

# Gener Based Recommendation System for Movies and TV Shows

Gaurav Bedse, Sanskar Sharma, Vijay Singh, Keshavraaj Chodhary

Department of Computer Engineering

Smt. Kashibai Navale College Engineering, Pune, Maharashtra, India

**Abstract:** *A web application to predict a personalized recommendation of a Movie/TV Show by performing machine learning approach such as EDA, Model election, Model Training and Validation using some classification algorithms based on their accuracy with the main parameter of genre.”.*

**Keywords:** Recommendation System

## I. INTRODUCTION

In the boom of OTTs, who doesn't like to watch TV shows and Movies? Every weekend there's a question of "What to watch? What's Trending?" and to give you the result according to your preference is our main goal at focus in this project.

In the current world, movies and tv shows are the best source of entertainment. Not only for entertainment but also, they are a major source of commerce, marketing, and benefits in the education. With the growth of different technologies, online streaming of movies and TV shows has become widely popular and there exist several streaming platforms like Netflix, Amazon Prime Video, and YouTube. With a lot of content to consume every platform promotes their most popular shows to users to consume with the help of ads and marketing. Simple and yet attractive UI makes the users get more gaged on the screen on the daily basis. Not only the urban areas in India are consuming the content but also according to the news due to the internet movement of getting a super cheap internet, the rural crowd is even more attracted to the OTTs. Also, the platforms are even promoting regional content like Marathi movie viz "sairat", Popular MX Original "Samantar", South Indian Blockbusters like "Jai Bhim" which was the most popular movie rated 9.6 on IMDb. Not only the local public are consuming local content but also the global content too. Korean Dramas like "Squid Game", Spanish content like "Money Heist", were in the top watched tv shows in India. Therefore, considering the above scope and the graph of binge watching the title on your tv screens, we got the idea of building a machine learning model that will provide user's content Selecting a recommendations. The model will get the user's input by asking them several questions regarding their favourite last watched movie or a series, then according to the users' preferences, the model will provide the best 5 recommendations of tv shows or movies from that specific user.

## II. METHODOLOGIES OF PROBLEM SOLVING

The problem defined above was tackled with the help these 5 points. First was to ask the correct questions. The questions were taken into account to give the best possible recommendations to the users according to their last watched titles. Second was the method to collect the data, from where we can get the best data for the recommendation. The answer was IMDb. It is a library of thousands and thousands of titles which helped us a lot to get the recommendations. Then we needed to alter these data and clean the data for our project liking. Preprocessing was an essential step for us because to get the accurate data and remove irrelevant data from it made our recommendations even better. Then we trained it better with the parameter of genre and the average rating of the title. The Final result comprised of the Top 10 recommendations of TV Shows and Movies according to the choice made by the users. We made sure that our methodology must be agile in nature. If that, any changes are meant, then proper discussion and implementation is done. Proper decision making was done to evaluate what methods should we use at the time of implementation.

The What to Watch has got lots of problems as we all know, but the way we handle the problem was critical and according to us, better impulsive decisions were made.

### **III. LITERATURE SURVEY**

To identify the Machine Learning algorithms that can be used for the recommendations of movies/tv shows, we choose Systematic Literature Survey as our research procedure that helps to find the works in the same field and analyse them using standard procedures.

Taushif Anwar and Uma Vijayasundaram [1] uses the method of collaborative filtering and used the mean absolute error for accuracy metric. The proposed approach is compared with well-known machine learning approaches namely k nearest neighbor (K-NN), singular value decomposition (SVD) and Co-clustering. This limit based only on the fact that it uses the collaborative study and comparison of recommender system. D Sisodia, DS Sisodia [2] describes classification algorithm like Naïve Bayes, SVM and Decision tree To predict the diabetes among the pregnant women. Naïve Bayes has obtained 76% accuracy. But only used for the context related to the medical field with small amount of data. N Ivan, E Ahishakiye, EO Omulo, D Taremwa [3] gave the research about comparing two classification algorithms like Decision tree and Naïve Bayes for the prediction of crime category. Decision Tree algorithm performs well with 84% accuracy. The research is related only towards one category which is not sufficient for project scope. EHA Rady, AS Anwar [4] focused on predicting the kidney disease using classification algorithm like Naïve Bayes and SVM. SVN algorithm has performed best in the prediction. The prediction format for the same is of a Y/N type which is not to be included in the project. AM Shahiri, W Husain [5] researched about predicting the students' performance using datasets of student information using Rule learner, a Decision tree classifier, a Neural Network and a K-Nearest Neighbor classifier. The Neural Network model predicted with high accuracy. Performance is based on the task of the students which has no relation in the project concept, but the techniques were useful to study. MH Latif, H Afzal [6] tried to predict the movie popularity using IMDB movie source as the dataset and machine learning techniques. The model trained using Logistic Regression perform well with the accuracy of 84%. Used the ML technique only for predicting the popularity of the movie. Not any other parameter considered. R Sharda, D Delen [7] in their research concluded that "Random Forest Classifier" predicts with high accuracy. Prediction based on the fact Y/N for the success of the movie and no ratings, genres considered. M Govindarajan [8] shows that hybrid approach of Naïve Bayes and Genetic algorithm performs best for analysing the sentiment from movie reviews. Analysis is done on the basis of review included by the users which should not be a possible parameter for recommendations.

YF Huang [9] he proposed a model that takes the 277 features from different audio and video sources using self-adaptive harmony search algorithm and these features fit into a model with SVM classifier that classifies the movies into different genres with 92% accuracy. Classification based on audio and video but did not include users' preferences as a deciding variable. K Mouthami, KN Devi [10] In this research, they gavetextual reviews are classified into positive, negative and neutral reviews using Bag of Words model with SVM classifier and Fuzzy classifications algorithm to improve the efficiency in classification. Using overall reviews dividing into positive and negative would not be suitable is recommendations to be personalized. Warda Ruheen Bristi, Zakia Zaman, Nishat Sultana [11] gave rating prediction used rather than the prediction of a show/ movie according to users' requirement. Also focused only on one source of IMDB ratings. The research is done to predict IMDB rating of the movies from Wikipedia as the dataset and model id trained using J48, Random Forest classifier, Naïve Bayes classifier and bagging. RFC predicts the rating with 99% accuracy.

From the study of above research papers using Systematic Literature Study, we have identified that Support Vector Machine (SVM), Naive Bayes classifier, Random Forest classifier, and Logistic Regression algorithms may be used for training the model to get the recommendations predictions.

#### **3.1 Gap Analysis**

The GAP from "Taushif Anwar and Uma Vijayasundaram [1] Comparative study of recommender system approaches and movie recommendation using collaborative filtering" can be restricted to-

- Predicting only the IMDB ratings – The IMDB ratings were predicted rather than the recommendations of the Movies and TV shows. In our project we use the help of the IMDB ratings for a nearby title with the genre being the same.
- Focused only on the Hollywood content restricted only to movies – In India, the local crowd prefers to get more recommendations in their regional content rather that foreign, Our Project makes the effort to provide more local content thereafter.

- Worked only on the multiplex-released movies, not OTT platforms in India. - The research done by the author focuses only on the movie which are making in the box office. Theatrical releases are affecting due to the pandemic and this results in the recommendations which calculates the gross amount collected by the movie and not gors for the personalized thing
- The Data is trained only for the year 2018. - Movie and Tv shows are releasing in the alarming rates with more than 1000s of releasing every month. So, this gap can be critical if the end-user chooses to get an recommendation of latest type, he/she will not be able to do so.
- And these gaps overcome by -
- Prediction of the title of a TV Show / Movie using the data provided by the user. - Only the user will get his/her recommendation according to the choices provided by them which will not consider the external factors but only personalized recommendations
- Focused on Hollywood as well as content from North (Bollywood) and south India. - Hindi, as well as regional languages will be given the more importance rather than the focused Hollywood content
- Focus on Both TV Show and Movie Prediction based on OTTs present in India - Indian based OTTs like Netflix India, prime video India, Disney+ Hotstar, Sony Liv, Zee5 etc. will be giving an edge to the respective subscribed user to leverage their subscriptions.
- The Data to be used for training would be of 2 decades based on popularity from the years and Genre would be the main attribute in focus. - Apart from this there would be also recommendations like top 10 movies and TV, Trending Tv and movies, Series of the day, Movie of the day and recommendations according to the trend.

#### **IV. SOFTWARE REQUIREMENTS SPECIFICATION ASSUMPTIONS AND DEPENDENCIES**

##### **4.1 Introduction**

A document or set of documentation that describes the features and behavior of a system or software application is understood as System Requirements Specification (SRS). SRS describes the entire software to be developed. The two main requirements in SRS are Functional and Non-Functional requirements. It may also contain the user cases that the software must provide.

##### **4.2 Purpose of SRS**

The main purpose of the SRS documentation is to provide detailed survey of this software. The hardware, software and the end-user requirements are reported in this SRS.

##### **4.3 Requirement of SRS**

The SRS is included as a requirement in any project/software development as a main feature because of the fact this documentation helps to overview the project beforehand and helps in development of another important documentation. The goals and the aim of the project can be easily classified with the help of SRS. A perfect design is implemented because of the detailed and sufficient requirements provided in the SRS.

#### **V. USER CLASSES AND CHARACTERISTICS**

Page/App Manager: The whole of the app/page and the ecosystem is handled by this main component. All the queries/feedbacks and recommendations are handled by this individual/s. It is responsible for the operations/model maintenance for the recommendations.

Customer: The end-user to receive all the recommendations by the software and providing the feedback according to it. Detailed information is provided to the user by the user provided some insightful data.

Recommender (Future advancement) - Individual/s who is responsible for daily trends recommendation of titles of Movies/TV not controlled by Machine.

##### **5.1 Assumptions**

- Assumption of maximum 5 recommendation per use
- Assumption of users having streaming subscriptions and had watched a title before.

- Assumption of recommendation without having a current OTT Platform

### 5.2 Dependencies

- The Recommendation relies on the user to provide an actual last watched TV/Movie Title
- The application will depend on frequent ML training due to ever increasing data.
- Dependency only on the region restricted to Indian OTT platform space.

## VI. FUNCTIONAL REQUIREMENTS

The official definition of 'a functional requirement' is that it essentially specifies something the system should do. The Functional Requirements Specification given below documents the operations and activities that a system must be able to perform according to the given requirements

- FR1 – Inputs
  - Enter the string containing correct last watched title. Also, there should be also a check whether the title is actually present
- FR2 – Data Retrieval.
  - Retrieve correct genre from the user's title or from list of genres.
- FR3 – Data Processing
  - Reload the process if any information is not relevant/incorrect/no data
  - Process the attributes like genre to give precise recommendations
- FR4 – Data Prediction and Testing
  - Predict recommendations in the form of top 5's and they must be relevant to the data produced by the user.
- FR5 – Data Maintenance
  - Model implementation according to the most recent data.
  - Data to be maintained frequently because of newly added titles and trained thereafter
- FR 6 – Administrative functions
  - Providing general recommendations according to the trends happening in the recent times.

### 6.1 External Interface Requirements

#### A. User Interface

Any Web Browser Preferably PC can be used for an end user device interface.

The logical characteristics of each user interface that the system needs. Some possible items to include are-

- References to GUI standards with yellow and black as the main themes
- Font Family to include is Roboto/Sans Serif
- Images of every title to be web scraped and applied to concern.
- Resolution on 1920×1080 (Desktop browser)
- Buttons on every page to give a dynamic look.

#### B. Software Interfaces

- LANGUAGES: Python (with libraries mentioned below)
- Platform: Windows
- Backend: MySQL, Pycharm/spyder

### 6.2 Non Functional Requirements

The Requirements which are not adhered to the Functional requirements can be represented by non-functional requirements. They give the overview of how the things are done rather than specific mandatory tasks. "If user doesn't want TV recommendations, then don't use tv in database" is an example of this requirements.

Some typical non-functional requirements are:

**Performance Requirements**

- PERFORMANCE – The system gives recommendations based on users' choice rather than overall data.
- SCALABILITY – Large amount of data is used for model training, testing purposes.
- MAINTAINABILITY – Quite a maintenance is required because of new data produced every day.
- USABILITY – system can be used frequently to get new recommendations, also very easy to use app for user experience.

**Safety Requirements**

- Authentication – Users and the Admins Authentication would be done on the prior basis to prevent from any data and leak of users personal privacy.

**Security Requirements**

- Data Privacy– Data which is provided by the user is only for the purpose of getting recommendations for that user only and no information is shared to ads.

**Software Quality Attributes**

- RELIABILITY- Predictions are done instantly in the form of 5 titles and also the OTT platform where user can stream if present.
- EFFICIENCY – model is trained quite well with the specific ML algorithm to get efficient results as far as possible.

**System Requirements**

This section provides information to ensure that the system will communicate properly with another system components. All the user/software/hardware requirements are provided with a specific description.

**Database Requirements**

All the data shall be stored in text-based flat files. For the Machine Learning Database title which is to be of type string, title of type int, genre with type list, rating of type int is provided.

The Requirement of Database for the given project according to the size of data is-

- 2 Databases viz – Train/validation database and user database
- MYSQL Server and SQLite will be the respective data base servers according to the size

**Software Requirements**

Connections between this product and other software components (identified by name and version). The given specifications are the updated version as of now, we have provided with version numbers and description below.

- Windows 10/11 64-bit operating system.
- Python -V 3.9.2, an open-source programming language that is dynamically programmed and supports multiple programming including functional, object-oriented programming.
- Anaconda -V 1.7.2, an open-source environment consists of data science packages and available for windows, macOS, Linux operating systems. It is used to build and run the machine learning models

**Hardware Requirements:**

- For experimentation, we have used a laptop. The minimum hardware requirement was in accordance - A Laptop/PC
- Disk Space – Minimum 5 GB (Movie and TV Data)
- Processor - AMD RYZEN 5 3550H with Radeon Vega Mobile GFX 2.10Ghz / Intel core i5-9300H CPU
- RAM – 4 GB



Analysis Models

The software development life cycle (SDLC) is a framework defining tasks performed at each step in the software development process. SDLC is a structure followed by a development team within the software organization. It consists of a detailed plan describing how to develop, maintain and replace specific software.



Fig. Agile Model

The agile model is being implemented in our prototype implementation.

VII. SYSTEM IMPLEMENTATION PLAN

The following is the brief implementation process that the team will carry forward for a smooth project development. In this plan, starting from the planning, problem discussion, risks, assessments etc. are thoroughly debated and discussed. The plan is discussed in earlier stages. The design of this implementation plan is done in the design phase. Further evaluation and the improvements needed are modified in the implementation phase. The Table below shows the short overview of the implementation plan.

Table with 3 columns: Cycle, Process, and Timeline. It lists 6 cycles from 'Decoding the problem' to 'Finalization' with corresponding dates.

Table: System Plan Implementation

7.1 System Design

A. System Architecture

The process of the system architecture shown below starts from the user prompting to start getting the personalized recommendation. The user is given the option to choose between the TV show and Movie. If the user doesn't choose between the option, the user is sent to the surprised recommendations which contain the recommendation according to the trends. If the user chooses TV Show, then he/she/they are asked to provide an input of their last watched TV show. The Same process is for the user who choose the movie option. The last watched input provided by the user in their respective Tv or Movie section is then sent to the system and with the help of web scraping and EDA with the help of Python, the information like Title, IMDb title id, Genre, IMDb website, Ratings, and Summary about the title is sent back to the user. Also, where the title is streaming (OTT) is also sent as an output to the user. Now the respective user is asked to input a decision about whether they want to watch the title according to their last watched title. If yes, then the input is sent to the system which contains the machine learning model which according to the suitable classification algorithm with the main parameter of genre, scans the data base and with the help of the model training and validation, the end user is provided with the best 10 recommendations of respective Tv shows and Movie according to their last watched title's genre



parameter. If the user decides against the last watched, then the system provides the saved list of all available genres to choose from. Then the user is asked to input the number associated with the genre or genres present in the list gave by the system. The chosen genre/genres then are sent to the machine learning containing system which executes the same process which was executed in the last watched sub process. The system the predicts/provides the best 10 recommendations of their TV show and Movies according to the genre which was chosen from the list. The System asks the user if are the satisfied or do they wish to continue. If positive, the entire process is repeated. If not, the stop takes place and return to the home page

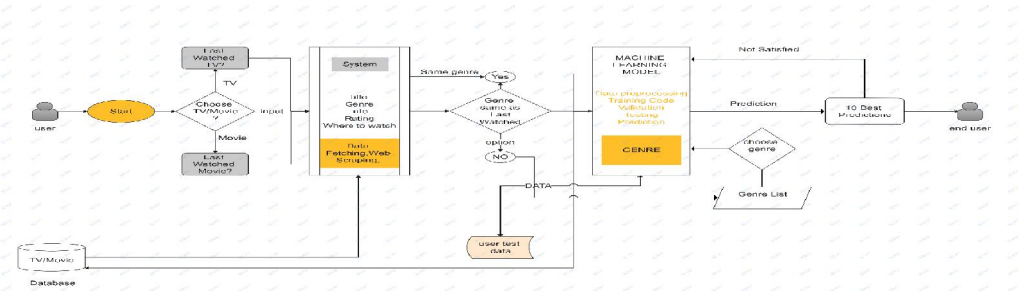
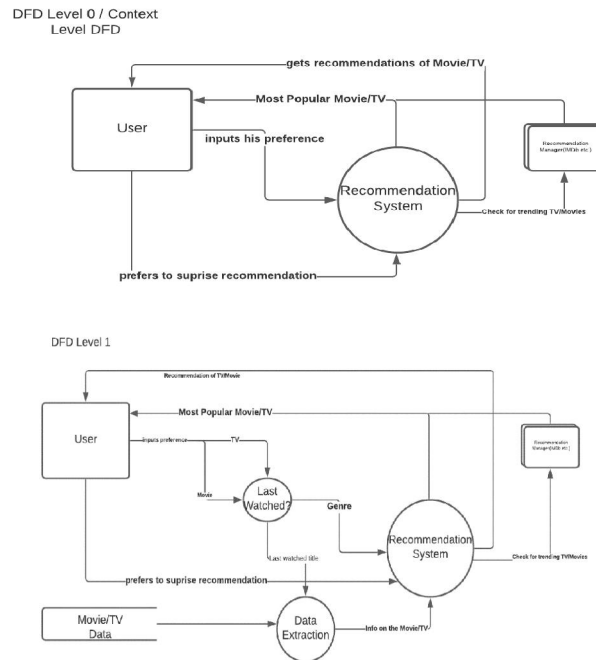


Figure: System Architecture

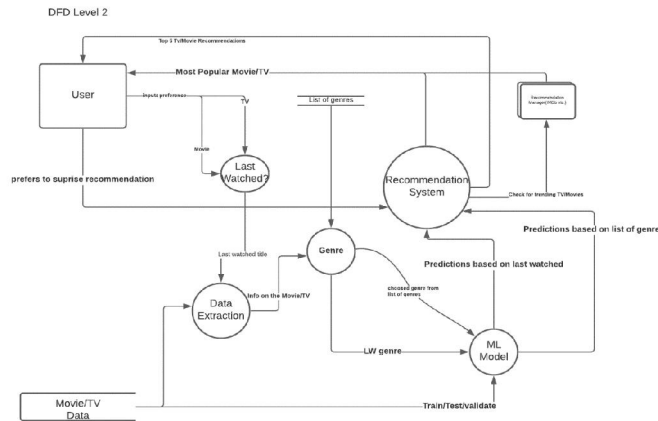
The above fig, shows the system architecture of the recommendation system.

B. Data Flow Diagrams

The mapping of the flow of the information among the application diagrammatically is said to be a data flow diagram. The flow given below is helped with symbols such as rectangles, arrows, circles etc. This diagram helps us to define the flow of our problem even more easily and expressive than the words.

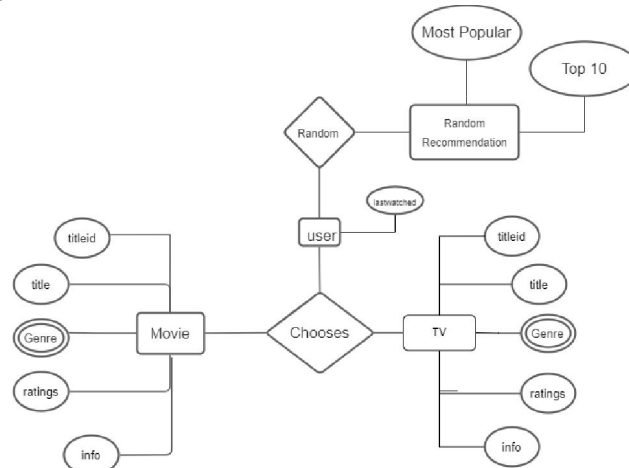


The above figure, explains the dataflow diagram Level 0 AND Level 1 of the recommendation system. It represents the flow of data from different components



The above figure, describes the Data Flow Diagram Level 1 & Level2, of the system.

**C. Entity Relationship Diagram**

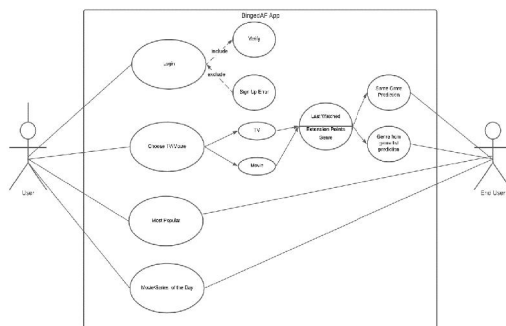


ER Diagram

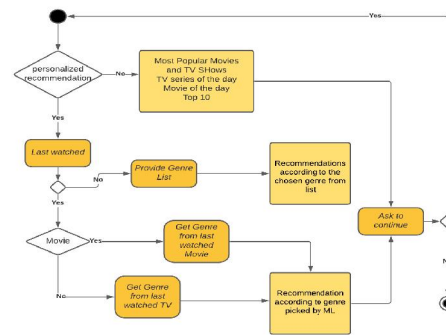
The above figure, explains the Entity Relationship Diagram of proposed recommendation system. The inter-connected components are shown with their attributes and relationship with each other.

**7.2 UML Diagrams**

**A. Use Case Diagram**



UML Diagram

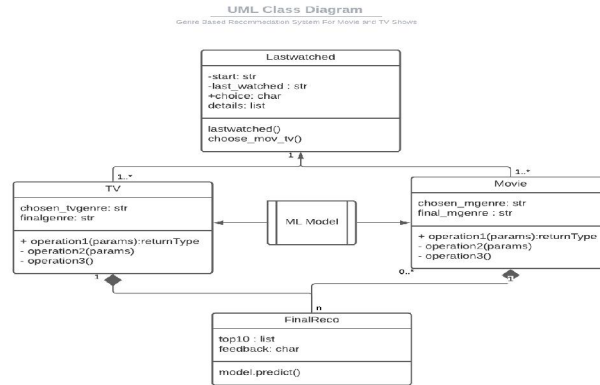


Activity Diagram





The above figure, explains the UML diagram of our recommendation system The diagram explains the components and their functions.



Activity Diagram

The above figure explains the activity diagram of our recommendation system It explains all the activities involved from the start to the end.

### 7.3 Project Plan

Every project tells a story about its goals, team, timing, and deliverables—and it requires detailed project planning and management to get the story right. Project planning is the process of defining the project scope, objectives, and steps needed to get the work done. It's one of the most important processes in project management. The output of the project planning process is a project management plan

#### A. Project Estimate

As our project focuses on the academic perspective, determining project estimates for costs, resources, and time needed was not that difficult job. A project estimate is an estimation of the time, cost, and resources needed to complete a project.

#### B. Reconciled Estimates

Estimate Reconciliation means the process of revising a previously Completed Package, such as through discussions with team members. Here, estimates were discussed whether to change the project if the cost factor was considered to buy the official API of the IMDb. The Time factor was reconciled according to the non-availability of the team members physically.

#### C. Project Resources

A resource is a necessary asset whose main role is to help carry out a certain task or project. A resource can be a person, a team, a tool, finances, and time. Most projects require many different resources to be completed. Resources should be assessed and allocated before a project begins. Poor resource planning can result in running out of resources midway through a project or delaying deadlines in delivering the final product or service.

The main resource for our project to execute was the data. The data resource was gathered with the help of data provide by the IMDb and their websites for scraping. The VCS like git was also another helpful resource in our project to make changes in a distributed way. We got the human resources as the team and the college guide along with the head of department providing us the resources needed. For the smooth running of our project we needed machines as a resource and good devices with proper suitable ram and rom. This material resource was met by all the team members. The time resources were also met because of the delay in examinations and the pandemic delay helped us to give more time for the project.

#### **D. Risk Management**

In project management, risk management is the practice of identifying, evaluating, and preventing or mitigating risks to a project that have the potential to impact the desired outcomes. Project managers are typically responsible for overseeing the risk management process throughout the duration of a given project at an industry level projects.

#### **E. Risk Identification**

Risk identification is the process of determining risks that could potentially prevent the program, enterprise, or investment from achieving its objectives. It includes documenting and communicating the concern. The Risks which were documented and identified are

1. The project is dependent on the data of IMDb which is not free. The average cost for API is 5000\$ per year
2. The Server of the websites which are used to scrap the data
3. The IMDb unofficial library Cinemagoer is not legally registers, so risk factor of termination a concern.
4. Skill level of team for this project shall be put to test.
5. Memory and Space for the user machine to run the project

#### **F. Risk Analysis**

Risk analysis is the process of identifying and analyzing potential issues that could negatively impact our projects. This process is done in order to help avoid or mitigate those risks. The above risks identified were analyzed and mitigated by-

1. How the data can be used if not from API, analysis of the IMDb dataset was done
2. Analysis of websites which were used for web scraping to check whether they were always running.
3. Cinemagoer python library was scanned if there was any limitations/termination issues
4. Skills of all the team members were analyzed through proper discussion.
5. Analysis of how much minimum memory and space does our project require.

### **VIII. OVERVIEW OF RISK MITIGATION, MONITORING, MANAGEMENT**

This can be divided into Risk Mitigation, Monitoring, and Management Plan (RMMM). In this plan, all works are done as part of risk analysis. As part of the overall project

#### **8.1 Risk Mitigation**

- Meet the current team members to determine causes for how can we get the data from relevant sources like IMDb.
- Mitigate those causes that are under our control before the project starts.
- Once the project commences, assume a dataset provided and develop techniques to ensure continuity by updating the data frequently
- Organize our team so that information about each development activity is widely dispersed.

#### **8.2 Risk Monitoring**

- General attitude of team members based on project pressures and how are they working on the specified risks.
- Interpersonal relationships among team members.
- Potential problems with duplicated data, websites down, not member availability etc.
- The availability of Cinemagoer and when does it tends to respond more.

#### **8.3 Risk Management**

Risk management and contingency planning assumes that mitigation efforts have failed and that the risk has become a reality. Continuing, the project is well underway, and data source for our project is costing us. If the mitigation strategy has been followed, backup dataset is available, information is documented, and knowledge has been dispersed across the team. In addition, temporarily refocus resources (and readjust the project schedule) to those functions that are fully independent of data, enabling genres to be listed for project to “get up to the speed “.



8.4 Project Schedule

Due to the extensive time slot available because of the pandemic, project got an ample amount of time to be completed. Here are some of the charts and timelines which were adhered to while performing the project

A. Project Task Set

PROJECT TASK LIST

TASK	PRIORITY	START	END	% COMPLETE	DONE	EST. HOU	ACTUAL HOURS	NOTES
Project Start 02-07-2021								
						Totals	Est.Hours 1500	Act.Hours 0
Genre Based Recommendation								
Problem Discussion	HIGH	05-07-2021	15-07-2021	100%	✓	120		
Right Questions	MEDIUM	15-07-2021	30-07-2021	100%	✓	120		
Data Gathering	MEDIUM	03-08-2021	02-09-2021	100%	✓	120		
Data Preprocessing	HIGH	05-09-2021	05-10-2021	100%	✓	120		
Data Visualization	MEDIUM	19-10-2021	18-11-2021	100%	✓	120		
Coding and Implementation	HIGH	25-11-2021	23-02-2022	100%	✓	720		
Testing	MEDIUM	06-03-2022	21-03-2022	100%	✓	120		
Maintenance	LOW	07-04-2022		30%	○	60		

Project Timeline



Project Task List

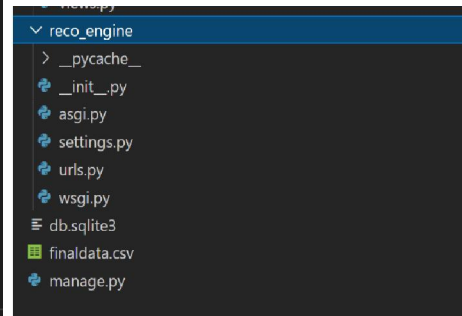
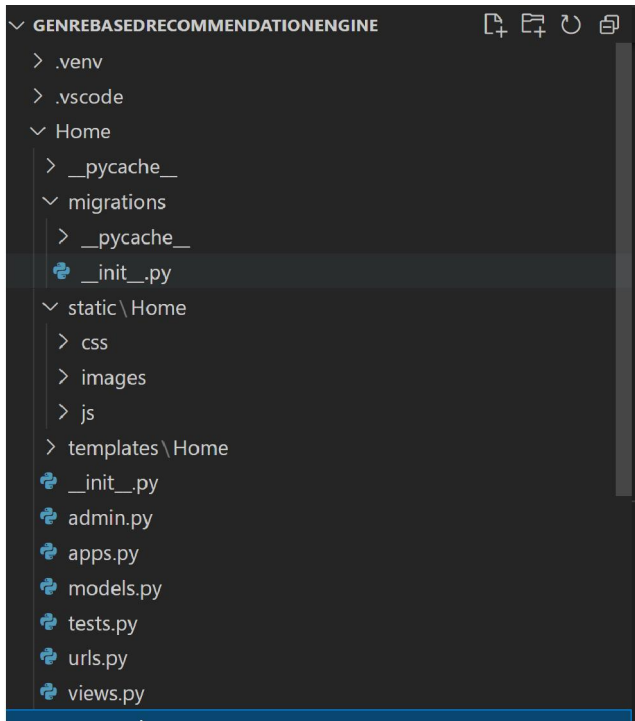
B. Team Organization

As guided by the professor we were asked to form a team of 4. We undoubtedly found four people with similar interests and skills for our project to succeed. Here is the overview of our team structure.

IX. PROJECT IMPLEMENTATION

9.1 Overview of Project Modules

A module is a collection of source files and build settings that allow you to divide your project into discrete units of functionality. Your project can have one or many modules, and one module may use another module as a dependency. You can independently build, test, and debug each module. Here is the image showing all the project files and modules which we have used





Here we have created a Django web app using python as the main language. Virtual environment is created and provided with all the suitable libraries. There are two apps. i.e. The main app which is the reco\_engine containing all the necessary server side codes. The configuration files like settings.py, url.py are present in this app. The project contains the main runnable which is manage.py which start the application

The ap within the main app is Home, which contains the necessary codes in the view.py file. All the url routing is done toward the main app by the urls.py in the secondary app. All the front end codes are present in the template folder containing html and static file in the static folder. The Data which is trained and validated is used as finaldata.csv which contains the data required for the model to get recommendations

## 9.2 Tools and Technologies Used

We have used python as the programming language for the implementation and used the following python libraries:

1. numPy -V 1.20.2, an open-source python package for N-dimensional arrays and numerical computations. It consists of several collection of classes which can be used to perform different mathematical operations
2. pandas -V 1.0.5, an open-source python library for fast, flexible operations and good tool for data manipulation and analysis. It is used to create data frames and perform operations on data frames.
3. scikit-learn -V 0.23.1, an open-source python tool used for simple and efficient tools for predictive analysis. It contains most of the machine learning algorithms in it.
4. regex -V 2020.6.8, an open-source library used to clean the text that contains the unwanted information in it.
5. matplotlib -V 3.2.2 is a comprehensive python library for creating interactive and animated visualizations in python.
6. requests -V 2.24.0 is used to get URL requests, no need to do manually.
7. pickle -V 0.7.5 is used for serializing and de-serializing binary protocols
8. seaborn, a python library for the visualization of the data.
9. Beautiful Soup 4.9.0 - Beautiful Soup is a Python library for pulling data out of HTML and XML files.
10. CINEMAGOER(Previously IMDbPY):is a Python package for retrieving and managing the data of the IMDb movie database about movies and people.

## 9.3 Software Testing

### A. Types of Testing

Software testing is the process of evaluating and verifying that a software product or application does what it is supposed to do. The benefits of testing include preventing bugs, reducing development costs and improving performance. Proper testing for the project is done with the help of unit testing. Types of Testing used are-

#### Integration testing:

Ensuring that software components or functions operate together.

Although each software module is unit tested, defects still exist for various reasons like

- A Module, in general, is designed by an individual software developer whose understanding and programming logic may differ from other programmers. Integration Testing becomes necessary to verify the software modules work in unity
- At the time of module development, there are wide chances of change in requirements by the clients. These new requirements may not be unit tested and hence system integration Testing becomes necessary.
- Interfaces of the software modules with the database could be erroneous
- External Hardware interfaces, if any, could be erroneous
- Inadequate exception handling could cause issues.

#### Unit testing:

Validating that each software unit performs as expected. A unit is the smallest testable component of an application. UNIT TESTING is a type of software testing where individual units or components of a software are tested. The purpose is to validate that each unit of the software code performs as expected. Unit Testing is done during the development (coding phase) of an application by the developers. Unit Tests isolate a section of code and verify its correctness. A unit



may be an individual function, method, procedure, module, or object. In SDLC, STLC, V Model, Unit testing is first level of testing done before integration testing. Unit testing is a WhiteBox testing technique that is usually performed by the developer. Though, in a practical world due to time crunch or reluctance of developers to tests, QA engineers also do unit testing.

#### **X. CONCLUSION & FUTURE WORK**

This project aims to give an agile system for recommending the users their own personalized recommendations. If this product is implemented, it will save a million-dollar worth time for the users who take out their precious time. Many Countries have already implemented this concept but this product will be a magic in India and for our own people. Our proposed system will be able to overcome the gap between the local and the global crowd.

For future works, this product could turn into a smooth-running startup with lots more to offer. The Plan would include Live Human-Based Recommendations, Weekly Watchlist, running ads on the product etc. The hope and the efforts will continue to make this project even more amazing than it ever was.

#### **APPENDIX A**

NP-hardness (non-deterministic polynomial-time hardness), in computational complexity theory, is the defining property of a class of problems that are informally "at least as hard as the hardest problems in NP". A simple example of an NP-hard problem is the subset sum problem.

A more precise specification is: a problem  $H$  is NP-hard when every problem  $L$  in NP can be reduced in polynomial time to  $H$ ; that is, assuming a solution for  $H$  takes 1 unit time,  $H$ 's solution can be used to solve  $L$  in polynomial time. As a consequence, finding a polynomial algorithm to solve any NP-hard problem would give polynomial algorithms for all the problems in NP, which is unlikely as many of them are considered difficult.

NP-complete problem, any of a class of computational problems for which no efficient solution algorithm has been found. Many significant computer-science problems belong to this class—e.g., the traveling salesman problem, satisfiability problems, and graph-covering problems. Hence, our problem is np-complete.

#### **REFERENCES**

- [1]. Comparative study of recommender system approaches and movie recommendation using collaborative filtering - Article in International Journal of Systems Assurance Engineering and Management · April 2021
- [2]. Prediction of diabetes disease among women using classification algorithm - D Sisodia, DS Sisodia - Procedia computer science, 2018
- [3]. Crime category prediction using classification algorithm N Ivan, E Ahishakiye, EO Omulo, D Taremwa – 2017
- [4]. Classification algorithm for prediction of kidney disease EHA Rady, AS Anwar - Informatics in Medicine Unlocked, 2019
- [5]. Performance of student prediction using data mining techniques AM Shahiri, W Husain - Procedia Computer Science, 2015
- [6]. Movie popularity prediction using machine learning techniques MH Latif, H Afzal - International Journal of Computer Science and ..., 2016
- [7]. Movie success prediction using machine techniques and comparisons R Sharda, D Delen - Expert Systems with Applications, 2006
- [8]. Sentiment analysis of movies using Genetic and Naives Bayes algorithm by M Govindarajan · Cited by 61
- [9]. Classification of movie into different genres based on audio and video features using Support Vector machine algorithm by YF Huang · 2012
- [10]. Classification of textual reviews using sentiment analysis.K Mouthami, KN Devi... - ... conference on Information ..., 2013
- [11]. Predicting the IMDB ratings of movies using ensemble learning and classification algorithm. 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)