Movie Recommender System

Name : Aditya Tanaji Patil
Date : Sunday, 31 December

Abstract :

This movie recommender system recommends movies to the users based on similarities between movies keywords. System is build such that it will show top 5 similar movies. This recommender system will give personalized recommendations.

This project introduces a movie recommender system focused on delivering personalized recommendations to enhance user engagement. Leveraging a collaborative filtering algorithm, the system analyzes similarities between movies, generating accurate suggestions based on similar users' choices. Implemented in Python with a user-friendly interface, the system demonstrates improved accuracy and efficiency in content discovery. To make this project even more appealing and engaging this project is hosted on a local server which helps user to select movies from dropdown menu and get suggestion of similar movies. Project is hosted using pycharm streamlit library. The report discusses challenges faced during implementation and suggests avenues for future enhancements, contributing to the evolving landscape of personalized recommendation systems in the entertainment industry.

Introduction:

In the contemporary era of digital media consumption, the sheer volume of available content necessitates sophisticated systems to aid users in discovering relevant options. Movie recommender systems have emerged as pivotal tools to address this challenge, employing algorithms that analyze user preferences and behaviors to provide tailored suggestions/recommendations. This project aims to contribute to this field by developing a movie recommender system designed to enhance the user experience. The proliferation of streaming platforms and diverse cinematic offerings underscores the significance of effective recommendation systems, making the pursuit of personalized content discovery a key objective.

Table Of Content:

Problem Statement:

The goal of the project is to recommend a movie to the user.
Providing related content out of relevant and irrelevant collection of items to users of online service providers.

Objective of the Projects

- Improving the Accuracy of the recommendation system.
- Improve the Quality of the movie Recommendation system.
- Improving the Scalability.
- Enhancing the user experience.

Scope of the Project:

The objective of this project is to provide accurate movie recommendations to users. The goal of the project is to improve the quality of movie recommendation system, such as accuracy, quality and scalability of system than the pure approaches. This is done using Hybrid approach by combining content based filtering and collaborative filtering, To eradicate the overload of the data, recommendation system is used as information filtering tool in social networking sites .Hence, there is a huge scope of exploration in this field for improving scalability, accuracy and quality of movie recommendation systems Movie Recommendation system is very powerful and important system. But, due to the problems associated with pure collaborative approach, movie recommendation systems also suffers with poor recommendation quality and scalability issues.

Methodology:

Data collection:
Data is downloaded from kaggle.
Data is in .csv format
Dataset link is mentioned in Github Repository.

Content Based Algorithm:
content-based algorithms play a crucial role by focusing on the intrinsic features of items and users. In the context of our movie recommender system, the content-based approach centers on extracting relevant features from movies and aligning them with user preferences

CountVectorizer :
Here words are converted into vectors and plotted on 2d plot
Vectors are plotted on basis of frequency of occurrence of a word in the entire dataset
Code : from sklearn.feature_extraction.text import CountVectorizer

Cosine- Similarity:
Here distances and similarity between 2 vectors is found
Code : from sklearn.metrics.pairwise import cosine_similarity

Natural Language Processing:
NLTK library is used for text processing. Which includes removal of stopword, stemming
Code for stemming: from nltk.stem.porter import PorterStemmer

Training and Testing the model:

once the implementation of algorithm is completed . we have to train the model to get the result. We have tested it several times the model is recommend different set of movies to different users.

Python libraries:

For the computation and analysis we need certain python libraries which are used to perform analytics. Packages such as SKlearn, Numpy, pandas, Matplotlib, Flask framework, etc are needed.

SKlearn:

It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

NumPy:

NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python. Pandas: Pandas is one of the most widely used python libraries in data science. It provides high-performance, easy to use structures and data analysis tools. Unlike NumPy library which provides objects for multi-dimensional arrays, Pandas provides in-memory 2d table object called Data frame.

Pandas:

The Pandas library in Python is a powerful and popular tool for data manipulation and analysis. It provides data structures like DataFrame and Series, which are essential for handling structured data.

NLTK:
The Natural Language Toolkit (NLTK) is a powerful library for working with human language data in Python. It provides a wide range of tools and resources for tasks such as tokenization, stemming, lemmatization, part-of-speech tagging, named entity recognition, and more.

Streamlit:
Streamlit is an open-source Python library that simplifies the process of creating web applications for data science and machine learning. It allows developers to build interactive and data-centric applications with minimal effort. Here's an overview of Streamlit's key features and how it can be used:

SYSTEM ANALYSIS AND DESIGN:

System Architecture of Proposed System:
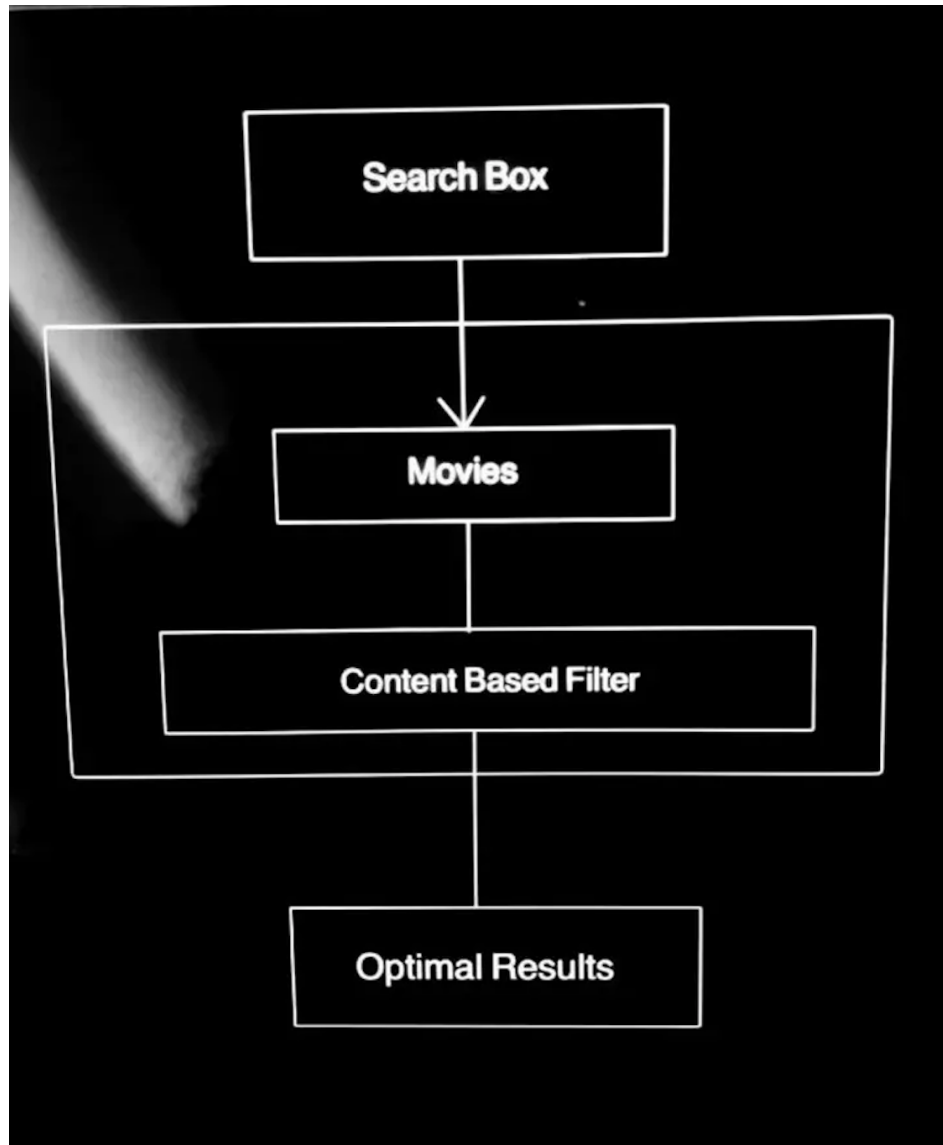
Implementation:

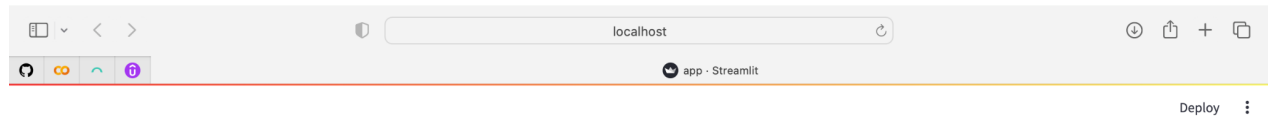The Proposed System Make Use Different Algorithms and Methods for the implementation of Hybrid Approach

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.

Formula:

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$
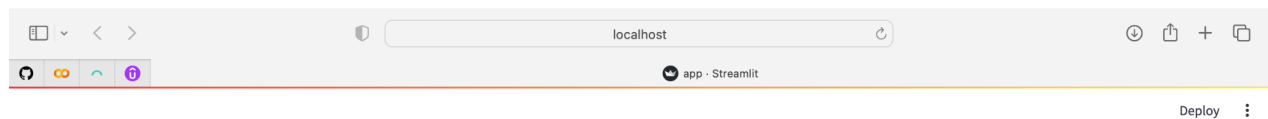
RESULTS:

Movie Recommender System

Movies list

Harry Potter and the Goblet of Fire

Recommend

Harry Potter and   Harry Potter and   Harry Potter and   Harry Potter and   Harry Potter and

REFERENCES:

[1] Hirdesh Shivhare, Anshul Gupta and Shalki Sharma (2015), "Recommender system using fuzzy c-means clustering and genetic algorithm based weighted similarity measure", IEEE International Conference on Computer, Communication and Control.

[2] Manoj Kumar, D.K. Yadav, Ankur Singh and Vijay Kr. Gupta (2015), "A Movie Recommender System: MOVREC", International Journal of Computer Applications (0975 – 8887) Volume 124 – No.3.

[3] RyuRi Kim, Ye Jeong Kwak, HyeonJeong Mo, Mucheol Kim, Seungmin Rho,Ka Lok Man, Woon Kian Chong (2015),"Trustworthy Movie Recommender System with Correct Assessment and Emotion Evaluation", Proceedings of the International MultiConference of Engineers and Computer Scientists Vol II.

[4] Zan Wang, Xue Yu*, Nan Feng, Zhenhua Wang (2014), "An Improved Collaborative Movie Recommendation System using Computational Intelligence",Journal of Visual Languages & Computing,Volume 25, Issue 6.

[5] Debadrita Roy, Arnab Kundu, (2013), "Design of Movie Recommendation System by Means of Collaborative Filtering", International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 4.