MOVIE RECOMMENDER SYSTEM USING MACHINE LEARNING

Submitted by –

SHUBHAM KISHOR KOTHAWADE (R19112022)

RUTUJA HARISCHANDRA DALVI (R19112007)

SUDESHNA DAS (R19112008)

WHAT ARE RECOMMENDER SYSTEMS?

The goal of a recommender system is to generate meaningful recommendations to a collection of users for items or products that might interest them. Suggestions for books on Amazon, or movies on Netflix, are real-world examples of the operation of industry-strength recommender systems. Here, we have implemented the same for movies where the user can search a movie and according to that he is recommended with 5 different movie names but with similar type.

TYPES OF RECOMMENDER SYSTEMS

1) **Content-based recommender system –**

In a content-based recommender system the algorithms used are such that it recommends users similar items that the user has liked in the past or is examining currently.

2) **Collaborative Recommender System –**

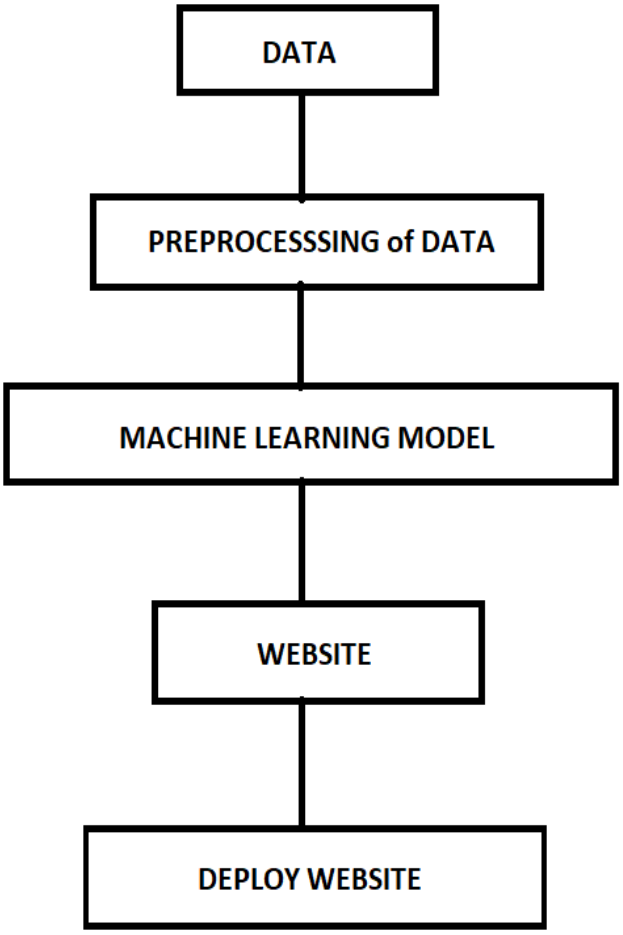
Collaborative filtering is based on the assumption that people who agreed in the past will agree in the future and that they will like similar kind of objects as they liked in the past.

3) **Hybrid Recommender System –**

Combining the two systems in a manner that suits a particular industry is known as Hybrid Recommender system. This is the most used Recommender system that many companies look after, as it combines the strengths of two or more Recommender system and also eliminates any weakness which exist when only one recommender system is used.

For our project, we have used Content-based recommender system.

PROJECT FLOW



- It consists of dataframes with the following columns like movie_id, title, genre, keywords, cast, crew and so on.

- Processing of data is done to create tags from overview, genres, keywords, cast, crew and so on.

- After processing the data, a data model is created using Machine Learning.

- Using the Machine Learning model, a website is created using PyCharm.

- Once the website is created and is functional enough, we can deploy it.

APPLICATIONS USED

1) **Jupyter Notebook –**

Jupyter Notebooks is used for all sorts of data science tasks including data cleaning and transformation, numerical simulation, exploratory data analysis, data visualization, statistical modelling, machine learning, deep learning, and much more. Since we are using Machine Learning, Jupyter notebook makes it easy to use.

2) **PyCharm –**

PyCharm provides smart code completion, code inspections, on-the-fly error highlighting and quick-fixes, along with automated code refactoring's and rich navigation capabilities. We have used for the frontend development.

PACKAGES USED

1) **CountVectorizer –**
   CountVectorizer is a great tool provided by the scikit-learn library in Python. It is used to transform a given text into a vector on the basis of the frequency (count) of each word that occurs in the entire text.

2) **cosine_similarity** – cosine_similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them. It is present in the sklearn.metrics.pairwise package.

3) **streamlit** - Streamlit is an open-source app framework in Python language. It helps us create web apps for data science and machine learning in a short time. It is compatible with major Python libraries such as scikit-learn, Keras, PyTorch, SymPy(latex), NumPy, pandas, Matplotlib etc.

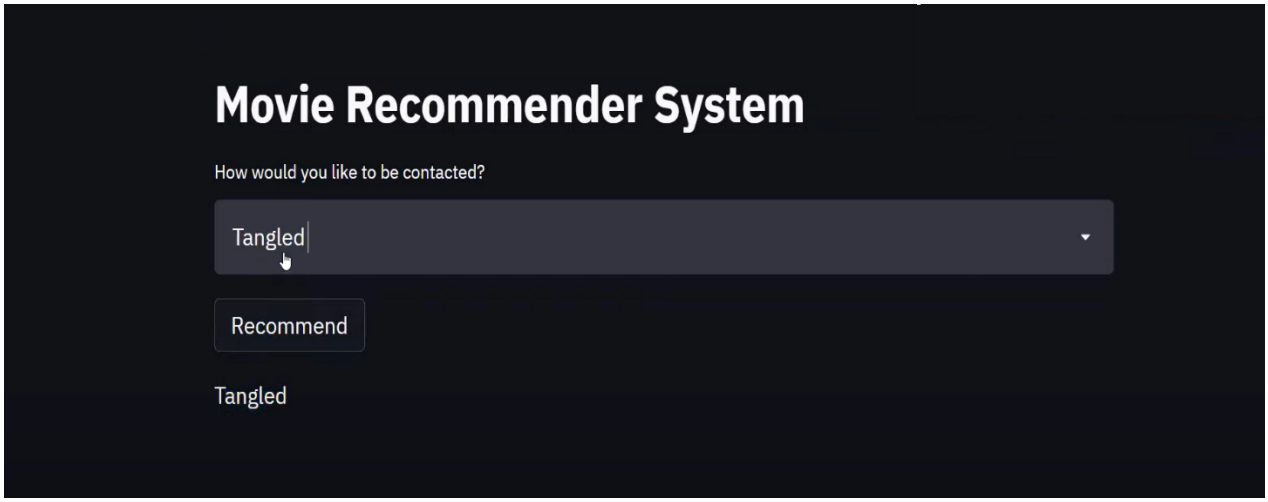
4) **ast** - The ast module helps Python applications to process trees of the Python abstract syntax grammar. The abstract syntax itself might change with each Python release; this module helps to find out programmatically what the current grammar looks like. An abstract syntax tree can be generated by passing ast.

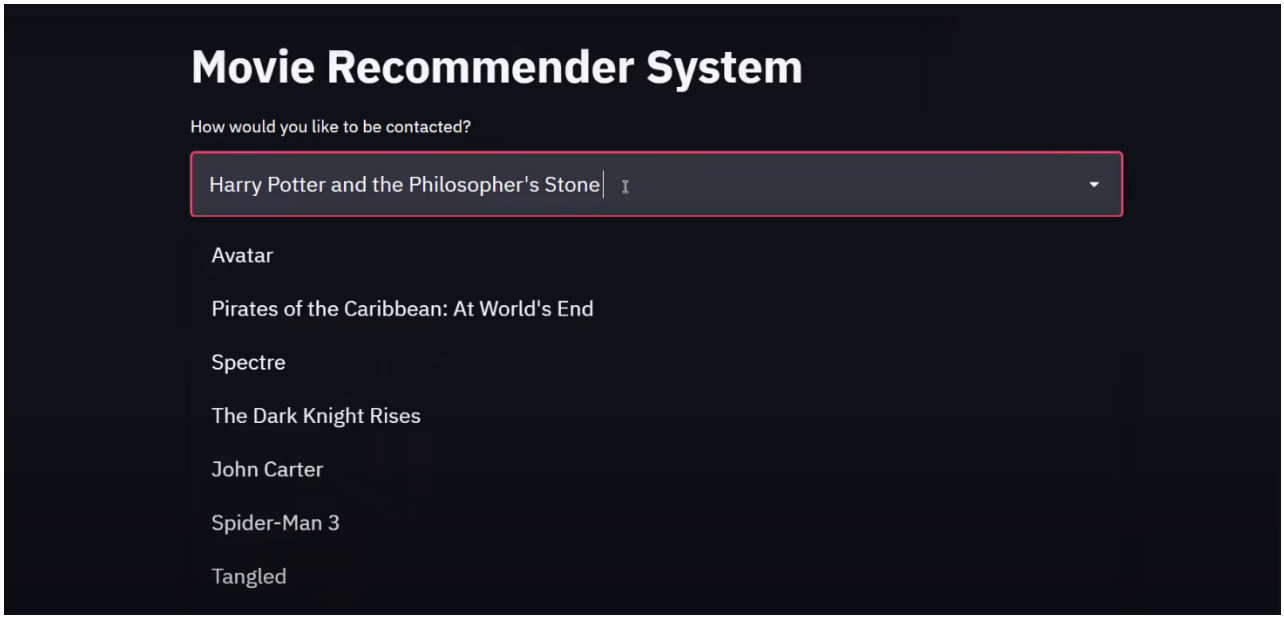
5) **pickle -** The pickle module can transform a complex object into a byte stream and it can transform the byte stream into an object with the same internal structure.

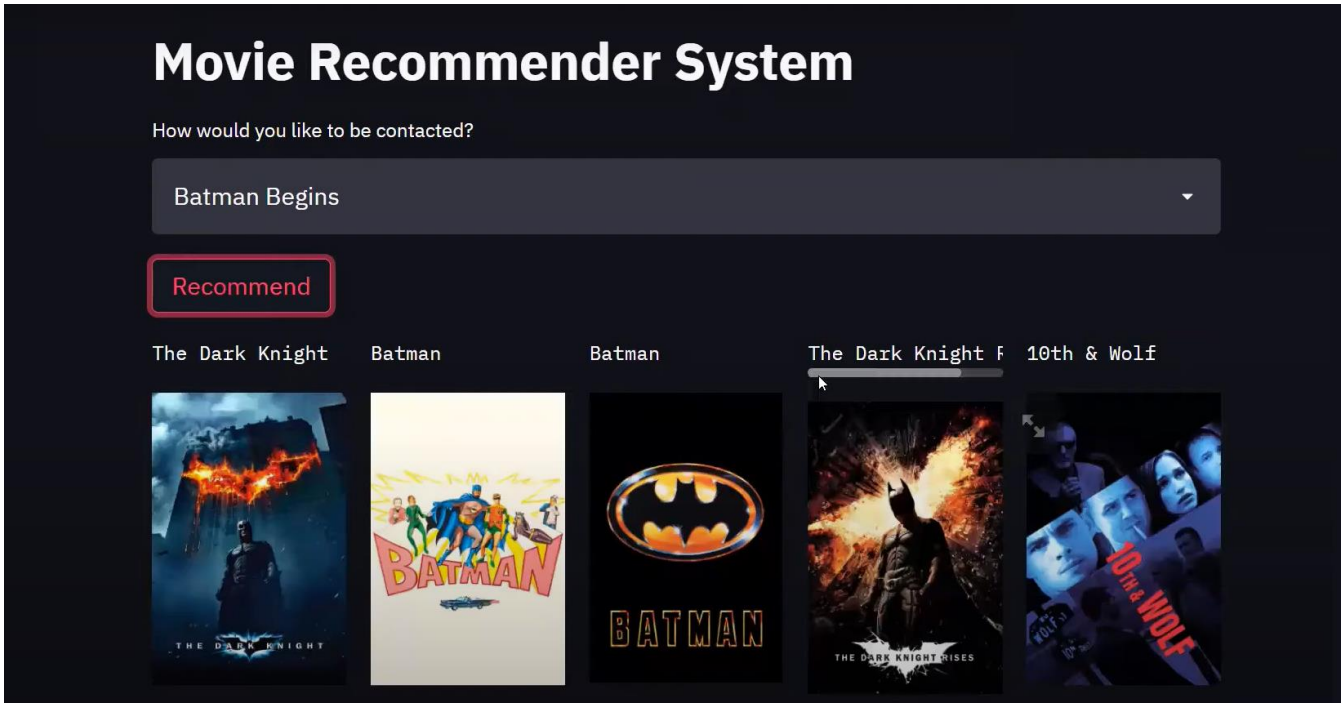
6) **NumPy** - NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices.

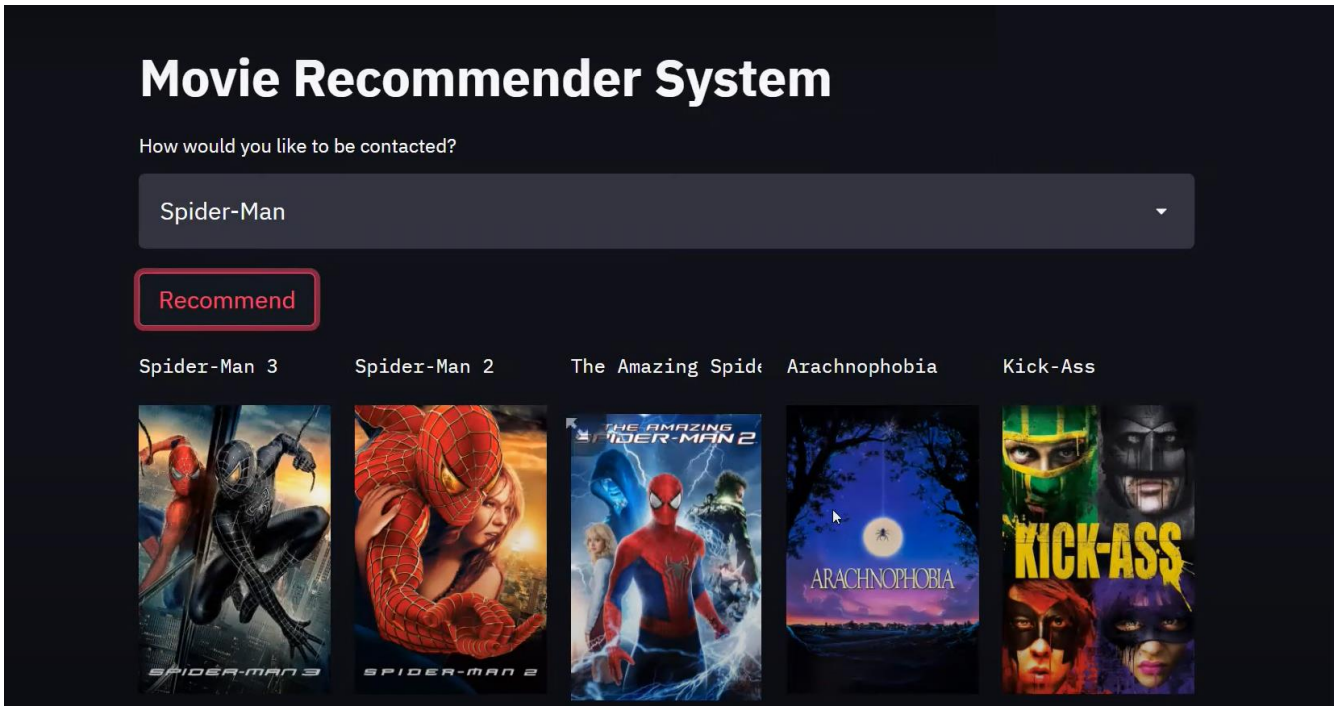
7) **pandas -** Pandas is an open-source Python package that is most widely used for data science/data analysis and machine learning tasks.

OUTPUT VIEW


Movie Recommender System
How would you like to be contacted?
Tangled
Recommend
Tangled


Movie Recommender System
How would you like to be contacted?
Harry Potter and the Philosopher's Stone
Avatar
Pirates of the Caribbean: At World's End
Spectre
The Dark Knight Rises
John Carter
Spider-Man 3
Tangled


Movie Recommender System
How would you like to be contacted?
Batman Begins
Recommend
The Dark Knight
Batman
Batman
The Dark Knight R
10th & Wolf
BATMAN
THE DARK KNIGHT RISES
10TH & WOLF


Movie Recommender System
How would you like to be contacted?
Spider-Man
Recommend
Spider-Man 3
Spider-Man 2
The Amazing Spide
Arachnophobia
Kick-Ass
SPIDER-MAN 3
SPIDER-MAN 2
THE AMAZING SPIDER-MAN 2
ARACHNOPHOBIA
KICK-ASS

THANK YOU