University of Asia Pacific

Team B1-G3
Movie Recommendation Engine

Md Shamaun Nabi (18201050)
Md. Musfiqur Rahman (18201054)
Al Amin (18201063)

Project Title

The title of our project is “**Movie Recommendation Engine**”

Motivation

The very basic idea behind this system literally is that movies that basically are much more popular and critically acclaimed will definitely have a kind of higher probability of being actually liked by the actually average audience.For example, the type of music one would like to really hear while exercising differs greatly from the type of music he’d for the most part listen to when cooking dinner, very contrary to popular belief. Netflix similarly recommends DVDs that may definitely be of interest, and famously prizes to researchers who could actually improve the quality of their recommendations, particularly further showing how netflix similarly recommends DVDs that may definitely be of interest, and famously prizes to researchers who could basically improve the quality of their recommendations in a really major way. Social networking sites like Facebook use variants on recommender, very further showing how netflix similarly recommends DVDs that may literally be of interest, and famously prizes to researchers who could literally improve the quality of their recommendations, very further showing how netflix similarly recommends DVDs that may kind of be of interest, and famously prizes to researchers who could particularly improve the quality of their recommendations, or so they specifically thought.

Problem Statement

This is a web based system where there is a movie web service which provides services to users to rate movies, see recommendations, put comments and see similar movies. There are systems which deal with the self-recommendation rather than considering the likes and dislikes of users, we thereby build a system that intakes the users wishes and then recommends a watch-list of movies which is based on their selected genre. And thus makes the watch more preferable and enjoyable to the user. Given a set of users with their previous ratings for a set of movies, can we predict the rating they will assign to a movie they have not previously rated? Ex. “Which movie will you like” given that you have seen ‘Harry Potter and the Sorcerer's Stone’, ‘Harry Potter and the Chamber of Secrets’, ‘Harry Potter and the Prisoner of Azkaban’ and users who
saw these movies and also liked “Harry Potter and the Goblet of Fire?”

Objective, Solution and Project Output

- Improve retention
- Caters to the user’s preferences and keeps them hooked to the application.
- Increase sales
- Can improve business by a great margin by giving various recommendations of
- different items.
- Form habits
- Influencing usage patterns in users.
- Accelerate work
- Helps the analysts for further research and reduces their work.

Impact on Society:

The aspect of study literally is to check the level of acceptance of the system by the user, or so they for all intents and purposes thought. This includes the process of training the user to use the system efficiently, or so they basically thought. The user must not for all intents and purposes feel threatened by the system, instead must accept it as a necessity, which generally is fairly significant. The level of acceptance by the users solely depends on the methods that essentially are employed to educate the user about the system and to for the most part make him familiar with it in a actually major way. His level of confidence must be raised so that he actually is also able to make some constructive criticism, which really is welcomed, as he is the final user of the system in a subtle way.

Related works and Background study :

Literature review:

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

This recommendation engine will calculate the similarities between the different users. On the basis of that similarities calculated, this engine will recommend movie to a user. Based on the selected genre movies are recommended and the selected genre is stored for further recommendations.

Feasibility Study:

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

Three key considerations involved in the feasibility analysis are,

- economical feasibility
- technical feasibility
- social feasibility

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

Technical Feasibility: This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

Social Feasibility: The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

Knowledge profile:

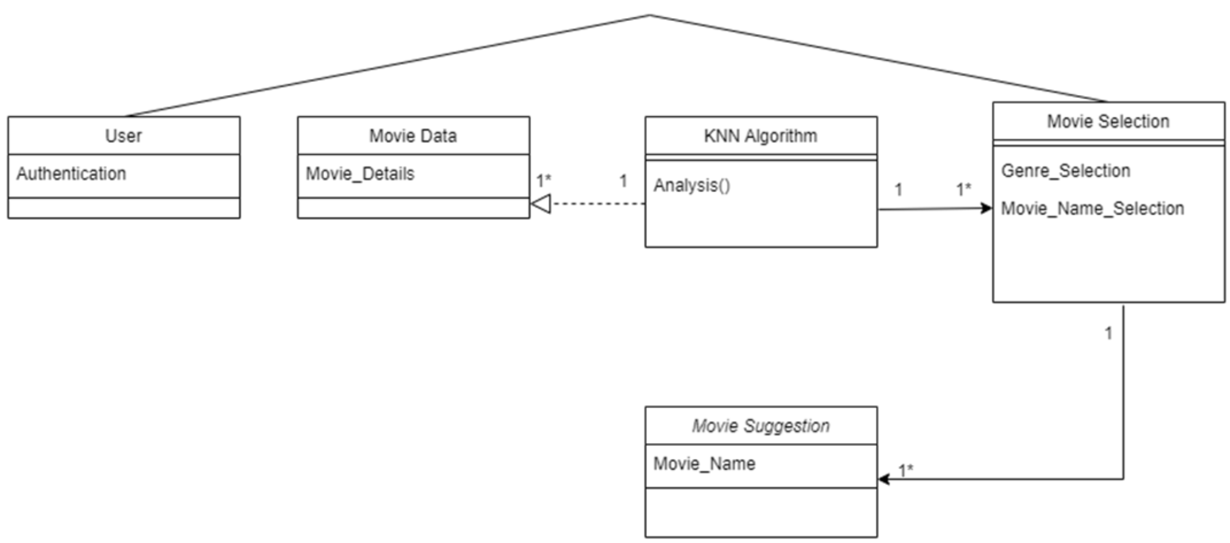
- K — short name
- K1 — natural sciences
- K2 — mathematics
- K3 - engineering fundamentals
- K4 — specialist knowledge
- K5 — engineering design
- K6 — engineering practice

- K7 — comprehension
- K8 — research literature

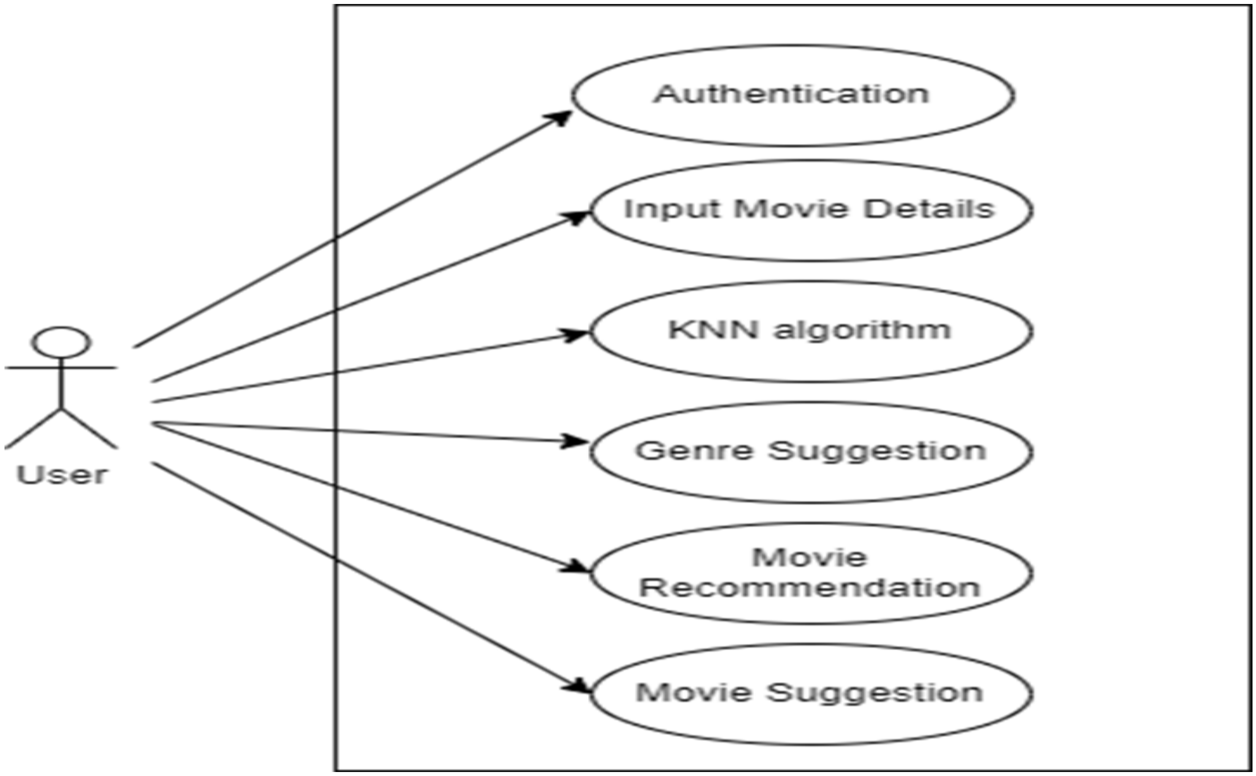
Solution methodology:

We need to perform preprocessing on the dataset and combine the relevant features into a single feature. Later, we need to convert the text from that particular feature into vectors. Later, we need to find the similarity between the vectors. Finally, get the recommendations from the vector result which would give the nearest node as recommends.

Class-Diagram:



Use Case-Diagram:



.

Algorithm: We have used K-Nearest Neighbor(KNN) for our dataset. KNN makes inference about a movie, KNN will calculate the “distance” between the target movie and every other movie in its database, then it ranks its distances and returns the top K nearest neighbor movies as the most similar movie recommendations.

Dataset: In this project, we are using The TMDb Dataset. We have collected it from Kaggle.. Our dataset is in two parts of a csv file. They are -

- tmdb_5000_credits.csv
- tmdb_5000_movies.csv

There are almost 5000 movies in our dataset in which we will train our model.

Framework: Python, https://streamlit.io/, etc.

Critical Challenges

Based on user surveys and evaluations, recommendation systems can be characterized into two parts-

1.Content-based recommendation system

2. Collaborative filtering recommendation system

Content-based recommendation system: Content-based filtering is an approach that uses the descriptions of what users viewed or bought in the past, and then an item is recommended based on the similarities of previously used items.

Challenges in developing recommendation system:

Cold start:

This problem arises when new users or new items are added to the system, a new item can't recommend to users initially when it is introduced to the recommendation system without any rating or reviews and hence it is hard to predict the choice or interest of users which leads to less accurate recommendations.

Sparsity:

It happens many times when most of the users do not give ratings or reviews to the items they purchased and hence the rating model becomes very sparse which could lead to data sparsity problems, it decreases the possibilities of finding a set of users with similar ratings or interest.

Synonymy:

Synonymy arises when a single item is represented with two or more different names or listings of items having similar meanings; in such conditions, the recommendation system can't recognize whether the terms show various items or the same item.

Privacy:

Generally, an individual needs to feed his personal information (have an experience with hyper-personalization) to the recommendation system for more beneficial services but it causes the issues of data privacy and security, many users feel hesitation to feed their personal data into recommendation systems that suffer from data privacy issues.

Scalability:

One biggest issue is the scalability of algorithms having real-world datasets under the recommendation system, a huge changing data is generated by user-item interactions in the form of ratings and reviews and consequently, scalability is a big concern for these datasets.

Latency:

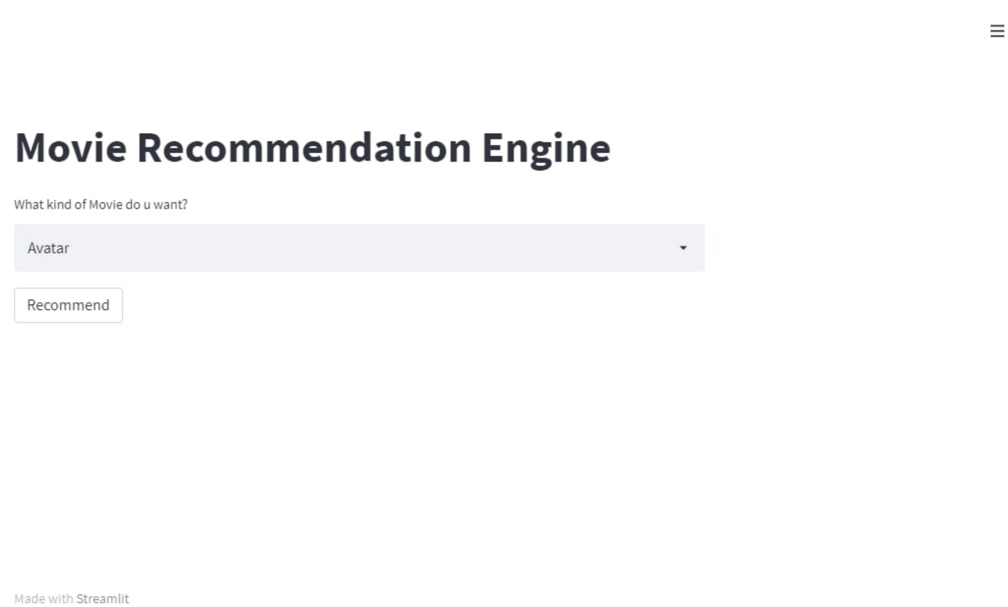
We observe many products are added more frequently to the database of recommendation systems, only already existing products are recommended to users as newly added products are not rated yet.

So an issue of Latency arises.

Final Result:

Here are some snapshot of our recommendation system:

Front-end Snapshots:



Movie Recommendation Engine

What kind of Movie do u want?

The Dark Knight Rises

Recommend

The Dark Knight	Batman Returns	Batman	Batman Forever	Batman Begins
				

Movie Recommendation Engine

What kind of Movie do u want?

spider

Spider

Spider-Man

Spider-Man 2

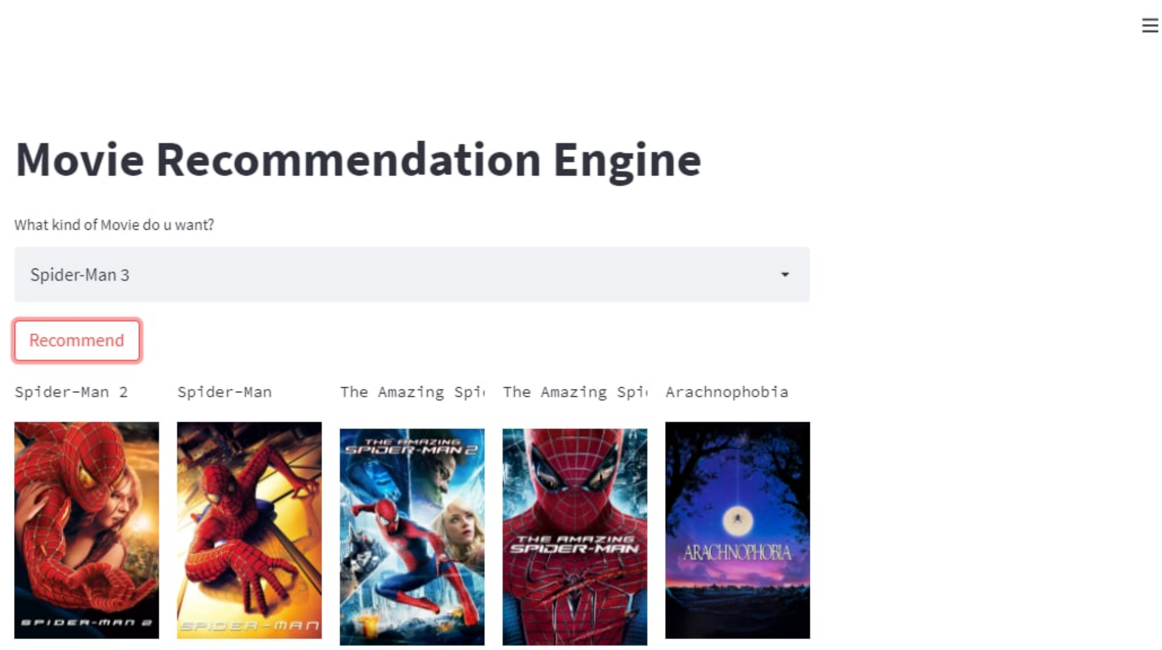
Spider-Man 3

Along Came a Spider

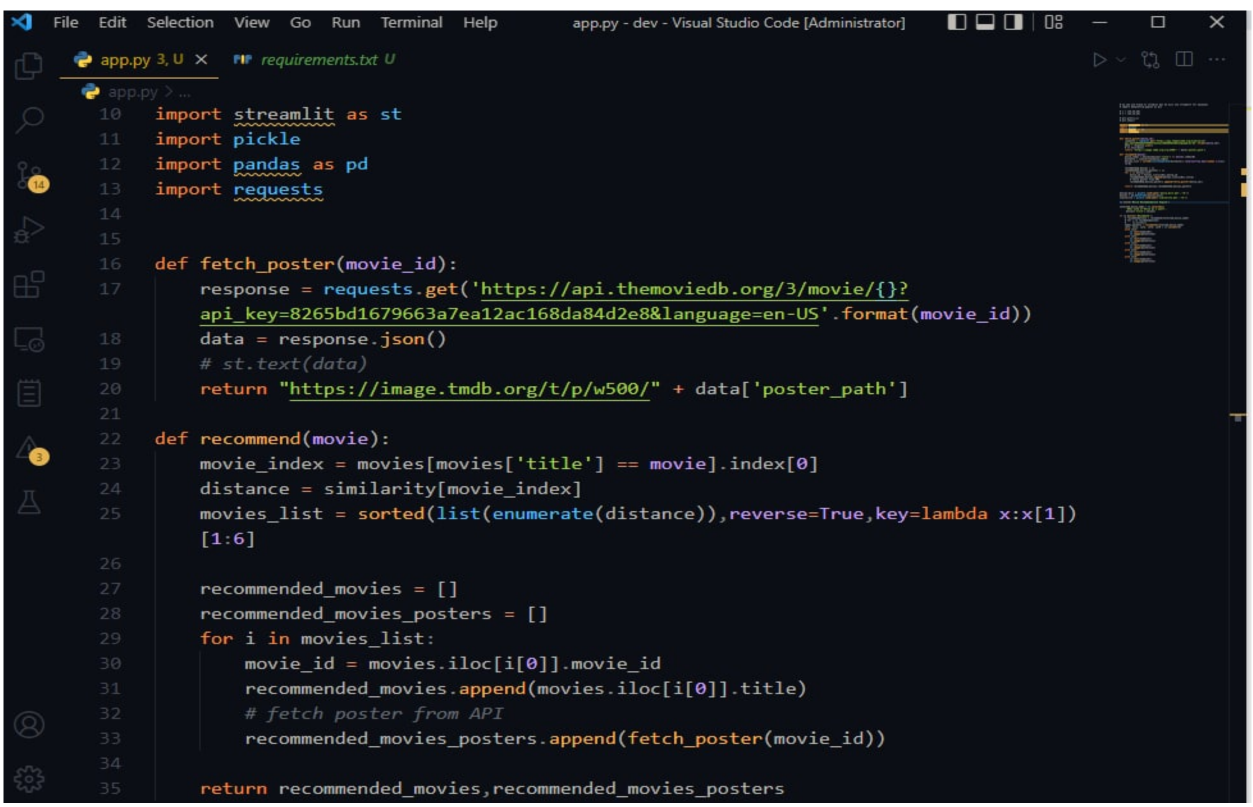
The Amazing Spider-Man

Kingdom of the Spiders

The Amazing Spider-Man 2



Back-end Snapshots:




```python
import streamlit as st
import pickle
import pandas as pd
import requests


def fetch_poster(movie_id):
    response = requests.get('https://api.themoviedb.org/3/movie/{}?
    api_key=8265bd1679663a7ea12ac168da84d2e8&language=en-US'.format(movie_id))
    data = response.json()
    # st.text(data)
    return "https://image.tmdb.org/t/p/w500/" + data['poster_path']

def recommend(movie):
    movie_index = movies[movies['title'] == movie].index[0]
    distance = similarity[movie_index]
    movies_list = sorted(list(enumerate(distance)),reverse=True,key=lambda x:x[1])
    [1:6]

    recommended_movies = []
    recommended_movies_posters = []
    for i in movies_list:
        movie_id = movies.iloc[i[0]].movie_id
        recommended_movies.append(movies.iloc[i[0]].title)
        # fetch poster from API
        recommended_movies_posters.append(fetch_poster(movie_id))

    return recommended_movies,recommended_movies_posters
```

```python
movies_dict = pickle.load(open('movie_dict.pkl','rb'))
movies = pd.DataFrame(movies_dict)
similarity = pickle.load(open('similarity.pkl','rb'))

st.title('Movie Recommendation Engine')

selected_movie_name = st.selectbox(
    'What kind of Movie do u want?',
    movies['title'].values)

if st.button('Recommend'):
    # recommendations = recommend(selected_movie_name)
    # for i in recommendations:
    #     st.write(i)
    names,posters = recommend(selected_movie_name)
    col1, col2, col3, col4, col5 = st.columns(5)
    with col1:
        st.text(names[0])
        st.image(posters[0])
    with col2:
        st.text(names[1])
        st.image(posters[1])
    with col3:
        st.text(names[2])
        st.image(posters[2])
    with col4:
        st.text(names[3])
        st.image(posters[3])
```

In [87]:

```python
def recommend(movie):
  movie_index = new_df[new_df['title'] == movie].index[0]
  distance = similarity[movie_index]
  movies_list = sorted(list(enumerate(distance)),reverse=True,key=lambda x:x[1])[1:6]

  for i in movies_list:
    print(new_df.iloc[i[0]].title)
    #print(i[0])
```

In [88]:

```python
recommend('Avatar')
```

```
Aliens vs Predator: Requiem
Aliens
Falcon Rising
Independence Day
Titan A.E.
```

In [89]:

```python
recommend('Batman Begins')
```

```
The Dark Knight
Batman
Batman
The Dark Knight Rises
10th & Wolf
```

```
Remove the column that actually don't need to our dataset to predict our goal
```

So, We keep the below columns in our dataset

```
# genres
# id
# keywords
# title
# overview
# cast
# crew
```

```
# This column is totally imbalanced in our dataset
# so, we need to remove this column [en-4510>others]
# movies['original_language'].value_counts()
```

So, we keep the belows columns

```
movies = movies[['movie_id', 'title','overview','genres','keywords','cast','crew']]
```

Github Link:

Source code of these project are given below:

https://github.com/Shamaun-Nabi/Movie-Recommendation-Engine-ML

How Ps are addressed through the project and mapping

Ps	Attribute	How Ps are addressed through the project	CO	PO

P1	Depth of Knowledge Requirement	The project requires study of research on Machine Learning system mainly supervised learning's KNN model, Digital data processing **(K8)** Data collection from user sites like online movies specially movies, series, web series, documentation film etc.**(K7)** Engineering design (multi-layer model design on analyzing phase) **(K5)** and user interface development **(K6)** knowledge of software engineering and data pre-processing (feature extension), various algorithms **(K3, K4, k2).**	CO1, CO2, CO3, CO7, CO8	PO(a), PO(b), PO(c), PO(e), PO(j)
P2	Range of Conflicting Requirement	Create an appropriate (K8) machine-learning model (k4) to detect (k2) a movie from similar type movies.	CO1, CO2, CO7	PO(a), PO(b), PO(d)
P3	Depth of Analysis Required	Use supervised machine learning model (**K4, k2**), a type of machine learning model instead of tensor-flow or other API based movie detection algorithm.	CO1, CO2	PO(a), PO(b)
P4	Familiarity of Issues	Based on searching detection (**K5, k2**), choice of genre related online result, decision or movie tag (k4).	CO1, CO2, CO7	PO(a), PO(b), PO(c)
P5	Extent of applicable codes	Using KNN supervised learning model (K5) and other standard library functions (K6, K3), build the proper solution model and user interface with data synchronization (K4, K2) of the proposed system following the engineering way.	CO1, CO2, CO3, CO4, CO6, CO7	PO(a), PO(b), PO(c), PO(e)
P7	Interdependence	Creating model (algorithmic part) (K2,K4), classifying data characteristic using modern tools(K6), proposed user application interface.	CO1, CO2, CO3, CO7	PO(a), PO(b), PO(e)

How As are addressed through the project

As	Attribute	How As are addressed through the project
A1	Range of Resources	In development stage, the project requires the use of diverse resources including different type of material : Movie dataset Information’s: Movie meta-data, Movie knowledge Technologies: machine learning(ML) model for supervised, People: Developers. Designer, Analyzer
A2	Level of Interaction	A better interaction is required among the researchers and developers (student), Viewers and participants (system users).
A3	Innovation	A degree of innovation is required to develop the machine-learning based supervised learning model using the available data set.
A4	Consequences for society and the environment	Because of this, our viewers are going to have a more convenient platform. Even in the condition of having physical distance, people can choose their desired movie without wasting time and follow the updated trend.

A5	Familiarity	The project deals with a recommendation system based on users searching, watching, viewing rating and so on analysis of participants.

CO-PO mapping for this project

CO No.	**CO Statements**	**Corresponding POs (Appendix-1)**
CO1	Identify a real-life problem **(Recommendation System)** that can be translated to an engineering and/or computing solution through design, development and validation	12 1
CO2	Identify outcomes and functional requirements of the proposed solution **(Movie Recommendation System)** considering web based and/or hardware specification and standards	1,2 a,b
CO3	Identify sub components of a complex problem, prepare timeline and appropriate budget using the project management skill	11 k
CO4	Identify and validate the impact of environmental considerations and the sustainability of a system/subsystem of a complete project	7 g
CO5	Assess professional, ethical, and social impacts and responsibilities of the design project	6,8 f,h

CO6	Function effectively in a multi-disciplinary team	9 i
CO7	Analyze, design, build, and evaluate engineering/computing system/subsystem with given specifications and requirements	3, 4 ,5 c,d,e
CO8	Present design project results through oral presentations	10 j

Program outcomes (PO) for engineering programs

No	**PO**	**Differentiating Characteristic**
1	Engineering Knowledge	Breadth and depth of education and type of knowledge, both theoretical and practical
2	Problem Analysis	Complexity of analysis
3	Design/ development of solutions	Breadth and uniqueness of engineering problems i.e. the extent to which problems are original and to which solutions have previously been identified or codified
4	Investigation	Breadth and depth of investigation and experimentation
5	Modern Tool Usage	Level of understanding of the appropriateness of the tool
6	The Engineer and Society	Level of knowledge and responsibility
7	Environment and Sustainability	Type of solutions.

8	Ethics	Understanding and level of practice
9	Individual and Team work	Role in and diversity of team
10	Communication	Level of communication according to type of activities performed
11	Project Management and Finance	Level of management required for differing types of activity
12	Lifelong learning	Preparation for and depth of Continuing learning.

Project management (Time-Table) :

Lab	Working Progress
Lab 1	Introduction
Lab 2	Project Proposal Permission
Lab 3	Project Proposal Permission
Lab 4	Planning and Requirement Analysis
Lab 5	Requirement Analysis and data collecting
Lab 6	UI/Ux Planning
Lab 7	Data pre-processing
Lab 8	Data pre-processing
Lab 9	Model Building
Lab 10	Model Building
Lab 11	Model Building
Lab 12	Connect with Web
Lab 13	Deployment

Lab 14	Final Presentation and Submission