A Project Report

on

MOVIE RECOMMENDATION SYSTEM USING COMBINATION OF CONTENT-BASED AND COLLABORATIVE APPROACHES

Submitted in partial fulfillment of the

requirement for the award of the degree of

Bachelor of Technology in Computer Science & Engineering

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Under The Supervision of
Dr. Abdul Aleem
Professor
Department of Computer Science and Engineering

Submitted By:

Shubham Sharma-18SCSE1010103

Chetan Sharma-18SCSE1010026

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING / DEPARTMENT OF COMPUTERAPPLICATION
GALGOTIAS UNIVERSITY, GREATER NOIDA
INDIA

January - 2022

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the project, entitled **"MOVIE RECOMMENDATION SYSTEM USING COMBINATION OF CONTENT-BASED AND CALLOBRATIVE APPROACHES "** in partial fulfillment of the requirements for the award of the **BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING** submitted in the **School of Computing Science and Engineering of Galgotias University, Greater Noida,** is an original work carried out during the period of **Jan-2022 to May-2022,** under the supervision of **Dr. Abdul Aleem , Professor, Department of Computer Science and Engineering** of School of Computing Science and Engineering, Galgotias University, Greater Noida

The matter presented in the project has not been submitted by me/us for the award of any other degree of this or any other places.

Shubham Sharma - 18SCSE1010103

Chetan Sharma- 18SCSE1010026

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Dr.Abdul Aleem

Professor

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of **Shubham Sharma-18SCSE1010103, Chetan Sharma- 18SCSE1010026** has been held on **_________________** and his/her work is recommended for the award of **BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING.**

Signature of Examiner(s) **Signature of Supervisor(s)**

Signature of Project Coordinator **Signature of Dean**

Date: May, 2022

Place: Greater Noida

Table of Contents

ABSTRACT

Recommendation systems have made discovering the things simple that one need. Movie recommendation systems target helping movies devotees by recommending what movies to observe without having to go through the long process of choosing from a large set of movies that go up to thousands and millions that is time-consuming and confusing. This research aims to lessen human exertion by proposing motion pictures dependent on the client's advantages. To handle such problems, a model is introduced that combines both content-based and collaborative approaches. It will give logically express results contrasted with various frameworks that depend on a content-based approach. Content-based proposal frameworks are compelled to individuals, these frameworks don't endorse things out of the container, along these lines restricting your decision to investigate more. In short, the proposed model focus on a framework that is able to address the issues of existing approaches.

INTRODUCTION

A proposal framework is a sort of data sifting framework which endeavors to anticipate the inclinations of a client, and make recommends dependent on these inclinations. There are a wide assortment of utilization for suggestion frameworks. These have become progressively well known in the course of the most recent couple of years and are currently used in most internet based stages that we use. The substance of such stages differs from films, music, books and recordings, to companions and stories via online media stages, to items on web based business sites, to individuals on expert and dating sites, to indexed lists returned on Google. Frequently, these frameworks can gather data about a clients decisions, and can utilize this data to improve their ideas later on.For instance, Facebook can screen your cooperation with different stories on your feed to realize what sorts of stories appeal to you. Here and there, the recommender frameworks can make upgrades dependent on the exercises of an enormous number of individuals. For instance, if Amazon sees that countless clients who purchase the most recent Apple Macbook additionally purchase a USB-C-to USB Adapter, they can suggest the Adapter to another client who has recently added a Macbook to his truck. Because of the advances in recommender frameworks, clients continually anticipate great proposals. They have a low edge for administrations that can't make suitable ideas. On the off chance that a music streaming application isn't ready to anticipate and play music that the client likes, then, at that point, the client will basically quit utilizing it.This has prompted a high accentuation by tech organizations on further developing their proposal frameworks. Be that as it may, the issue is surprisingly complicated. Each client has various inclinations also, likes. Also, even the flavor of a solitary client can fluctuate contingent upon an enormous number of factors, like state of mind, season, or sort of action the client is doing. For instance, the sort of music one might want to hear while practicing varies significantly from the kind of music he'd pay attention to when preparing supper. Another issue that suggestion frameworks need to tackle is the investigation versus abuse issue.They should investigate new areas to find more concerning the client, while as yet capitalizing on what is as of now thought about of the client. Three fundamental methodologies are utilized for our recommender frameworks. One is Demographic Filtering i.e They submit summed up proposals to each client, in light of film ubiquity and additionally 7 type. The System prescribes similar motion pictures to clients with comparative segment highlights. Since every client is unique , this methodology is viewed as excessively straightforward. The fundamental thought behind this framework is that motion pictures that are more well known and widely praised will have a higher likelihood of being enjoyed by the normal crowd. Second is content-based sifting, where we attempt to profile the clients intrigues utilizing data gathered, and suggest things in view of that profile. The other is communicant sifting, where we attempt to bunch comparable clients together and use data about the gathering to make suggestions to the client.

1.1 Objectives

Hybrid Movies Recommendation System goal is to recommended movies by merging two recommendation system i.e, content based and collaborative filtering and create a system which possesses both system properties.

The main goal of the our system is:

I. To recommend movies.

II. To recommend movies based on user history i.e, the next movies recommended based on the number of times the user watches a certain genre of movie.

III. To recommend movies based on overall rating i.e, movies which are famous and these movies can have different genre also.

1.2 Problem Statement

Giving related substance out of an important and unimportant assortment of things to clients of online specialist organizations.

Hybrid movies recommendation system aims to recommend movies to users based on user-movie (item) ratings.

Given a bunch of clients with their past appraisals for a bunch of motion pictures, would we be able to anticipate the rating they will allot to a film they have not recently evaluated?

LITERATURE SURVEY

With regards to a survey of the writing, a suggestion framework utilizing a substance based collective and mixture approach by a past specialist is an alternate way to deal with the improvement of proposal based motors. In 2007 a web based and information based knowledge film suggestion framework has been offered utilizing the crossover sifting technique . In 2017, a film proposal framework upheld style and rating coefficient of connection reason by the creators. In 2013 a Bayesian organization and trust model based film suggestion motor have been prescribed to anticipate evaluations for clients and things, essentially from datasets to suggest clients their decision as well as the other way around.In 2018, the creators constructed a suggestion motor by investigating the evaluations data set gathered from Kaggle to suggest films for a client chosen from Python. In 2018 film suggestion motors give a cycle to assist clients classify clients with comparable k-mean cuckoo esteems and support learning based recommender frameworks, which are utilizing bicycling procedures . Starting exploration mostly focused on the substance of the suggestion framework that inspected the provisions of the item to finish the proposal task. Tests checked that their methodologies were more flexible and exact. Bayesian networks are utilized for model-put together inclinations based with respect to their specific situation.At the point when customers embrace new conduct, it is hard for community sifting to respond in a flash. Accordingly, the two analysts and professionals want to adjust community sifting strategy and content-based technique to address the issue. Ternary executed Unplugged Learning of Machine Learning to inspect the extremity of machine reflectivity.

2.1 Background

Over the previous decade, countless proposal frameworks for an assortment of areas have been created and are being used. These proposal frameworks utilize an assortment of strategies, for example, content-based methodology, community approach, information-based methodology, utility-based methodology, half breed approach, and so forth The vast majority of the web-based proposal frameworks for an assortment of things use appraisals from past clients to make suggestions to current clients with comparative interests. One such framework was planned by Jung, Harris, Webster, and Her storage (2004) for further developing indexed lists. The framework urges clients to enter longer and more educational pursuit questions, and gathers appraisals from clients with regards to whether or not indexed lists meet their data need. These appraisals are then used to make proposals to later clients with comparative necessities.

2.2 Existing product and system

Current Social Networking World Internet long-range interpersonal communication locales, which started in 1995 with Classmates.com, have flooded in ubiquity and use through verbal promoting. From that point forward, a wide scope of virtual networks have framed filling various needs and focusing on differing specialty audiences: recommendation framework and haven't exploited interpersonal interaction networks or group intelligence. A few sites, like Blockbuster, do give individualized proposals dependent on a client's evaluations however do exclude any long range interpersonal communication part. Yippee! Motion pictures goes further and utilize individual evaluations to recommend films right now playing in theater, on TV, and out on DVD. It likewise draws upon its tremendous client base to give arrangements of comparative film fans, their appraisals, and surveys. Other film destinations, as Flixster, adopt an alternate strategy. Flixster structures online networks around films and proposes motion pictures to watch dependent on what your companions have evaluated.

2.3 Movie Recommendation System by K-Means Clustering AND K-Nearest Neighbor.

A suggestion framework gather information about the client's inclinations either certainly or unequivocally on various things like films. An implied securing in the advancement of film suggestion framework utilizes the client's conduct while watching the motion pictures. Then again, an express securing in the advancement of film suggestion framework utilizes the client's past evaluations or history. The other supporting method that are utilized in the advancement of suggestion framework is bunching. Bunching is a cycle to bunch a bunch of items so that articles in similar groups are more like each other than to those in different groups. K-Means Clustering alongside K-Nearest Neighbor is carried out on the film focal point dataset to get the best-improved outcome. In existing method, the information is dispersed which brings about countless groups while in the proposed procedure information is accumulated and brings about a low number of bunches. The course of suggestion of a film is enhanced in the proposed conspire. The proposed recommender framework predicts the client's inclination of a film based on various boundaries. The recommender framework chips away at the idea that individuals are having normal inclination or decision. These clients will effect on one another viewpoints. This cycle improves the interaction and having lower RMSE.

2.4 Movie Recommendation System Using Collaborative Filtering:

Communitarian separating frameworks break down the client's conduct and inclinations and anticipate what they would like dependent on similitude with different clients. There are two kinds of collaborative filtering systems; user-based recommender and item-based recommender.

A. Use-based filtering: Client based inclinations are extremely normal in the field of planning customized frameworks. This methodology depends on the client's likings. The interaction begins with clients giving evaluations (1-5) to certain films. These evaluations can be certain or unequivocal. Express appraisals are the point at which the client unequivocally rates the thing on some scale or shows approval/disapproval to the thing. Regularly unequivocal evaluations are difficult to assemble as only one out of every odd client is tremendously keen on giving criticism's. In these situations, we assemble certain evaluations dependent on their conduct. For example, on the off chance that a client purchases an item at least a couple of times, it demonstrates a positive inclination. In setting to film frameworks, we can infer that assuming a client watches the whole film, he/she has some like capacity to it. Note that there are no unmistakable principles in deciding verifiable evaluations. Then, for every client, we first discover some characterized number of closest neighbors. We ascertain relationship between's clients' appraisals utilizing Pearson Correlation calculation. The suspicion that assuming two clients' appraisals are exceptionally corresponded, then, at that point, these two clients should appreciate comparable things and items is utilized to prescribe things to clients.

B. Item-based filtering: Not at all like the client based sifting technique, thing put together concentrations with respect to the comparability between the thing's clients rather than the actual clients. The most comparative things are figured early. Then, at that point, for proposal, the things that are generally like the objective thing are prescribed to the client.

2.5 System Study

Feasibility Study

The plausibility of the task is examined in this stage and strategic agreement is advanced with an extremely broad arrangement for the venture and some quotes. During framework investigation the achievability investigation of the proposed framework is to be completed. This is to guarantee that the proposed framework isn't a weight to the organization. For attainability examination, some comprehension of the significant necessities for the framework is fundamental.

Three key contemplations engaged with the plausibility examination are:

Economical Feasibility

Technical Feasibility

Social Feasibility

Economical Feasibility:

This review is completed to check the monetary effect that the framework will have on the association. How much asset that the organization can fill the innovative work the framework is restricted. The consumptions should be advocated. Consequently the created framework also affordable and this was accomplished on the grounds that the greater part of the advancements utilized are openly accessible. Only the customized products had to be purchased.

Technical Feasibility:

This review is completed to check the specialized practicality, that is, the specialized prerequisites of the framework. Any framework created should not have a popularity on the accessible specialized assets. This will prompt high requests on the accessible specialized assets. This will prompt high requests being put on the customer. The created framework should have an unassuming prerequisite, as just insignificant or invalid changes are needed for executing this framework.

Social Feasibility:

The part of study is to really take a look at the degree of acknowledgment of the framework by the client. This incorporates the method involved with preparing the client to utilize the framework productively. The client should not feel undermined by the framework, rather should acknowledge it as a need. The degree of acknowledgment by the clients exclusively relies upon the strategies that are utilized to teach the client about the framework and to make him acquainted with it. His degree of certainty should be raised so he is additionally ready to make some helpful analysis, which is invited, as he is the last client of the framework.

2.6 Hybrid Recommendation And Approach

One common occurrence in RSS research is the demand to combine recommendation techniques to achieve peak performance. All of the known recommendation techniques have advantages and disadvantages, and many researchers have chosen to combine techniques in different ways in order to leverage their advantages. This session surveys the different hybrid recommendation approaches.

Hybrid systems combine two or more techniques in order to gain better performance with fewer limitations of each approach. Many hybrid systems have been applied to travel and tourism applications. For instance F. Ricci et al. represent a movement planning.recommender framework that is case-based, subsequently is information based, yet additionally Collaborative-based since it suggests travel benefits that have been assessed emphatically by others.

Fab is a proposal framework intended to assist clients with investigating the tremendous measure of data accessible on the web. This half breed framework joins the Content-based and Collaborative strategies for suggestion such that takes advantage of the benefits of the two methodologies while keeping away from their weaknesses. Fab's mixture structure takes into consideration programmed acknowledgment of new issues applicable to different gatherings of clients. It additionally empowers two scaling issues relating to the rising number of clients and reports, to be tended to.

One major tactic for improving recommendation is to combine Collaborative filtering with Content-based recommenders. We can illustrate the benefits of such hybrid systems with a simple example; suppose one user has rated the NBA page from CBSSports.com favorably, while another has rated the NBA page from CNNSL.com favorably, pure Collaborative filtering would find no correlation between the two users. However, Content analysis can show that the two items are in fact quite similar, thus indicating.

2.7 Survey of Technologies

Hybrid Movies Recommendation System goal is to recommended movies by merging two recommendation system i.e, content based and collaborative filtering and create a system which possesses both system properties.

PYTHON :

- □Python is a well known programming language.
- □Python is a deciphered significant level universally useful programming language.
- □Python can be utilized on a server to make web applications.
- □Python can be utilized close by programming to make work processes.
- □Python can interface with data set frameworks. It can likewise peruse and adjust records.
- □Python can be utilized to deal with enormous information and perform complex arithmetic.
- □Python can be utilized for quick prototyping, or for creation prepared programming advancement.
- □Python chips away at various stages (Windows, Mac, Linux, Raspberry Pi, and so on)
- □Python has a straightforward linguistic structure like the English language.

- Python has sentence structure that permits engineers to compose programs with less lines than some other programming dialects.
- Python runs on a mediator framework, implying that code can be executed when it is composed. This implies that prototyping can be extremely fast.
- Python can be treated in a procedural manner, an article situated way or a useful way.
- Python was intended for clarity, and has a few likenesses to the English language with impact from arithmetic.
- Python utilizes new lines to finish an order, rather than other programming dialects which regularly use semicolons or enclosures.
- Python depends on space, utilizing whitespace, to characterize scope; like the extent of circles, capacities and classes. Other programming dialects frequently utilize wavy sections for this reason.
- Python Comments can be utilized to clarify Python code.
- Python Comments can be utilized to make the code more intelligible.
- Python Comments can be utilized to forestall execution when testing code.
- It upholds practical and organized programming strategies just as OOP.
- It tends to be utilized as a prearranging language or can be gathered to byte-code for building huge applications.
- It gives extremely undeniable level unique information types and supports dynamic sort checking.
- It upholds programmed trash assortment.
- It very well may be effectively incorporated with C, C++, COM, ActiveX, CORBA, and Java.

- Python supports multiple programming design, including object-situated, basic, and utilitarian or procedural programming styles.
- Python isn't planned to work in a specific region, for example, web programming. To that end it is known as multipurpose programming language since it tends to be utilized with web, undertaking, 3D CAD, and so on

REQUIREMENTS AND ANALYSIS

Problem Definition

This paper depends on proposal framework that prescribes various things to clients. This framework will prescribe motion pictures to clients. This framework will give more exact outcomes when contrasted with the current frameworks. The current framework deals with individual clients' evaluating. This might be at some point futile for the clients who have diverse taste from the suggestions shown by the framework as each client might have various preferences. This framework computes the similitudes between various clients and afterward prescribe film to them according to the appraisals given by the various clients of comparative preferences. This will give an exact proposal to the client.

Giving related substance out of pertinent and unimportant assortment of things to clients of online specialist co-ops.

Half breed motion pictures proposal framework expects to prescribe films to clients dependent on client film (thing) evaluations.

Given a bunch of clients with their past appraisals for a bunch of motion pictures, would we be able to foresee the rating they will relegate to a film they have not recently evaluated?

Requirement Specification

Hybrid Movies Recommendation System provides the Customers with better and more efficient means of movies recommendation system and from all over the world. Technologies such as python.

- Python is a famous programming language.
- Python is a deciphered undeniable level universally useful programming language.
- Python can be utilized on a server to make web applications.
- Python can be utilized close by programming to make work processes.
- Python can associate with data set frameworks. It can likewise peruse and change records.
- Python can be utilized to deal with large information and perform complex math.

SOFTWARE AND HARDWARE REQUIREMENTS

Hardware Used

1. Intel Core 2 Quad Q9550,E0 revision
2. 4 GB PC2-6400 800Mhz RAM
3. 500 GB Hard disk drive

Software Requirements

1. Windows 7 Ultimate
2. IE8/Firefox/Safari

Preliminary Product Description

We indulged ourselves into a lot of research before we started the actual work on the — “HYBRID MOVIES RECOMMENDATION SYSTEM” application. We Studied about the different kinds of movies recommendation system such as content based and collaborative filtering that were needed to build our application. hybrid movies recommendation system is a web based application to support user friendliness and make the recommendation more efficient.

CONCEPTUAL MODELS

DATA FLOW DIAGRAM

1-Data Flow

An arrow represents data flow; it represents the path over which data travels in the system? A data flow can move between processes, flow into or out of data stores, to and from external entities.

2-Bubbles (Process)

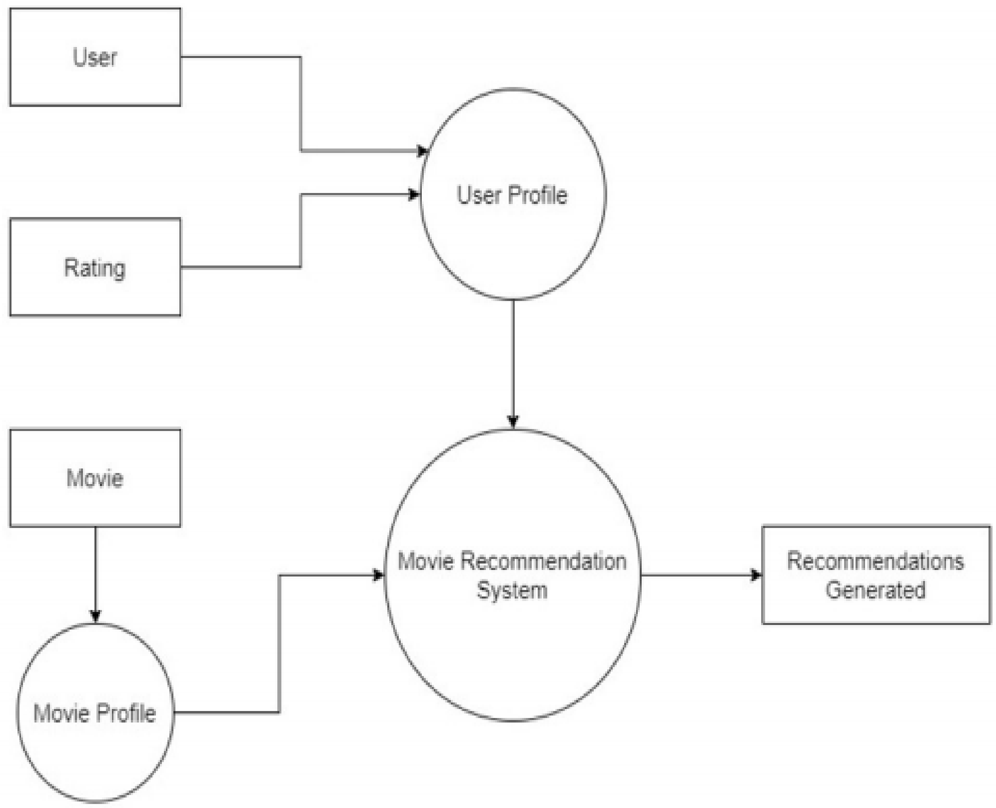
A circle or bubble represents that transforms data from once form to another by performing some tasks with the data.

3-Data store

A data store is a place where data is held temporarily from one transaction to the next or is stored permanently.

4-Entity

Outer Entity image addresses wellsprings of information to the framework or objections of information from the framework.


User
Rating
User Profile
Movie
Movie Recommendation System
Recommendations Generated
Movie Profile

E-R DIAGRAM

1-Entity

Entity has a set of attribute whose value is uniquely, identify the entity or distinguish the entity from the other thing or entity in the world.

2-Attribute

Are used to define an entity that is the property that describe an entity. That is the properly that describe an entity.

3-Relationship

Is association among different/several,entities. It connect to one or more entities.

4-Derived

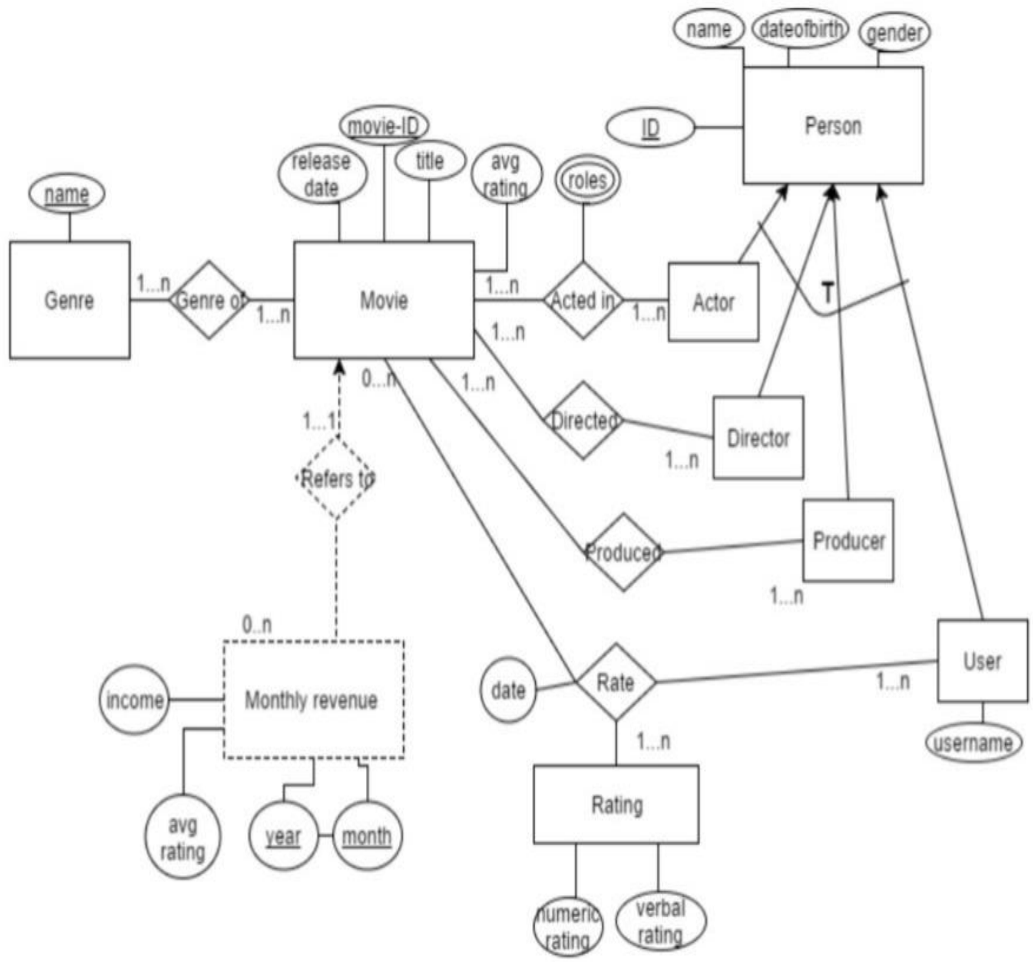
Attribute Any attribute which is derived from the relationship is called derived attribute.

5-Cardinality

Carnality refers to the multiplicity of the entities. How many entity are actually engaged in the relationship.

Types of cardinality:-

1. Many to 1
2. 1 to 1
3. 1 to Many
4. Many to Many



name
dateofbirth
gender
ID
Person
movie-ID
release date
title
avg rating
roles
name
Genre
1...n
Genre of
1...n
Movie
1...n
Acted in
1...n
Actor
T
1...n
0...n
1...n
Directed
Director
1...n
1...1
Refers to
Produced
Producer
1...n
0..n
User
Monthly revenue
income
date
Rate
1...n
username
1...n
Rating
avg rating
year
month
numeric rating
verbal rating

CLASS DIAGRAM

Class graph in the Unified Modeling Language (UML) is a kind of static design outline that portrays the construction of a framework by showing the framework's classes, their characteristics, tasks (or techniques), and the connections among objects.

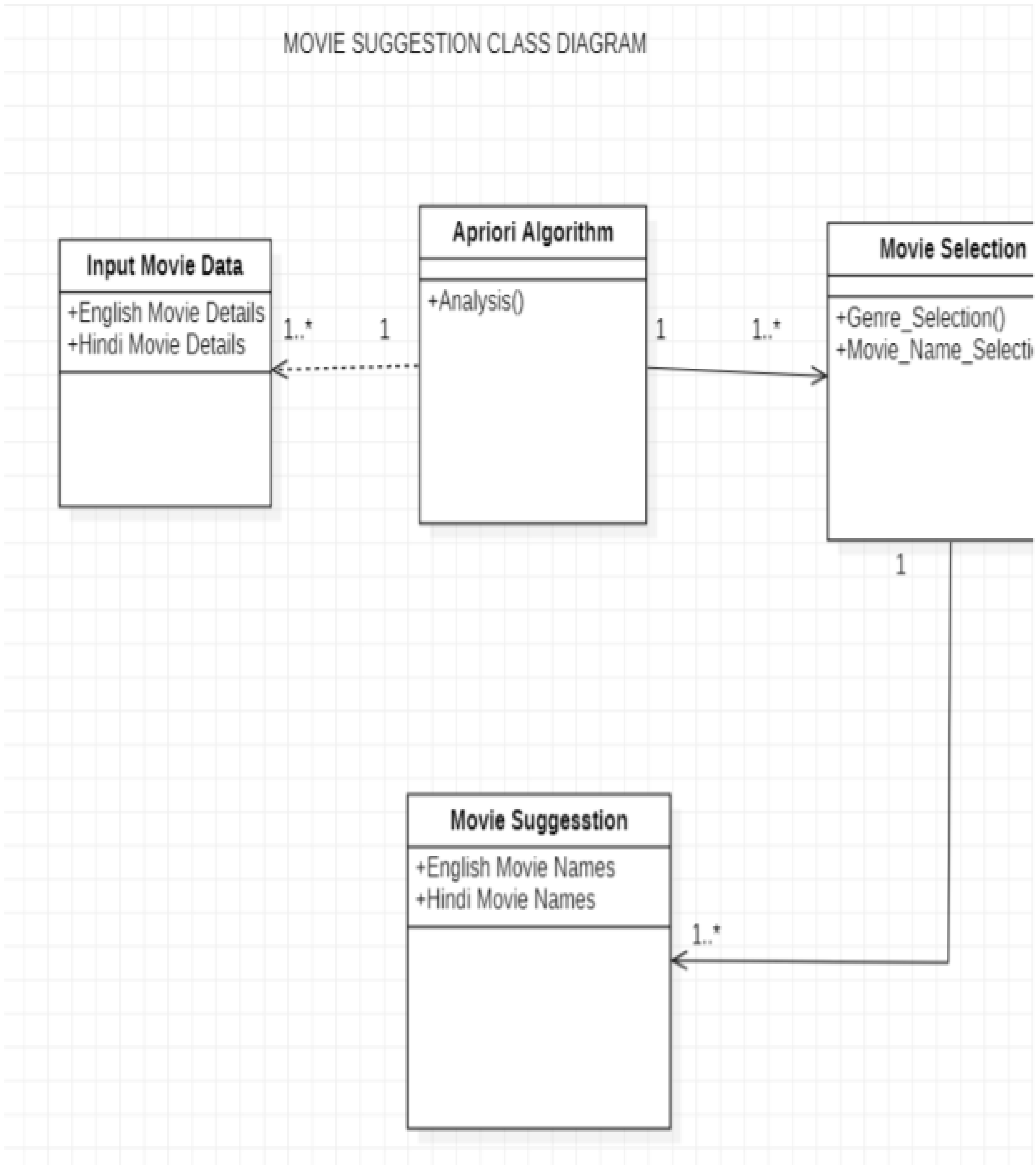
- The above class graph depicts the classes, traits and strategies for film suggestion framework.

- Input film information: It is class which takes the information of the film subtleties.

- Deduced calculation: This class will do investigations dependent on the given information and recommends film.

- Film determination: This class helps in choosing film dependent on the class and the film evaluations.

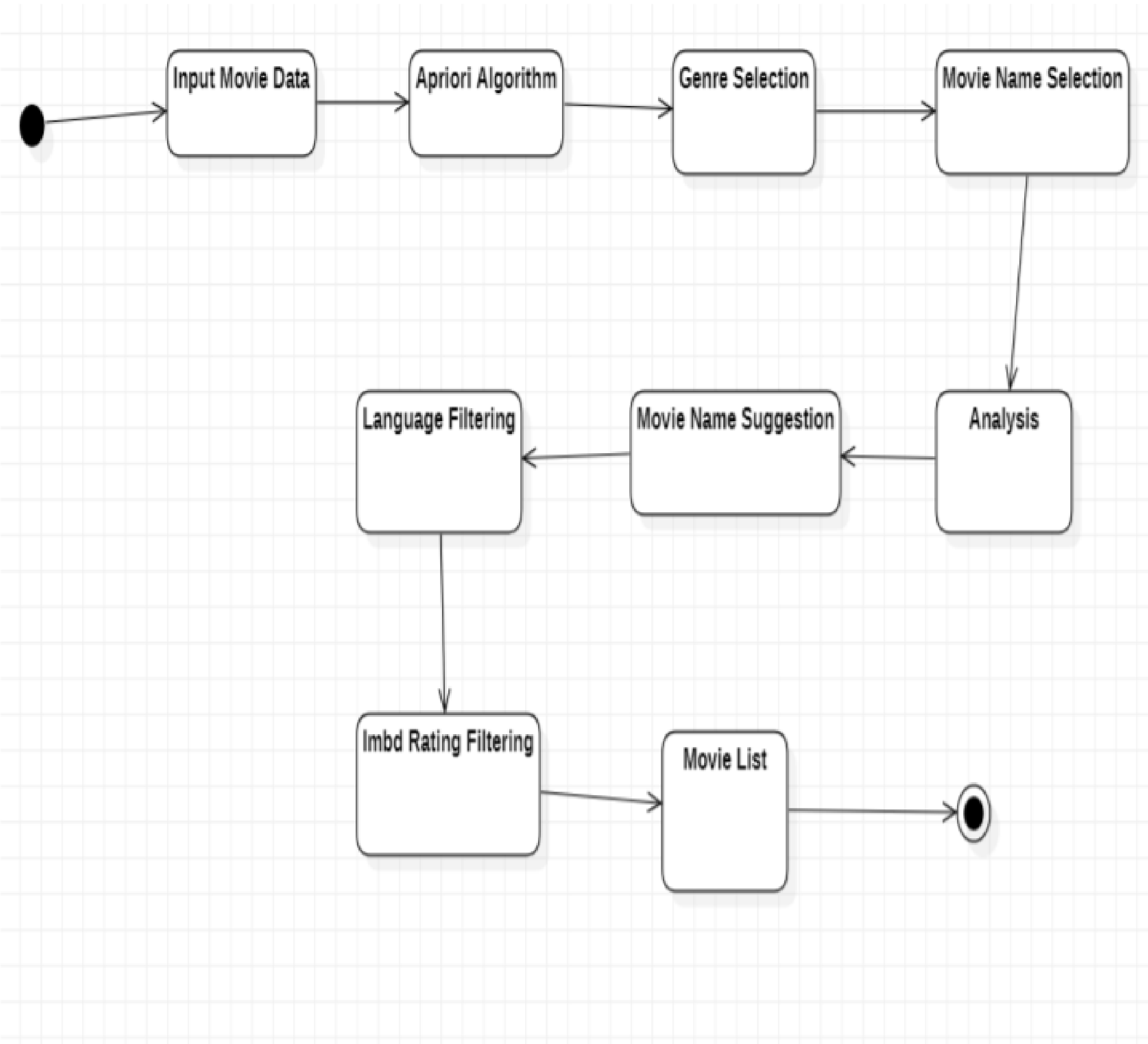
□Film idea: This class recommends the film to client dependent on the habitually watched movie(genre).



STATE DIAGRAM

A state graph is a chart utilized in software engineering to portray the conduct of a framework considering every one of the potential conditions of an article when an occasion happens. State charts graphically address limited state machines. They are simply used to comprehend object conduct all through the entire framework.

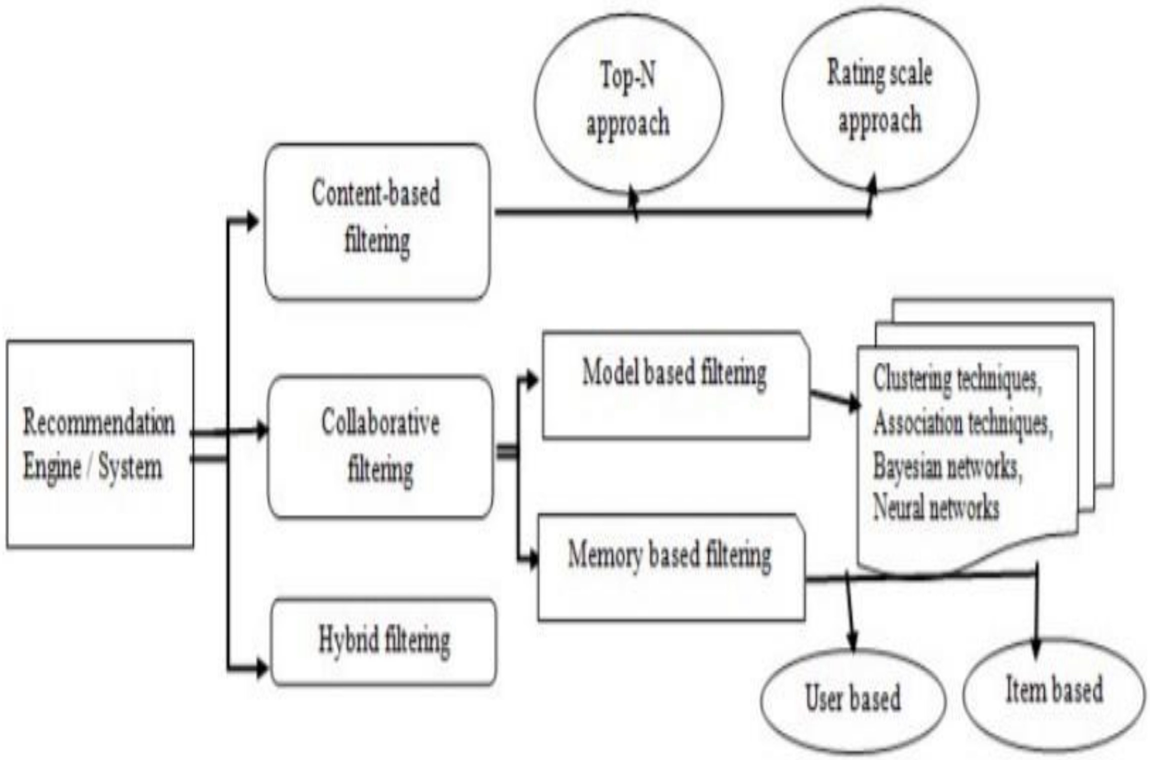
- In the above graph it portrays the conduct of film suggestion framework.
- At the point when the information of film is given it checks for the comparative class by utilizing proposal calculation and recommends the film to client.
- Idea depends on the film appraisals, class, language and oftentimes watched motion pictures.
- In the event that the recently watched film's evaluating is three, then, at that point, the recommended film's appraising ought to be equivalent or more prominent than three.



STATE DIAGRAM OF MOVIE RECOMENDATION

ARCHITECTURE DIAGRAM

An architecture diagram is a graphical representation of a set of concepts, that are part of an architecture, including their principles, elements and components. ... An Example Architecture Diagram of an Enterprise Architecture to create a Modern Smart and Green Company, using various concepts and principles.



SEQUENCE DIAGRAM

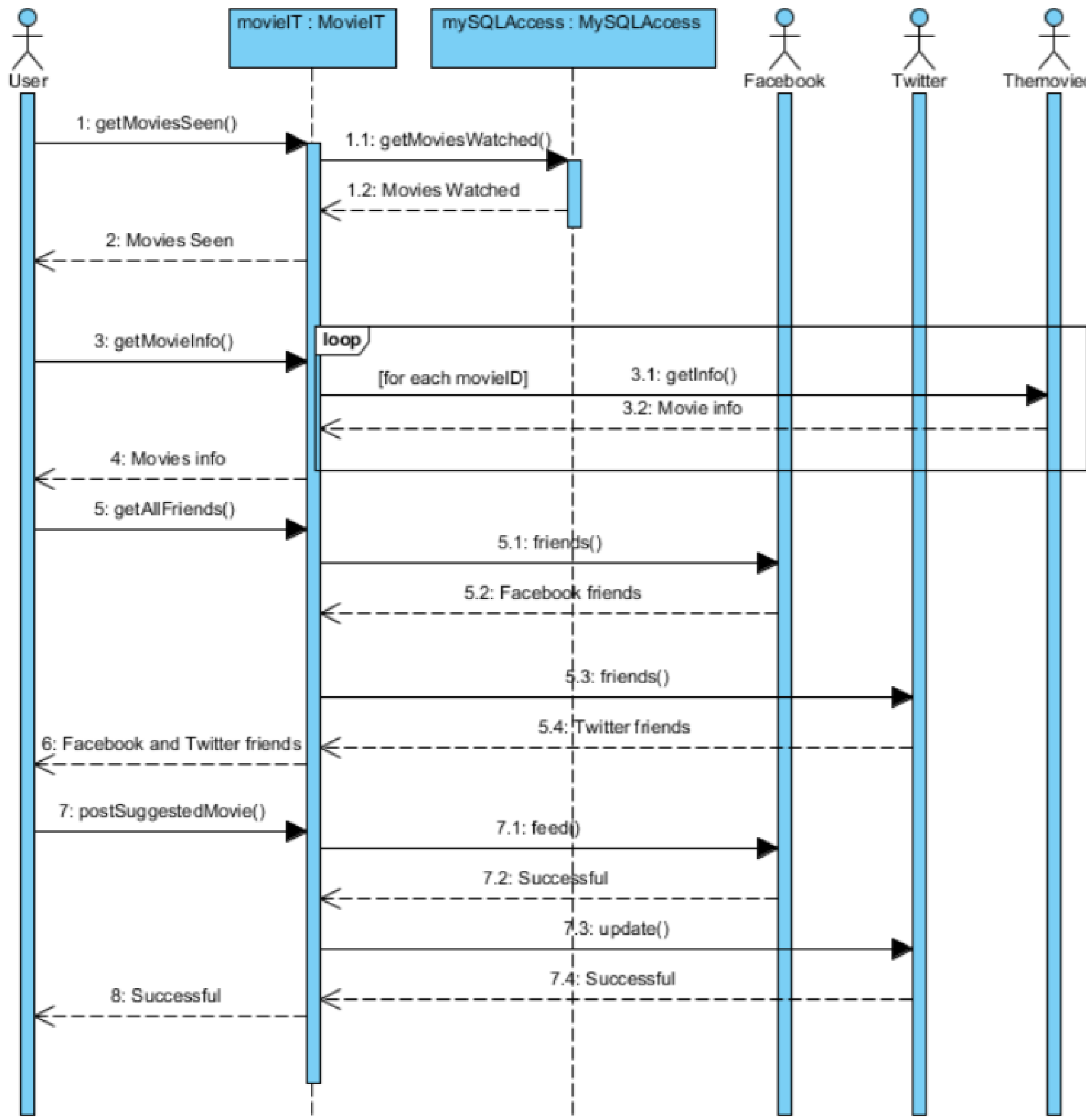
An arrangement chart just portrays association between objects in a successive request for example the request where these associations happen. We can likewise utilize the terms occasion graphs or occasion situations to allude to a succession outline. Succession graphs depict how and in what request the items in a framework work.

- When a film is chosen by the client the server gets the subtleties of his recently watched motion pictures.

- The data is put away and to see the data again we can simply check for the client history utilizing the client id.

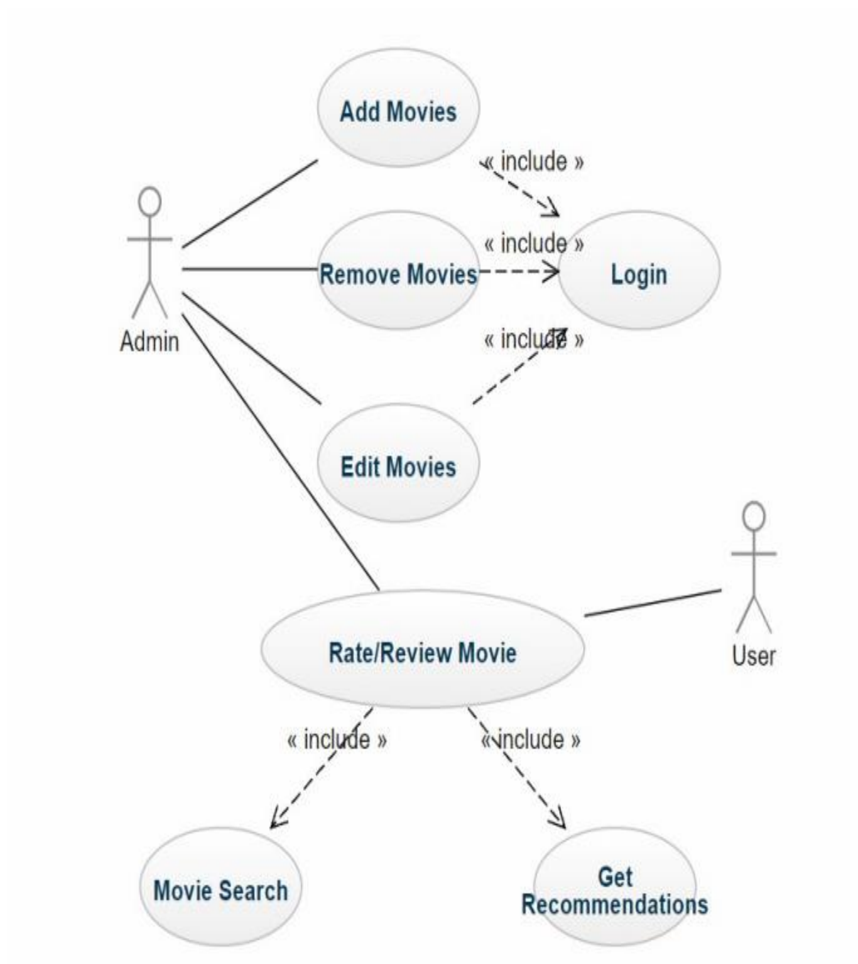
- For additional the suggestion can be imparted to companions on various stages.

The age of suggestions continues in circle until the client chooses one for his watch.



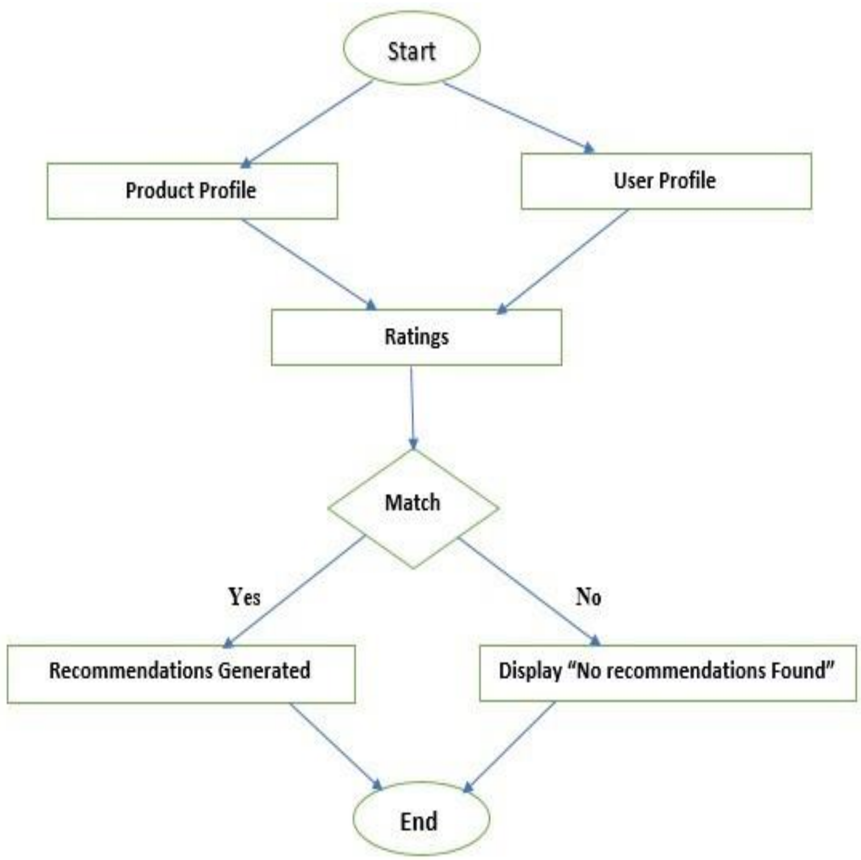
USE CASE DIAGRAM

A utilization case shows a bunch of utilization cases, entertainers and their relationship. The use case diagram make system and classes approachable by presenting an outside view of how the elements may be used in context.



FLOW CHART

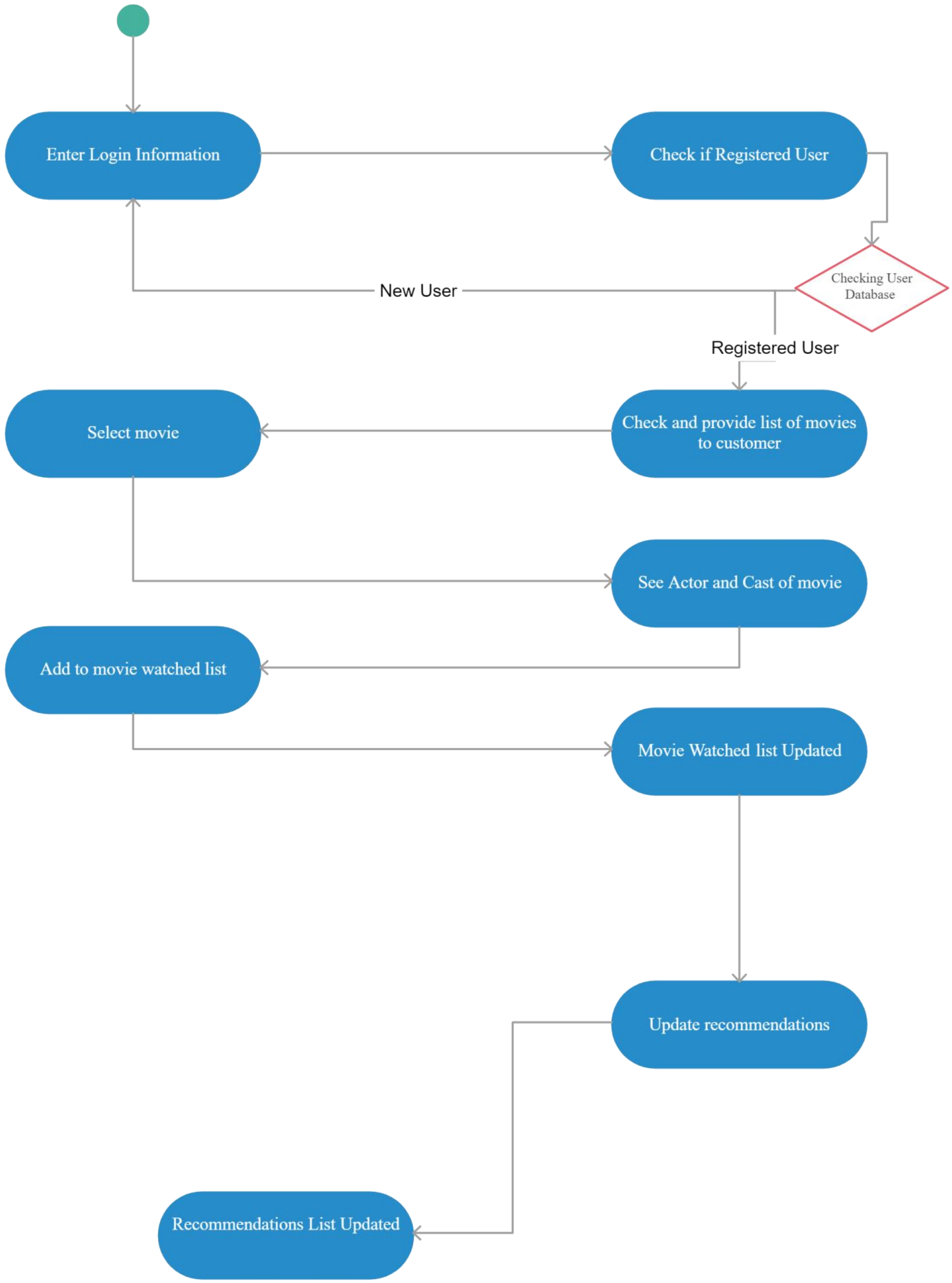
A flowchart is a graph that portrays an interaction, framework or PC calculation. They are generally utilized in different fields to report, study, plan, improve and convey frequently complex cycles in clear, straightforward charts. Flowcharts, once in a while spelled as stream diagrams, use square shapes, ovals, precious stones and possibly various different shapes to characterize the kind of step, alongside interfacing bolts to characterize stream and succession. They can go from basic, hand-attracted graphs to far reaching PC drawn charts portraying various advances and courses.



ACTIVITY DIAGRAM

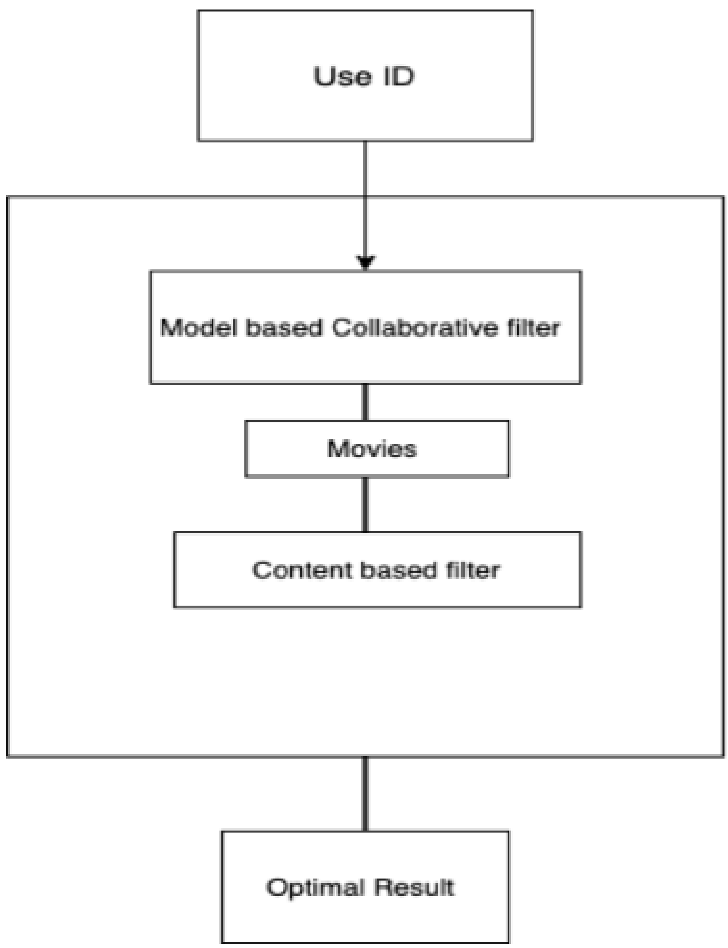
Movement chart is characterized as an UML outline that spotlights on the execution and stream of the conduct of a framework rather than execution. It is likewise called object-situated flowchart. Movement graphs comprise of exercises that are comprised of activities which apply to social displaying innovation.

- First and foremost, need to enter the client certifications.
- On the off chance that another client needs to enlist.
- Then, at that point, select the pertinent classification and the motion pictures are suggested.
- Distinguish applicant use cases, through the assessment of business work processes.
- Distinguish pre-and post-conditions (the unique circumstance) for use cases.
- Model work processes between/inside use cases.
- Model complex work processes in procedure on objects.
- Model exhaustively complex exercises in a general movement Diagram
- For this the framework utilizes the deduced calculation, and along these lines the cycle proceeds and consistently another proposal list is created.


Enter Login Information
Check if Registered User
Checking User Database
New User
Registered User
Select movie
Check and provide list of movies to customer
See Actor and Cast of movie
Add to movie watched list
Movie Watched list Updated
Update recommendations
Recommendations List Updated

SYSTEM DESIGN

For each unique individual utilize distinctive rundown of motion pictures are suggested ,as client login or enters the client id dependent on two unique methodologies utilized in the venture each will prescribe the arrangement of films to the specific client by consolidating the both the arrangement of film dependent on the client the half and half model will prescribe the single rundown. .



IMPLEMENTATION AND TESTING

Python Code:

```
import pandas as pd
import numpy as np
column_names = ['user_id', 'item_id', 'rating', 'timestamp']
df = pd.read_csv('/content/u.data', sep='\t', names=column_names)
movie_titles = pd.read_csv("/content/Movie_Id_Titles")
movie_titles.head()
df = pd.merge(df,movie_titles,on='item_id')
df.head()
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style('white')
%matplotlib inline
df.groupby('title')['rating'].mean().sort_values(ascending=False).head()
df.groupby('title')['rating'].count().sort_values(ascending=False).head()
ratings = pd.DataFrame(df.groupby('title')['rating'].mean())
ratings.head()
ratings['num of ratings'] = pd.DataFrame(df.groupby('title')['rating'].count())
ratings.head()
plt.figure(figsize=(10,4))
ratings['num of ratings'].hist(bins=70)
plt.figure(figsize=(10,4))
ratings['rating'].hist(bins=70)
sns.jointplot(x='rating',y='num of ratings',data=ratings,alpha=0.5)
```

```
moviemat = df.pivot_table(index='user_id',columns='title',values='rating')
moviemat.head()
ratings.sort_values('num of ratings',ascending=False).head(10)
ratings.head()
starwars_user_ratings = moviemat['Star Wars (1977)']
liarliar_user_ratings = moviemat['Liar Liar (1997)']
starwars_user_ratings.head()
similar_to_starwars = moviemat.corrwith(starwars_user_ratings)
similar_to_liarliar = moviemat.corrwith(liarliar_user_ratings)
corr_starwars = pd.DataFrame(similar_to_starwars,columns=['Correlation'])
corr_starwars.dropna(inplace=True)
corr_starwars.head()
corr_starwars.sort_values('Correlation',ascending=False).head(10)
corr_starwars = corr_starwars.join(ratings['num of ratings'])
corr_starwars.head()
corr_starwars[corr_starwars['num of ratings']>100].
sort_values('Correlation',ascending=False).head()
corr_liarliar = pd.DataFrame(similar_to_liarliar,columns=['Correlation'])
corr_liarliar.dropna(inplace=True)
corr_liarliar = corr_liarliar.join(ratings['num of ratings'])
corr_liarliar[corr_liarliar['num of ratings']>100].
sort_values('Correlation',ascending=False).head()
```

OUTPUT

```
import pandas as pd
import numpy as np
```

```
[ ] column_names = ['user_id', 'item_id', 'rating', 'timestamp']
    df = pd.read_csv('/content/u.data', sep='\t', names=column_names)
```

```
[1] df.head()
```

	user_id	item_id	rating	timestamp
0	0	50	5	881250949
1	0	172	5	881250949
2	0	133	1	881250949
3	196	242	3	881250949
4	186	302	3	891717742

```
[ ] movie_titles = pd.read_csv("/content/Movie_Id_Titles")
    movie_titles.head()
```

	item_id	title
0	1	Toy Story (1995)
1	2	GoldenEye (1995)
2	3	Four Rooms (1995)
3	4	Get Shorty (1995)
4	5	Copycat (1995)

```
[ ] df = pd.merge(df,movie_titles,on='item_id')
    df.head()
```

	user_id	item_id	rating	timestamp	title
0	0	50	5	881250949	Star Wars (1977)
1	290	50	5	880473582	Star Wars (1977)
2	79	50	4	891271545	Star Wars (1977)
3	2	50	5	888552084	Star Wars (1977)
4	8	50	5	879362124	Star Wars (1977)

```
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style('white')
%matplotlib inline
```

```
df.groupby('title')['rating'].mean().sort_values(ascending=False).head()
```

```
title
Marlene Dietrich: Shadow and Light (1996)     5.0
Prefontaine (1997)                            5.0
Santa with Muscles (1996)                     5.0
Star Kid (1997)                               5.0
Someone Else's America (1995)                 5.0
Name: rating, dtype: float64
```

```
df.groupby('title')['rating'].count().sort_values(ascending=False).head()
```

```
title
Star Wars (1977)              584
Contact (1997)                509
Fargo (1996)                  508
Return of the Jedi (1983)     507
Liar Liar (1997)              485
Name: rating, dtype: int64
```

```
ratings = pd.DataFrame(df.groupby('title')['rating'].mean())
ratings.head()
```

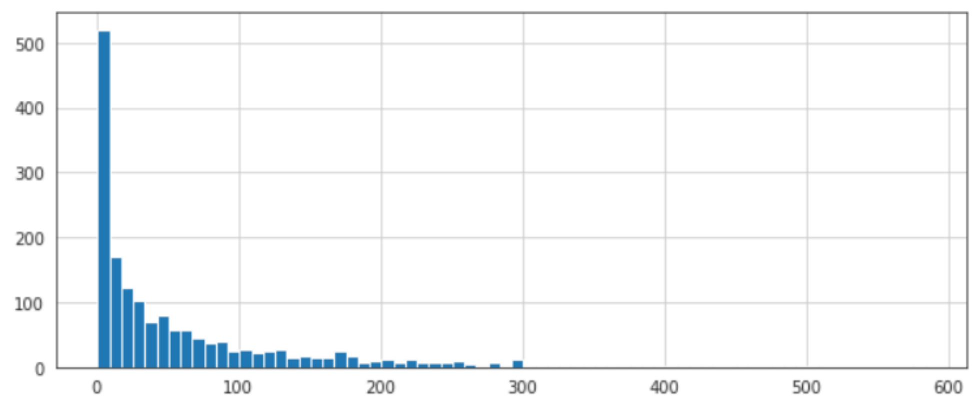
	rating
title	
'Til There Was You (1997)	2.333333
1-900 (1994)	2.600000
101 Dalmatians (1996)	2.908257
12 Angry Men (1957)	4.344000
187 (1997)	3.024390

```
[ ] ratings['num of ratings'] = pd.DataFrame(df.groupby('title')['rating'].count())
    ratings.head()
```

title	rating	num of ratings
'Til There Was You (1997)	2.333333	9
1-900 (1994)	2.600000	5
101 Dalmatians (1996)	2.908257	109
12 Angry Men (1957)	4.344000	125
187 (1997)	3.024390	41

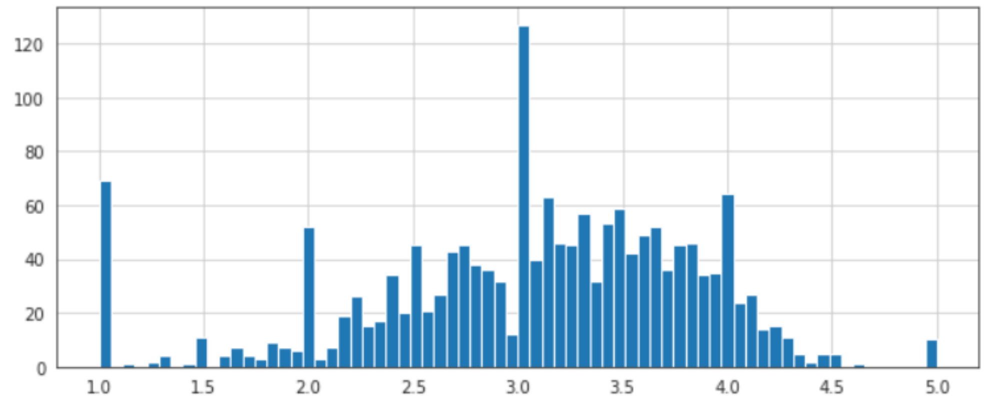
```
[ ] plt.figure(figsize=(10,4))
    ratings['num of ratings'].hist(bins=70)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb81ff0fb10>
```



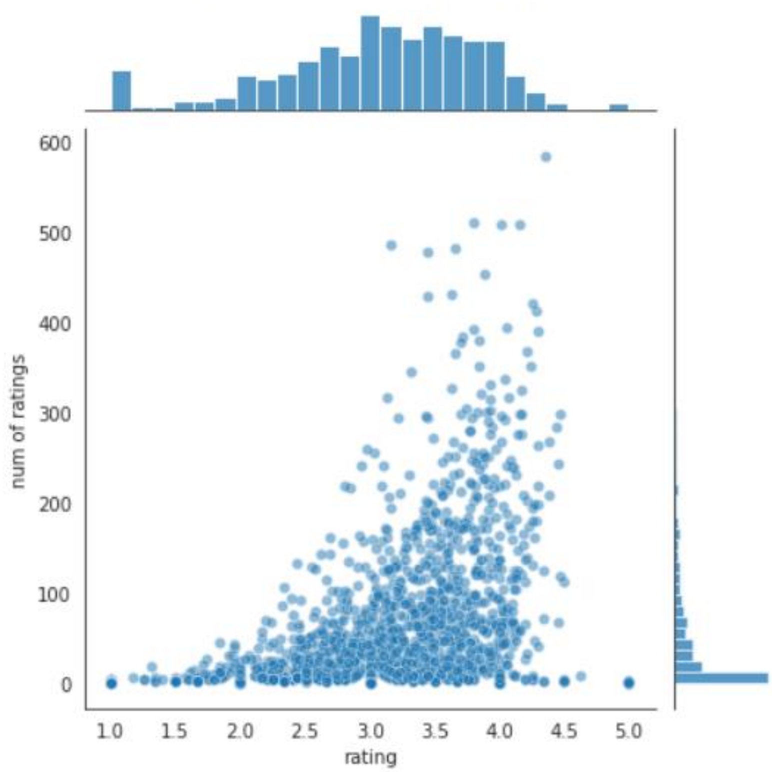
```
[ ] plt.figure(figsize=(10,4))
    ratings['rating'].hist(bins=70)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb81feb8790>
```




```
[ ] sns.jointplot(x='rating',y='num of ratings',data=ratings,alpha=0.5)
```

```
<seaborn.axisgrid.JointGrid at 0x7fb81fd18a90>
```



```
[ ] moviemat = df.pivot_table(index='user_id',columns='title',values='rating')
    moviemat.head()
```

title	'Til There Was You (1997)	1-900 (1994)	101 Dalmatians (1996)	12 Angry Men (1957)	187 (1997)	2 Days in the Valley (1996)	20,000 Leagues Under the Sea (1954)	2001: A Space Odyssey (1968)	3 Ninjas: High Noon At Mega Mountain (1998)	39 Steps, The (1935)	8 1/2 (1963)	8 Heads in a Duffel Bag (1997)	8 Seconds (1994)	A Chef in Love (1996)	Above the Rim (1994)	Absolute Power (1997)	Abyss, The (1989)	Ace Ventura: Pet Detective (1994)	Ace Ventura: When Nature Calls (1995)
user_id																			
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	2.0	5.0	NaN	NaN	3.0	4.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	3.0	3.0	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	NaN	NaN	NaN	NaN	NaN	3.0	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

5 rows × 1664 columns

```
[ ] ratings.sort_values('num of ratings',ascending=False).head(10)
```

title	rating	num of ratings
Star Wars (1977)	4.359589	584
Contact (1997)	3.803536	509
Fargo (1996)	4.155512	508
Return of the Jedi (1983)	4.007890	507
Liar Liar (1997)	3.156701	485
English Patient, The (1996)	3.656965	481
Scream (1996)	3.441423	478
Toy Story (1995)	3.878319	452
Air Force One (1997)	3.631090	431
Independence Day (ID4) (1996)	3.438228	429

```
[ ] ratings.head()
```

title	rating	num of ratings
'Til There Was You (1997)	2.333333	9
1-900 (1994)	2.600000	5
101 Dalmatians (1996)	2.908257	109
12 Angry Men (1957)	4.344000	125
187 (1997)	3.024390	41

```
[ ] starwars_user_ratings = moviemat['Star Wars (1977)']
    liarliar_user_ratings = moviemat['Liar Liar (1997)']
    starwars_user_ratings.head()
```

```
user_id
0    5.0
1    5.0
2    5.0
3    NaN
4    5.0
Name: Star Wars (1977), dtype: float64
```

```
[19] corr_starwars = pd.DataFrame(similar_to_starwars,columns=['Correlation'])
     corr_starwars.dropna(inplace=True)
     corr_starwars.head()
```

title	Correlation
'Til There Was You (1997)	0.872872
1-900 (1994)	-0.645497
101 Dalmatians (1996)	0.211132
12 Angry Men (1957)	0.184289
187 (1997)	0.027398

```
[20] corr_starwars.sort_values('Correlation',ascending=False).head(10)
```

title	Correlation
Hollow Reed (1996)	1.0
Commandments (1997)	1.0
Cosi (1996)	1.0
No Escape (1994)	1.0
Stripes (1981)	1.0
Star Wars (1977)	1.0
Man of the Year (1995)	1.0
Beans of Egypt, Maine, The (1994)	1.0
Old Lady Who Walked in the Sea, The (Vieille qui marchait dans la mer, La) (1991)	1.0
Outlaw, The (1943)	1.0

```
[21] corr_starwars = corr_starwars.join(ratings['num of ratings'])
     corr_starwars.head()
```

title	Correlation	num of ratings
'Til There Was You (1997)	0.872872	9
1-900 (1994)	-0.645497	5
101 Dalmatians (1996)	0.211132	109
12 Angry Men (1957)	0.184289	125
187 (1997)	0.027398	41

```
[22] corr_starwars[corr_starwars['num of ratings']>100].sort_values('Correlation',ascending=False).head()
```

title	Correlation	num of ratings
Star Wars (1977)	1.000000	584
Empire Strikes Back, The (1980)	0.748353	368
Return of the Jedi (1983)	0.672556	507
Raiders of the Lost Ark (1981)	0.536117	420
Austin Powers: International Man of Mystery (1997)	0.377433	130

```
[23] corr_liarliar = pd.DataFrame(similar_to_liarliar,columns=['Correlation'])
     corr_liarliar.dropna(inplace=True)
     corr_liarliar = corr_liarliar.join(ratings['num of ratings'])
     corr_liarliar[corr_liarliar['num of ratings']>100].sort_values('Correlation',ascending=False).head()
```

title	Correlation	num of ratings
Liar Liar (1997)	1.000000	485
Batman Forever (1995)	0.516968	114
Mask, The (1994)	0.484650	129
Down Periscope (1996)	0.472681	101
Con Air (1997)	0.469828	137

CONCLUSION

In this venture, to work on the precision, quality and adaptability of film suggestion framework, a Hybrid methodology by bringing together substance based separating and community sifting; utilizing Singular Value Decomposition (SVD) as a classifier and Cosine Similarity is introduced in the proposed approach. Existing unadulterated methodologies and proposed half and half methodology is carried out on three distinctive Movie datasets and the outcomes are looked at among them. Relative outcomes portrays that the proposed approach shows an improvement in the precision, quality and versatility of the film suggestion framework than the unadulterated methodologies. Additionally, registering season of the proposed approach is lesser than the other two unadulterated methodologies.

Future Scope of the Project

In the proposed approach, It has thought about Genres of films at the same time, in future we can likewise consider period of client as per the age film inclinations additionally changes, as for instance, during our youth we like enlivened motion pictures more when contrasted with different motion pictures. There is a need to chip away at the memory necessities of the proposed approach later on. The proposed approach has been carried out here on various film datasets as it were. It can likewise be carried out on the Film Affinity and Netflix datasets and the presentation can be processed later on.

REFERENCES

[1] C.Aggarwal, Recommender Systems. Springer International Publishing, 2016.

[2] D. Liben-Nowell, and J. Kleinberg, "The link-prediction problem for social networks", Journal of the American Society for Information Science and Technology, Vol. 58, No. 7, pp. 1019-1031, 2007.

[3] Z. Wu, and Y. Li, "Link Prediction Based on Multi-steps Resource Allocation", in Proceedings of the 2014

[4] G. Adomavicius, and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-ofthe-art and possible extensions," IEEE transactions on knowledge and data engineering, 17(6), pp.734-749, 2005.

[5] T. Zhou, L. Lü, and Y. Zhang, "Predicting missing links via local information," The European Physical Journal B, Vol. 71, No. 4, pp. 623-630, 2009.

[6] A. Javari, J. Gharibshah, and M. Jalili, "Recommender systems based on collaborative filtering and resource allocation," Social Network Analysis and Mining, Vol. 4, No. 1, pp. 234, 2014.

[7] M. Roth, A. Ben-David, D. Deutscher, G. Flysher, I. Horn, A. Leichtberg, and R. Merom, "Suggesting friends using the implicit social graph," in Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM: Washington, DC, USA., pp. 233-242, 2010.

[8] M. Li, BM. Dias, I. Jarman, W. El-Deredy, and P.J. Lisboa, "Grocery shopping recommendations based on basketsensitive random walk," in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM: Paris, France, pp. 1215-1224, 2009.

[9] H. Ma, and H. Liu, "Design of Clothing Clustering Recommendation System on SOM Neural Network," In 8th International Conference on Social Network, Communication and Education, Atlantis Press, 2018.

[10] H. Ma, H. Yang, M.R. Lyu, and I. King, "SoRec: social recommendation using probabilistic matrix factorization," in Proceedings of the 17th ACM conference on Information and knowledge management, ACM: Napa Valley, California, USA., pp. 931-940, 2008.

[11] A. Aleem, and A. Kumar, and M. M. Gore, “A Study of Manuscripts Evolution to Perfection”. In Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE-2019), pp:278- 282, 2019