

MOVIE RECOMMENDATION SYSTEM USING MACHINE LEARNING

**A PROJECT REPORT SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF**

**BACHELOR'S IN TECHNOLOGY
IN COMPUTER SCIENCE AND INFORMATION TECHNOLOGY**

BY

MOHD MUTTALIB & MD YASIR RASHID

(18BTCS003HY & 18BTCS055HY)

UNDER THE GUIDANCE OF

DR. MUQEEM AHMAD

ASSISTANT PROFESSOR OF DEPT. OF CS &IT



DEPARTMENT OF CS & IT

MAULANA AZAD NATIONAL URDU UNIVERSITY

GACHIBOWLI, HYDERABAD-500032

2022



Certificate

We hereby declare that the project work presented in this report entitled “**MOVIE RECOMMENDATION SYSTEM USING MACHINE LEARNING**” towards the partial fulfillment of the requirement for the award of the degree of **Bachelor's of Technology in Computer Science & Information Technology (B.Tech CS&IT)** submitted in the **Department of CS&IT**, Maulana Azad National Urdu University, Hyderabad, Telangana, India, is an authentic record of our own work carried out under the guidance of **Dr. Muqem Ahmad (Assistant Prof. MANUU, Hyderabad), Department of CS&IT**, Maulana Azad National Urdu University, Hyderabad (Telangana).

We have not submitted the matter embodied in this project report for the award of any other degree or diploma to any other University or Institute.

.....

Guide

.....

External Examiner

.....

**Head
Department of CS & IT**

CANDIDATE’S DECLARATION

We hereby declare that the project work presented in this report entitled **“MOVIE RECOMMENDATION SYSTEM USING MACHINE LRARNING”** towards the partial fulfilment of the requirement for the award of the degree of **Bachleor’s of Technology in Computer Science & Information Technology (B.Tech CS&IT)** submitted in the **Department of CS&IT**, Maulana Azad National Urdu University, Hyderabad, Telangana, India, is an authentic record of our own work carried out under the guidance of **Dr. Muqeen Ahmad (Assistant Prof. MANUU, Hyderabad), Department of CS&IT**, Maulana Azad National Urdu University, Hyderabad (Telangana).

We have not submitted the matter embodied in this project report for the award of any other degree or diploma to any other University or Institute.

Date:

Place: MANUU Campus Hyderabad

(MOHD MUTTALIB & MD YASIR RASHID)

ACKNOWLEDGEMENT

First of all we would like to express our sincere gratitude towards **(Dr. Abdul Wahid, Prof. & Dean School of Computer Science & Information Technology, Dr. Syed Imtiyaz Hassan, Prof. & Head Department of CS&IT, Project co-ordinator Mrs. Geeta Pattun, Prof. Department of CS&IT)**, our project guide Internal Guide **Dr. Muqem Ahmad Assistant professor, Department of CS&IT, MANUU Hyderabad**, for consistently providing us with the required guidance to help us in the timely and successful completion of this report.

We are deeply indebted to our **Project co-ordinator Mrs. Geeta Pattun, Assistant Professor, Department of CS&IT, MANUU** for her valuable suggestions and support. In spite of her extremely busy schedules in the Department, she was always available to share with us, her deep insights, wide knowledge and extensive experience.

We sincerely thank our **Dean School of Computer Science & Information Technology, Head Department of CS&IT** for giving sufficient guidance for completing the project in time.

We express our whole hearted gratitude to our **VC sir Syed Ainul Hassan, MANUU**, for providing the excellent environment for carrying through our academic schedules and project with ease.

We would like to thank all our friends and especially our classmates for all the thoughtful and mind stimulating discussions we had, which prompted us to think beyond the obvious. We have enjoyed their company so much during our stay at MANUU.

MOHD MUTTALIB | MD YASIR RASHID

18BTCS003HY | 18BTCS055HY

ABSTRACT

In the world where things have taken a vast transition extremely rapidly in the phase of technology, our life has so much been affected in various productive ways. From morning till the day ends spending time over the network helps us in one way or another. Whether reading books or buying books or listening to books, shopping online, watching dramas & film, reading newspaper or blog or any piece of intellectual document. All these now are so easily available by just a click of mouse or a touch, providing us the information that we just need, want and potentially desire for which is impacting our way of sustenance. However, there is something that is playing a phenomenal role from the back of the screen; it's known as recommendation system. The system of recommendation or recommendation system or recommender system has the ability of searching the things on what our interests are or perhaps what we may opt for. And in this project, our proposition is of the movie recommendation system that assists those who are movie lovers by recommending them the movies they want to watch without spending much time searching for over the internet. This will lead to less for just a little work. For this problem, handling, we have worked on with the model of content-based filtering technique and collaborative-based technique. It is going to provide us an increasingly precise result, by the provision of data of the users. The movie then is shown stacked based on the choices and, the experiences of the others and the ratings of the other users in a very efficient and effective way. And these or recommendation system has changed the style of searching the things of our interest. The movie then is shown in stacked form horizontally after the search according to the genres, actors and ratings provided by the other users. And these information of the users then is saved and using various types of knowledge. And thus the user is being recommended the movie very easily after the browsing, from which they find a movie of their own choice.

TABLE OF CONTENTS

DESCRIPTION	PAGE NO.
COVER PAGE	i
CERTIFICATION	ii
DECLARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
CONTENTS	vi
1. Introduction	8-12
1.1 Introduction	
1.2 Objective	
1.3 Scope of the project	
1.4 Report layout	
2. Feasibility Study	13-15
2.1 Economical Feasibility	
2.2 Technical Feasibility	
2.3 Social Feasibility	
2.4 Feasible system Implementation	
3. Proposed Work	16-28
3.1 Literature Survey	
3.2 Existing System / Problem	
3.3 Proposed System / Solution	
3.4 Methodology for movie recommendation	
3.5 Tools and techniques	
4. System Implementation	29-37
4.1 Python and its features	
4.2 Coding	
4.3 Installation & Execution	
5. Result & Discussion	38-46
5.1 Data Description	
5.2 Category	
5.3 Document	
5.4 Result	
6. Conclusion & Future Work	47-49
6.1 Conclusion and Future Work	
6.1.1.1 Summary of the research	
6.1.1.2 Future Work	
References	50

ABBREVIATIONS AND ACRONYMS (TEMPLATE)

ML	MACHINE LEARNING
AI	ARTIFICIAL INTELLIGENCE
SKL	SCIKIT LEARN
CF	COLLBORATIVE FILTERING
ANN	ARTIFICIAL NEURAL NETWORK
CSV	COMMA SEPARATED VALUE
KNN	K NEAREST NEIGHBOUR
EC	ELECTRONIC COMMERCE
API	APPLICATION PROGRAM INTERFACE

CHAPTER - 1

INTRODUCTION



1. INTRODUCTION

1.1 INTRODUCTION TO PROJECT

In the modern period where internet plays a very important part of sustenance to all of us where the people who are making a use of technology are seldom having a complication. Whether it is booking a rest room for a night in a hotel or for the economic adventures. The amount of all those knowledge are so prevalent and a lot. In order to not encounter the same problem relentlessly, we have come up with an idea of recommendation system for the assistance of the users going on, because of how interesting it is and will be after the people are going to use this in day to day life. Many a systems are being discovered in the past and working at present like Amazon.com where people does the shopping of all kinds. And systems like this are scaling up increasing the financial support.[2]

Recommender Systems have added to the economy of the some of the e-commerce websites (like Amazon.com) and Netflix which have made these systems a salient parts of their websites. A glimpse of the profit of some websites is shown in table below:[2]

Netflix	2/3 rd of the movies watched are recommended
Google News	recommendations generate 38% more click-throughs
Amazon	35% sales from recommendations
Choicestream	28% of the people would buy more music if they found what they liked

Table 1.1 Profit of companies

1.2 OBJECTIVE

This project talks about the recommendation system that recommends countless things to users. This system will recommend movies to users with a provision of more precise results as compared to the existing systems. The existing system works on individual users' rating. This may be sometime useless for the users who have different taste from the recommendations shown by the system as every user may have different tastes. This system calculates the similarities between different users and then recommend movie to them as per the ratings given by the different users of similar tastes. This will provide a precise recommendation to the user. This is a web based as well as android system where there is a movie web service which provides services to user to rate movies, see recommendations put comments and see similar movies. [1]

So,

1. To study the system of recommendation of movies
2. Propose the method of implementation
3. Validation of the system
 - a. Improving the Accuracy of the recommendation system
 - b. Improve the Quality of the movie Recommendation system
 - c. Improving the Scalability
 - d. Enhancing the user experience

1.3 SCOPE OF THE PROJECT

A recommendation system has been implemented based on hybrid approach of collaborative filtering engine and context based engine. We have tried to combine the existing algorithms for recommendation to come up with a hybrid one. It improves the performance by overcoming the drawbacks of traditional recommendation systems. It describes the conventional Content, Collaborative Filtering and Context Filtering recommendation approaches along with their precision, recall and accuracy parameters. This paper has presented a number of utilized evaluation metrics, from which some were used to measure quality, while others to measure performance. Recommender systems make the selection process easier for the users.[3] Hybrid recommendation engine is a competent system to recommend Movies for e-users, whereas the other recommender algorithms are quite slow with inaccuracies. This recommender system will assuredly be a great web application, which can be clubbed with today's high demanding online purchasing web sites. Our approach can be extended to various domains to recommend books, music, etc.[4]

1.4 REPORT LAYOUT

This project is research based project and content is summarized view. So in this project, we have given five chapters according to content of my project.

Chapter 1: Here, in this chapter we have given the introduction about this project and as well as objective, scope also.

Chapter 2: Here in this chapter we have explained about feasibility study.

Chapter 3: Here in this chapter we have elaborated about problem analysis and related work. We have explained everything about existing problem and proposed solution and what methodology used is in this project and system requirement is also explained.

Chapter 4: Here in this chapter we have elaborated the execution, the visualization of our software, project, tools used, and eventually the input/output.

Chapter 5: Here in this chapter we have given the system implementation about our project. Every point we have explained about the technology used and in the language we have used completing the coding. And also we have explained the algorithm used in this project and how work that implementation etc. everything will explain.

Chapter 6: In this chapter we talk about the conclusion and the summarization and future work of this project. And also in the end after this chapter we have included the references, i.e. where I have captured and get my deep knowledge and content of the project from.

That's all about our project report layout.

CHAPTER - 2

FEASIBILITY STUDY

2.0 FEASIBILITY STUDY

The feasibility of the project is analysed in this phase and business proposal is put forward with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden. For feasibility analysis, some understanding of the major requirements for the system is essential.[4]

Three key considerations involved in the feasibility analysis are:

- ECONOMIC FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

2.1 Economical Feasibility

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system is well within the budget and this was achieved, because most of the technologies used are freely available. Only the customized products had to be purchased.[5]

2.2 Technical Feasibility

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client.

The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.[6]

2.3 Feasible System Implementation

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus, it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective.[4]

The implementation stage involves careful planning, investigation of the existing system and its constraints on the implementation, designing of methods to achieve changeover and evaluation of changeover methods.[4]

Implementation is the process of converting a new system design into operation. It is the phase that focuses on user training, site preparation and file conversion for installing a candidate system. The important factor that should be considered here is that the conversion should not disrupt the functioning of the organization.[4]

CHAPTER - 3

PROPOSED WORK

3. PROPOSED WORK

3.1 LITERATURE SURVEY

Over the long period of developing time, a large number of recommendation systems for a variety of domains have been developed and introduced and are in use. These recommendation systems use a variety of methods such as content based approach, collaborative approach, knowledge based approach, utility based approach, hybrid approach, etc. Most of the online recommendation systems for a variety of items use ratings from previous users to make recommendations to current users with similar interests. One such system was designed by Jung, Harris, Webster and Herlocker (2004) for improving search results. The system encourages users to enter longer and more informative search queries, and collects ratings from users as to whether search results meet their information need or not. These ratings are then used to make recommendations to later users with similar needs. [2].[1].

Recommendation systems create recommendations. The people who are using, they probably would acknowledge as per their needs and desires and concluding the results of what their experience would be.

A recommendation system or recommendation engine is a model used for information filtering where it tries to predict the preferences of a user and provide suggestions based on these preferences. These systems have become increasingly popular nowadays and are widely used today in areas such as movies, music, books, videos, clothing, restaurants, food, places and other utilities. These systems collect information about a user's preferences and behaviour, and then use this information to improve their suggestions in the future. Movies are a part and parcel of life. There are different types of movies like some for entertainment, some for educational purposes, some are animated movies for children, and some are horror movies or action films. Movies can be easily differentiated through their genres like comedy, thriller, animation, action etc. Other way to distinguish among movies can be either by releasing year, language, director etc. Watching movies online, there are a number of movies to search in our most liked movies. [3]

Movie Recommendation Systems helps us to search our preferred movies among all of these different types of movies and hence reduce the trouble of spending a lot of time searching our favourable movies. So, it requires that the movie recommendation system should be very reliable and should provide us with the recommendation of movies which are exactly same or most matched with our preferences. A large number of companies are making use of recommendation systems to increase user interaction and enrich a user's shopping experience. Recommendation systems have several benefits, the most important being customer satisfaction and revenue. 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.[2]

3.1.1 Types of Recommendation Systems:

There are three major vital types of recommendation engines or systems: collaborative filtering, content-based filtering – and a hybrid of the two.[2]

3.1.1.1 Content-based:

The content-based filtering works on the principle that if you like a particular item, you will also like this other item. To make recommendations, algorithms use a profile of the customer's preferences and a description of an item (genre, product type, color, word length) to work out the similarity of items using cosine and Euclidean distances. [2]

The limitation of content-based filtering is that the system is limited to recommending products or content similar to what the person is already buying or using. It can't go beyond this to recommend other types of products or content. For example, it couldn't recommend products beyond a utensil if the customer had only brought a utensil.[1]

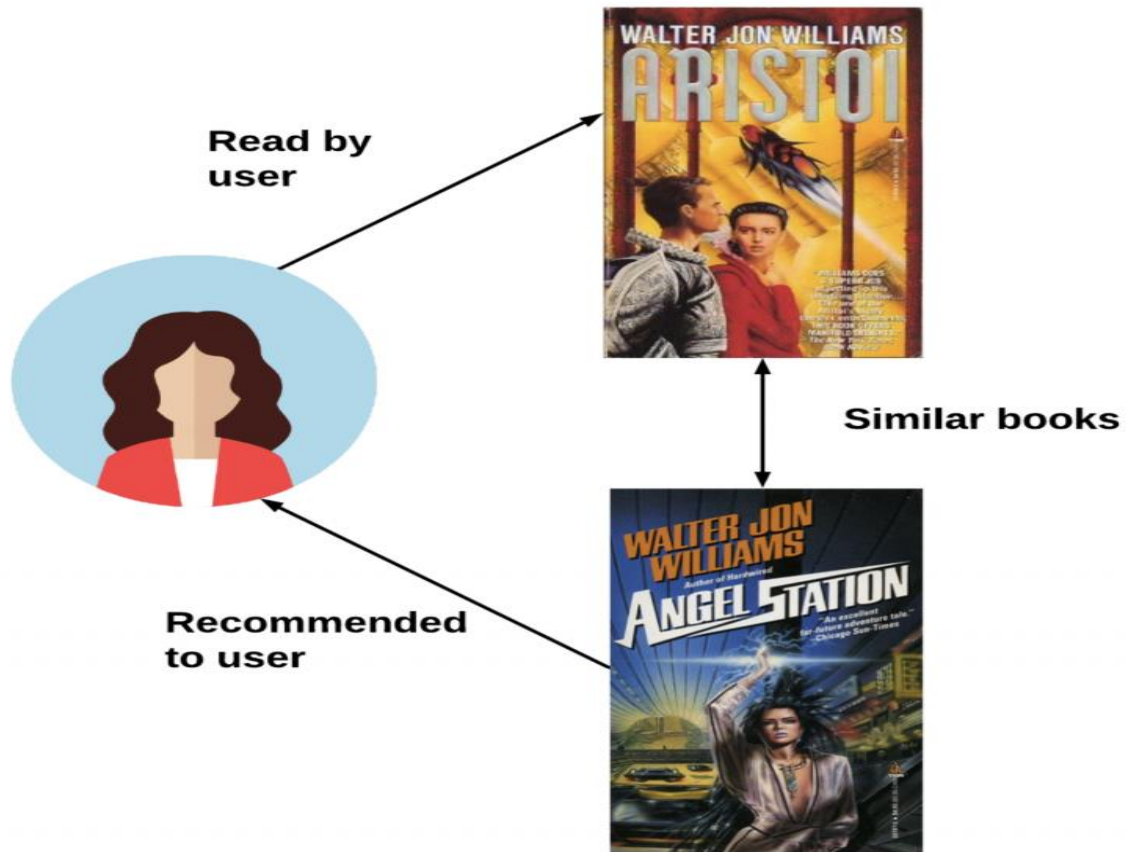


Image 1: Content based filtration

3.1.1.2 Collaborative-based:

The idea of collaborative filtering is simple: User group behaviour is used to make recommendations to other users. Since the recommendation is based on the preferences of other users it is called collaborative.[3]

There are two types of collaborative filtering: memory-based and model based.

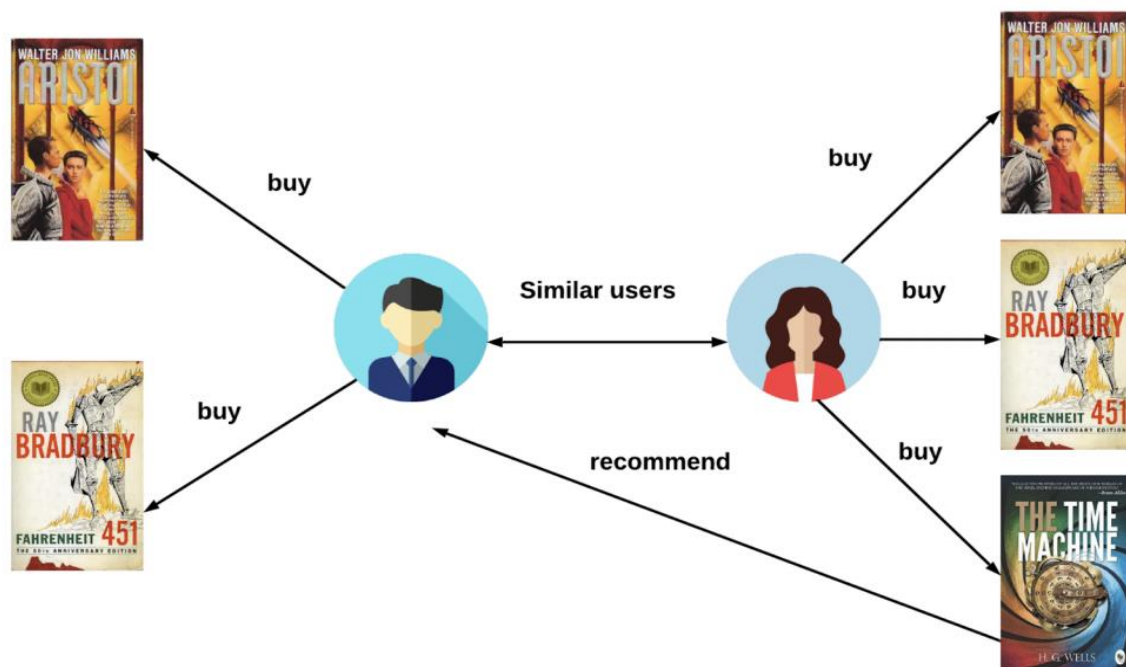
3.1.1.2.1 Memory based:

Memory based techniques are applied to raw data without pre-processing. They are easy for implementation and the resulting recommendations are generally easy to explain. Each time it is necessary to make predictions over all the data which slows down the recommender.[3]

There are two types: user based and item based collaborative filtering.

3.1.1.2.1

User based – “Users who are similar to you also liked...” Products are recommended to the user based on the fact that they were purchased / liked by users who are similar to the observed user. If we say that users are similar what does that mean? For an instance Harry and Sheyla love mystery books. When a new mystery book appears and Sheyla buys that book, since Harry also likes mystery books then we can recommend the book that Sheyla bought.[2]



3.1.1.2.2

Item based – “Users who liked this item also liked...” If John, Robert and Jenny highly rated sci-fi books Fahrenheit 451 and The time machine, for example gave 5 stars, then when Tom buys the book Fahrenheit 451 then the book The time machine is also recommended to him because the system identified books as similar based on user ratings.[3]

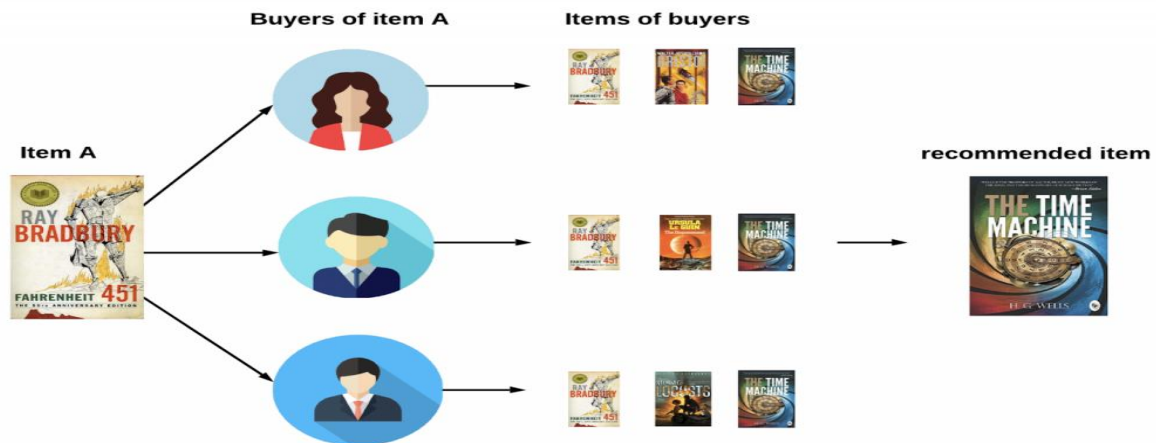


Image 2: Collaboration based filtration

3.1.1.2.3

Model based:

These models were developed using machine learning algorithms. A model is created and based on it, not all data, gives recommendations, which speeds up the work of the system. This approach achieves better scalability. Dimensionality reduction is often used in this approach. The most famous type of this approach is matrix factorization.[3]

3.1.1.3

Hybrid recommendation technique:

A hybrid recommendation engine looks at both the meta (collaborative) data and the transactional (content-based) data. Because of this, it outperforms both.

In a hybrid recommendation engine, natural language processing tags can be generated for each product or item (movie, song), and vector equations used to calculate the similarity of products. A collaborative filtering matrix can then be used to recommend items to users depending on their behaviours, activities, and preferences. Netflix is the perfect example of a hybrid recommendation engine. It takes into account both the interests of the user (collaborative) and the descriptions or features of the movie or show (content-based).[3]

3.1.1.4

Comparison:

Each approach has its advantage and disadvantage, and the effects are different as well for different dataset. The approach may not suitable for all kinds of problems because of the algorithm itself. For example, it is hard to apply automate feature extraction to media data by content-based filtering method. And the recommendation result only limits to items the user ever chose, which means the diversity is not so good. It is very hard to recommend for users who never choose anything. Collaborative filtering method overcomes the disadvantage of mentioned before somehow. But CF based on big amount of history data, so there are problems of sparsity and cold start. In terms of cold start, as collaborative filtering is based on the similarity between the items chosen by users, there are not only new user problem, but also new item problem, which means it is hard to be recommended if the new item has never been recommended before.[3]

Recommendation algorithms	Advantages	Disadvantages
Content based	Recommendation result is intuitive and easy to interpret; No need for users' access history data; No new item problem and no sparsity problem; Supported by the mature technology of classification learning.	Limited by the features extraction methods; New user problem; The training of classifier needs massive data; Poor scalability.
Collaboration filtering	No need for professional knowledge; Performance improving as the increasing of the user number; Automatic; Easy to find user's new interesting point; Complex unstructured item can be processed. eg. Music, Video, etc.	Sparsity problem; Poor scalability; New user and new item problem; The recommendation quality limited by the history data set.

Table 2. Comparison between content and collaboration

3.1.1.5

Famous Recommender Systems:

E-Commerce, the most famous e-commerce website, Amazon, is the active application and promoter of recommender system. The recommender system of Amazon reaches deeper into all kinds of products in the image is the recommendation list of Amazon. Apart from personalized recommendation list, another important application of recommender system is relevant recommendation list. When you buy something in Amazon, the relevant goods will be shown below. Amazon has two kinds of relevant recommendation, one is customers who bought this item also bought in the given image. Another is what other items do customers buy after viewing this item in the given image. The difference between the two recommendations is the calculation of the different user behaviours. The most important application of relevant recommendation is cross selling. When you are buying something, Amazon will tell you what other customers who bought this item also bought and let you decide whether buy it at the same time. If you do, the goods will be packed and provide a certain discount.[3]

A. Movie and Video website

Personalized recommender system is a very important application for movie and video website, which can help users to find what they really like among the vast of videos. Netflix is the most successful company in this field. Amazon and it are the two most renowned companies in the recommender system.[2]

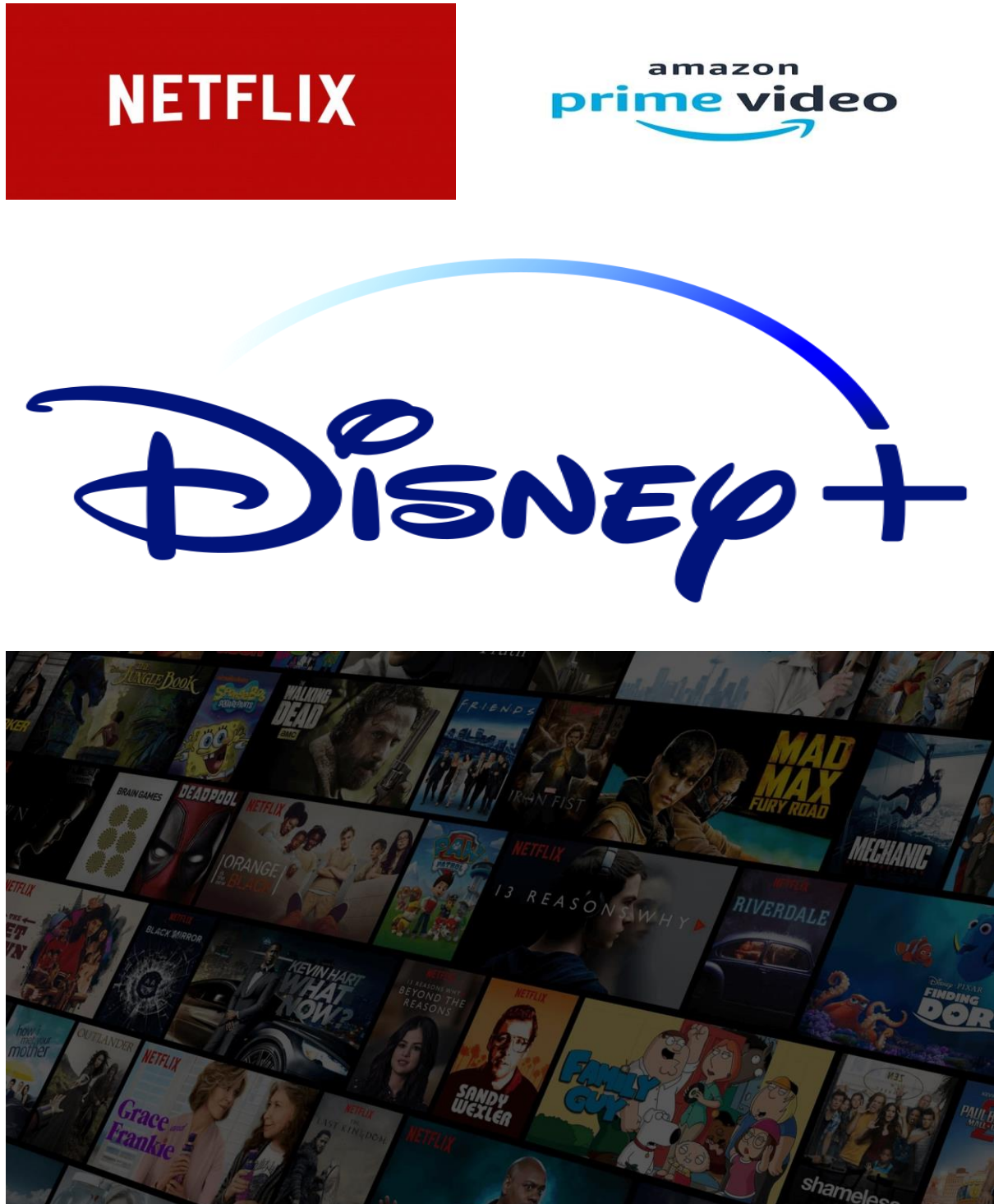
We can find that the recommendation result consists of the following parts.

- The title and poster of the movie.
- The feedback of the user, including Play, Rating and Not Interested.
- Recommendation reason.

It can be illustrated that the recommendation algorithm is similar with Amazon according to the recommendation reason of Netflix. Netflix declared that 60% of their users find movies that they are interested in by the recommender system. Similarly, YouTube with countless users and the videos uploaded by them is a serious complication, the YouTube then researchers created comparing the click rate of personalized suggest list and popular list.[2]

3.1.1.6

Recommendation platforms in real world:



Images from [12],[13],[14],[15]

3.2 EXISTING SYSTEM PROBLEM

3.2.1 Challenges Faced:

In developing any system the biggest challenge is to satisfy the end users for which the system is being developed. We also faced certain challenges while developing our system. Some of them are:[4]

- To have a system that is user friendly and easy to understand and use.
- To create a data set that has all relevant information about a particular movie.
- The biggest challenge was to have the most appropriate movie recommended list.
- To make our system diversifiable so that it can satisfy users of different geographical locations.
- To give weights to different attributes.

Current Social Networking World

Internet social networking sites, which began in 1995 with Classmates.com, have surged in popularity and use through word-of-mouth advertising. Since then, a wide range of virtual communities have formed serving different purposes and targeting varying niche audiences:[4]

Social Movie Platforms

In particular, we've chosen to explore the movie niche as this is an area where our project can provide significant improvements compared to existing products and systems. Traditional movie websites (IMDB, AOL Movies) function by providing global user ratings on movies in their database. Movies are categorized by metadata such as genre, era, directors, and so on. Users can search for movies, browse lists and read reviews written by critics or other users. However, most of these services lack any personal recommendation system and haven't taken advantage of social-networking communities or crowd wisdom. Some websites, such as Blockbuster, do provide individualized recommendations based on a user's ratings but do not include any social networking component. Yahoo! Movies goes further and

uses personal ratings to suggest movies currently playing in theatre, on TV, and out on DVD. It also draws upon its vast user base to give lists of similar movie fans, their ratings, and reviews. Other movie sites, like Flixster, take a different approach. Flixster forms web-based communities around movies and suggests movies to watch based on what your friends have rated.[1],[4].

3.3 PROPOSED SYSTEM SOLUTION

The system is built on windows 2007 operating system. The system uses advanced java technology along with machine learning concepts. MySQL is used for storing data. This system uses three-tier architecture. The web service layer provides the android user to rate movies, view similar recommendations given by the system and comment on it. The proposed system is a better system than any other existing systems. This system has added the positive features of existing systems and has overcome the drawbacks of existing systems. The system uses all the existing algorithms i.e. content based, context based and collaborative based algorithms. All these algorithms are combined to give more precise result. The following modules are developed as:[4]

- Admin :The system admin will add movie in a database, view movies and update it
- Recommendation Engine: This recommendation engine will calculate the similarities between the different users. On the basis of that similarities calculated, this engine will recommend movie to a user .
- Movie Web Service: This will allow user to rate movies, comments on movies. This service will also show the movie recommendation to the users.
- Android User: The android user can rate a movie, can comment on any movie, and can see similar movies recommended by other users who are similar to this user.

3.4 METHODOLOGY FOR MOVIE RECOMMENDATION

The hybrid approach proposed is an integrative method by merging fuzzy means clustering method and genetic algorithm based weighted similarity measure to construct a movie recommendation system. The proposed movie recommendation system gives finer similarity metrics and quality than the existing Movie recommendation system but the computation time which is Movie Recommendation System taken by the former proposed recommendation system is more than the existing recommendation system. This problem can be fixed by taking the clustered data points as an input dataset. The proposed approach is for improving the scalability and quality of the movie recommendation system .We use a Hybrid approach, by unifying Content-Based Filtering and Collaborative Filtering, so that the approaches can be profited from each other. For computing similarity between the different movies in the given dataset efficiently and in least time and to reduce computation time of the movie recommender engine we used cosine similarity measure. [6]

3.4.1 Agile Methodology:

1. Collecting the data sets: Collecting all the required data set from Kaggle website. In this project we require movie.csv, ratings.csv, users.csv.

2. Data Analysis: To make sure that the collected data sets are correct and analysing the data in the csv files i.e. checking whether all the column Fields are present in the data sets.

3. Algorithms: Inside our project we have only two algorithms one is cosine similarity and other is single valued decomposition that are used to build the machine learning recommendation model.

4. 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 to recommend different set of movies to different users.

5. Improvements: In the later stage we can implement different algorithms and methods for better recommendation.

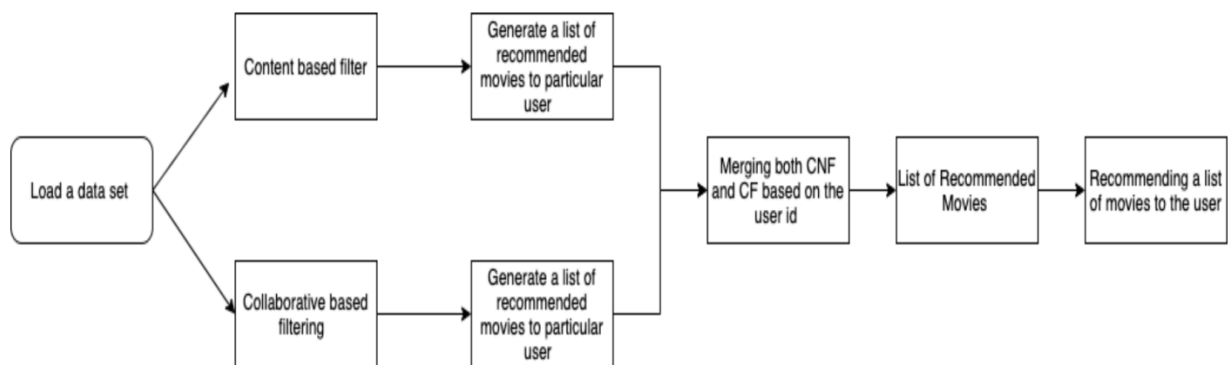


Image 4: Data flow diagram

First of all we take data to check in what form it is. If it is in content based or collaboration form, then we proceed the next process. According to the types of data we filter the data which in return generates the movie lists. And then when we apply the content based algorithm we get the result of movie recommendation. However, in the collaborative, it depends on the user's rating.

3.5 TOOLS AND TECHNIQUES

Operating System:	Windows10/11
Coding Language:	Python
IDE:	Anaconda/Jupyter
Database:	Dataset (Kaggle)

CHAPTER - 4

SYSTEM

IMPLEMENTATION

4. SYSTEM IMPLEMENTATION

4.0 SYSTEM IMPLEMENTATION

4.1 PYTHON AND ITS FEATURES

4.1.1 Python

Python is a general purpose high level programming language. It was developed by a scientist named Guido Van Rossam in 1989 while working at National Research Institute at Netherlands. [7]

Guido developed Python language by taking almost all feature from different language.

1. Functional Programming Feature from C
2. Object Oriented Programming feature from C++
3. Scripting Language Feature from Perl and Shell Script
4. Modular Programming features from Modula-3
5. Most of syntax in python Derived from C and ABC language

Where we can use Python:

We can use everywhere. The most common important areas are:

1. For developed Desktop Application
2. For developed web Application
3. For developed database Application
4. For developed games
5. For Data Analysis Application
6. For Machine Learning
7. For developing Artificial Intelligence Applications

4.1.1.1 Features of Python:

1. Simple and easy to learn
2. Freeware and Open Source
3. High Level Programming Language
4. Platform Independent
5. Portability
6. Dynamically Typed
7. Both Procedure Oriented and Object Oriented
8. Interpreted
9. Extensible
10. Embedded
11. Extensive Library

For Machine Learning, Python Language is the best programming language and from using of python language this machine learning project developed easily. Because python language has rich library and many functions are predefined. So For Developing of this project we have used Python Programming Language. [7]

We have used many libraries in python like pandas, numpy, matplotlib, seaborn, time, warning and sklearn etc.

4.1.1.2 Pandas

Pandas is a python library which is open source that is made mainly for working with labelled data both intuitively and easily. It provides various data structures and operations for manipulating numerical data and time series. It is fast and it has high performance and productivity for users.[7]

4.1.1.3 Numpy

Numpy is a library in python which is used for working with arrays. It is open source project and we can use freely. It also has working in domain of algebra and matrices. [7]

4.1.1.4 Matplotlib

Matplotlib is a python library which is comprehensive library for creating static, animated and interactive visualization in python. [7]

4.1.1.5 Seaborn

Seaborn is a python library which is open source built on top of matplotlib. It is used for data visualization and exploratory data analysis. Seaborn works with data frame and pandas library. [7]

4.1.1.6 Warning

Warning is a python library. It is used when warning messages are typically issued in situations where it is useful to alert the user of some condition in a program, where that condition normally doesn't warrant raising an exception and terminating the program. [7]

4.1.1.7 Sklearn || SciKit Learn

SKlearn is an open source python library is the most useful and robust for machine learning in python. It provides a selection of efficient tools for machine learning and statically modelling including classification, regression, clustering and dimensionality reduction via a consistence interface in python. [7]

4.2 CODING

We did all code in python language using dataset from Kaggle website. We used machine learning technology that is the reason we used python language. Python has rich library. It has many inbuilt functions modules and libraries which was very helpful during the coding of the project. [7]

IMPLEMENTATION OF CODING

- Step1: Read CSV file of dataset.
- Step2: Select features of datasets
- Step3: Create a column in DF which contains all selected features
- Step 4: Create count matrix from this new combined column
- Step5: Compute the cosine similarity based on count matrix
- Step6: Get index of this movie from title
- Step7: We will get the list of similar movies in descending order of similarity score
- Step8: Lastly, Print title of first 15 movies

4.3 INSTALLATION AND EXECUTION GUIDELINES

We installed Anaconda editor, because in this editor all the libraries installed automatically also Jupiter lab and Jupiter notebook all works automatically. All coding has been done in Jupiter Lab and in Jupiter lab execution is a lot more easier.

Following are all the guidelines of installation and execution:

Step1: Download Anaconda editor

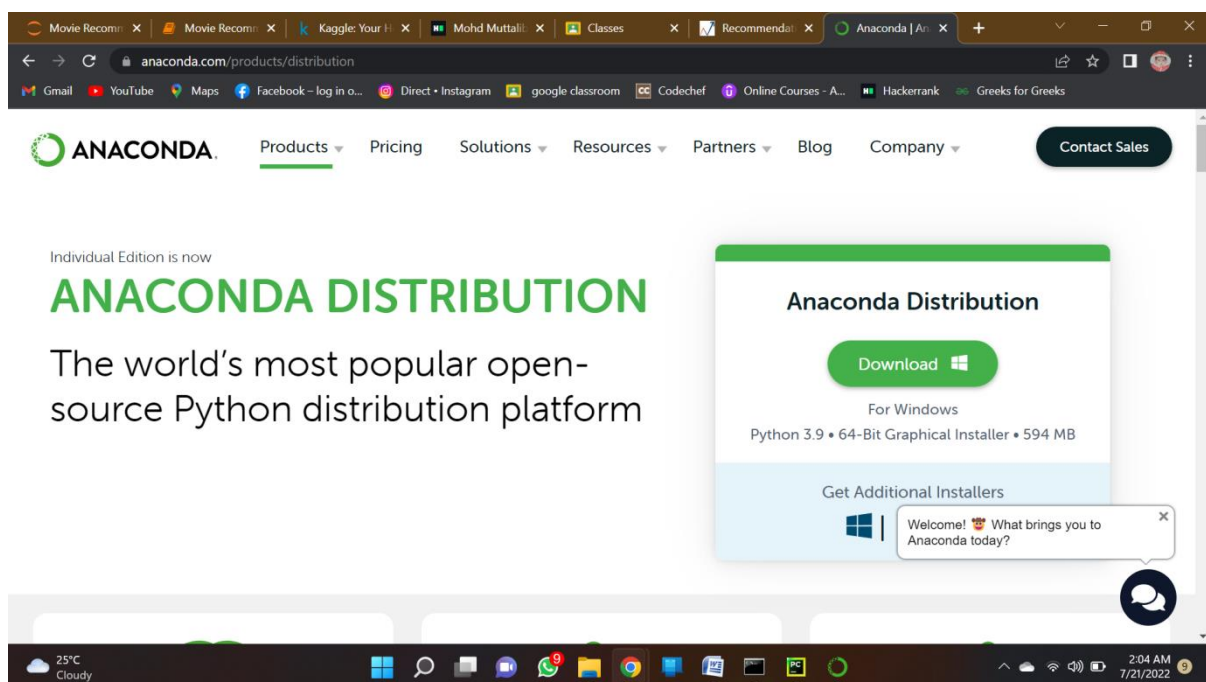


Image 4.1: Download editor

Step 2: Click and open Anaconda

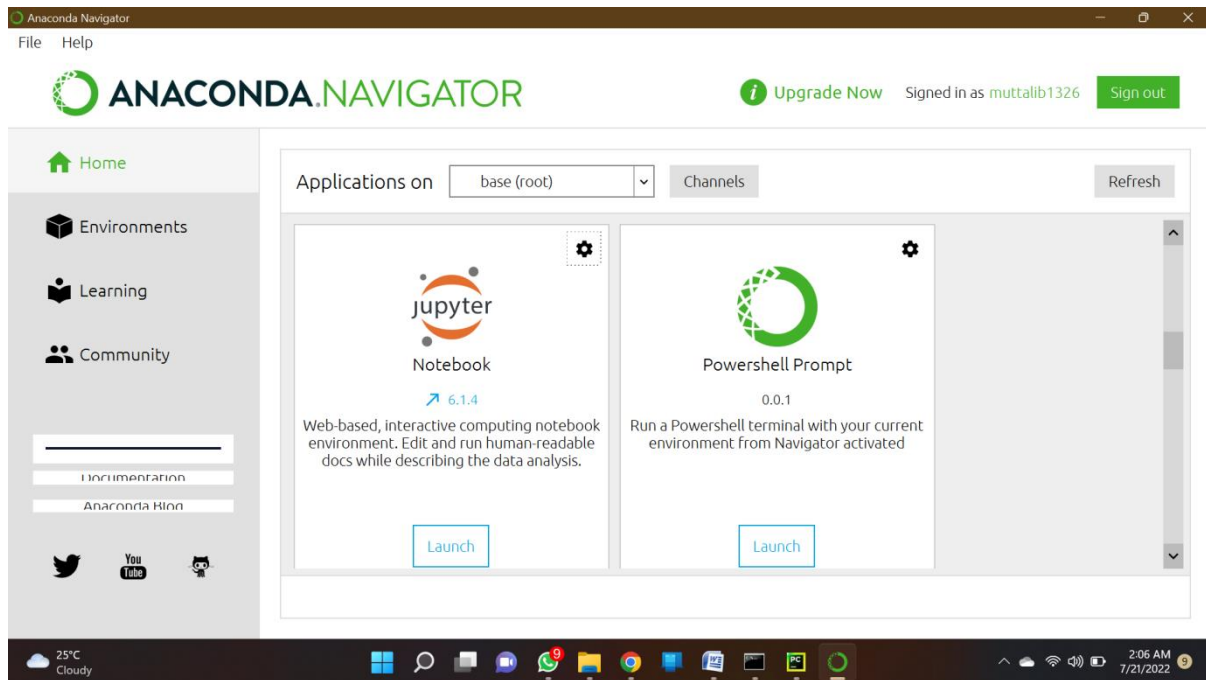


Image 4.2: Anaconda Editor

When Anaconda is downloaded and installed, then open Anaconda editor. Click Jupiter Lab. When clicking the Jupiter lab, Jupiter lab would open and then coding process begins. All coding would be in cells; there are variety of cells in Jupiter lab. We can run all coding cells wise and get the output in console screen.

Coding can be done in another editor also but we used anaconda editor.

Step 3: Click/Open Jupiter Lab

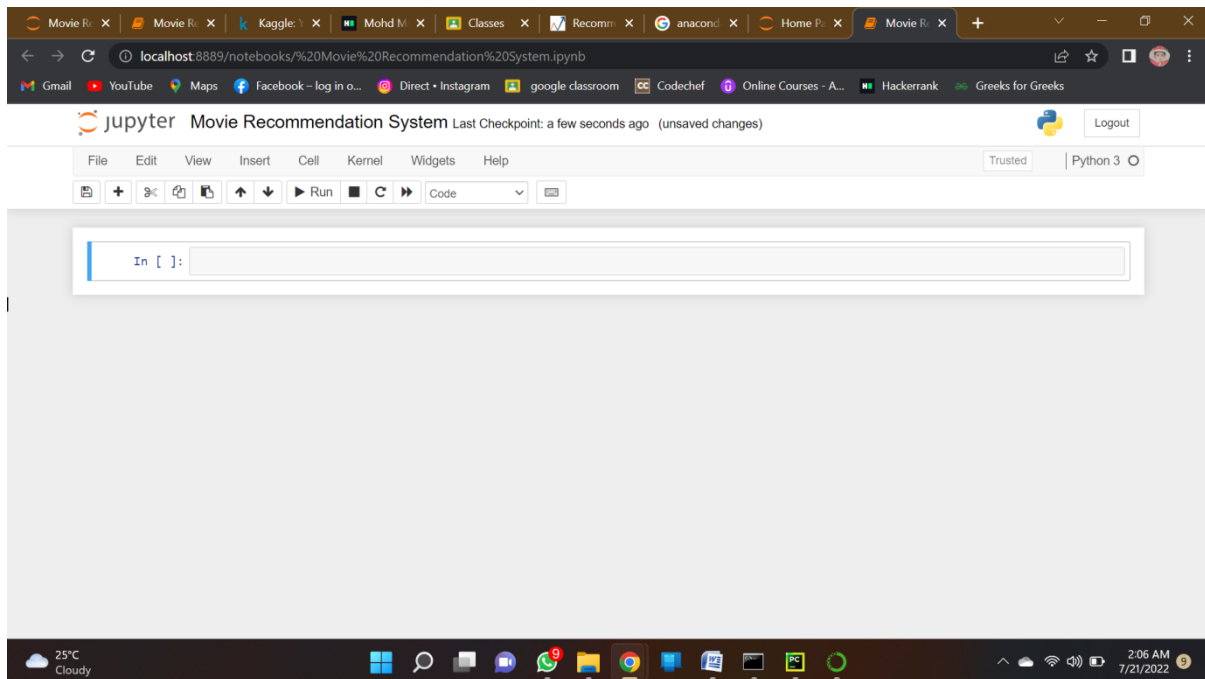


Image 4.3: Jupiter Lab

Step 4: Start coding

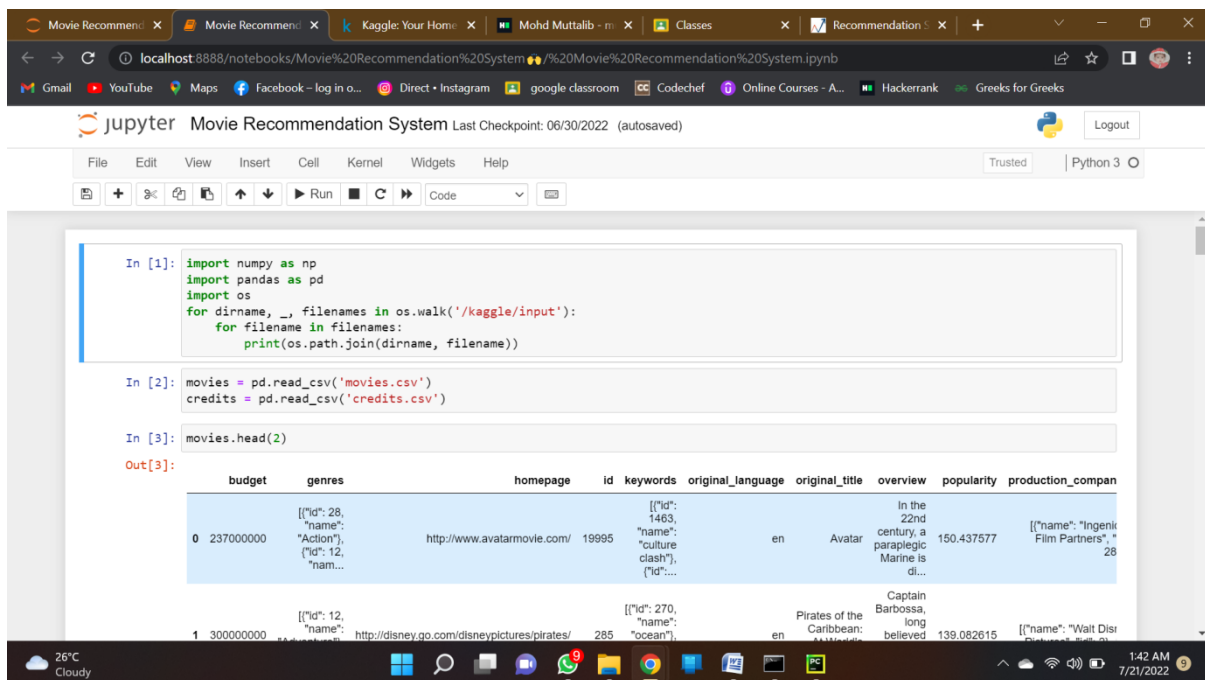


Image 4.4: Jupiter Lab coding screen

Step 5: Saved files

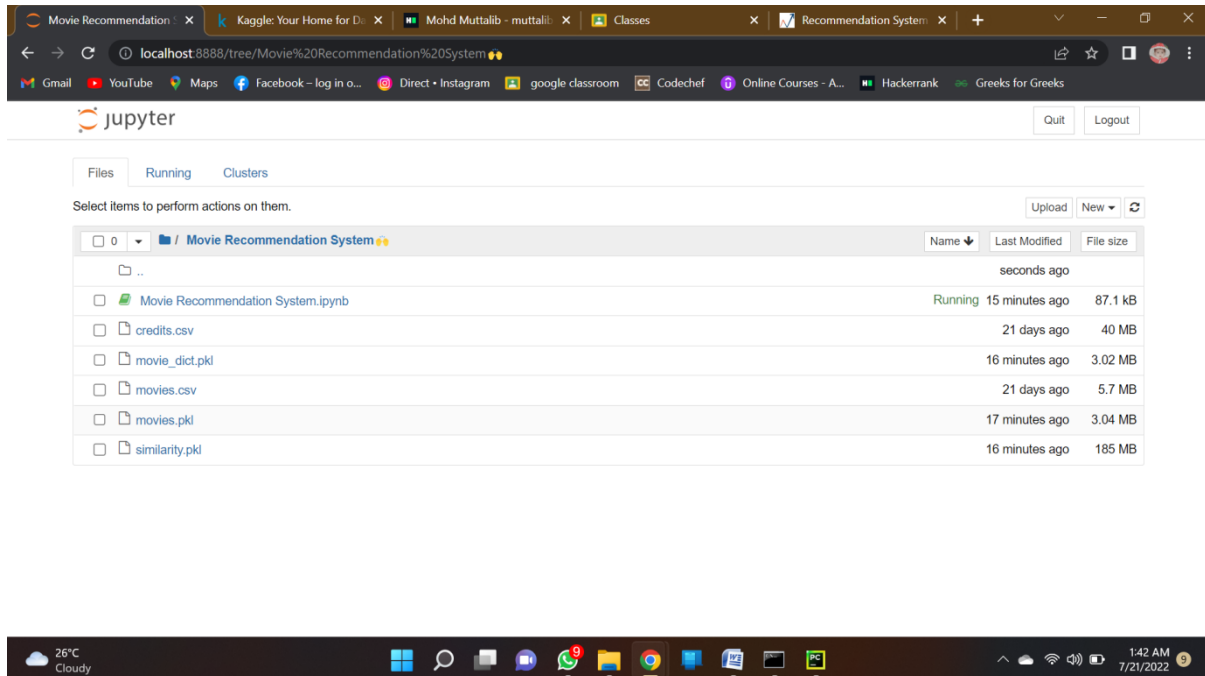


Image 4.4: Jupyter Lab saved files screen

Step8: Get the output on console screen

In this step finally we get the expected result on the screen with a relatively high amount of accuracy that we wanted based on the tools and techniques implementation.

CHAPTER 5:

RESULT AND DISCUSSION



5. RESULT AND DISCUSSION

5.1 DATA DESCRIPTION:

In this chapter, we have introduced how to implement the content-based recommender system based on the principle or objective mentioned. After that, we have tested the system and saw the result to prove the improvement of our system. [4]

Dataset: All the movie data we used is from, our own database. In the end we get numerous movies and related information. For the perspective of recommender system, a movie can be described by a collection of features, which can be genres, actors, directors and so on.

- Director: The director, most of movies only have one director, but some of them have two or more.
- Actor: A movie normally has a lot of actors, but most of them is useless for recommender system and bring disadvantageous effects. So we only get three main actors for one movie.
- Keyword: We use the names of the movies.
- Release Year: This is when the movie is released.
- Vionel Theme: Theme is a kind of keyword that describes movies in a different perspective, such as Time Travel and Comic Book.
- Language: Language is the language that occurs in the movie.
- Location: Location is which is where the movie happens.
- Vionel Scene: Scene of the movie is analysed as well. We will recognize the background of every frame in the movie by machine learning. For example, bar, hall room, store are what we recognized. [4]

5.2 CATEGORY

In the scenario of movie, we will divide the movies into numerous categories by the normal genres. Table below shows the categories we used. Each movie in the case is a document, which is represented by the eight features described. As we said before, the movie is represented by a specific model; each feature for the movie is a term in the document. [5],[4].

Sci-Fi	Crime	Romance	Animation	Music
Comedy	War	Horror	Adventure	News
Biography	Thriller	Western	Mystery	Short
Drama	Action	Documentary	Musical	History
Family	Fantasy	Sport		

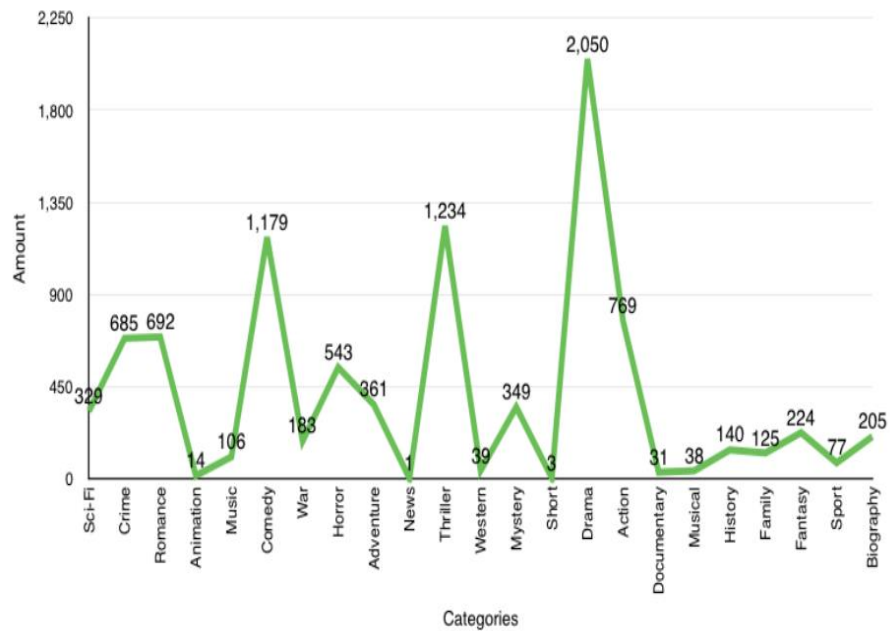
Table 3: Categories of movies

In many other content-based recommender systems, the genre is used as vector to calculate similarity. But this is only one aspect of the movie and there are a lot of other features of movie such as background, actor, etc. So we add more features and some of them are very unique because they are extracted by our own research. But we didn't simply add features together to for calculation; there are plenty of reasons. The genre is the natural feature that we can use as category. There are several distribution of genres in the movie database as well. Compared with the principle mentioned before, each genre is a category for documents. The number of document in each category is shown. Each document contains many terms which are features in our case, that is how they are described.[4]

5.3 DOCUMENT

As we discussed before, the document in our case is the movie which contains many features. The movie will be represented by a model in the experiment. Previously, we introduced the features that are used to model the movie. The model is like this kind of format: [4]

MovieModel = [Directors, Actors, Keywords, Release Year, VionelThemes, Languages, Locations, VionelScenes]



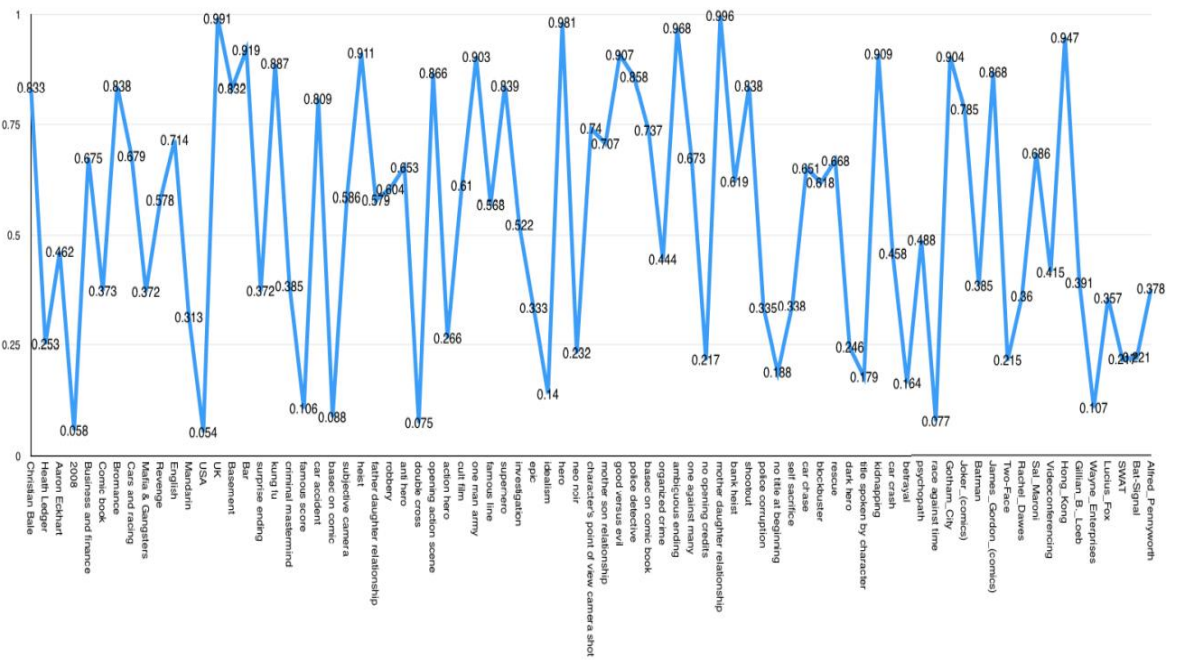
Graph 1: Categories of movies in graph

A movie can have multiple directors and actors, so the vector is pretty long generally. Here we use movie The Dark Knight to illustrate the model.

Directors	Christopher Nolan
Actors	Christian Bale Heath Ledger Aaron Eckhart
Keywords	Keywords
Release Year	2008
Vionel Themes	Business and finance Comic book Bromance Cars and racing Mafia & Gangsters Revenge
Language	English, Mandarin
Location	USA, UK
Vionel Scene	Basement, Bar

Table 4: Movies Documents

Above is the basic information of movie The Dark Knight. We have not listed keywords in the table because there are plenty of keywords for the movie, which is difficult to show them in the table. We can get a very long vector after the calculation according to Equation above, which is the model for the movie. Each weight represents the importance of a feature for the movie.



Graph 2: Features of Movie; The Dark Knight

There are 80 features totally for movie The Dark Knight, where the number is calculated, which shows the importance of each feature. From this perspective, we can see that if features of a movie have similar distribution, it means that the two movies are similar. Actually, one feature is one dimension for the model; we list all features in one dimension in the figure just because multiple-dimension is hard to show by figure.

5.4 RESULT

In our case, feature to movie is term to document. We can easily convert movie to the model which can be used to calculate the similarity. After previous calculation, every movie in the database can be represented by a vector. Then we use cosine similarity discussed previously, to calculate similarity for each movie. This figure shows the final recommendation of The Dark Knight, which is a screenshot of our current demo system.

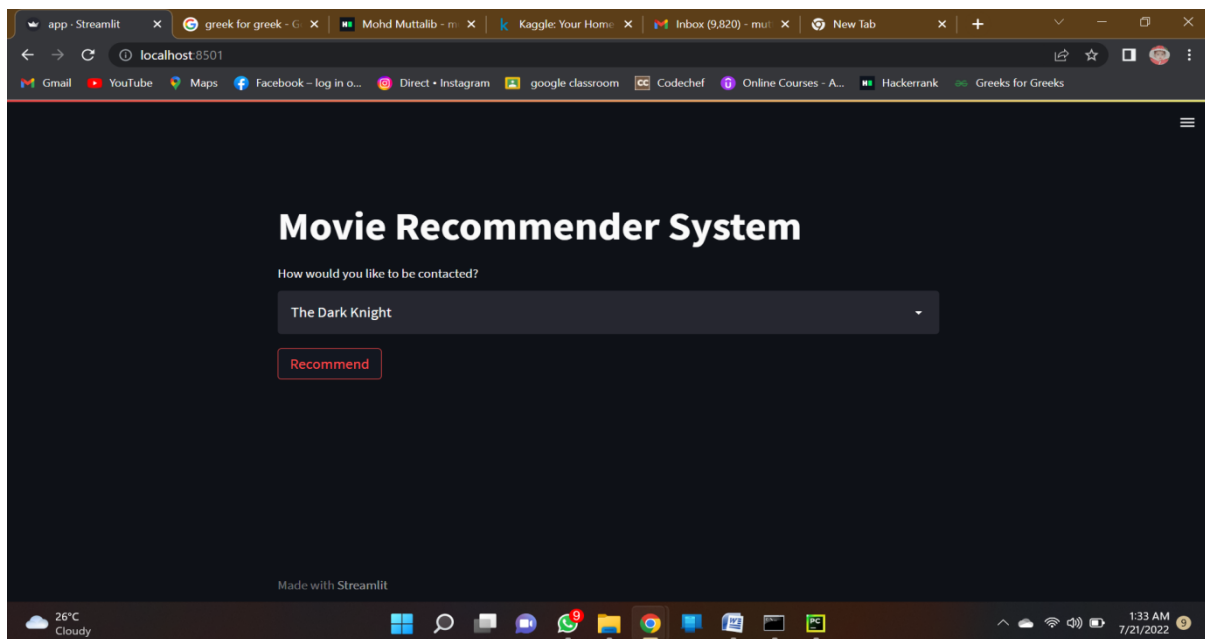


Image 5.1: The main screen of output

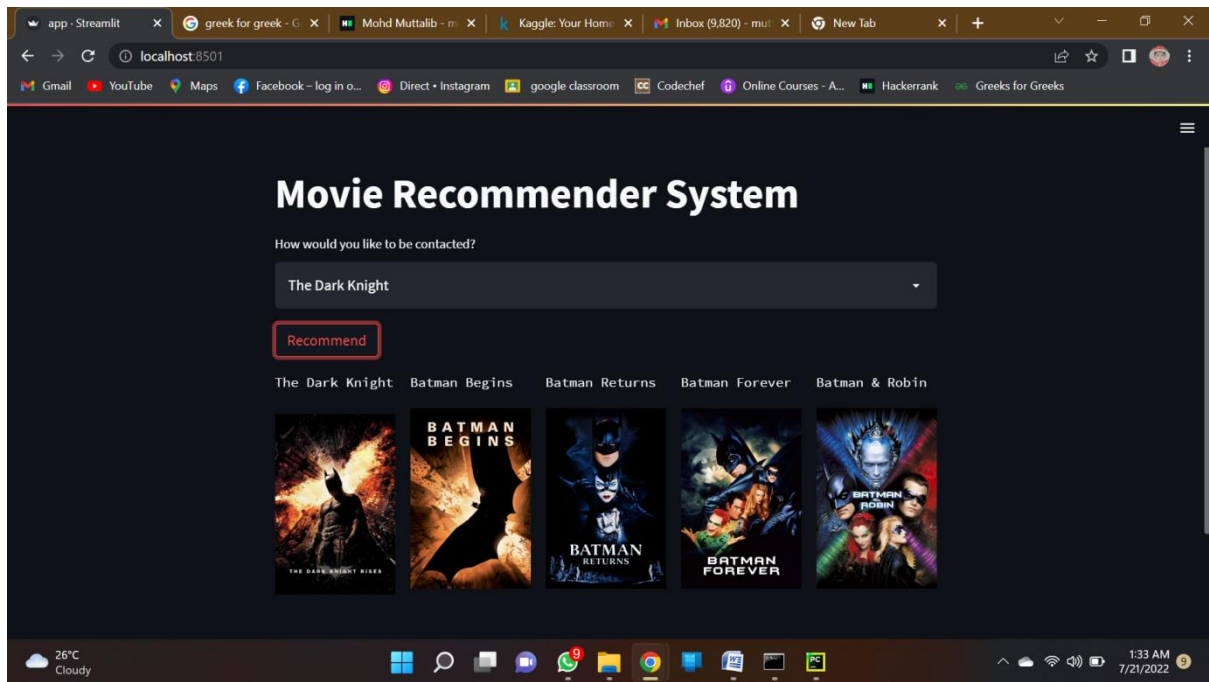


Image 5.2: The recommendation output

MORE EXAMPLES FROM EXECUTION

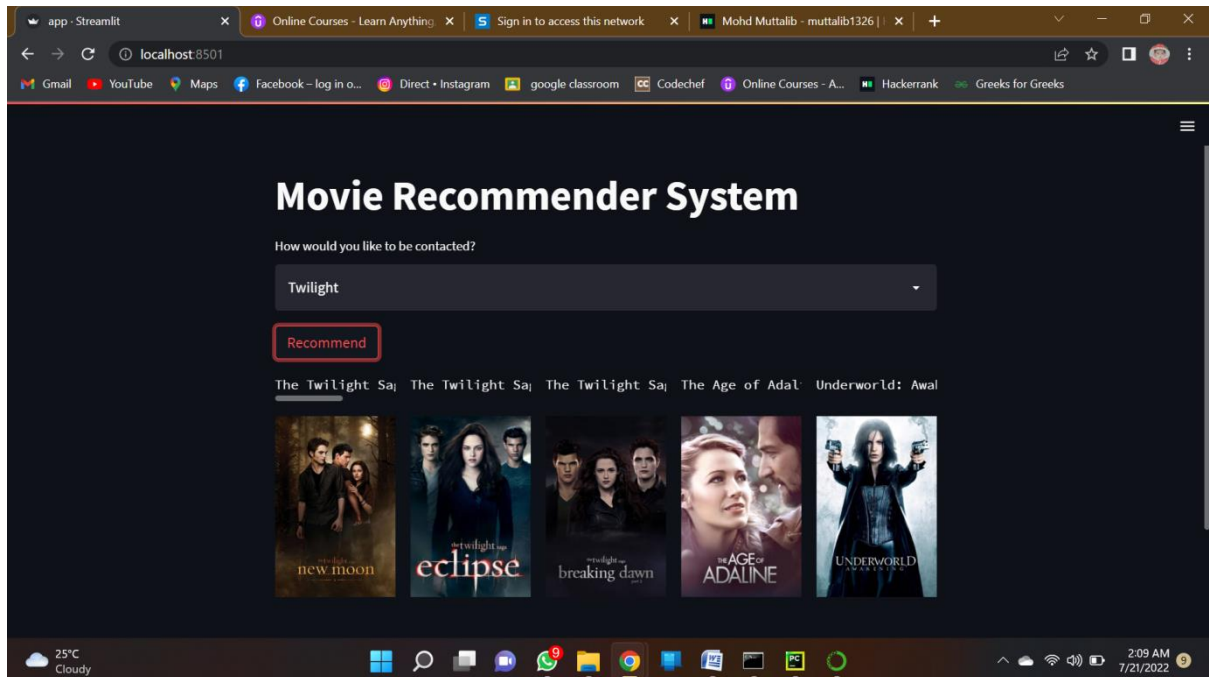
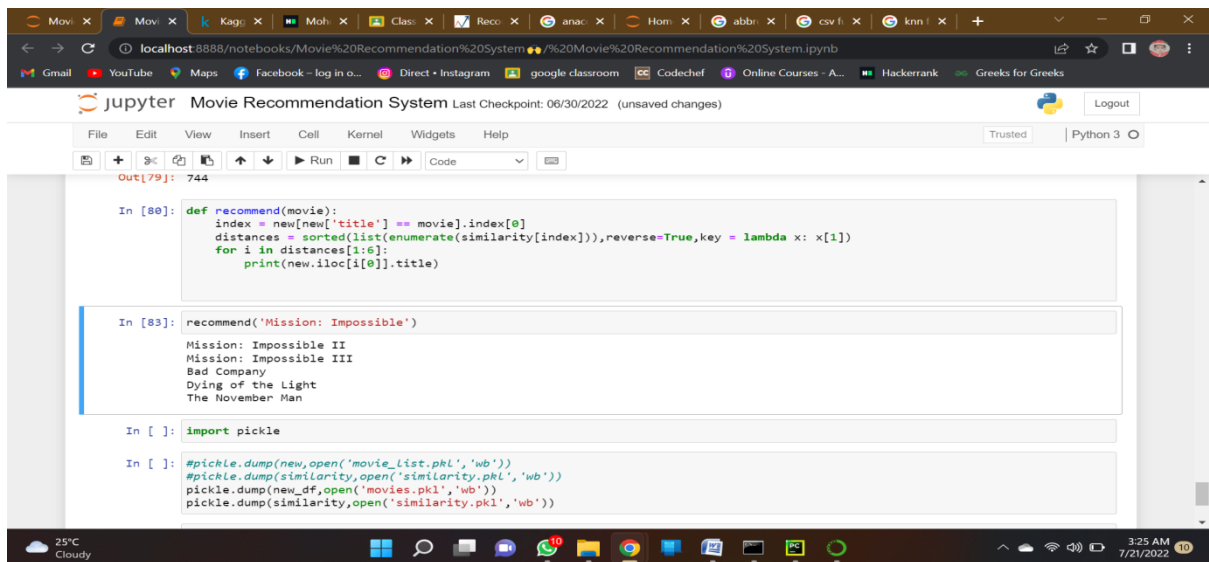


Image 5.3: More results

INPUT/ OUTPUT SCREEN



```
Out[79]: 744

In [80]: def recommend(movie):
         index = new[new['title'] == movie].index[0]
         distances = sorted(list(enumerate(similarity[index])),reverse=True,key = lambda x: x[1])
         for i in distances[1:6]:
             print(new.iloc[i[0]].title)

In [83]: recommend('Mission: Impossible')
Mission: Impossible II
Mission: Impossible III
Bad Company
Dying of the Light
The November Man

In [ ]: import pickle

In [ ]: #pickle.dump(new,open('movie_list.pkl','wb'))
        #pickle.dump(similarity,open('similarity.pkl','wb'))
        pickle.dump(new_of,open('movies.pkl','wb'))
        pickle.dump(similarity,open('similarity.pkl','wb'))
```

Image 5.4: Input screen

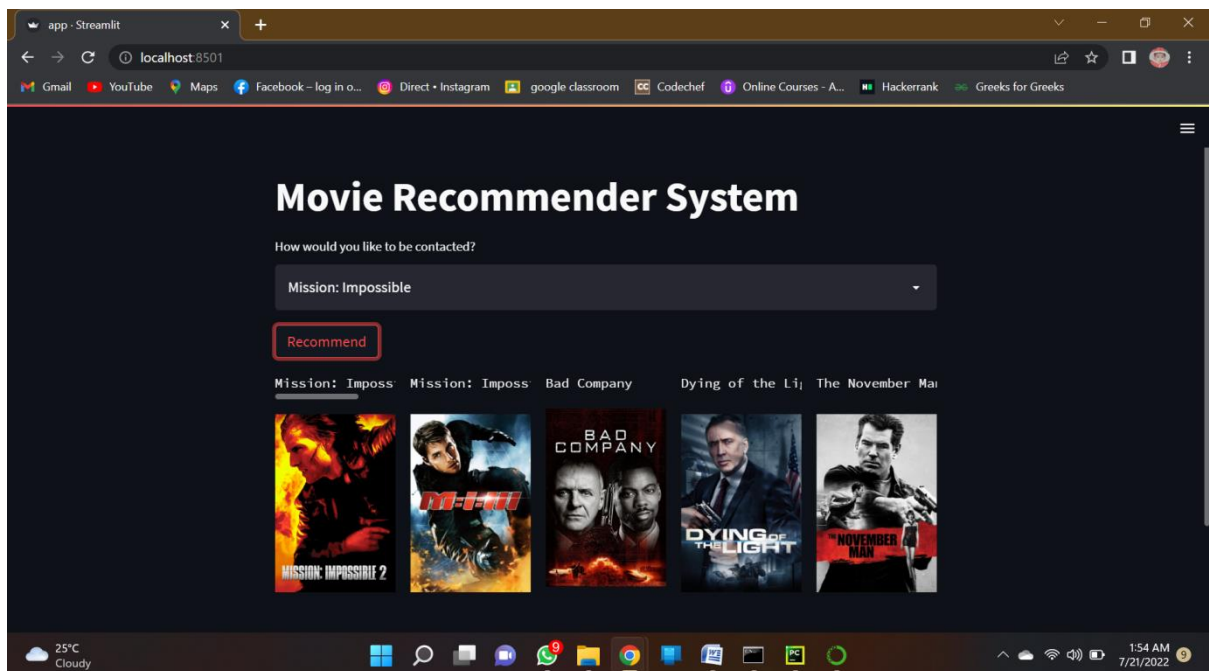


Image 5.4: Output screen

CHAPTER - 6
CONCLUSION
&
FUTURE WORK

6. CONCLUSION & FUTURE WORK

6.0 CONCLUSION & FUTURE WORK

6.1 SUMMARY OF THE RESEARCH

Recommender system has become more and more important because of the information overload. For content-based recommender system specifically, we attempt to find a new way to improve the accuracy of the representative of the movie. For the problems we mentioned in the beginning, we implement content-based recommender algorithm which means there is no cold start problem. Also, we list all the features in our recommender system. Some of them are from other researches that we did, so the features are diverse and more accurate than others. Then we introduced the cosine similarity which is commonly used in industry. For the weight of features, we introduced other features which improve the representative of the movie. Our thesis introduces a content-based recommender system for the movie website. The features used in the system are extracted from various aspects of the movie, which are diverse and unique. We introduce a new approach for setting weight for these features. In the end of the project, we use various metrics to evaluate the improvement of the new approach. It is illustrated that the new approach contributes positively according to the evaluation.

6.2 FUTURE WORK

Recommender system has been developing for a long period of time, which ever entered a low point. In the past few years, the development of machine learning, large-scale network and high performance computing is promoting new development in this field. In the coming days work we will consider the following aspects in.

- Use collaborative filtering recommendation. After getting enough user data, collaborative filtering recommendation will be introduced. As we have discussed collaborative filtering is based on the social information of users, which will be analyzed in the future research.

- Introduce more precise and proper features of movie. Typical collaborative filtering recommendation uses the rating instead of object features. In the future we certainly would extract features such as color and subtitle from movie which can provide a more accurate description for movie.

- Introduce user dislike movie list. The user data is always useful in recommender systems. In the future we will collect more user data and add user dislike movie list. We will input dislike movie list into the recommender system as well generate scores that will be added to previous result. By this way we can improve the result of recommender system.

- Introduce machine learning. For future study, dynamic parameters will be introduced into recommender system, we will use machine learning to adjust the weight of each feature automatically and find the most suitable weights.

- Make the recommender system as an internal service. In the future, the recommender system is no longer an external website that will be just for testing. We will make it as an internal APIs for developers to invoke. Some movie lists in the website will be sorted by recommendation.

REFERENCES

1. [1]. Sharma, R. (2020, May) Movie Recommendation System using Machine Learning, *Retrieved*
2. [2]. MA, K. 2016. Content-based Recommender System for Movie Website, *Retrieved*
3. [3]. PAVAN KUMAR P B, NITESH S, MURALIDHARA REDDY SY, PRAVEEN V. 2020. "MOVIE RECOMMENDATION SYSTEM" *Retrieved*
4. [4]. Kumar, M. (August 2015). A Movie Recommender System MOVREC, *International Journal of Computer Applications*.
5. [5]. Z Wang, X Yu, N Feng, Z Wang. 2014. Journal of Visual Languages & Computing, Elsevier.
6. [6]. M Goyani, N Chaurasiya - ELCVIA: electronic letters on computer vision ..., 2020
7. [7]. V Subramaniaswamy, R Logesh... - ... Journal of High ..., 2017 - inderscienceonline.com
8. [8]. G Arora, A Kumar, GS Devre... - International journal of ..., 2014 - academia.edu
9. [9]. S Agrawal, P Jain - 2017 International Conference on I-SMAC ..., 2017 - ieeexplore.ieee.org
10. [10]. EA Eyjolfssdottir, G Tilak, N Li - UC Santa Barbara: Technical Report, 2010 – Citeseer
11. [11]. CSM Wu, D Garg, U Bhandary - 2018 IEEE 9th International ..., 2018 - ieeexplore.ieee.org
12. [12]. Netflix logo, Retrieved from wallpapercave.com
13. [13]. Disney star, Retrieved from google images
14. [14]. Netflix background image, Retrieved from Netflix.com
15. [15]. Amazon prime log, Retrieved from google images

