

**A Mini Project**  
ON  
**“Movie Recommendation model”**

SUBMITTED BY  
**Kohraj Gajanan Kakade**  
OF  
**TE (Computer Engineering) Academic Year: 2022-2023**

GUIDED BY  
**Prof. Swati S. Bhosale**



**DEPARTMENT OF COMPUTER ENGINEERING**

**HSBPVT'S PARIKRAMA COLLEGE OF ENGINEERING, KASHTI**

**SAVITRIBAI PHULE PUNE UNIVERSITY**

**2023-24**

## **Acknowledgment**

I am profoundly grateful to **Prof. Bhosale Swati S.** for her expert guidance and continuous encouragement throughout to see that this project reached its target from its commencement to its completion.

Finally, we must express our sincere heartfelt gratitude to all the staff members of the Computer Engineering Department who helped me directly or indirectly during this course of work.

## **ABSTRACT**

Using the Movie Recommendation model using the scikit-learn library in Python and we perform following analytics on the given dataset.

- a. Describe the dataset
- b. Find out the top 5 movies

## CONTENTS

CHAPTER	TITLE	PAGE NO
1	Problem Statement Definition	6
2	Software Requirements Specification	6
3	Introduction	6
4	Data Analytics	7
5	Code Snippets	8
6	Conclusion	11
7	References	12

## **1. Problem Statement Definition:**

Develop a movie recommendation model using the sci-kit-learn library in Python.

## **2. Software Requirements Specification**

### **TECHNOLOGY USED:**

#### **Libraries:**

- Pandas
- NumPy
- Scikit-learn

#### **Requirements:**

- Anaconda
- Jupyter Notebook

### 3. Introduction:

- **skit-learn Library:**

Sklearn, also known as scalar, is a widely-used open-source machine learning library designed for Python. It provides an array of tools for data mining, data analysis, and machine learning tasks, including regression, classification, clustering, and dimensionality reduction.

Sklearn is built on top of other popular scientific computing libraries such as NumPy, SciPy, and Matplotlib, and offers a consistent API for using a range of machine learning models and algorithms. Additionally, it includes many useful features such as cross validation, model selection, and pre-processing. Some of the most commonly used machine learning algorithms that can be implemented using sklearn include linear regression, logistic regression, support vector machines (SVM), k-nearest neighbors (KNN), decision trees, random forests, and neural networks.

To use sklearn, you must first install the SciPy (Scientific Python) stack, which includes NumPy for base n-dimensional array support, SciPy for fundamental scientific computing library support, Matplotlib for comprehensive 2D/3D plotting, Python for an enhanced interactive console, SymPy for symbolic mathematics, and Pandas for data structures and analysis.

### **Recommendation System Mechanism:**

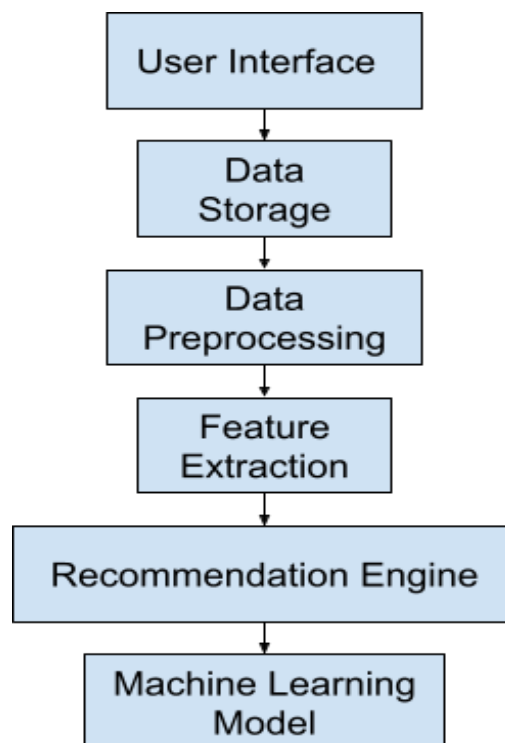
The mechanism underlying a recommendation system generally involves a series of steps:

- **Data Collection:** The first step is to collect user data, including browsing history, purchase history, ratings, user behavior, and other pertinent information.
- **Data Preprocessing:** Once the data is gathered, it must be preprocessed to remove

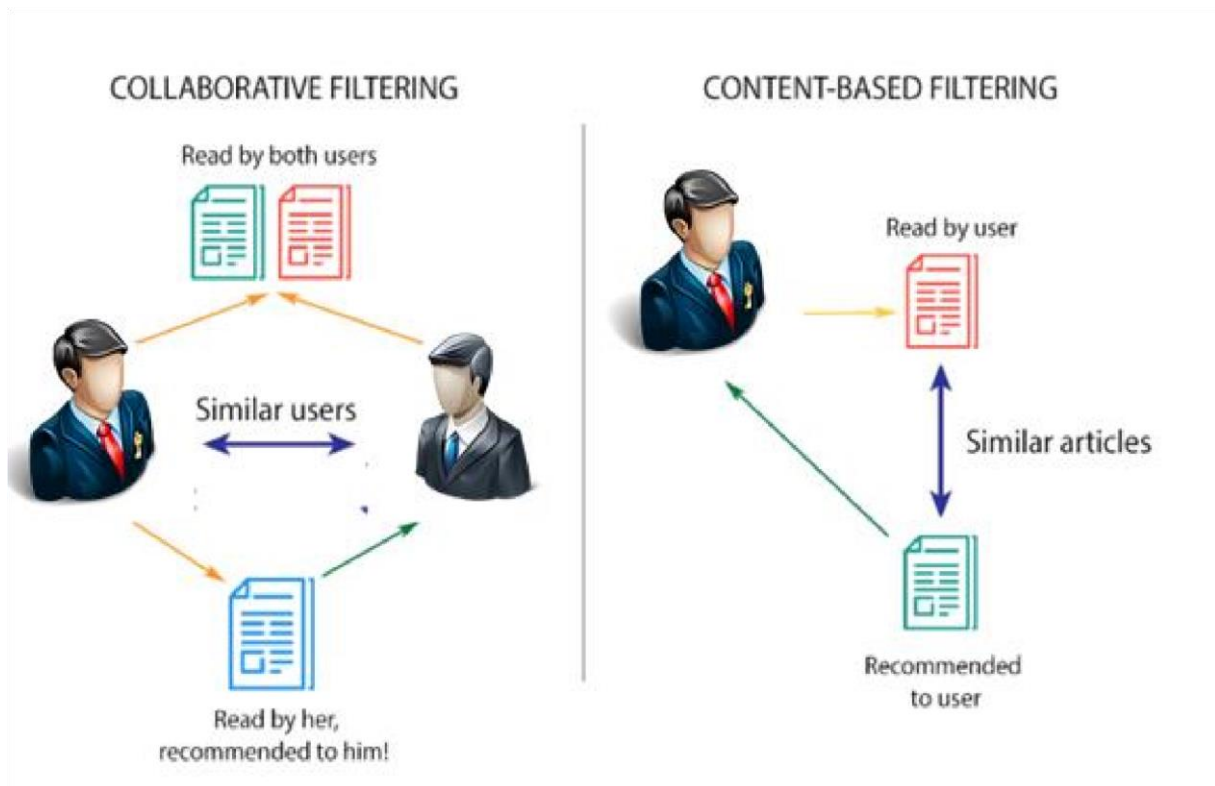
duplicates, manage missing values, and transform it into a suitable format for analysis.

- **Feature Extraction:** In this step, the system extracts relevant features from the data, such as genre, actors, or keywords, which can be utilized to identify patterns and similarities between items.
- **Similarity Calculation:** The system computes the similarity between items based on their features, using various techniques such as cosine similarity or Euclidean distance.
- **Recommendation Generation:** Based on the similarity scores, the system generates a list of recommended items that are comparable to the ones the user has interacted with previously.
- **Evaluation:** Finally, the system evaluates the recommendations by measuring their accuracy and relevance to the user, using various metrics such as precision, recall, and F1-score.

#### Architecture:



## Types of recommendation systems:



## Code & Output:

```
In [1]: from sklearn.metrics.pairwise import cosine_similarity
import pandas as pd
import numpy as np
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics.pairwise import cosine_similarity

In [2]: df = pd.read_csv("https://raw.githubusercontent.com/rashida048/Some-NLP-Projects/master/movie_dataset.csv")

In [3]: df.head()
```

	index	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	runtime	spoken_languages	status	tagline	title	vote_ave
0	0	237000000	Action Adventure Fantasy Science Fiction	http://www.avatarmovie.com/	19995	culture clash future space war space colony so...	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	162.0	[{"iso_639_1": "en", "name": "English"}, {"iso_639_1": "fr", "name": "French"}]	Released	Enter the World of Pandora.	Avatar	
1	1	300000000	Adventure Fantasy Action	http://disney.go.com/disneypictures/pirates/	285	ocean drug abuse exotic island east india trad...	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	139.082615	169.0	[{"iso_639_1": "en", "name": "English"}]	Released	At the end of the world, the adventure begins.	Pirates of the Caribbean: At World's End	
2	2	245000000	Action Adventure Crime	http://www.sonypictures.com/movies/spectre/	206647	spy based on novel secret agent sequel mi6	en	Spectre	A cryptic message from Bond's past sends him o...	107.376788	148.0	[{"iso_639_1": "fr", "name": "French"}, {"iso_639_1": "en", "name": "English"}]	Released	A Plan No One Escapes	Spectre	
3	3	250000000	Action Crime Drama Thriller	http://www.thedarkknighttrises.com/	49026	dc comics crime fighter terrorist secret ident...	en	The Dark Knight Rises	Following the death of District Attorney Harve...	112.312950	165.0	[{"iso_639_1": "en", "name": "English"}]	Released	The Legend Ends	The Dark Knight Rises	

Windows taskbar: Type here to search, 36°C Sunny, 14:41, 11-04-2024



```
In [13]: i=0
print("Top 5 similar movies to "+movie_user_likes+" are:\n")
for element in sorted_similar_movies:
    print(get_title_from_index(element[0]))
    i=i+1
    if i>5:
        break
```

Top 5 similar movies to Avatar are:

Guardians of the Galaxy  
 Aliens  
 Star Wars: Clone Wars: Volume 1  
 Star Trek Into Darkness  
 Star Trek Beyond  
 Alien

In [ ]:

DSBDA\_Mini\_Project.pdf x DSBDA Mini project 1 n... x Home Page - Select or ci x DSBDA Mini project - Ju... x aboutblank x DSBDA Mini project x

File | C:/Users/DELL/Downloads/DSBDA%20Mini%20project.html

4	4	260000000	Action Adventure Science Fiction	http://movies.disney.com/john-carter	49529	based on novel mars medallion space travel pri...	en	John Carter	John Carter is a war- weary, former military ca...	43.926995	-	132.0	[{"iso_639_1": "en", "name": "English"}]	Released	Lost in our world, found in another.	John Carter
---	---	-----------	---	--------------------------------------	-------	---	----	-------------	--	-----------	---	-------	---	----------	--	----------------

5 rows x 24 columns

```
In [4]: features = ['keywords', 'cast', 'genres', 'director']

In [6]: def combine_features(row):
        return row['keywords']+" "+row['cast']+" "+row['genres']+" "+row['director']

In [7]: for feature in features:
        df[feature] = df[feature].fillna('')

        df["combined_features"] = df.apply(combine_features, axis=1)

In [8]: cv = CountVectorizer()
count_matrix = cv.fit_transform(df["combined_features"])

In [9]: cosine_sim = cosine_similarity(count_matrix)

In [10]: def get_title_from_index(index):
        return df[df.index == index]["title"].values[0]
def get_index_from_title(title):
        return df[df.title == title]["index"].values[0]

In [11]: movie_user_likes = "Avatar"
movie_index = get_index_from_title(movie_user_likes)
similar_movies = list(enumerate(cosine_sim[movie_index]))

In [12]: sorted_similar_movies = sorted(similar_movies, key=lambda x: x[1], reverse=True)[1:]

In [13]: i=0
print("Top 5 similar movies to "+movie_user_likes+" are:\n")
for element in sorted_similar_movies:
    print(get_title_from_index(element[0]))
```

Type here to search 36°C Sunny 14:46 11-04-2024

## **Conclusion**

Thus, we have successfully processed and visualized data with new datasets as it is suitable when data is available.

## References

- <https://www.studocu.com/in/document/savitribai-phule-pune-university/data-science-honors-course/mini-project-dsbd-miniproject/58609858>