# **Experiment No. 5**

**<u>Aim:</u>** Develop Content based social media analytics model for business.

## **Theory:**

This project is a Movie Recommender System which recommends movies to the user:-Based on

1. Popularity: - Correlation

2. Collaborative Filtering:- KNN

3. Content Based: - NLP and Cosine Similarity

Using Classification algorithm, NLP and Correlation.

#### **Methods Used:**

## • Feature Engineering:-

Feature engineering is the process of transforming raw data into features that are suitable for machine learning models. In other words, it is the process of selecting, extracting, and transforming the most relevant features from the available data to build more accurate and efficient machine learning models.

## • Machine Learning:-

Machine learning (ML) is a discipline of artificial intelligence (AI) that provides machines with the ability to automatically learn from data and past experiences while identifying patterns to make predictions with minimal human intervention.

#### • Data Visualization:-

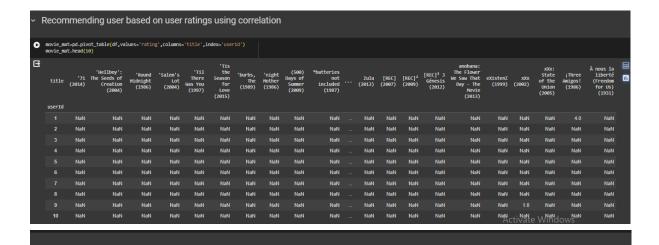
Data visualization is the representation of data through use of common graphics, such as charts, plots, infographics, and even animations. These visual displays of information communicate complex data relationships and data-driven insights in a way that is easy to understand.

#### • NLP:-

Natural language processing (NLP) is a machine learning technology that gives computers the ability to interpret, manipulate, and comprehend human language. Organizations today have large volumes of voice and text data from various communication channels like emails, text messages, social media news feeds, video, audio, and more. They use NLP software to automatically process this data, analyze the intent or sentiment in the message, and respond in real time to human communication.

#### **Result:**

```
import pandas as pd
    import numpy as np
    import seaborn as sns
    import matplotlib.pyplot as plt
    %matplotlib inline
[6] movie=pd.read_csv('movies.csv')
    movie.head()
                                                                                   屈
       movieId
                                     title
                                                                          genres
     0
                             Toy Story (1995) Adventure|Animation|Children|Comedy|Fantasy
                                                                                   11.
                             Jumanji (1995)
                                                          Adventure|Children|Fantasy
     2
                     Grumpier Old Men (1995)
                                                                  Comedy|Romance
     3
             4
                      Waiting to Exhale (1995)
                                                            Comedy|Drama|Romance
             5 Father of the Bride Part II (1995)
     4
                                                                          Comedy
[7] df=pd.read_csv('ratings.csv')
    df.head()
                                           翩
       userId movieId rating timestamp
     0
                           4.0 964982703
                                           ıl.
                           4.0 964981247
     2
                     6
                           4.0 964982224
     3
                    47
                           5.0 964983815
                           5.0 964982931
                    50
     4
rate_count=df.groupby('title')['rating'].count().sort_values(ascending=False)
     rate_count
📑 title
    Forrest Gump (1994)
                                                  329
    Shawshank Redemption, The (1994)
                                                  317
    Pulp Fiction (1994)
                                                  307
    Silence of the Lambs, The (1991)
                                                  279
    Matrix, The (1999)
                                                  278
    King Solomon's Mines (1950)
                                                   1
    King Solomon's Mines (1937)
                                                    1
    King Ralph (1991)
                                                   1
    King Kong Lives (1986)
                                                    1
    À nous la liberté (Freedom for Us) (1931)
    Name: rating, Length: 9719, dtype: int64
```



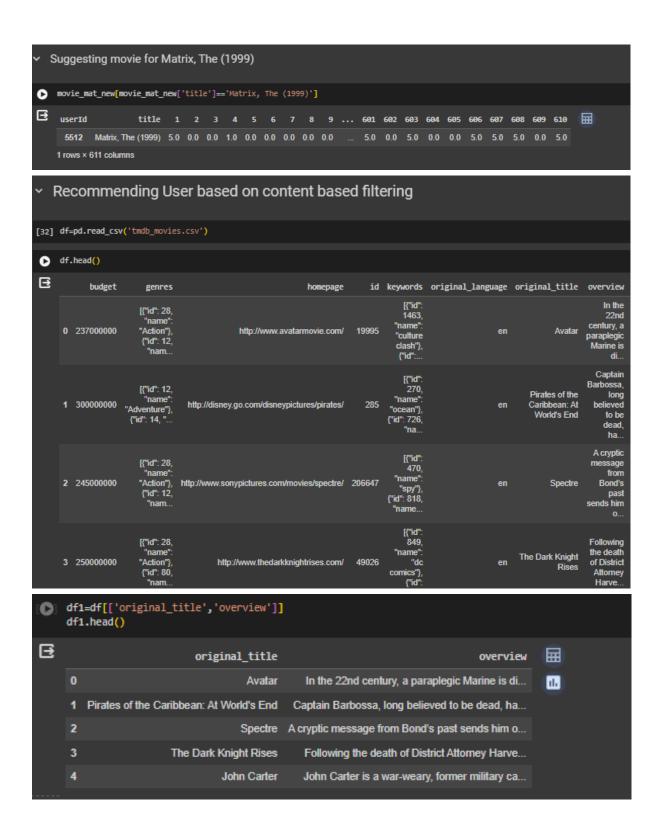
# Suggesting movie for Matrix, The (1999)

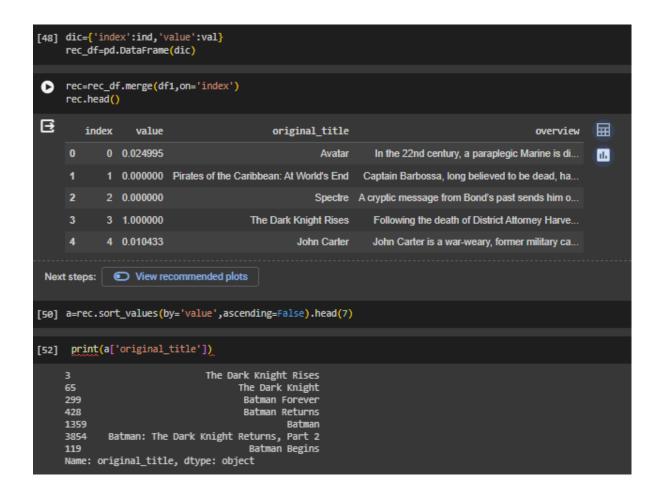
```
[D] x=corr['Matrix, The (1999)'].sort_values(ascending=False)
x
```

```
⋳
    title
    Haywire (2011)
                                                                1.0
    Back to the Beach (1987)
                                                                1.0
    Demolition (2016)
                                                                1.0
    Gerry (2002)
    Wolf Children (Okami kodomo no ame to yuki) (2012)
    Zoom (2015)
                                                                NaN
    Zulu (2013)
                                                                NaN
    [REC] 3 3 Génesis (2012)
                                                                NaN
    anohana: The Flower We Saw That Day - The Movie (2013)
                                                                NaN
    À nous la liberté (Freedom for Us) (1931)
                                                                NaN
    Name: Matrix, The (1999), Length: 9719, dtype: float64
```

# Recommending user using Collaborative Filtering







### **Conclusion:**

The Content-Based Social Media Analytics Model provides a powerful tool for businesses to analyze various content types like text, emoticons, images, audio, and video. Through advanced techniques such as NLP and machine learning, the model identifies topics, detects issues, tracks trends, and analyzes sentiment. It also delves into multimedia content, offering insights into consumer preferences and engagement patterns. Ultimately, this model enables businesses to make informed decisions, optimize marketing strategies, and enhance customer experiences in the ever-evolving digital world.