MOVIE RECOMMENDATION SYSTEM USING COLLABORATIVE FILTERING

A project synopsis submitted in fulfilment of the Academic requirements
for the award of the Degree of

BACHELOR OF ENGINEERING
IN
INFORMATION TECHNOLOGY

By

KONDURU SAI KIRAN   (2451-19-737-011)
CHIPPA HARIKA   (2451-19-737-013)
THATIPELLI SAI KIRAN   (2451-19-737-049)

Under the guidance of
Dr.A.V.Krishna Prasad,
Associate Professor, Dept. of I.T
Maturi Venkata Subba Rao(MVSR) Engineering College
Nadergul, Hyderabad.

DEPARTMENT OF INFORMATION TECHNOLOGY
MATURI VENKATA SUBBA RAO (MVSR) ENGINEERING COLLEGE
(An Autonomous Institution)
(Affiliated to Osmania University, Hyderabad. Recognized by AICTE)
Nadergul, Saroornagar Mandal, Hyderabad-501510
2022-23

i

MATURI VENKATA SUBBA RAO (MVSR) ENGINEERING COLLEGE

(An Autonomous Institution)
(Affiliated to Osmania University, Hyderabad. Recognized by AICTE)
Nadergul, Saroornagar Mandal, Hyderabad-501510

DEPARTMENT OF INFORMATION TECHNOLOGY

CERTIFICATE

This is to certify that the project work entitled "MOVIE RECOMMENDATION SYSTEM USING COLLABORATIVE FILTERING" is a bonafide work carried out by **K.Sai Kiran (2451-19-737-011), Ch.Harika (2451-19-737-013), T.Sai Kiran (2451-19-737-049)** in partial fulfilment of the requirements for the award of degree of **Bachelor of Engineering** in **Information Technology** from **Maturi Venkata Subba Rao Engineering College,** affiliated to OSMANIA UNIVERSITY, Hyderabad, during the Academic Year 2022-23. under our guidance and supervision.

The results embodied in this report have not been submitted to any other university or institute for the award of any degree or diploma.

Signature of the Coordinator   Signature of Guide

Signature of Head, ITD   Signature of External Examiner

DECLARATION

We hereby declare that the contents presented in the Project Thesis titled **"MOVIE RECOMMENDATION SYSTEM USING COLLABORATIVE FILTERING"** submitted in partial fulfillment for the award of Degree of Bachelor of Engineering *in INFORMATION TECHNOLOGY (IT), MATURI VENKATA SUBBA RAO (MVSR) ENGINEERING COLLEGE* affiliated to *OSMANIA UNIVERSITY*, *Hyderabad* is a record of the original work carried out by us under the supervision Of **Dr. A.V. Krishna Prasad.** Further this is to state that the results embodied in this project report have not been submitted to any University or Institution for the award of any Degree.

Signature of the Student | Signature of the Student | Signature of the Student
K.SAI KIRAN | **CH. HARIKA** | **T.SAI KIRAN**
2451-19-737-011 | **2451-19-737-013** | **2451-19-737-49**

ACKNOWLEDGEMENT

Vision & Mission

MVSR Engineering College Department of Information Technology

COURSE NAME: MINI PROJECT
WORK IICOURSE CODE: PW961IT

VISION

To impart technical education to produce competent and socially responsible engineers in the field of Information Technology.

MISSION

M1. To make teaching learning process effective and stimulating.

M2. To provide adequate fundamental knowledge of sciences and Information Technology with positive attitude.

M3. To create an environment that enhances skills and technologies required for industry.

M4. To encourage creativity and innovation for solving real world problems.

M5. To cultivate professional ethics in students and inculcate a sense of responsibility towards society.

PROGRAM EDUCATIONAL OBJECTIVES (PEOs)

The Bachelor's program in Information Technology is aimed at preparing graduates who will:

I. Apply knowledge of mathematics and Information Technology to analyze, design and implement solutions for real world problems in core or in multidisciplinary areas.
II. Communicate effectively, work in a team, practice professional ethics and apply knowledge of computing technologies for societal development.
III. Engage in Professional development or postgraduate education to be a life-long learner.

PROGRAM OUTCOMES (POs)

1. Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
4. Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesisof the information to provide valid conclusions.
5. Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.
6. The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
7. Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
11. Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
12. Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PROGRAM SPECIFIC OUTCOMES (PSOS):

(1) Hardware design: An ability to analyze, design, simulate and implement computer hardware / software and use basic analog/digital circuits, VLSI design for various computing and communication system applications.

(2) Software design: An ability to analyze a problem, design algorithm, identify and define the computing requirements appropriate to its solution and implement the same.

COURSE OBJECTIVES:

1. To enhance practical & Professional skills.
2. To familiarize the tools and techniques of symmetric literature survey and documentation.
3. To expose students to industry practices and teamwork.
4. To encourage students to work with innovative and entrepreneurial ideas.

COURSE OUTCOMES:

On successful completion of this course students will be able to:
1. Define a problem of the recent advancements with applications towards society.
2. Outline requirements and perform requirement analysis for solving the problem.
3. Design and develop a software and/or hardware based solution within the scope of project using contemporary technologies and tools.
4. Test and deploy the applications for use.
5. Develop the Project as a team and demonstrate the application, with effective written and oral communications.

ABSTRACT

Movies are one of the sources of entertainment, but the problem is in finding the desired content from the ever-increasing millions of content every year. However, recommendation systems come much handier in these situations. The aim of this paper is to improve the accuracy and performance of a regular filtering technique. Although varieties of methods are used to implement a recommendation system, Content-based filtering is the simplest method. Which takes input from the users, rechecks his/her history/past behavior, and recommends a list of similar movies. In this paper, to prove the effectiveness, K-NN algorithms and collaborative filtering are used to mainly focus on enhancing the accuracy of results as compared to content-based filtering. This approach is based on cosine similarity using k-nearest neighbor with the help of a collaborative filtering technique, at the same time removing the drawbacks of the content-based filtering. Although using Euclidean distance is preferred, cosine similarity is used as the accuracy of cosine angle and the equidistance of movies remain almost the same.

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

1. INTRODUCTION

There are many different types of recommender systems to recommend hotels,clothing, food, songs, movies and so on. They all work on different scenarios but the main concept behind them all are same. We made movie recommender system using collaborative system.

Collaborative Filtering- This system matches persons with similar interests and provides recommendations based on this matching.

This is achieved by making a matrix of items each user has rated/viewed/liked/clicked depending upon the task at hand, and then computing the similarity score between the users and finally recommending items that the concerned user isn't aware of butusers similar to him/her are and liked it.

1.1. PROBLEM STATEMENT

Recommendation systems are often implemented as a service applicable to every user. This study will analyze the advantages of these recommender systems and discuss whether the resources spent give a satisfactory outcome for companies. There are several machine learning algorithms that can be used for this purpose. To delimit this report, it will focus on the K nearest neighbors algorithm. KNN is one of the more intuitive algorithms for the purpose of movie rating prediction, which made it interesting to analyze its' usefulness. A baseline method of using a movies' mean rating as an estimate for all users' rating predictions will be used for comparison.

1.2. OBJECTIVES

The objectives of this project are:
- To list out the similar type of movies.
- Take the movie rating from the user if user has watched the movie .

1.3. MOTIVATION

Over the past years, the internet has broadened the horizon of various domains to interact and share meaningful information. As it is said that everything has its pros and cons therefore, along with the expansion of domain comes information overload and difficulty in extraction of data especially the movies. Movies has become a medium to connect all people of the world. Hence we choose movie recommendation system to provide accurate recommendations.

1.4. EXISTING SYSTEM

The most popular recommendation systems now-a-days are:
- Amazon(e-commerce): It uses AI and ML algorithms for Hybrid recommendation system.
- Netflix(entertainment): It uses Content-based filtering for personalised   recommendations.
- Spotify: It uses Item based and User based filtering for songs recommendations
- IMDB: It uses user based collaborative filtering for movies

Drawbacks:

- Some of the collabarative filtering has a cold start problem.
- Doesn't work large set of data sets.

1.5.PROPOSED SYSTEM

To recommend movies to the user. To recommend the movies in the given range of the years by the user. Construction of neighborhood information.User can also rate the movie if he watched the movie.

ADVANTAGES:

It only recommends what the user is interested in watching. And here we have the target is labelled data. Unlike K-Means clustering algorithm, This works on similarity measure calculated using cosine similarity and comparatively gives more accurate result.

1.6. SCOPE AND PURPOSE

This report will focus on implementing a method to efficiently predict users' movie ratings. The method investigated will be the KNN algorithm . All data for this report will be taken from the public source MovieLens from Kaggle.

2. LITERATURE SURVEY

Mhetre, R. et al. focused on movie recommendation and the technique currently present for this is the collaborative filtering technique. Of the collaborative filtering techniques, the matrix factorization algorithms namely Alternative Least Square and Singular Value Decomposition are implemented to predict or recommend the movies. Further to improve the processing and time computation for a large dataset we have used Apache Spark along with Elastic search and the accuracy is compared between the two algorithms for different values of testing subsets.

Bei-Bei Cui : In the spread of information, how to quickly find one's favorite movie in a large number of movies become a very important issue. Personalized recommendation system can play an important role especially when the user has no clear target movie. In this paper, we design and implement a movie recommendation system prototype combined with the actual needs of movie recommendation through researching of KNN algorithm and collaborative filtering algorithm. Then we give a detailed principle and architecture of JAVAEE system relational database model. Finally, the test results showed that the system has a good recommendation effect.

Nagamanjula, R. and Pethalakshmi, A. proposed a movie recommender system based on new user similarity metric and opinion mining. The primary objective of this paper is to find the type of opinions (positive, negative or neutral) for movies and also suggest top-k recommendation list for users and extract aspect-based specific ratings from reviews and also recommend reviews to users depends on user similarity and their rating patterns. Finally, validating the proposed movie recommendation system for various evaluation criteria, and also the proposed system shows better result than conventional systems

LITERATURE SURVEY:

S.No	Author(s)	Problem Identified	Techonology Used	Drawbacks
1.	Wilkinson and Schreiber	data sparsity	a collaborative filtering algorithm based on the ALS	Cold start problem
2.	Wei	cold start problem	deep learning neural network	Low scalability
3.	Kupisz and Unold	Low processing ability of large-scale data	item-based collaborative filtering algorithm Hadoop's disk-based MapReduce para-digm and Spark's in-memory based RDD paradigm	Cannot provide accurate personalized results.

3. SYSTEM REQUIREMENT SPECIFICATIONS

3.1 SOFTWARE REQUIREMENTS:

- Operating System : Windows/Linux
- Programming Language : Python
- Software : PyCharm

3.2 HARDWARE REQUIREMENTS:

- Ram : 4GB and above
- Processor : Intel i3 and above or Ryzen 3+ Series
- Hard disk Space : Minimum 5GB of Space Required

4. TECHNOLOGIES USED

Python:

Python is an interpreted high-level general-purpose programming language. Its design philosophy emphasizes code readability with its use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented and functional programming. It is often described as a "batteries included" language due to its comprehensive standard library.

Guido van Rossum began working on Python in the late 1980s, as a successor to the ABC programming language, and first released it in 1991 as Python 0.9.0.Python 2.0 was released in 2000 and introduced new features, such as list comprehensions and a cycle-detecting garbage collection system (in addition to reference counting). Python 3.0 was released in 2008 and was a major revision of the language that is not completely backward-compatible. Python 2 was discontinued with version 2.7.18 in 2020. Python consistently ranks as one of the most popular programming languages.
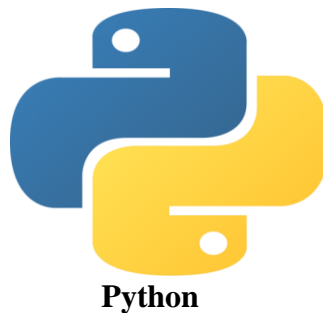
Python

PyCharm:

PyCharm is the most popular IDE used for the Python scripting language. This chapter will give you an introduction to PyCharm and explain its features.

PyCharm offers some of the best features to its users and developers in the following aspects-

- Code completion and inspection
- Advanced debugging
- Support for web programming and frameworks such as Django and Flask

Features of PyCharm

- Code Completion
- SQLAlchemy as Debugger
- Git Visualization in Editor
- Code Coverage in Editor
- Package Management
- Local History
- Refactoring

User Interface of PyCharm Editor
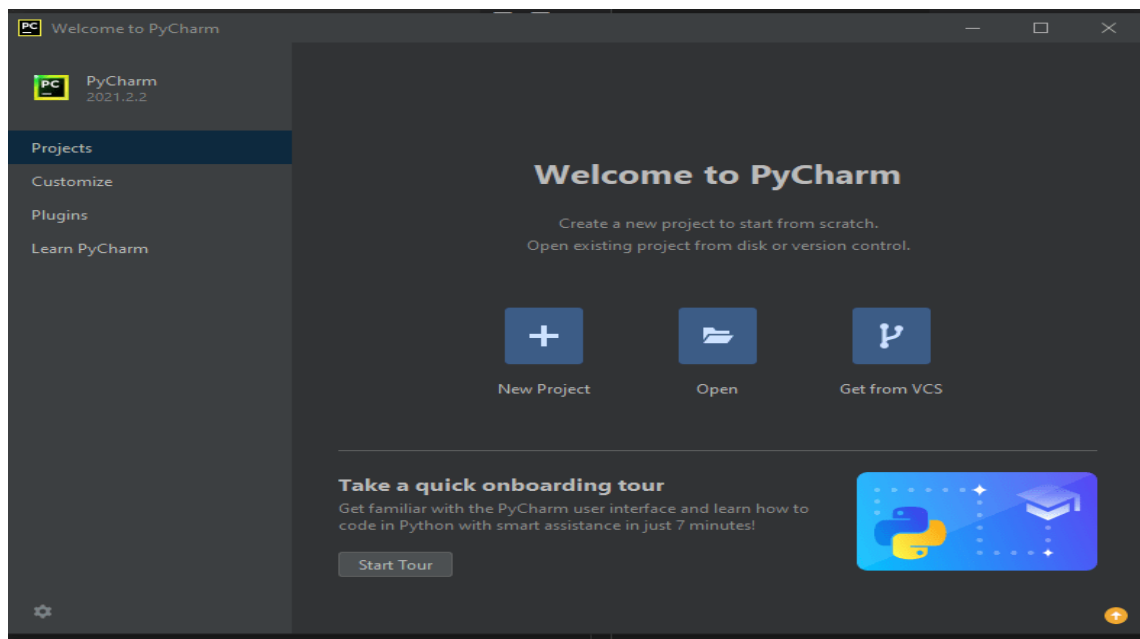
User Interface of PyCharm Editor

5.SYSTEM DESIGN

5.1 DIAGRAMS APPLICABLE

UML DIAGRAMS:

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML comprises two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software systems, as well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

GOALS:

The Primary goals in the design of the UML are as follows:

1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development processes.
4. Provide a formal basis for understanding the modeling language.
5. Encourage the growth of the OO tools market.
6. Integrate best practices.

USE CASE DIAGRAM:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.
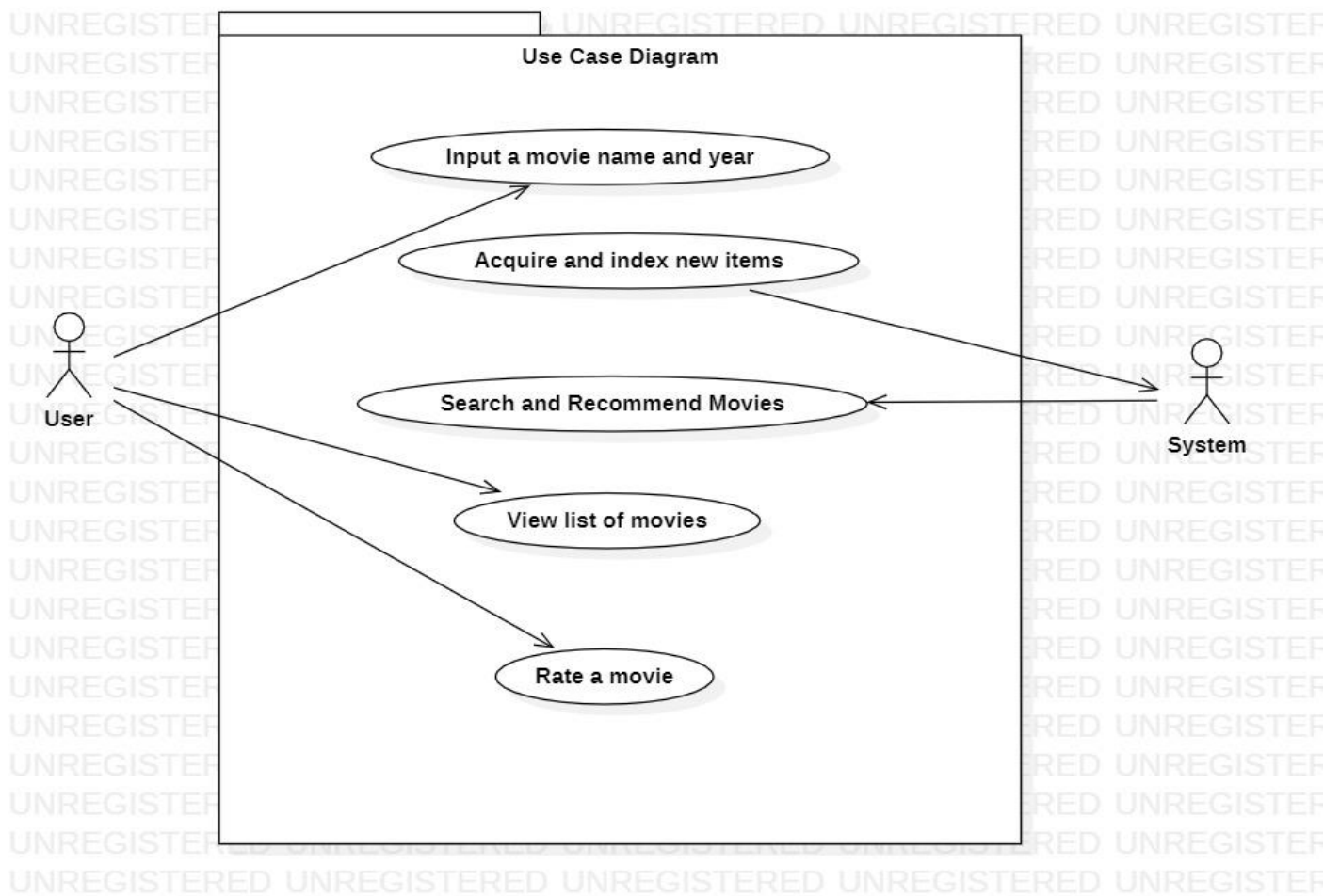

Fig.2.Use Case Diagram

CLASS DIAGRAM:

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.
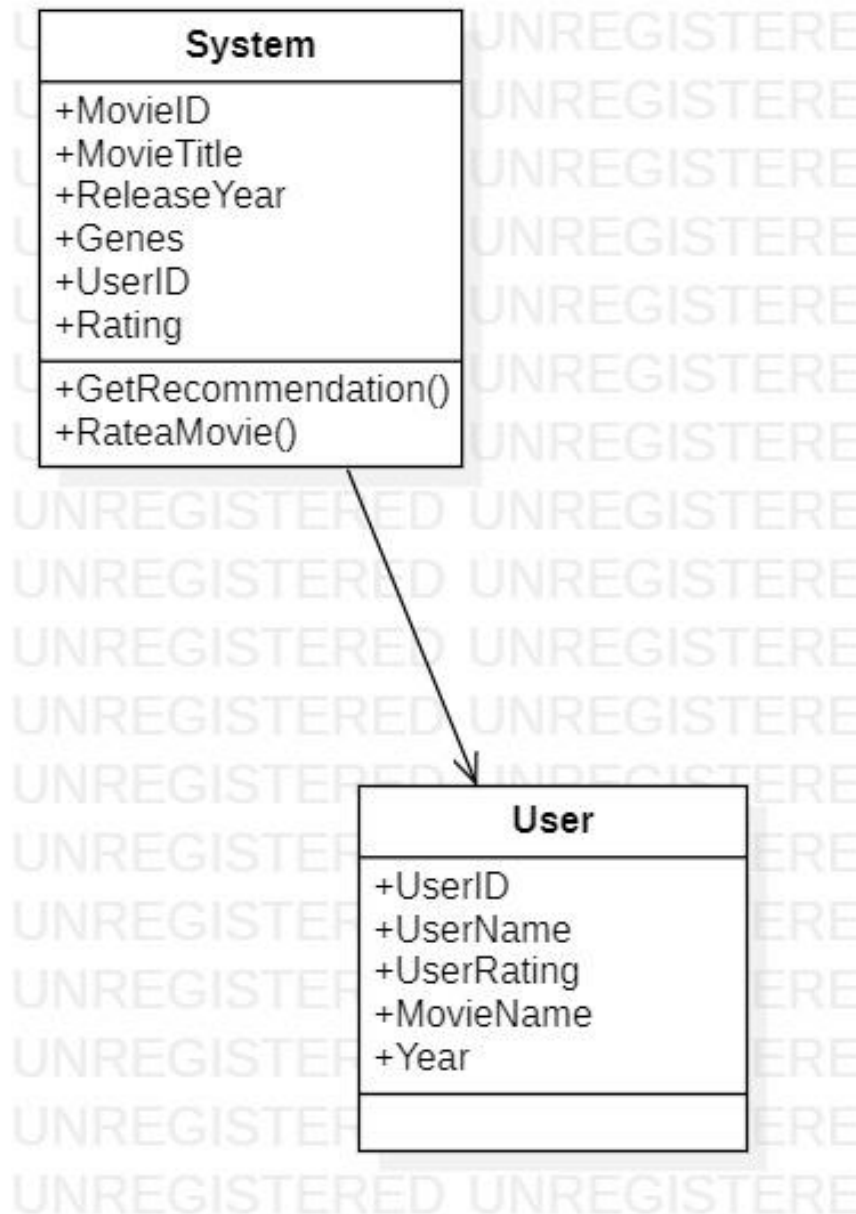

Fig 3:Class Diagram

SEQUENCE DIAGRAM:

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.
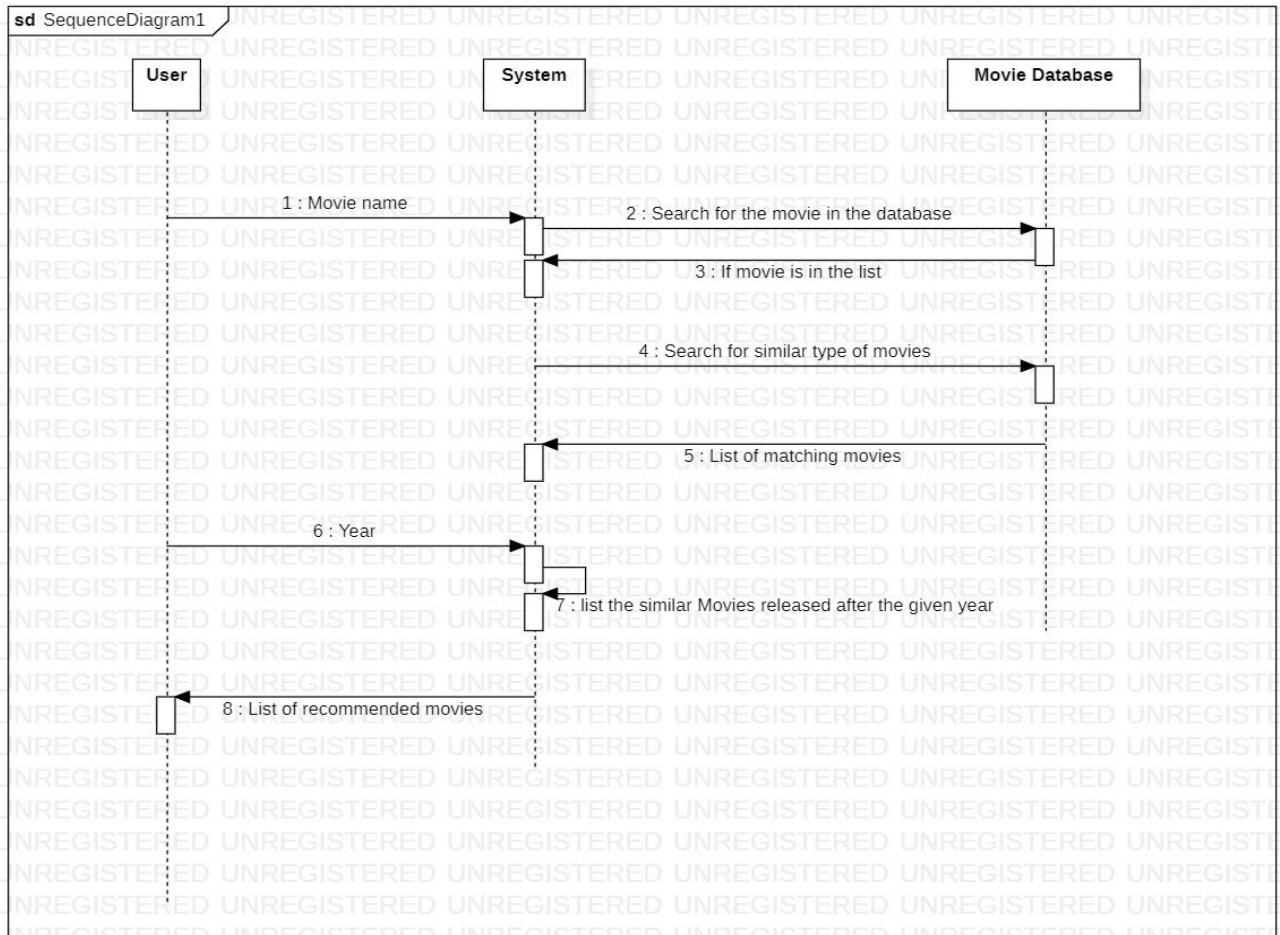

Fig 4:Sequence Diagram

ACTIVITY DIAGRAM:

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.
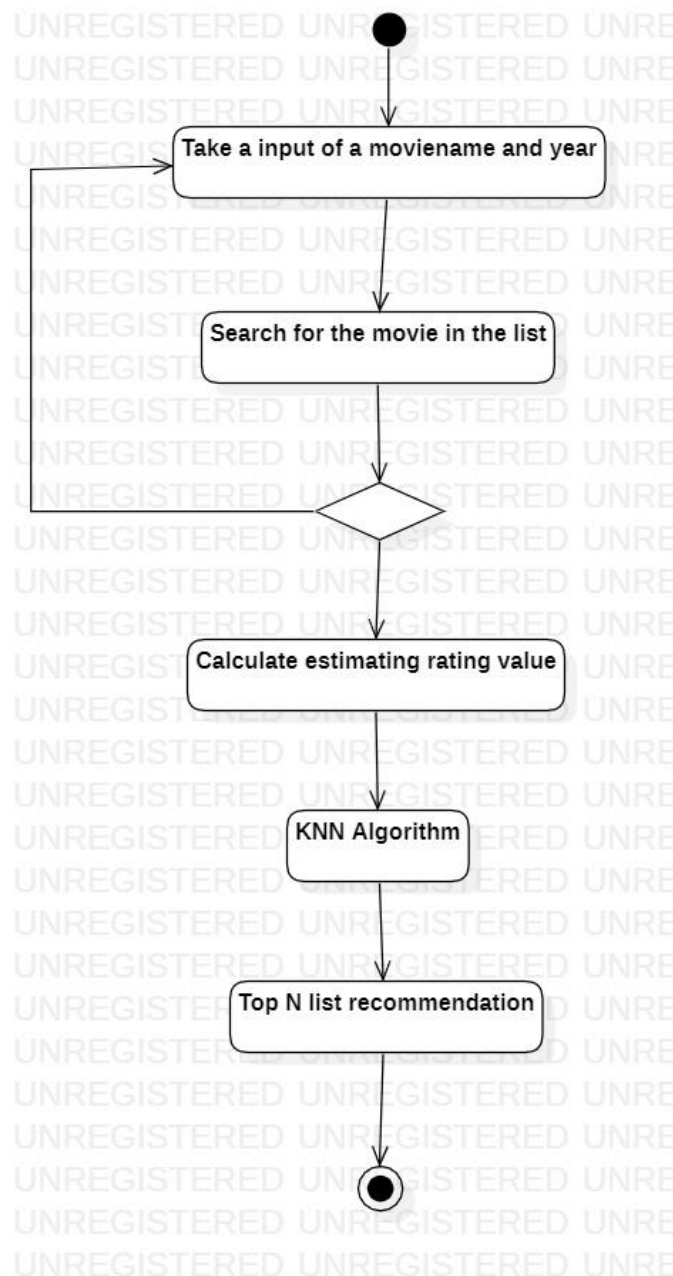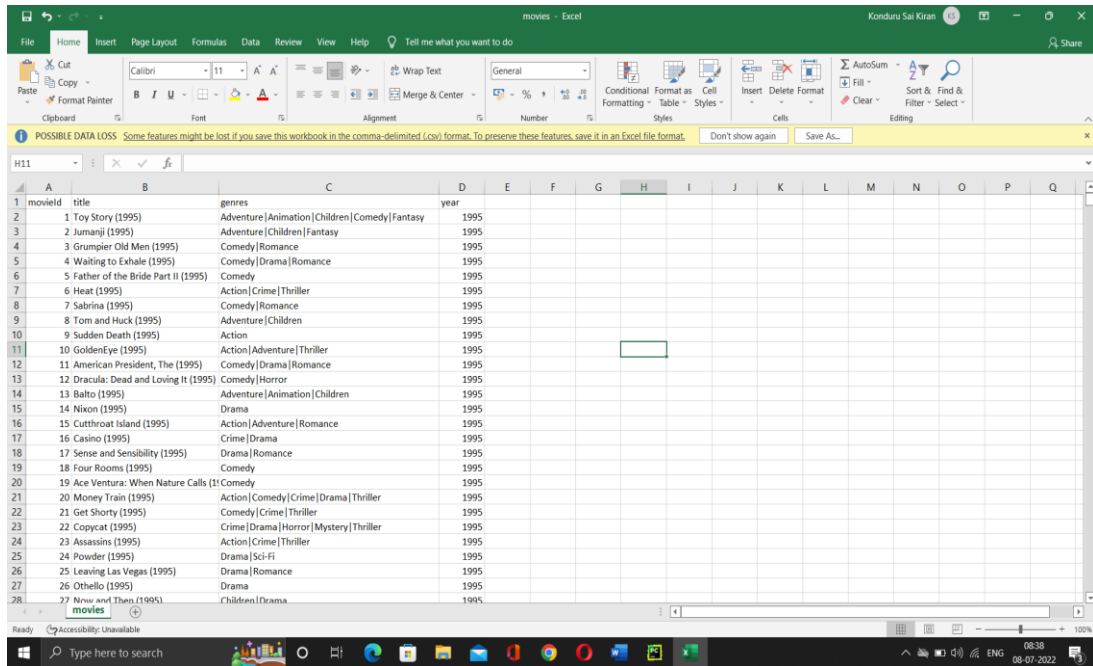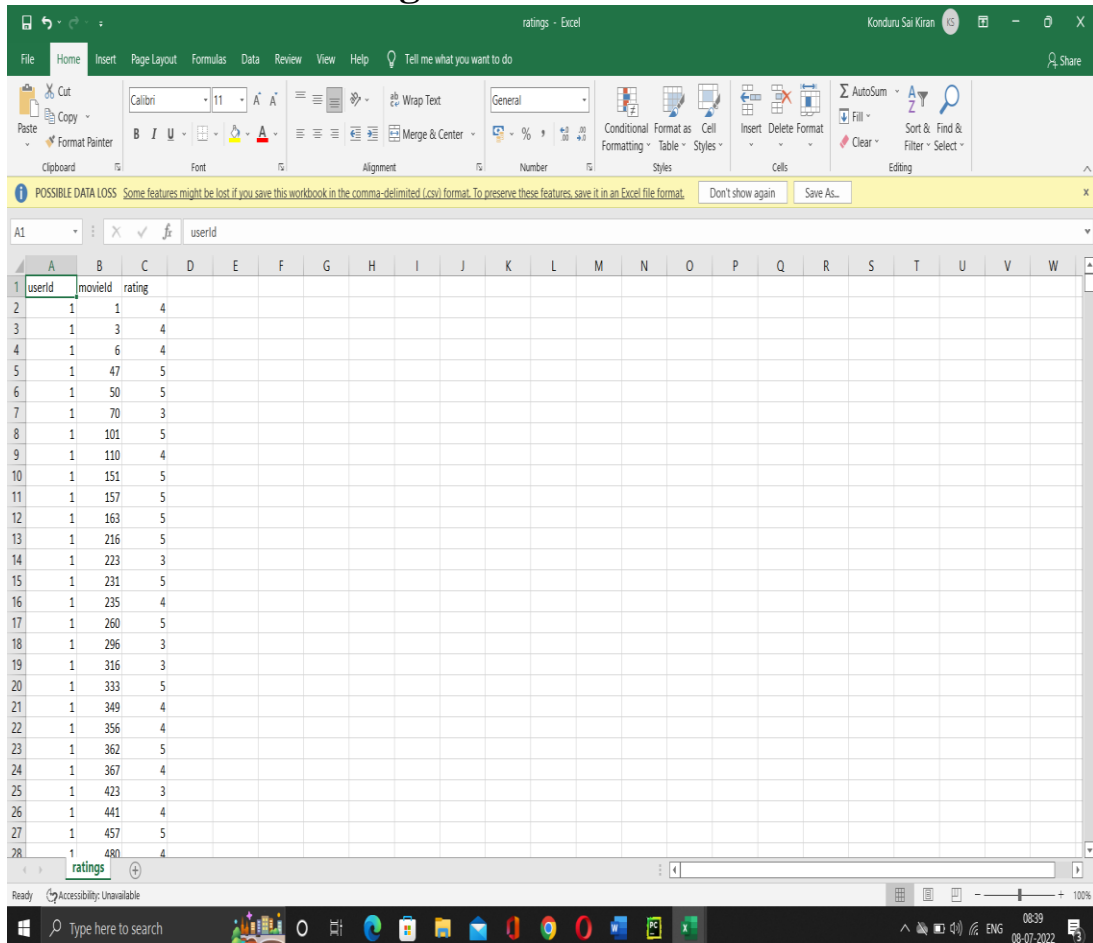

Fig 5:Activity Diagram

DATASET

Fig 6.Movie Data Set

Fig 7.User Data Set

6.IMPLEMENTATION

6.1. ENVIRONMENTAL SETUP

Installing Python:

1. To download and install Python visit the official website of Python https://www.python.org/downloads/ and choose your version.
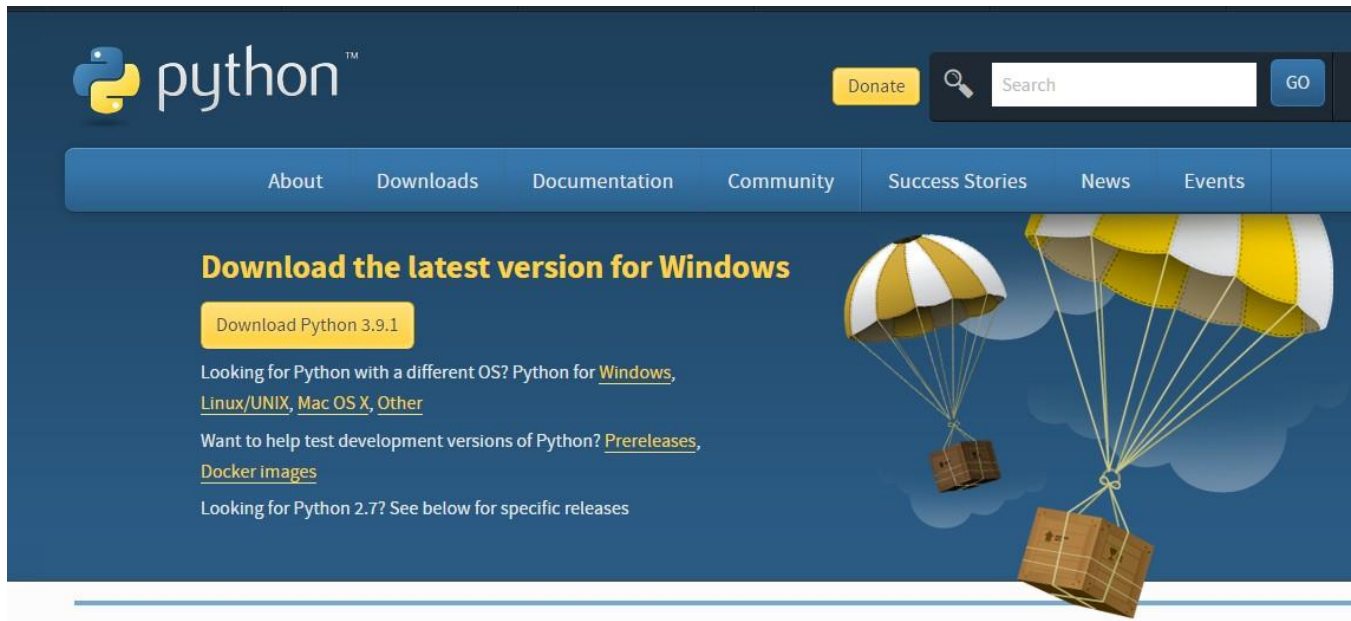
FIG-8 PYTHON INSTALLATION

2. Once the download is complete, run the exe for install Python. Now click on InstallNow.

3. You can see Python installing at this point.

4. When it finishes, you can see a screen that says the Setup was successful. Now clickon "Close".

Installing PyCharm:

1. To download PyCharm visit the
   website https://www.jetbrains.com/pycharm/download/

and Click the"DOWNLOAD" link under the Community

Section.

Download PyCharm

Windows   Mac   Linux

Professional

For both Scientific and Web Python
development. With HTML, JS, and SQL
support.

Download

Free trial

Community

For pure Python development

Download

Free, open-source

FIG-9 PYCHARM DOWNLOAD

2. Once the download is complete, run the exe for install PyCharm. The setup
   wizardshould have started. Click "Next".

3. On the next screen, Change the installation path if required. Click "Next".

4. On the next screen, you can create a desktop shortcut if you want and click
   on"Next".

5. Choose the start menu folder. Keep selected JetBrains and click on "Install".

6. Wait for the installation to finish.

7. Once installation finished, you should receive a message screen that PyCharm is
   installed. If you want to go ahead and run it, click the "Run PyCharm Community
   Edition" box first and click "Finish".

8. After you click on "Finish", the following screen will appear
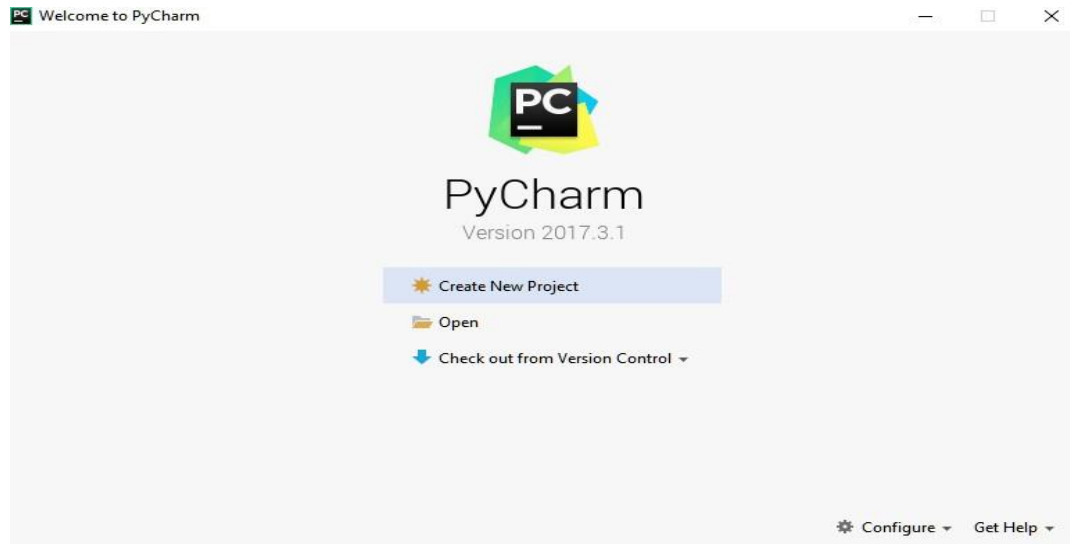
FIG-10 PYCHARM

9. You need to install some packages to execute your project in a proper way.

10. Open the command prompt/ anaconda prompt or terminal as administrator.

11. The prompt will get open, with specified path, type "pip install package name" which you want to install (like numpy, pandas, scikit-learn, matplotlib.pyplot)

    Ex: pip install numpy

To run project, follow below steps

1) Install python and create virtual environment.

2) Install all necessary requirements.

3) Create 'Python' folder in C directory and put 'Plant' folder in it

4) start Flask server and run in browser to get first page

6.2 PACKAGES:

PANDAS:

Pandas is an open source Python package that is most widely used for data science/data analysis and machine learning tasks. It is built on top of another package named <u>Numpy</u>, which provides support for multi-dimensional arrays.

NUMPY:

NumPy is a Python library used for working with arrays.It also has functions for working in domain of linear algebra, fourier transform, and matrices.NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely. NumPy stands for Numerical Python.

SCIPY.SPARSE:

NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices. NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely. NumPy stands for Numerical Python.

SKLEARN.NEIGHBOURS:

sklearn.neighbors provides functionality for unsupervised and supervised neighbors-based learning methods. Unsupervised nearest neighbors is the foundation of many other learning methods, notably manifold learning and spectral clustering.

MATPLOTLIB:

Matplotlib is a cross-platform, data visualization and graphical plotting library for Python and its numerical extension NumPy. As such, it offers a viable open source alternative to MATLAB. Developers can also use matplotlib's APIs (Application Programming Interfaces) to embed plots in GUI applications.

Movie Dataset:

  * MovieId – once the recommendation is done, we get a list of all similar movieId and get the title for each movie from this dataset.
  * genres – which is not required for this filtering approach.

	userId	movieId	rating	timestamp
0	1	1	4.0	964982703
1	1	3	4.0	964981247
2	1	6	4.0	964982224
3	1	47	5.0	964983815
4	1	50	5.0	964982931

User dataset :
The dataset before removing the null values

userId movieId	1	2	3	4	5	6	7	8	9	10	...	601	602	603	604	605	606	607	608	609	610
1	4.0	NaN	NaN	NaN	4.0	NaN	4.5	NaN	NaN	NaN	...	4.0	NaN	4.0	3.0	4.0	2.5	4.0	2.5	3.0	5.0
2	NaN	NaN	NaN	NaN	NaN	4.0	NaN	4.0	NaN	NaN	...	NaN	4.0	NaN	5.0	3.5	NaN	NaN	2.0	NaN	NaN
3	4.0	NaN	NaN	NaN	NaN	5.0	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.0	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	3.0	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	NaN	NaN	NaN	NaN	NaN	5.0	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	3.0	NaN	NaN	NaN	NaN	NaN	NaN

5 rows × 610 columns

Filling the NaN with 0.0 to male things understandable for the algorithm:

userId movieId	1	2	3	4	5	6	7	8	9	10	...	601	602	603	604	605	606	607	608	609	610
1	4.0	0.0	0.0	0.0	4.0	0.0	4.5	0.0	0.0	0.0	...	4.0	0.0	4.0	3.0	4.0	2.5	4.0	2.5	3.0	5.0
2	0.0	0.0	0.0	0.0	0.0	4.0	0.0	4.0	0.0	0.0	...	0.0	4.0	0.0	5.0	3.5	0.0	0.0	2.0	0.0	0.0
3	4.0	0.0	0.0	0.0	0.0	5.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	5.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0

5 rows × 610 columns

Removing Noise from the data

  * To qualify a movie, a minimum of **10** users should have voted a movie.

To qualify a user, a minimum of **50** movies should have voted by the user.
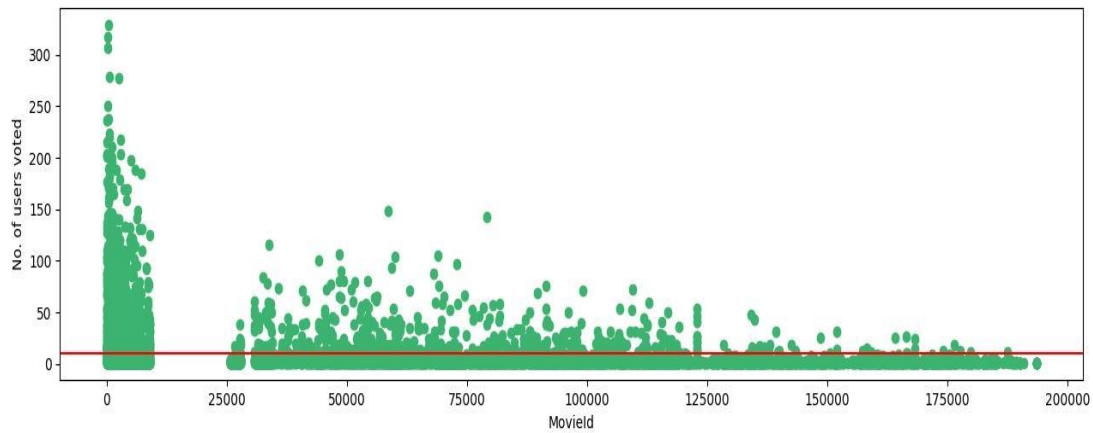
Making modification as per the threshold set.
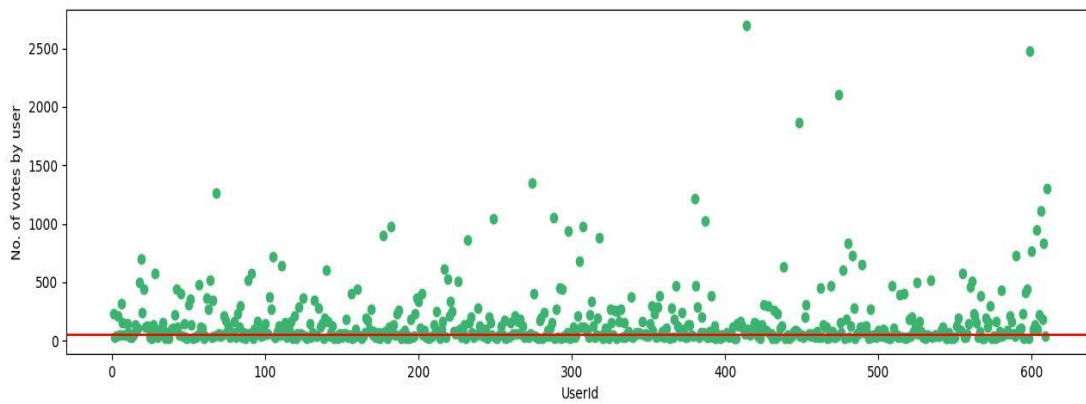
Cosine Similarity:The formula used to measure how similar the movies are based on their similarities of different properties. Mathematically, it shows the cosine of the angle of two vectors projected in a multidimensional space. The cosine similarity is very beneficial since it helps in finding similar objects.

$$CosSim(x, y) = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2} \sqrt{\sum_i y_i^2}}$$

Removing sparsity:
To reduce the sparsity we use the csr_matrix function from scipy library.
Making the movie recommendation system model :
Using the KNN Algorithm to compute similarity with cosine distance metric which is very fast .
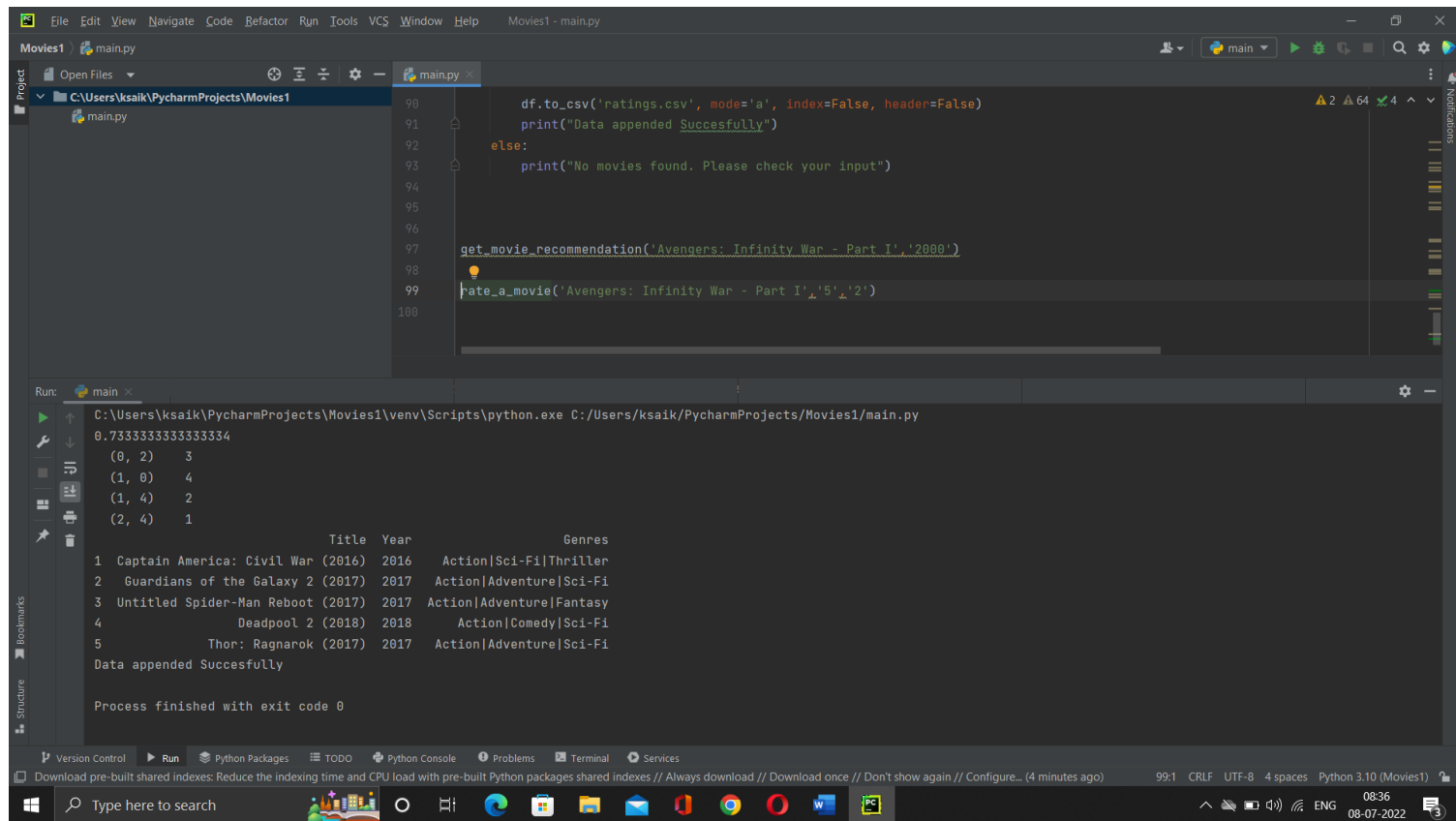
Recommending movies:
Gives top 5 movies as output.

RESULTS:

Result 1

7. TESTING

8.1 Testing

Unit Testing:

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.

- Pages must be activated from the identified link.

- The entry screen, messages and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format

- No duplicate entries should be allowed

- All links should take the user to the correct page.

Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.
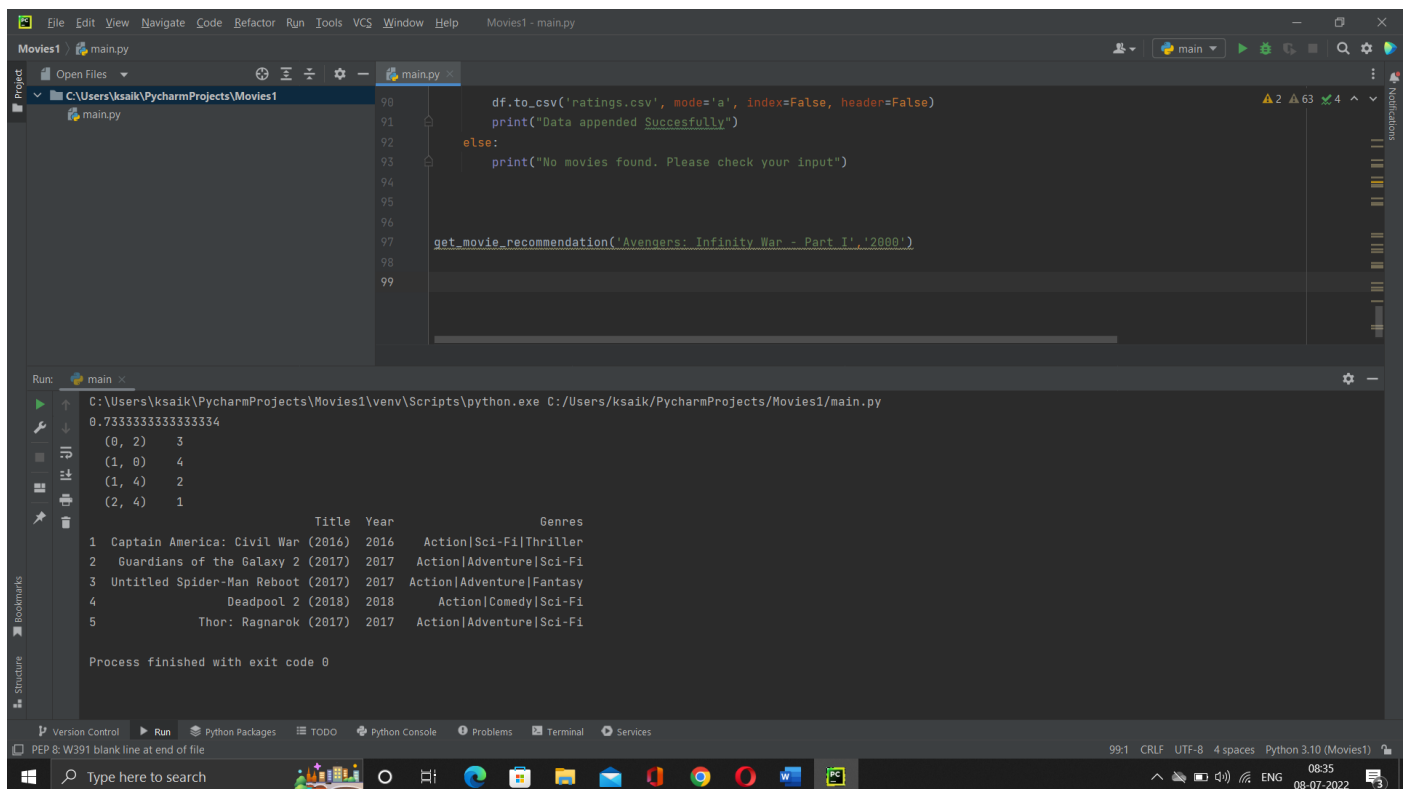
Test Results: All the test cases mentioned above passed successfully. No defects were encountered.

Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects were encountered.

8.2 Test Cases

Test Case

```
96
97  get_movie_recommendation('Toy Story 3','1960')
98
99
```

```
Run:      main ×
    C:\Users\ksaik\PycharmProjects\Movies1\venv\Scripts\python.exe C:/Users/ksaik/PycharmProjects/Movies1/main.py
    0.7333333333333334
      (0, 2)    3
      (1, 0)    4
      (1, 4)    2
      (2, 4)    1
                        Title  Year                          Genres
    1    Dark Knight, The (2008)  2008           Action|Crime|Drama|IMAX
    2            Iron Man (2008)  2008           Action|Adventure|Sci-Fi
    3  X-Men: First Class (2011)  2011  Action|Adventure|Sci-Fi|Thriller|War
    4      Sherlock Holmes (2009)  2009       Action|Crime|Mystery|Thriller
    5                 Up (2009)  2009   Adventure|Animation|Children|Drama

    Process finished with exit code 0
```
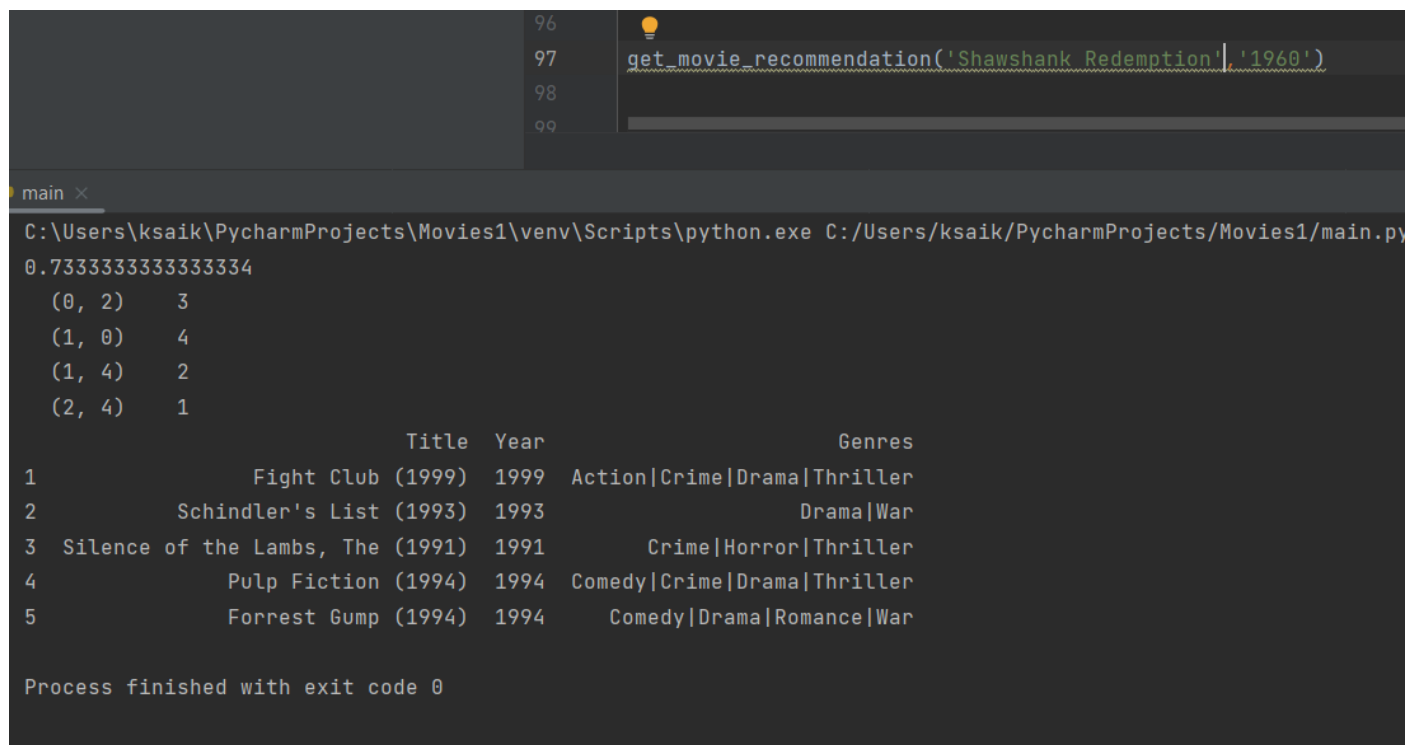
```
97  get_movie_recommendation('Dark Knight','1960')
98
99
```

```
Run:      main ×
    0.7333333333333334
      (0, 2)    3
      (1, 0)    4
      (1, 4)    2
      (2, 4)    1
                                        Title  ...                                  Genres
    1  Lord of the Rings: The Return of the King, The...  ...          Action|Adventure|Drama|Fantasy
    2                    Dark Knight Rises, The (2012)  ...             Action|Adventure|Crime|IMAX
    3                        Batman Begins (2005)  ...                      Action|Crime|IMAX
    4                           Iron Man (2008)  ...                  Action|Adventure|Sci-Fi
    5                          Inception (2010)  ...  Action|Crime|Drama|Mystery|Sci-Fi|Thriller|IMAX

    [5 rows x 3 columns]

    Process finished with exit code 0
```

CONCLUSION AND FUTURE ENHANCEMENTS

Under the condition of massive information, the requirements of movie recommendation system from film amateur are increasing. We implemented a complete movie recommendation system based on the KNN algorithm, collaborative filtering algorithm and recommendation system technology.

Future Enhancements:

We would like to improve the accuracy further expanding the dataset by including latest movies of all languages and develop a Android application.

REFERENCES

- https://ieeexplore.ieee.org/document/6382910

- Xiaoyuan Su and Taghi M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques," Advances in Artificial Intelligence, vol. 2009, Article ID 421425, 19 pages, 2009

- Cui, Bei-Bei. (2017). Design and Implementation of Movie Recommendation System Based on Knn Collaborative Filtering Algorithm. ITM Web of Conferences. 12. 04008. 10.1051/itmconf/20171204008

- Marovic, M., Mihokovic, M., Miksa, M., Pribil, S., & Tus,A. "Automatic movie ratings prediction using machine learning". In: 2011 Proceedings of the 34th International Convention MIPRO (2011,may), pp. 1640–1645.

- MovieLens. MovieLens(100k). 1998. URL: http : / / grouplens . org / datasets/movielens/100k/ (visited on 05/15/2020).

- Naumov, M.,and Mudigere, D. DLRM: An advanced, open source deep learning recommendation model. 2019. URL: https : / / ai . facebook . com / blog / dlrm - an - advanced - open - source - deep - learning - recommendation-model/. Accessed: 13 May 2020.

- Nguyen H.V., Bai L. "Cosine Similarity Metric Learning for Face Verification". In: Kimmel R., Klette R., Sugimoto A. (eds) Computer Vision – ACCV 2010. ACCV 2010. Lecture Notes in Computer Science, vol 6493. Springer, Berlin, Heidelberg (2011).

APPENDIX

Sample Code:
#Installing Modules
```python
import pandas as pd
import numpy as np
from scipy.sparse import csr_matrix
from sklearn.neighbors import NearestNeighbors
import matplotlib.pyplot as plt


movies = pd.read_csv("movies.csv")
ratings = pd.read_csv("ratings.csv")


movies.head()


ratings.head()


final_dataset = ratings.pivot(index='movieId',columns='userId',values='rating')
final_dataset.head()


final_dataset.fillna(0,inplace=True)
final_dataset.head()
```
#Removing Noise from Data
```python
no_user_voted = ratings.groupby('movieId')['rating'].agg('count')
no_movies_voted = ratings.groupby('userId')['rating'].agg('count')


f,ax = plt.subplots(1,1,figsize=(16,4))


plt.scatter(no_user_voted.index,no_user_voted,color='mediumseagreen')
plt.axhline(y=10,color='r')
plt.xlabel('MovieId')
plt.ylabel('No. of users voted')
plt.show()


final_dataset = final_dataset.loc[no_user_voted[no_user_voted > 10].index,:]


f,ax = plt.subplots(1,1,figsize=(16,4))
plt.scatter(no_movies_voted.index,no_movies_voted,color='mediumseagreen')
plt.axhline(y=50,color='r')
plt.xlabel('UserId')
plt.ylabel('No. of votes by user')
plt.show()


final_dataset=final_dataset.loc[:,no_movies_voted[no_movies_voted > 50].index]
```

#Removing Sparcity
```
sample = np.array([[0,0,3,0,0],[4,0,0,0,2],[0,0,0,0,1]])
sparsity = 1.0 - ( np.count_nonzero(sample) / float(sample.size) )
print(sparsity)

csr_sample = csr_matrix(sample)
print(csr_sample)

csr_data = csr_matrix(final_dataset.values)
final_dataset.reset_index(inplace=True)

knn = NearestNeighbors(metric='cosine', algorithm='brute', n_neighbors=20, n_jobs=-1)
knn.fit(csr_data)
```

#Recommendation_function
```
def get_movie_recommendation(movie_name,yr):
    yr=int(yr)
    n_movies_to_reccomend = 5
    movie_list = movies[movies['title'].str.contains(movie_name)]
    if len(movie_list):
        movie_idx= movie_list.iloc[0]['movieId']
        movie_idx = final_dataset[final_dataset['movieId'] == movie_idx].index[0]
        distances , indices =
knn.kneighbors(csr_data[movie_idx],n_neighbors=n_movies_to_reccomend+1)
        rec_movie_indices =
sorted(list(zip(indices.squeeze().tolist(),distances.squeeze().tolist())),key=lambda x: x[1])[:0:-1]
        recommend_frame = []
        for val in rec_movie_indices:
            movie_idx = final_dataset.iloc[val[0]]['movieId']
            idx = movies[movies['movieId'] == movie_idx].index
            if(yr<=int(movies.iloc[idx]['year'].values[0])):

recommend_frame.append({'Title':movies.iloc[idx]['title'].values[0],'Year':movies.iloc[idx]['year'].value
s[0],'Genres':movies.iloc[idx]['genres'].values[0]})
        df = pd.DataFrame(recommend_frame,index=range(1,n_movies_to_reccomend+1))
        print(df)
    else:
        print("No movies found. Please check your input")
```

#Rating a movie
```
def rate_a_movie(movie_name,rating,user_id):
    movie_list = movies[movies['title'].str.contains(movie_name)]
    rating=int(rating)
    user_id=int(user_id)
    if len(movie_list):
```

```python
        movie_idx = movie_list.iloc[0]['movieId']


        df = pd.DataFrame({'userId': [user_id],
                    'movieId': [movie_idx],
                    'rating': [rating]})

        df.to_csv('ratings.csv', mode='a', index=False, header=False)
        print("Data appended Succesfully")
    else:
        print("No movies found. Please check your input")
```

O-PO/PSO MAPPING

Course code	Statement Student will be able to	Cognitive Level	PO / PSO addressed
PW961.1	Define a problem of the recent advancements with applications towards society.	An	PO1,PO2,PO3,PO4,PO5,PO6,PO7,PO8,PSO1,PSO2
PW961.2	Outline requirements and perform requirement analysis for solving the problem.	An	PO1,PO2,PO3,PO4,PO5,PO6,PO7,PO8,PSO1,PSO2
PW961.3	Design and develop a software and/or hardware based solution within the scope of project using contemporary technologies and tools.	AP, E, An	PO1,PO2,PO3,PO4,PO5,PO6,PO7,PO8,PSO1,PSO2
PW961.4	Test and deploy the applications for use.	AP,E,An	PO8,PO9,PO10,PO11,PO12,PSO1,PSO2
PW961.5	Develop the Project as a team and Demonstrate the application, with effective written and oral communications.	C	PO8,PO9,PO10,PO11,PO12,PSO1,PSO2

Table 1: Course Outcomes - Cognitive levels
Cognitive Levels: R-Remember; U-Understand; Ap-Apply; An=Analyze; E-Evluate;
C-Create

Table 2: Number of performance indicators addressed by course outcomes

Course Code	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
No. of PIs addressed by course for a given PO	4	4	4	4	4	4	4	5	5	7	4	6	5	6
CO1	2	2	2	1	3	1	1	1	3	3	1	2	1	3
CO2	3	2	3	2	3	1	1	1	3	3	3	2	2	3
CO3	3	3	3	2	3	1	1	3	3	3	3	2	2	3
CO4								3	3	3	3	2	2	2
CO5								1	3	3	3	2	2	3

Table 3: Calculation of CO-PO/PSO correlation levels

PW961IT	PO1 %	PO1 Level	PO2 %	PO2 Level	PO3 %	PO3 Level	PO4 %	PO4 Level	PO5 %	PO5 Level	PO6 %	PO6 Level	PO7 %	PO7 Level	PO8 %	PO8 Level	PO9 %	PO9 Level	PO10 %	PO10 Level	PO11 %	PO11 Level	PO12 %	PO12 Level	PSO1 %	PSO1 Level	PSO2 %	PSO2 Level
CO1	50	3	50	2	50	2	25	1	50	3	25	1	25	1	20	1	50	3	50	3	25	1	30	1	20	1	50	3
CO2	75	3	50	2	75	3	50	2	50	3	25	1	25	1	20	1	50	3	50	3	75	3	30	1	50	3	50	3
CO3	75	3	75	3	75	3	50	2	50	3	25	1	25	1	50	3	50	3	75	3	75	3	30	1	50	3	50	3
CO4	50														50	3	50	3	75	3	75	3	30	1	20	1	50	3
CO5	50														20	1	50	3	50	3	75	3	30	1	40	2	50	3
No. Mapped	3	9	3	7	3	11	3	8	3	9	3	3	3	3	5	9	5	15	5	15	5	13	5	5	5	10	5	15
Average of Level	9/3=3		7/3=2.3		8/3=2.6		8/3=2.6		9/3=3		3/3=1		3/3=1		9/5=1.8		15/5=3		15/5=3		13/5=2.6		5/5=1		10/5=2		15/5=3	
Rounded average level	3		2		3		3		3		1		1		2		3		3		3		1		2		3	

Table 4: Course Articulation Matrix

PW961IT	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	3	2	2	1	3	1	1	1	3	3	1	1	1	3
CO2	3	2	3	2	3	1	1	1	3	3	3	1	3	3
CO3	3	3	3	2	3	1	1	3	3	3	3	1	3	3
CO4								3	3	3	3	1	1	3
CO5								1	3	3	3	1	2	3

PW961IT	3	2	3	3	3	1	1	2	3	3	3	1	2	3

RUBRICS ANALYSIS

Table 5: PO/PSO addressed by the Project

Project Name	Domain	In-house / Industry	PO/PSO addressed	Name and Signature of Guide

Table 6: Rubrics Evaluation

PO/PSO	PO1,PO2,PO6,PO7				PO3	PO4,PO5, PSO1	PO4,PO5, PSO2	PO8	PO9					PO10			PO11	PO12
Rubrics	R1				R2	R3	R4	R5	R6					R7			R8	R9
Roll. No.	CI	CII	CIII	Total	CIV	CV	CVI	CVII	CVIII	CIX	CX	Total		CXI	CXII	Total	CXIII	CIV
	4	4	4	12	4	4	4	4	4	4	4	12		4	4	8	4	4

Rubrics

for project

Focus

Areas:

1. Problem Formulation (PO1,PO2, PO6,PO7)
2. Project Design (PO3)
3. Build (PO4,PO5,PSO1)
4. Test& Deploy (PO4,PO5,PSO2)
5. Ethical responsibility (PO8)
6. Team Skills (PO9)
7. Project Presentation (P10)
8. Project Management (PO11)
9. Lifelong Learning (PO12)

Focus Areas	Criterion [c]	Exemplary 4	Satisfactory 3	Developing 2	Unsatisfactory 1
	I - Identify/Define Problem Ability to identify a suitable problem and define the project objectives.	Demonstrates a skillful ability to identify /articulate a problem and theobjectives are well defined And prioritized.	Demonstrates ability to Identify / articulate a problem and All major objectives are identified.	Demonstrates some ability to identify / articulate a problem that is partially connected to the issues and most major objectives are identified but one or two minor ones are missingor priorities are not established.	Demonstrates minimal or no ability to identify / articulate a problem and many major objectives are not identified.
Problem	**II-Collection of Background Information:** Ability to gather backgroundInformation (existing knowledge, research,and/or indications of the problem)	Collect sufficient relevant background information from appropriate sources, and is able to identify pertinent /critical information;	Collects sufficient relevant background information from appropriate sources;	Collects some relevant background information from appropriate Sources.	Minimal or no ability to collect relevant background information

Formulation (PO1,PO2, PO6, PO7)	**III- Define scope of the problem** Ability to identify problem scope suitable to the degree considering the impact on society and environment	Demonstrates a skillful ability to define thescope of problem accurately mentioning the relevant fields of engineering precisely. Considers,explains and evaluates theimpact of engineeringinterventions on Society and environment.	Demonstrates ability to define problem scope mentioning the relevant fields of engineering broadly. Considers and explains the impact of engineering interventions on society and environment	Demonstrates some ability to define problem scope mentioning some of the relevantfields. Some consideration of the impact of engineering interventions on society and environment.	Demonstrates minimal or no ability to define problem scope and fails to mention relevant fields of engineering. Minimal or no consideration of the impact of engineering interventions on society and environment

Project Design (PO3)	**IV- Understanding the Design Process and Problem Solving:** Ability to explain the design process including the importance of needs, specifications, concept generation and to develop an approach to solve a problem.	Demonstrates a comprehensive ability to understand and explain a design process. Considers multiple approaches to solving a problem, and can articulate reason for choosing solution	Demonstrates an ability to understand and explain a design process. Considers multiple approaches to solving a problem, which is justified and considers consequences.	Demonstrates some ability to understand and explain a design process. Considers a few approaches to solving a problem; doesn't always consider consequences.	Demonstrates minimal or no ability to understand and explain a design process. Considers a single approach to solving a problem. Does not consider consequences.

35

Build (PO4,PO5, PSO1)	**V- Implementing Design Strategy:** Ability to execute a solution taking into consideration design requirements using appropriate tool (software/hardware);	Demonstrates skillful ability to execute a solution taking into consideration all design requirements using the most relevant tool.	Demonstrates an ability to executea solution taking into consideration designrequirements using relevant tool.	Demonstrates some ability to execute a solution but notusing most relevant tool.	Demonstrates minimal or no ability to execute a solution. Solution does not directly attend to the problem.
Test & Deploy (PO4, PO5, PSO2)	**VI- Evaluating Final Design:** To evaluate/confirm the functioning of the final design. To deploy the project on the target environment	Demonstrates skillful ability to evaluate/confirm the functioning of the finaldesign skillfully, with deliberation for further Improvement after deployment.	Demonstrates an ability to evaluate/confirm the functioning of the final design. Theevaluation is complete and has sufficient depth.	Ability to evaluate/confirm the functioning of the finaldesign, but the evaluation lacks depth and/or is incomplete.	Demonstrates minimal or no ability to evaluate/confirm the functioning of the final design.
Ethical responsibility (PO8)	**VII - Proper Use of Others' Work:** Ability to recognize, understand and apply proper ethical use of intellectual property, copyrighted materials,and research.	Always recognizes and applies proper ethical use of intellectual property, copyrighted materials, and others' research.	Recognizes and applies proper ethical use of intellectual property, copyrighted materials, and others' research.	Some recognition and application of proper ethical use of intellectual property, copyrighted materials, and others' research.	Minimal or no recognition and/or application of proper ethical use of intellectual property, Copyrighted materials, or others 'research.
	VIII - Individual Work Contributions and Time Management: Ability to carry out individual Responsibilities and manage time (estimate, prioritize, establish deadlines/ milestones, follow timeline, plan for contingencies, adapt to change).	Designated jobs are accomplished by deadline; completed work iscarefully and meticulously prepared and meets all requirements.	Designated Jobs are jobs completed accomplished work deadline; meets requirements.	Designated jobs are accomplished by deadline; completed work meets most requirements.	Some Designated jobs are accomplished by deadline;completed meets work some requirements.

36

Team Skills (PO9)	**IX - Leadership Skills**: Ability to lead a team. (i) Mentors and accepts mentoring from others. (ii) Demonstrates capacity for initiative while respecting others' roles. (iii) Facilitates others' involvement. (iv) Evaluates team Effectiveness and plans for improvements	Exemplifies leadership skills.	Demonstrates leadership skills.	Demonstrates some leadership skills at times.	Demonstrates minimal or no Leadership skills.

	X - Working with Others: Ability to listen to, collaborate with, and champion the efforts of others.	Skillfully listens to, collaborates with, and champions the efforts of others.	Listens to, collaborates with, and champions the efforts of others.	Sometimes listens to, collaborates with, and champions others' efforts.	Rarely listens to, collaborates with, or champions others' efforts.

Project Presentation (P10)	**XI - Technical Writing Skills** Ability to communicate the main idea with clarity. Ability to use illustrations properly to support ideas (citations, position on page etc)	Main idea is clearly and precisely stated. Materials are seamlessly arranged in a logical sequence Illustrations are skillfully used to support ideas	Main idea is understandable. Material moves logically forward, Illustrations are properly used to support ideas	Main idea is somewhat Understandable. Material has some logical order and is somewhat coherent or easy to follow. Illustrations are for the most part properly used to support ideas	Main idea is difficult to understand. Material has little logical order, and is often unclear, incoherent. Illustrations are used, but minimally support ideas. (not properly cited etc)
	XII - Communication Skills for Oral Reports Ability to present strong key ideas and supporting details with clarity and concision. Maintain contact with audience, and ability to complete in the allotted time	Presentation logically and skillfully structured. Key ideas are compelling, and articulated with exceptional clarity and concision. Introduction, supporting details and summary are clearly evident and memorable, and ascertain the credibility of the speaker Presentation fits perfectly within time constraint.	Presentation has clear structure and is easy to follow. Key ideas are clearly and concisely articulated, and are interesting. There is sufficient detail to ascertain speaker's authority, and presentation includes an introduction and summary. Presentation fits within time constraint, though presenter might have to subtly rush or slow down.	Presentation has some structure. Key ideas generally identifiable, although not very remarkable. Introduction, supporting details and/or summary may be too broad, too detailed or missing. Credibility of the speaker may be questionable at times. Presentation does not quite fit within time constraint; presenter has to rush or slow down atend	Presentation rambles. Not organized; key ideas are difficult to identify, and are unremarkable. No clear introduction, supporting details and summary. Speaker has no credibility. Presentation is unsuitablably short or unreasonably long.

Project management (PO11)	**XIII - Monitoring an dControlling the Project**	Monitors timelines and progress toward project goals on a daily basis. Provides accurate, completereports of project progress.	Monitors timelines and progress toward project goals most of the time. Provides relatively accurate, complete reports of project progress withonly mi nor errors or omissions	Seldom monitors timelines and progress toward project goals. Provides relatively acc urate, yet clearly incomplete, reports of project progress	Does not monitor timeline s andprogress toward project goals. Provides inaccurate, incomplete reports of project progress
Lifelong Learni ng (PO12)	**XI V - Extend Scope of Work:** Ability to extend the project through implementation other study areas	Demonstratesa skillful ability to explorea subject/topic thoroughly, discusses the road map to extend the project in other areas.	Demonstrates an ability to explore a subject/topic, and shows possible areas in which project can be extended	Demonstrates some abilityto explore a subject/topic, providing some knowledge of areas in whichproject can be extended	Demonstrates minimal or no ability to explore a subject/topic, and does not discuss future work clearly mentioning other areas