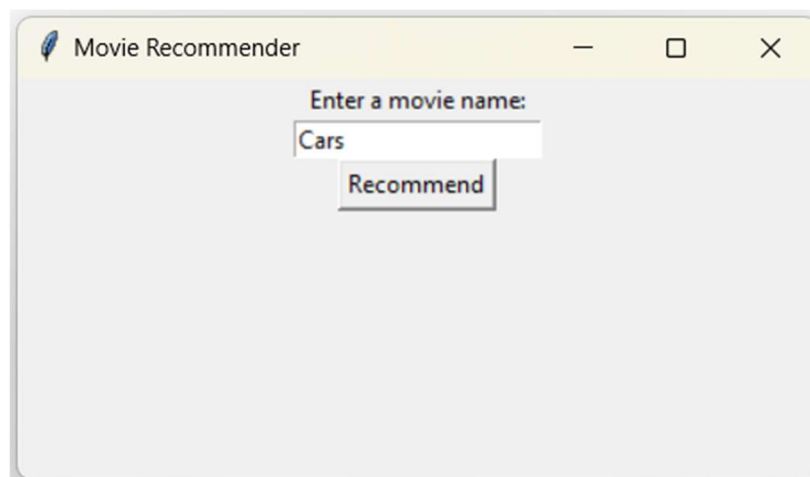
6. Input/Output:

The Movie Recommender System takes the name of a movie as input and provides the user with five movie recommendations. The user interface of the system is built using the Tkinter library in Python, making it easy for users to interact with the system. The input is provided through the GUI, and the output is displayed in a message box.

INPUT

'Cars' is the input received from the user.

OUTPUT
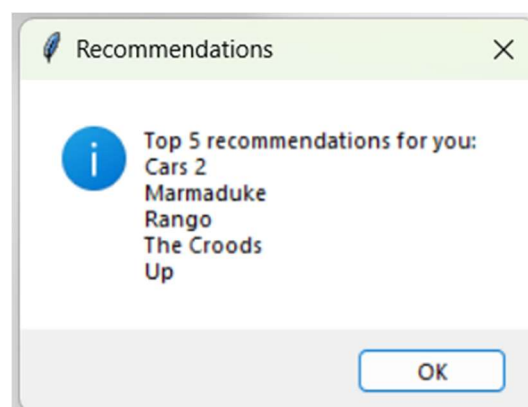
Movie Recommender provides the top 5 recommendations based on the input 'Cars'.

7. Scope of the Project:

The scope of this project is to develop a movie recommendation system that utilizes machine learning techniques to recommend movies to users based on their preferences. The system takes in user input in the form of a movie title and generates a list of similar movies that the user may enjoy. The recommendation is based on analyzing the metadata associated with the movies, such as their genres, cast, crew, and plot summaries. The system uses natural language processing techniques to analyze the text data and identify patterns and relationships between movies. The output of the system is a list of recommended movies that is displayed to the user. This project has a wide scope as it can be used by various online movie platforms to enhance the user experience by providing personalized recommendations.

The recommendation system can be integrated with other applications, such as a movie streaming platform or a movie review website. The project can be further developed to include collaborative filtering methods to provide even more accurate movie recommendations. The recommendation system can be improved through the addition of user feedback and ratings, allowing for a more personalized and tailored movie recommendation experience.

8. Development:

The development phase of the project involved several steps. Firstly, we imported the necessary libraries and loaded the datasets using pandas. Then I cleaned and pre-processed the data by removing unnecessary columns and rows, handling missing values and converting some columns into more useful formats. We also extracted important features from the data, such as cast, crew, genres, and keywords, and combined them into a single 'tags' column.

After the data was cleaned and processed, we used the CountVectorizer from sklearn to convert the tags into a matrix of word counts. We then applied stemming to the words in the tags to reduce them to their root form. Finally, we used cosine similarity to calculate the similarity between the movies based on their tags.

We also developed a GUI using tkinter, where the user can input the name of a movie and the system recommends five similar movies based on their tags. The GUI was designed to be user-friendly, with a simple interface and clear instructions.

Overall, the development phase of the project was focused on cleaning and preprocessing the data, feature extraction, and developing the recommendation system using cosine similarity. The final output was a functional GUI that allowed users to receive movie recommendations based on their input.

9. Conclusion:

The Movie Recommender project is a successful attempt to provide personalized movie recommendations to users. The project utilizes machine learning algorithms and natural language processing techniques to generate recommendations based on the user's preferences.

The project involved the collection of data from two different sources, data cleaning, and data preprocessing. The data was then analyzed using different natural language processing techniques, such as stemming, tokenization, and vectorization, to generate a similarity matrix. The similarity matrix was then used to generate movie recommendations for the user.
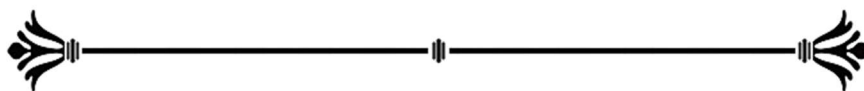
The project was successful in achieving its objective of creating a movie recommender system that recommends movies based on their similarity to the user's preferred movie. The project's scope was also discussed, which includes the potential to expand the system to include other factors such as user preferences, movie ratings, and user demographics.

The project's development process was described in detail, including the source code, input/output, and development methodology. The project used various Python libraries such as NumPy, Pandas, NLTK, and Scikit-learn, which were essential in achieving the project's objective.

In summary, the project's movie recommender system provides a useful tool for movie enthusiasts to discover new movies that they may enjoy. The system has the potential to save users time and effort that would otherwise be spent searching for movies manually. The project has potential for further development and expansion, and its success showcases the power of natural language processing techniques and machine learning algorithms in generating valuable insights and recommendations.

10. Bibliography:

- Overview - https://en.wikipedia.org/wiki/Recommender_system
- Datasets - https://grouplens.org/datasets/movielens/
- Python - https://www.python.org/
- tkinter - https://docs.python.org/3/library/tkinter.html
- scikit-learn - https://scikit-learn.org/stable/tutorial/index.html

Thank You!