

# **Movie Recommender System**

**A project Report**

*Submitted by*

**NITESH RAMESH MOREM**

**190770107550**

*In partial fulfilment for the award of the degree of*

**BACHELOR OF ENGINEERING**

**In**

**Computer Department**

**Silver Oak College of Engineering & Technology,**

**Ahmedabad-380061**



**Gujarat technological University,**

**Ahmedabad-382424**

**April 2023**



## Silver Oak College of Engineering & Technology

Ahmedabad-380061

### CERTIFICATE

This is to certify that the project report submitted along with the project entitled **Movie Recommender System** has been carried out by **Nitesh Morem (190770107550)** under my guidance in partial fulfilment for the degree of Bachelor of Engineering in Computer Engineering, 8th Semester of Gujarat Technological University, Ahmedabad during the academic year 2022-23.

Prof. Ramya Kumar

Dr. Satvik Khara

Internal Guide

Head of Department



Maxgen Technologies Pvt. Ltd.



Date: 26 Apr 2023

**INTERNSHIP COMPLETION LETTER**

This is to certify that **Mr. Nitesh Ramesh Morem**, Enroll no: 190770107550 studying in BE Computer Engineering Semester VIII, student of Silver Oak College of Engineering and Technology has completed his Internship in **Data Science** in our organization as a part of his curriculum from **30th Jan 2023 to 21th April 2023**.

He has successfully completed his Internship. During the period he was sincere, hardworking & fully devoted to project.

We wish him all success to his future career.

**Maxgen Technologies Pvt. Ltd.****Sadhana Kokate ( HR Manager )****Web Development | E-commerce | ERP | SEO Services | Industrial Training | Mobile App Development**[www.maxgentechnology.com](http://www.maxgentechnology.com) [www.maxgentechnologies.com](http://www.maxgentechnologies.com)

9099039845 | 9579972245

[info@maxgentechnologies.com](mailto:info@maxgentechnologies.com)



## Silver Oak College Of Engineering & Technology

Ahmedabad-380061

### DECLARATION

I hereby declare that the Internship report submitted along with the Internship entitled **Movie Recommender System** submitted in partial fulfilment for the degree of Bachelor of Engineering in Computer Engineering to Gujarat Technological University, Ahmedabad, is a bona fide record of original project work carried out by me at Maxgen Technologies under the supervision of Internal Guide:- Prof. Ramya Kumar & External Guide:- Mr. Arpan Patel and that no part of this report has been directly copied from any students' reports or taken from any other source, without providing due reference.

.

Nitesh Morem (190770107550)

Name of Student

Sign of Student

## ACKNOWLEDGEMENT

I would like to acknowledge the contributions of the following people without whose help and guidance the report would not have been completed. I acknowledge the counsel and support of our training coordinator, **Prof. Ramya Kumar**, with respect and gratitude, whose expertise, guidance, support, encouragement, and enthusiasm has made this report possible. Their feedback vastly improved the quality of this report and provided an enthralling experience. I am indeed proud and fortunate to be supported by her. I am also thankful to **Dr. Satvik Khara**, H.O.D of Computer Science Engineering Department, Silver Oak College of Engineering and Technology, Ahmedabad, for his constant encouragement, valuable suggestions and moral support and blessings.

Although it is not possible to name everyone individually, I shall ever remain indebted to the staff and Developers of Maxgen technologies their persistent support and cooperation extended during this work. This acknowledgement will remain incomplete if I fail to express my deep sense of obligation to my parents and God for their consistent blessings and encouragement

Nitesh Morem

(190770107550)

## ABSTRACT

*The movie recommendation system is an essential part of the entertainment industry, and it can be developed using various techniques such as collaborative filtering, content-based filtering, and hybrid filtering. In this project, we developed a content-based filtering movie recommendation system using machine learning algorithms and Python-based tools such as Pandas, NumPy, Scikit-learn, and Streamlit. The system takes user input in the form of movie titles and recommends similar movies based on features such as genre, cast, and director.*

*The main objective of this internship project is to provide hands-on experience in developing a data science project, starting from data pre-processing and feature engineering to the deployment of the final product. This internship also aims to enhance the participant's knowledge and skills in Python programming, machine learning algorithms, data pre-processing techniques, and deployment tools such as Streamlit.*

## List of Figures

Fig 1 Organization chart .....	2
Fig 2 Software development lifecycle .....	5
Fig 3 System snapshot.....	16
Fig 4 Types of system .....	17
Fig 5 Output of system.....	19
Fig 6 Input of system .....	20
Fig 7 Anaconda Navigator .....	21
Fig 8 Jupyter Notebook.....	22
Fig 9 Streamlit.....	22
Fig 10 Project Workflow .....	29

## **List of Tables**

Table 1 Internship Work Plan .....	12
------------------------------------	----



## **Abbreviations**

AI – Artificial Intelligence

CS- Cosine Similarity

ML- Machine learning

CLI – Command Line Interpreter

QA – Quality Assurance

API – Application Process Interface

TMDB – The Movie DataBase

## Table of Contents

Acknowledgement.....	i
Abstract.....	ii
List of Figures.....	iii
List of Tables.....	iv
List of Abbreviations.....	v
Table of Contents.....	vi
<b>Chapter 1 Overview of the company.....</b>	<b>1</b>
1.1 Overview.....	1
1.2 Scope of work.....	1
1.3 Organization chart.....	2
<b>Chapter 2 Overview of Different Process being carried out in company.....</b>	<b>3</b>
2.1 Work being carried out in each department.....	3
2.2 Technical Specification of Major Equipment in each department.....	4
2.3 Sequence of each stage of End product.....	5
2.4 Details of each stage of production.....	6
<b>Chapter 3 Introduction to Internship and Project.....</b>	<b>8</b>
3.1 Introduction to Internship and Project.....	8
3.2 Purpose.....	8
3.3 Objective.....	9
3.4 Scope.....	10
3.5 Technology and literature review.....	10
3.6 Project Planning.....	11
3.6.1 Project/Internship Development Approach and Justification.....	11
3.6.2 Roles and Responsibilities.....	12
<b>Chapter 4 System Analysis.....</b>	<b>13</b>
4.1 Study of current system.....	13

4.2 Problem and weakness of current system.....	13
4.3 Requirements of New system.....	14
4.4 System Feasibility.....	15
<b>Chapter 5 System design.....</b>	<b>16</b>
5.1 System Design and Methodology.....	16
5.2 Structure of system.....	17
5.3 Input/Output Interface design.....	19
<b>Chapter 6 Implementation.....</b>	<b>21</b>
6.1 Implementation Platform / Environment.....	21
6.2 Modules and technological specifications.....	23
6.3 Outcome.....	24
6.4 Result Analysis.....	24
<b>Chapter 7 Testing .....</b>	<b>25</b>
7.1 Testing strategy.....	25
7.2 Test Results and Analysis.....	26
<b>Chapter 8 Conclusion and Discussion.....</b>	<b>27</b>
8.1 Overall Analysis of Internship/Project viability.....	27
8.2 Dates of Continuous Evaluation.....	27
8.3 Problems Encountered and Possible Solutions .....	28
8.4 Summary of Internship/ Project Work.....	29
8.5 Limitations and future enhancement.....	30
<b>References.....</b>	<b>31</b>

## **CHAPTER 1 – OVERVIEW OF THE COMPANY**

### **1.1 OVERVIEW**

Maxgen Technologies is a software development company based in Ahmedabad, India. The company offers a range of IT services, including web development, mobile application development, e-commerce solutions, digital marketing, and more. Maxgen Technologies has a team of experienced developers and designers who use cutting-edge technologies to create high-quality software solutions for their clients. The company has worked with clients from various industries, including healthcare, finance, education, and retail. One of the key strengths of Maxgen Technologies is its focus on innovation and problem-solving.

The company works closely with its clients to understand their unique needs and challenges, and then develops customized solutions to address those needs. In addition to its software development services, Maxgen Technologies also offers training programs for students and professionals. The company provides training in various technologies, including Java, Python, PHP, and Android, among others. Overall, Maxgen Technologies is a leading software development company in Ahmedabad, known for its innovative solutions, high-quality work, and commitment to customer satisfaction.

Founded in 2014, Maxgen Technologies has grown to become one of the top software development companies in Ahmedabad, with a team of over 50 employees. The company has a strong focus on quality and customer satisfaction. It follows a rigorous development process that includes testing and quality assurance to ensure that its products meet high standards.

### **1.2 SCOPE OF WORK**

- E-commerce Development

They assist you in creating a flawless E-Commerce experience for your clients in your stores using knowledge in E-Commerce platforms like Magento and WooCommerce.

- Custom Application Development

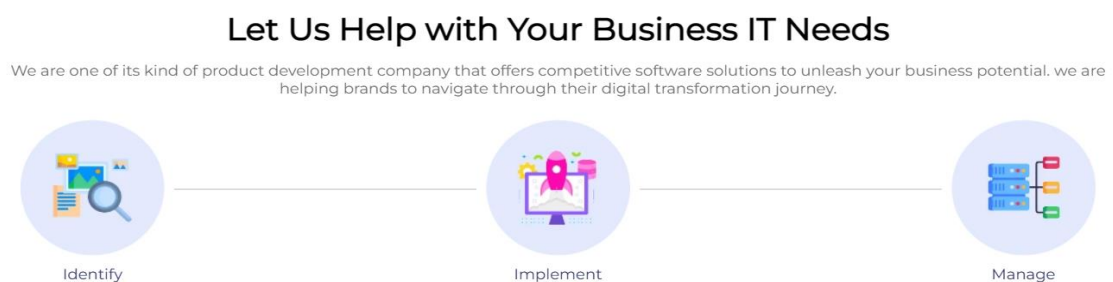
They can do it quickly & cost-effectively, whether you need to construct a bespoke web app from the start, move your old backend, or simplify current front-end functionality. We provide document management solutions, secure intranet, extranet, portal setup, application conversion to the cloud, collaboration and portal creation, corporate intelligence solutions, and maintenance, support, and upgrade services.

- Web Application Support and Maintenance

Also collaborate with you to manage web apps created by outside suppliers or by us. Our skilled engineers can easily handle any task, whether repairing problems or introducing new capabilities. We also track all of your future requirements and offer assistance. Provide, create and deploy web apps from our team of committed web application developers to help you increase operational effectiveness, hasten decision-making, and acquire a competitive edge. We constantly prioritize the applications scalability and security and modify Micro Services to follow the security standards.

We have created a benchmark in the web app development industry by utilizing the most recent web technologies to meet various project needs. Additionally, we devote ourselves entirely to the project's technical and commercial aspects. We often offer the finest inside the expected period by keeping communication open and adhering to the Agile strategy.

### 1.3 ORGANIZATION CHART



**Figure 1**  
**(maxgentechologies.com)**

## CHAPTER 2 – OVERVIEW OF DIFFERENT PROCESS BEING CARRIED OUT IN COMPANY

### 2.1 WORK BEING CARRIED OUT IN EACH DEPARTMENT

- **Business Analysis Department:** This department is responsible for analyzing clients' requirements and determining the feasibility of a project. They work closely with clients to understand their business needs and translate them into technical specifications. The department also prepares project.
- **UX/UI Design Department:**  
This department creates mockups, wireframes, and prototypes of the software user interface. They work closely with the Business Analysis team to understand the user requirements and design the interface that meets those requirements. The UX/UI design team also ensures that the interface is user-friendly, intuitive, and aesthetically pleasing.
- **Development Department:**  
The Development Department is responsible for building the software. They use various programming languages, frameworks, and tools to develop the software according to the specifications provided by the Business Analysis and UX/UI design teams. The Development team also ensures that the software is scalable, reliable, and meets the quality standards.
- **Development Department:**  
The Development Department is responsible for building the software. They use various programming languages, frameworks, and tools to develop the software according to the specifications provided by the Business Analysis and UX/UI design teams. The Development team also ensures that the software is scalable, reliable, and meets the quality standards.

- **Deployment and Hosting Department:**

This department is responsible for deploying the software to the client's server or hosting it on a cloud platform like AWS, Azure, or Google Cloud. The Deployment and Hosting team also configures the server, sets up the database, and ensures that the software is accessible to the end-users.

- **Digital Marketing Department:**

This department is responsible for promoting the client's software or website through various digital marketing channels like SEO, PPC, social media marketing, email marketing, and content marketing. They create marketing plans, strategies, and campaigns to drive traffic and increase conversions.

- **Training Department:**

Maxgen Technologies also has a Training Department that offers training programs in various technologies like Java, Python, PHP, Android, and more. They provide classroom training, online training, and corporate training to students and professionals to help them develop their skills and stay up-to-date with the latest trends in technology.

## **2.2 TECHNICAL SPECIFICATIONS OF MAJOR EQUIPMENTS IN EACH DEPARTMENT**

- Computers/laptops with software tools for project management, documentation, and analysis.
- Projector/screens for presentations and meetings.
- Communication tools such as telephones and video conferencing software.
- Computers/laptops with programming software such as Visual Studio, Eclipse and NetBeans.
- Testing and debugging tools such as JUnit, Selenium, and PyCharm.
- Version control software such as Git and SVN.

## 2.3 SEQUENCE OF EACH STAGE OF END PRODUCT

The process of developing a software is done via following the basic SDLC rule which is proved to be most efficient while developing a software. Hence, the company uses the same SDLC technique for developing a software

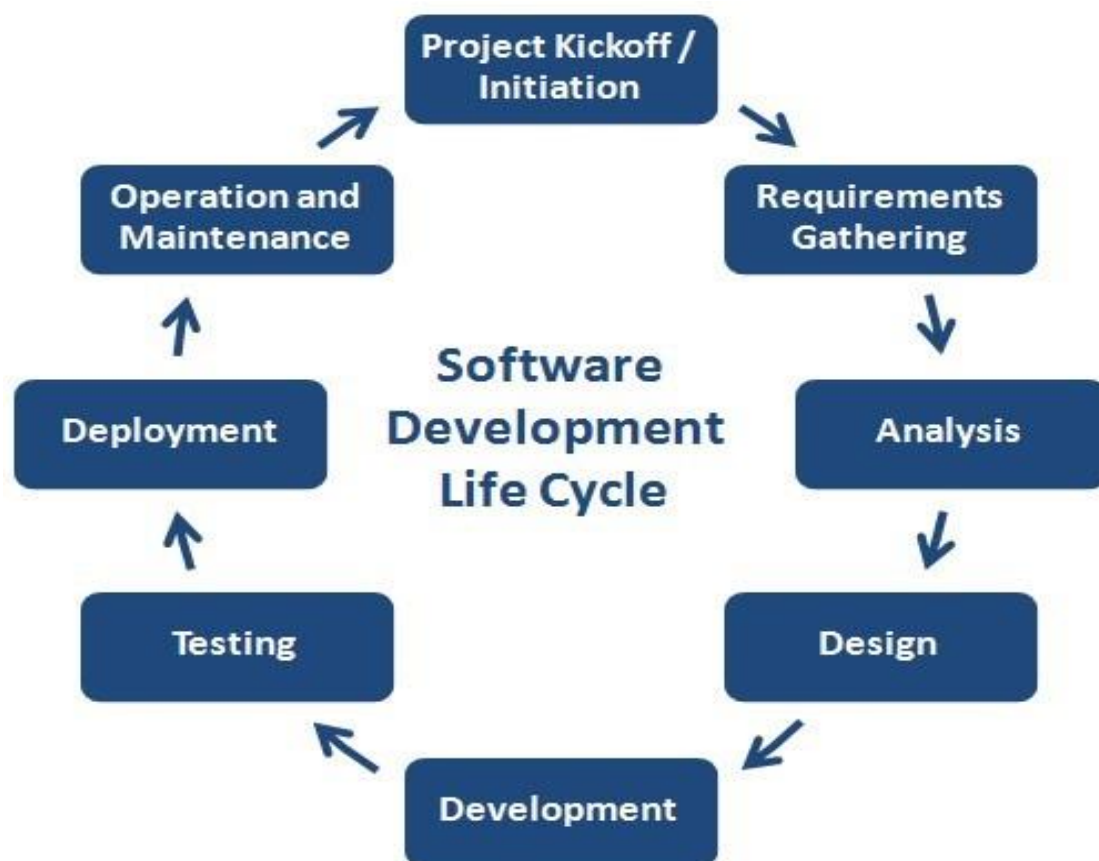


Figure 2

<https://medium.com/@96kavindugayantha/sdlc-b443e449b854>

The above given process is a schematic representation of the sequence of stages that the software development team undertakes to produce a software. It is basically a SDLC with a use of differently chosen words.



## 2.4 DETAILS ABOUT EACH STAGE OF PRODUCTION

- Planning and Analysis:

The first step in developing an end product is to understand the client's requirements, identify the scope of the project, and create a project plan. This involves conducting meetings with the client, gathering requirements, analysing the feasibility of the project, and creating a detailed project plan that includes timelines, budgets, and resource allocation.

- Design and Prototyping:

Once the project plan is in place, the design and prototyping phase begins. The UX/UI design team creates mock-ups, wireframes, and prototypes of the software user interface. The design team works closely with the Business Analysis team to ensure that the interface meets the user requirements.

- Development:

The next step is to develop the software. The Development team writes code using programming languages, frameworks, and tools to develop the software according to the specifications provided by the Business Analysis and UX/UI design teams. The Development team also ensures that the software is scalable, reliable, and meets the quality standards.

- Testing:

After the software is developed, the Quality Assurance team tests it to ensure that it meets the quality standards and is free of defects. The QA team creates test plans, test cases, and test scripts to verify the functionality of the software. They also perform load testing, stress testing, and security testing to ensure that the software performs well under different conditions.

- Deployment:

Once the software has passed the testing phase, it is ready for deployment. The Deployment and Hosting team deploys the software to the client's server or hosts it on a cloud platform like AWS, Azure, or Google Cloud. The Deployment and

- Maintenance and Support:

After the software is deployed, the Maintenance and Support team provides ongoing support and maintenance to ensure that the software continues to function properly. This involves monitoring the software, resolving any issues that arise, and providing updates and upgrades as needed.

- Training and Documentation:

The final step is to provide training and documentation to the end-users. The Training team provides classroom training, online training, and corporate training to users and professionals to help them use the software effectively. The Documentation team creates user manuals, technical documentation, and other materials to support the users of the software

## **CHAPTER 3 - INTRODUCTION TO INTERNSHIP AND PROJECT**

### **3.1 INTRODUCTION**

A movie recommendation system is a data science project that aims to provide personalized recommendations to users based on their viewing history, preferences, and behavior. This system uses machine learning algorithms and statistical models to analyze user data and generate a list of movies that are most likely to appeal to the user. The recommendation system typically uses collaborative filtering techniques, which involve analyzing user behavior to identify patterns and similarities between users, and using this information to make personalized recommendations. Other techniques that can be used include content-based filtering, which analyzes the attributes of movies to identify similar movies, and hybrid filtering, which combines collaborative and content-based filtering techniques.

To develop a movie recommendation system, data scientists need to collect and preprocess data, build and train machine learning models, and deploy the system to production. The system should be regularly updated and improved based on user feedback and changing trends in the movie industry. Movie recommendation systems are widely used by online movie streaming services, such as Netflix and Amazon Prime Video, to provide users with personalized recommendations and improve their overall viewing experience.

### **3.2 PURPOSE**

The purpose of the above system project is to develop a content-based filtering system that can recommend relevant movies to users based on their preferences. The system will utilize data processing and feature engineering techniques to extract meaningful features from the movie dataset. The extracted features will be used to build a machine learning model that can predict a user's movie preferences. The system will also employ label encoding to convert categorical data into numerical form for easy analysis. The objective of the internship project is to provide hands-on experience to the intern in developing a machine

learning-based recommendation system and to enhance their skills in data processing and feature engineering. This project also aims to provide valuable insights to businesses by analyzing user behavior and preferences, helping them to make informed decisions and optimize their product offerings. Through this project, I will gain practical experience in data science, machine learning, and software development, as well as improve problem-solving and critical thinking skills.

### **3.3 OBJECTIVE**

The scope of a movie recommendation system includes designing and implementing a personalized recommendation engine that can analyze user data, such as viewing history and behavior, to generate accurate and diverse movie recommendations. The system should be able to handle large amounts of data and incorporate additional information, such as social network data and movie content information, to enhance recommendation accuracy. Moreover, the movie recommendation system should have an intuitive and user-friendly interface that enables users to explore recommended movies easily. It should provide additional features such as movie trailers, reviews, and ratings to enhance user experience. Additionally, the system should ensure user privacy and data security by implementing appropriate measures to protect user data.

The objectives of a movie recommendation system are to enhance user engagement and satisfaction by providing personalized and diverse movie recommendations, increase user retention and loyalty, and improve revenue for movie streaming services through increased user activity and subscriptions. The system should also aim to overcome the cold start problem, where a new user has no or limited data, by incorporating additional information such as social network data and movie content information. Continuous monitoring and improvement of the system based on user feedback and data analysis should also be a key objective.

### **3.4 SCOPE**

The scope of a movie recommendation system includes designing and implementing a personalized recommendation engine that can analyze user data, such as viewing history and behavior, to generate accurate and diverse movie recommendations. The system should be able to handle large amounts of data and incorporate additional information, such as social network data and movie content information, to enhance recommendation accuracy.

Moreover, the movie recommendation system should have an intuitive and user-friendly interface that enables users to explore recommended movies easily. It should provide additional features such as movie trailers, reviews, and ratings to enhance user experience. Additionally, the system should ensure user privacy and data security by implementing appropriate measures to protect user data.

### **3.5 Technology and Literature review**

Movie recommendation systems have become an important area of research in recent years due to the increasing demand for personalized movie recommendations. Several studies have been conducted to develop and improve movie recommendation systems.

One of the most widely used approaches in movie recommendation systems is collaborative filtering. Content-based filtering analyses the attributes of movies, such as genre, actors, and directors, to recommend movies that are similar to the ones that the user has already watched. Content-based filtering analyses movie attributes to recommend similar movies. Hybrid filtering combines both approaches to improve recommendation accuracy.

Recent research has focused on improving the performance of recommendation systems by incorporating additional information, such as social network data and deep learning models. Deep learning models have shown significant improvement in recommendation accuracy. Overall, movie recommendation systems are an important area of research in data science, with several approaches being used to improve their performance.

### **3.6 Project Planning**

The project planning involved the identification of the problem of content filtering and the development of a system that uses basic and content-based filtering techniques to provide relevant recommendations to users. The project team created a detailed project plan with a timeline, milestones, and specific tasks to be completed. The project plan also identified the tools and technologies required for the project, such as Python, Streamlit, and Kaggle. The team also conducted a literature review to identify previous research on content filtering and recommendation systems, which informed the development of the system. The project planning process was crucial in ensuring that the project was completed within the specified timeline and budget, and that the final system met the specified requirements.

#### **3.6.1 Project/Internship Development Approach And Justification**

The project/internship development approach and justification involve outlining the process and reasoning behind the chosen methodology for the project. It aims to provide a clear understanding of the steps involved in the project and the justification for the chosen approach. This includes the selection of appropriate tools and technologies, a review of relevant literature, and the establishment of objectives and goals for the project.

The approach is centered on data pre-processing and feature engineering using Python and various libraries, followed by content-based filtering to create personalized recommendations for users. The justification for this approach lies in the potential for improved accuracy and relevance of recommendations, as well as the practical applications of content-based filtering in the real world.

The development approach for the content-based filtering system used in the movie recommendation project involved several steps. First, the team identified the key features of movies that would be used to generate recommendations. Then, they analysed the movie database to assign weights to each feature based on its relevance to the movie's overall content. Overall, the development approach for the content-based filtering system involved a combination of data analysis, software development, testing, and continuous improvement.

The Table below shows the work plan of the internship

	<b>FEBRUARY</b>	<b>MARCH</b>	<b>APRIL</b>
<b>WEEK 1</b>	Intro to python	Fundamentals of statistics	Backend
<b>WEEK 2</b>	Python modules	Data cleaning	Data Analysis
<b>WEEK 3</b>	Python libraries	Data Pre-processing	Streamlit
<b>WEEK 4</b>	Machine learning	Kaggle research	Deployment
<b>WEEKENDS</b>	Theory of Data science	Kaggle projects	Discussions
<b>DAILYWORK</b>	8 hours	8 hours	8 hours

Table 1 Internship Work Plan

### 3.6.2 Roles And Responsibilities

- **Communication:** Communicating with the project supervisor and team members to ensure timely completion of tasks and resolution of issues.
- **Reporting:** Preparing progress reports and presenting them to the project supervisor.
- **Requirement analysis:** Conducting a detailed analysis of project requirements and ensuring that they are met.
- **Coding:** Writing clean and efficient code to implement the machine learning algorithms.
- **Testing:** Testing the code to ensure that it meets the project requirements and is bug-free.
- **Documentation:** Documenting the code and the entire project to ensure that it can be easily maintained and updated in the future.

## **CHAPTER 4 - SYSTEM ANALYSIS**

### **4.1 STUDY OF CURRENT SYSTEM**

As the project is focused on developing a content-based filtering system for recommending online courses, a study of current systems was conducted. The study aimed to explore existing content-based filtering methods and systems used in the domain of online course recommendation. Through the review of relevant literature, it was found that content-based filtering is a widely used approach in online course recommendation systems due to its simplicity and effectiveness. Various techniques such as cosine similarity, and collaborative filtering were found to be commonly used in content-based filtering systems. However, some limitations were identified such as the need for a large amount of data and the inability to capture the user's evolving interests. Overall, the study provided valuable insights into the existing systems and techniques that could be incorporated into the development of the project.

### **4.2 PROBLEM AND WEAKNESS OF CURRENT SYSTEM**

- **Limited diversity:** Content-based filtering only recommends items that are similar to what a user has already liked or interacted with, which can limit the diversity of recommendations. It may not offer new or novel items to the user, leading to a lack of exploration and discovery.
- **Overfitting:** In some cases, content-based filtering may overfit to a user's previous preferences, resulting in recommendations that are too similar to previous items and not accounting for a user's changing tastes over time.
- **Cold start problem:** Content-based filtering may struggle to make recommendations for new users or new items that have limited data available about them. This is because it relies on analysing item attributes to make recommendations, and new items may not have enough data to be effectively analysed.



- Limited ability to capture context: Content-based filtering may not be able to capture the context in which an item is being recommended, such as the time of day or a user's current mood, which could impact their preference for a particular item
- Limited ability to capture user preferences: Content-based filtering is limited to the information that is available about an item's attributes, such as genre or topic. It may not be able to capture more nuanced preferences or tastes that a user may have, such as a preference for a specific actor or director.

### **4.3 REQUIREMENTS OF NEW SYSTEM**

- Improved accuracy: The new system should aim to improve the accuracy of the content-based filtering algorithm to ensure better recommendations.
- Scalability: The new system should be able to handle large amounts of data and be scalable to meet the growing needs of the platform.
- Real-time recommendations: The new system should be able to provide real-time recommendations to users based on their preferences.
- Improved user experience: The new system should focus on improving the user experience by providing a more personalized and relevant content recommendation system.
- Robustness: The new system should be robust and able to handle various types of data, including different formats and languages.
- Easy to maintain: The new system should be easy to maintain and update, with efficient coding and documentation for future development and upgrades.
- Cost-effective: The new system should be cost-effective and have a reasonable cost-to-benefit ratio for the organization.

## 4.4 SYSTEM FEASIBILITY

System feasibility refers to the ability of a proposed system to meet the needs of the users and the organization in terms of technical, operational, economic, and schedule feasibility. Feasibility study is carried out based on many purposes to analyse whether software product will be right in terms of development, implantation, contribution of project to the organization etc.

- Technical feasibility determines whether the proposed system can be implemented using the available technology and resources. This includes assessing the compatibility of the new system with existing systems and infrastructure.
- Operational feasibility examines whether the new system can be used efficiently and effectively in the current operational environment. This includes evaluating the skills of the users, the impact on the workflow, and the ease of use of the new system.
- Economic feasibility evaluates the costs and benefits of implementing the new system. This includes considering the initial investment, ongoing maintenance and operational costs, as well as the potential benefits in terms of increased productivity, reduced errors, and other factors.
- Schedule feasibility determines whether the project can be completed within the proposed timeline. This includes considering factors such as resource availability, project scope, and potential delays or risks.

Overall, system feasibility is an important consideration in determining the viability of a proposed system and ensuring its success in meeting the needs of the organization and its users

## CHAPTER 5 - SYSTEM DESIGN

### 5.1 SYSTEM DESIGN AND METHODOLOGY

System design plays a very crucial role in an application. It is the process of defining the components, modules, interfaces, and data for a system to satisfy the specified requirements. The desktop application (current system) provided by the company had a very distinctive as well as a user-friendly design. Some of the snapshots of the current system are as follow:

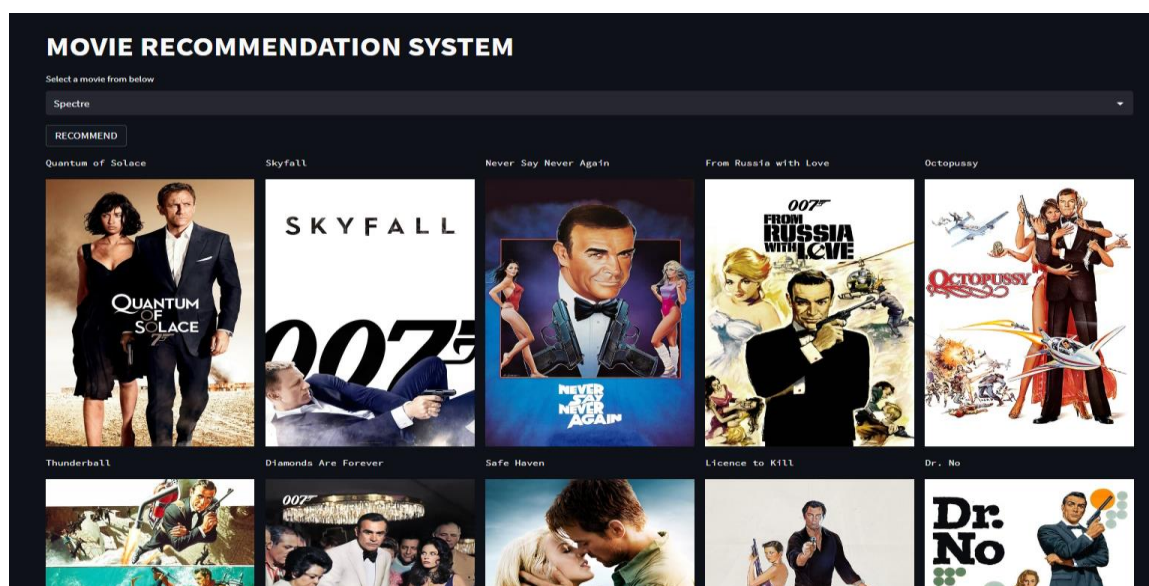


Figure 3 System snapshot

As it can be seen in the above figure, the movie with the poster and its name are being retrieved successfully.

I used a API from the TMDB website from where the posters along with their names are fetched to the current system. Initially there were a lot of errors due to the poster while the names were being retried successfully but posters were a problem as once the site goes under maintenance the posters will not be fetched which is a big limitation provided which is a small-scale system development project for specific data.

## 5.2 STRUCTURE OF SYSTEM

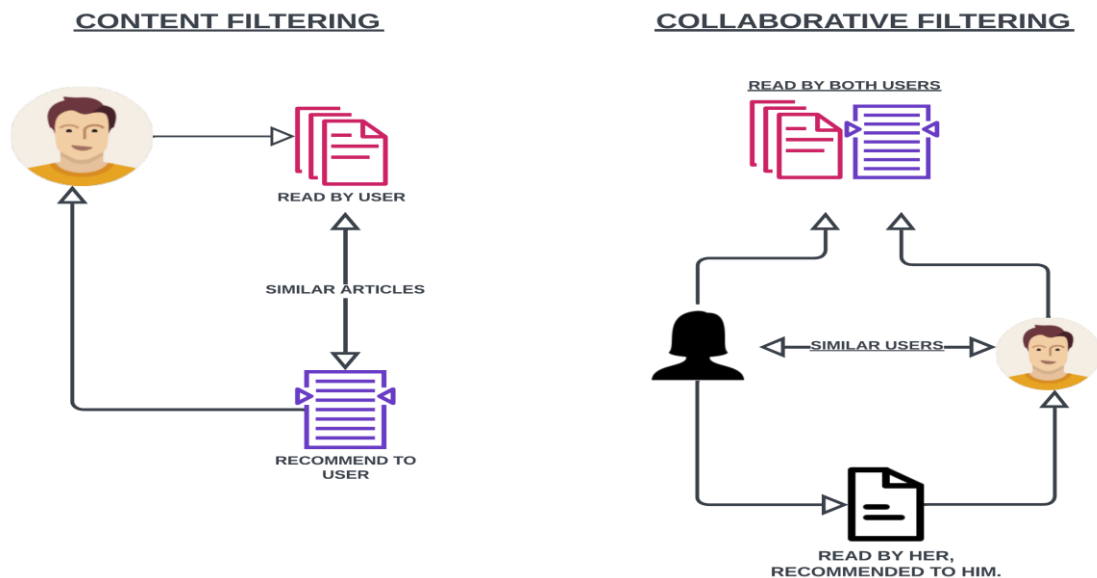


Figure 4 Type of system

Content-based filtering and collaborative filtering are two popular techniques used in recommender systems to provide personalized recommendations to users.

- **Content-based filtering:**

Content-based filtering is a technique where recommendations are based on the attributes or properties of the items being recommended. In this approach, the system recommends items that are similar to the items the user has interacted with in the past. The attributes used for recommendation can include features such as genre, author, artist, director, and so on.

For example, in a movie recommendation system, if the user has rated action movies highly in the past, the system will recommend other action movies with similar attributes such as the same director, actors, or genre.

- **Collaborative filtering:**

Collaborative filtering is a technique where recommendations are based on the preferences of other users who have similar preferences to the target user.

For example, in a movie recommendation system, if the target user has rated some movies highly, the system will find other users who have rated the same movies highly and recommend other movies that those users have also rated highly. This approach relies on the assumption that users who have similar preferences in the past are likely to have similar preferences in the future.

Both techniques have their own advantages and disadvantages, and their effectiveness depends on the type of data available and the specific application domain. Content-based filtering is generally better suited for recommending items that have clear and well-defined features or attributes, while collaborative filtering is better suited for recommending items that are highly subjective and difficult to define in terms of attributes.

I used content-based filtering system because it is best for small scale projects due to limited data and metadata. The decision to use content-based filtering or collaborative filtering, or a combination of both, in a recommender system depends on the specific requirements of the system and the nature of the data available. If the system has a large amount of item-related data, such as attributes or properties of the items, then content-based filtering may be a good choice. Content-based filtering is effective in recommending items that are similar to those that the user has interacted with in the past. It can be particularly useful when the user's preferences are well-defined and based on specific features or attributes of the items.

For example, in a movie recommendation system, if the system has a lot of data on the movie's attributes such as the director, genre, cast, and plot, then content-based filtering can be a good choice. The system can recommend movies with similar attributes to those that the user has liked in the past. However, if the system has limited item-related data or if the user's preferences are not well-defined, then collaborative filtering may be a better choice. Collaborative filtering can be effective in recommending items that are popular among users who have similar tastes to the target user.

In summary, the decision to use content-based filtering or collaborative filtering, or a combination of both, depends on the specific requirements of the system and the nature of the data available. In some cases, content-based filtering may be a good choice, while in others, collaborative filtering may be more effective.

### 5.3 INPUT/OUTPUT INTERFACE DESIGN

The movie recommendation system can be divided into several modules, each of which performs a specific function. The following are the key modules in a movie recommendation system:

- Data Collection Module: This module is responsible for collecting and storing user data, including viewing history, ratings, and user behavior. It can also gather additional data from social media platforms and other sources to enhance the recommendation accuracy.
- Preprocessing Module: The preprocessing module cleans and preprocesses the user data before feeding it to the recommendation engine. This module may involve data cleaning, normalization, and feature extraction.
- Recommendation Engine: The recommendation engine is the core module of the system, which generates movie recommendations based on user data and behavior. It uses various filtering techniques, such as collaborative filtering, content-based filtering, and hybrid filtering, to generate personalized recommendations for each user.
- Post-processing Module: This module is responsible for fine-tuning the recommendation results generated by the recommendation engine. It may involve filtering out irrelevant recommendations, enhancing recommendation diversity, and optimizing recommendation delivery.

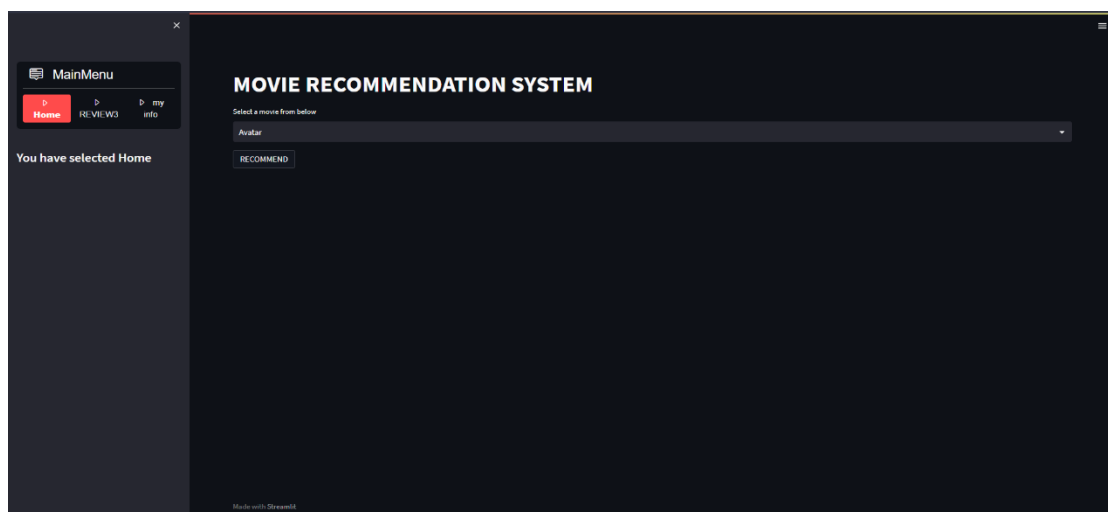


Figure 5 Output of system

- User Interface Module: The user interface module provides an intuitive and user-friendly interface for users to interact with the system. It may include features such as search, movie trailers, reviews, and ratings.

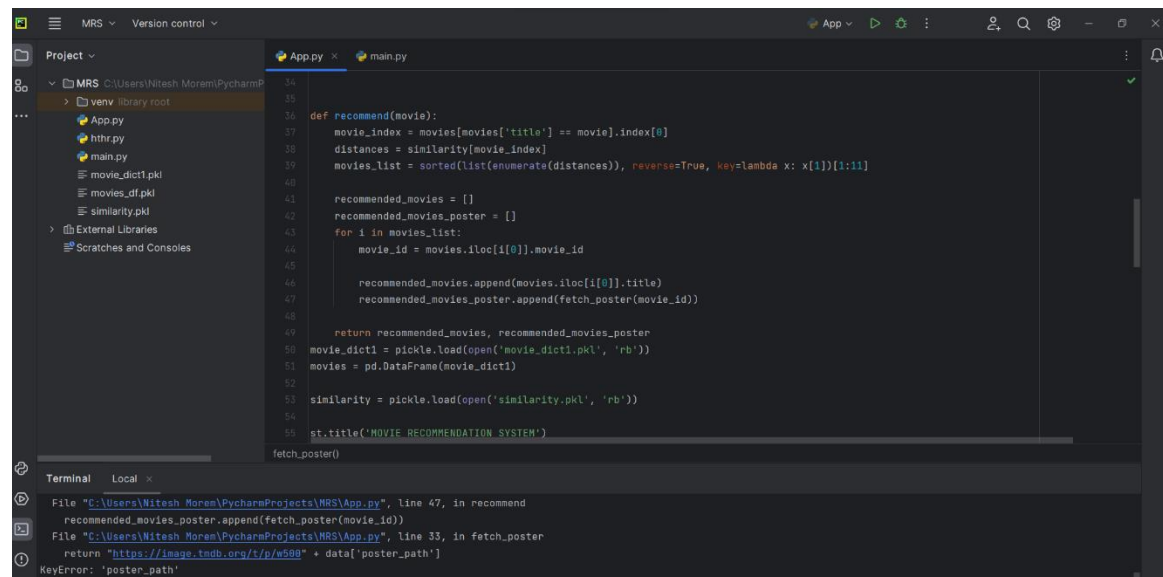


Figure 6 Input of System

The system is deployed and created on streamlit which is a very popular web application creator for data science projects. Streamlit is a free and open-source framework to rapidly build and share beautiful machine learning and data science web apps. It is a Python-based library specifically designed for machine learning engineers. Data scientists or machine learning engineers are not web developers and they're not interested in spending weeks learning to use these frameworks to build web apps. Instead, they want a tool that is easier to learn and to use, as long as it can display data and collect needed parameters for modelling. Streamlit allows you to create a stunning-looking application with only a few lines of code.

I used PyCharm as an IDE which is integrated development environment is a software application that provides comprehensive facilities to computer programmers for software development. An integrated development environment is a software application that provides comprehensive facilities to computer programmers for software development

## CHAPTER 6 - IMPLEMENTATION

### 6.1 IMPLEMENTATION PLATFORM/ENVIRONMENT

Python is a popular programming language that is commonly used for data science projects. It has several libraries and frameworks, such as Pandas, NumPy, Scikit-learn, that are useful for implementing machine learning algorithms and data analysis.

I used Anaconda navigator for most of the Backend processes which is a navigator many data science tools

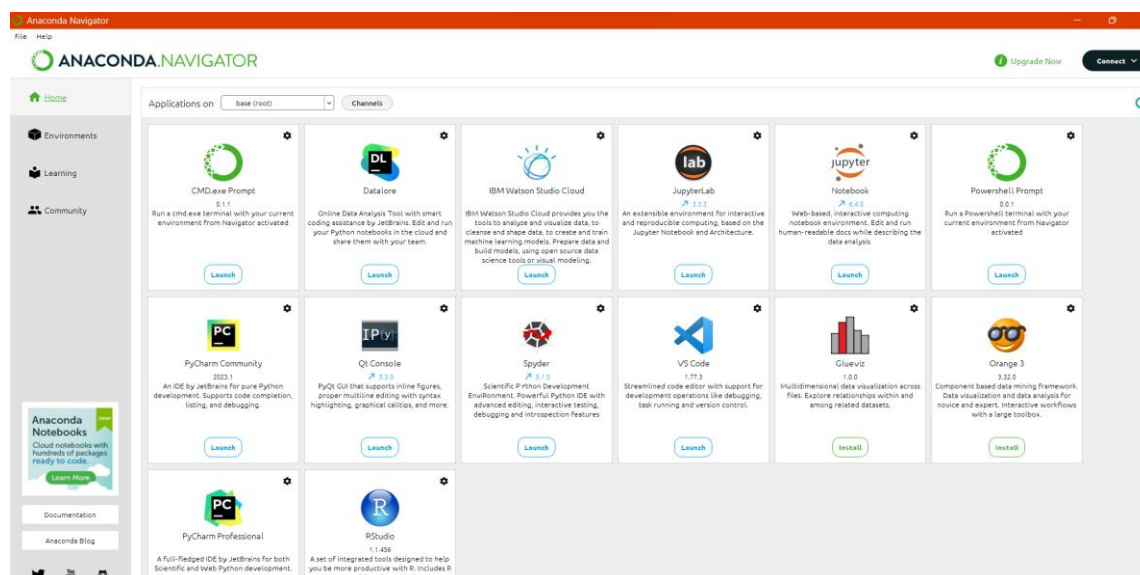


Figure 7 Anaconda Navigator

Anaconda Navigator is a desktop graphical user interface included in Anaconda that allows you to launch applications and easily manage conda packages, environments and channels without the need to use command line commands. In order to run, many scientific packages depend on specific versions of other packages. Data scientists often use multiple versions of many packages and use multiple environments to separate these different versions. The CLI program conda is both a package manager and an environment manager. This helps data scientists ensure that each version of each package has all the dependencies it requires and works correctly.



Jupyter Notebook is an interactive notebook environment that is useful for prototyping, testing, and debugging the system. The Jupyter Notebook is an open-source web application that you can use to create and share documents that contain live code, equations, visualizations, and text. Jupyter Notebook is maintained by the people at Project Jupyter.

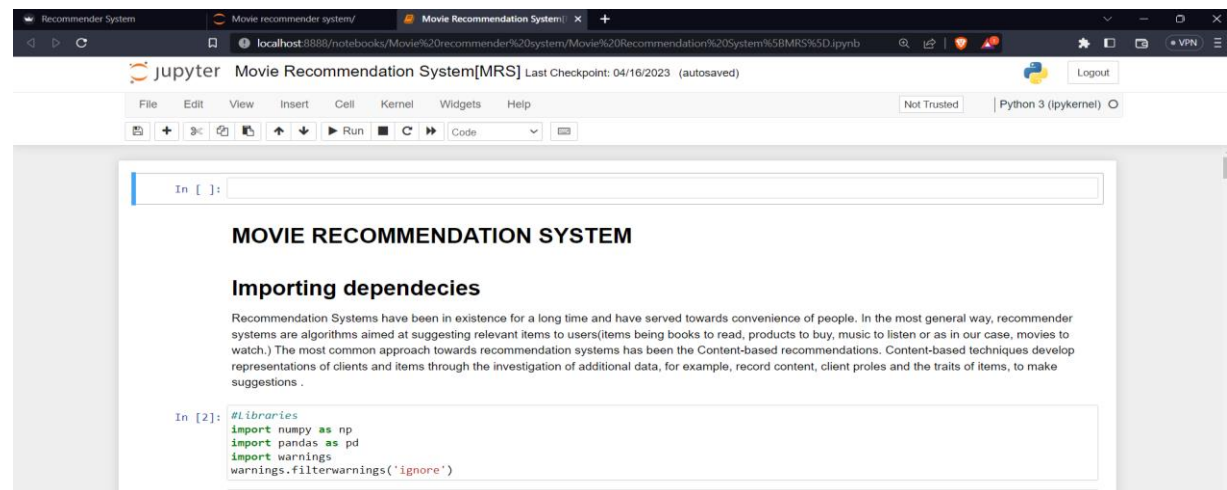


Figure 8 Jupyter Notebook

Streamlit is an open-source web application framework that is useful for creating interactive and customizable user interfaces. It can be used for implementing the user interface module of the system and providing users with a seamless experience.

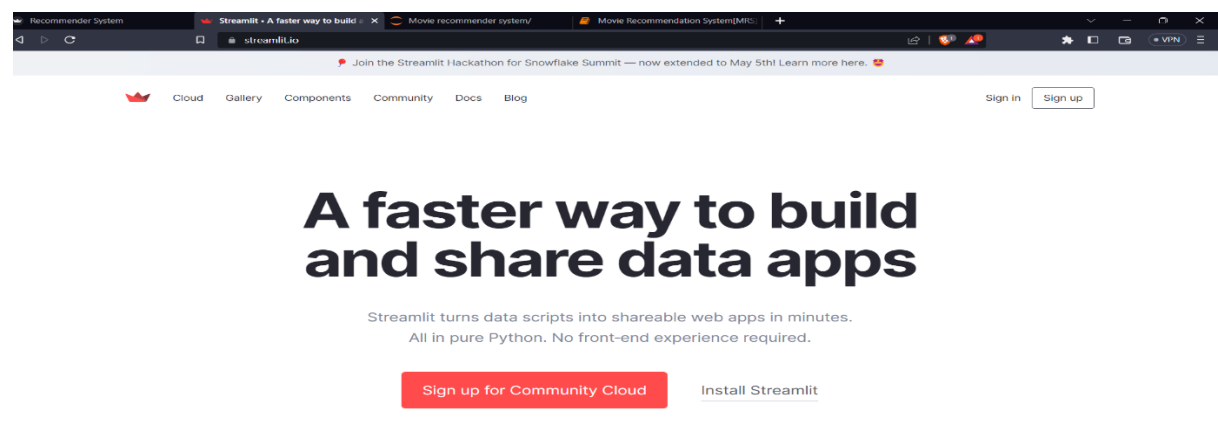


Figure 9 Streamlit

<https://streamlit.io/>

## 6.2 MODULES AND TECHNOLOGICAL SPECIFICATIONS

The movie recommendation system can be divided into several modules, each of which performs a specific function. The following are the key modules in a movie recommendation system:

Data Collection Module: This module is responsible for collecting and storing user data, including viewing history, ratings, and user behaviour. It can also gather additional data from social media platforms and other sources to enhance the recommendation accuracy.

Pre-processing Module: The pre-processing module cleans and pre-processes the user data before feeding it to the recommendation engine. This module may involve data cleaning, normalization, and feature extraction.

Recommendation Engine: The recommendation engine is the core module of the system, which generates movie recommendations based on user data and behaviour. It uses various filtering techniques, such as collaborative filtering, content-based filtering, and hybrid filtering, to generate personalized recommendations for each user.

Post-processing Module: This module is responsible for fine-tuning the recommendation results generated by the recommendation engine. It may involve filtering out irrelevant recommendations, enhancing recommendation diversity, and optimizing recommendation delivery.

User Interface Module: The user interface module provides an intuitive and user-friendly interface for users to interact with the system. It may include features such as search, movie trailers, reviews, and ratings.

User Profiling: This module involves creating user profiles based on their preferences and feedback. The technological specifications for this module include data mining techniques, clustering algorithms, and machine learning models.

Similarity calculation: This module involves calculating the similarity between the user's preferences and the features of the movies. The technological specifications for this module include similarity measures such as cosine similarity, and machine learning algorithms for feature selection and weighting.

## 6.3 OUTCOMES

The project outcome for a movie recommendation system using content-based filtering would be a functional and effective system that provides personalized and relevant movie recommendations to users based on their preferences and feedback. The system would have the following outcomes:

- **Improved User Experience:**

The movie recommendation system would provide an improved user experience by presenting personalized and relevant movie recommendations to users. This would lead to increased user engagement, satisfaction, and loyalty.

- **Increased User Engagement:**

The movie recommendation system would encourage users to engage more with the system by providing them with personalized recommendations that match their preferences. This would lead to increased usage and adoption of the system.

- **Improved Revenue Generation:**

The movie recommendation system could potentially increase revenue generation by encouraging users to watch more movies, purchase movie tickets, or subscribe to streaming services.

- **Enhanced Business Insights:**

The movie recommendation system would provide valuable business insights by analysing user preferences and feedback. This could be used to improve the product offering, marketing strategies, and overall business performance.

## 6.4 RESULT ANALYSIS

As this project is still under development, stating the actual outcomes is difficult, it is impractical to submit a proper result summary. However, it is projected to compete with existing systems leading this category, our system is going to be open-source and will rely on fundings and donations. As the system has major limitation which is data and there are very few companies which shares quality data freely which makes it difficult to do any analysis

## CHAPTER 7 – TESTING

### 7.1 TEST STRATEGY

Testing is a crucial aspect of ensuring the effectiveness and functionality of a movie recommendation system using content-based filtering. Here's a summary of the testing process for the system:

- **Unit Testing:**  
The system's individual modules and functions are tested to ensure they are working as expected. This involves testing the data collection and pre-processing, feature extraction, user profiling, similarity calculation, recommendation generation, and evaluation and feedback modules.
- **Integration Testing:**  
The different modules of the system are tested together to ensure they are integrated and working as a whole. This involves testing the data flow between the modules and verifying that the system is generating accurate and relevant recommendations.
- **Performance Testing:**  
The system's performance is tested to ensure it can handle large volumes of data and provide recommendations in real-time. This involves testing the system's response time, scalability, and reliability.
- **Maintenance Testing:** The system is tested after any updates or changes are made to ensure that the system continues to function correctly and provide accurate recommendations.

Overall, testing ensures that the movie recommendation system using content-based filtering is functioning correctly, providing accurate and relevant recommendations, and meeting user expectations. Through rigorous testing, the system can be optimized for performance and reliability, ensuring a better user experience and improved business outcomes.

.

## 7.2 TEST RESULT AND ANALYSIS

In general, the test results and analysis would involve running each test case and evaluating whether the system passed or failed the test. If a test fails, the system would need to be debugged and the test re-run to ensure that the system passes

The analysis of the test results would involve identifying any patterns or trends in the data, such as identifying which types of tests the system tends to perform well on and which types of tests it struggles with. This information can be used to improve the system's design and functionality. Additionally, the test results and analysis would involve gathering feedback from users to evaluate the system's performance in real-world scenarios. This feedback can be used to identify areas where the system can be improved and to identify any issues that may have been missed during the testing process.

Overall, the test results and analysis would be used to ensure that the movie recommendation system is reliable, accurate, and meets the needs of its users in a traditional way and gives the recommendation accurately proportional to the quality of data.

The analysis of test results for a movie recommendation system:

- The system should have a high rate of accuracy in generating movie recommendations based on user preferences and movie features.
- The response time of the system should be quick enough to provide movie recommendations to users in a timely manner.
- The system should be scalable enough to handle large volumes of data and provide accurate recommendations without experiencing a decrease in performance.
- The user feedback should be positive, indicating that the movie recommendations provided by the system are relevant and useful to the user.
- The system should be able to adapt to user feedback and provide updated movie recommendations based on changes in user preferences or feedback.
- The test results should identify any areas of weakness or limitations in the system, such as issues with data pre-processing or feature extraction.

## **CHAPTER 8 – CONCLUSION AND DISCUSSION**

### **8.1 OVERALL ANALYSIS OF INTERNSHIP/PROJECT VIABILITY**

During my time of internship in Maxgen technologies taught me a huge spectrum of things which enabled me to understand what, why, learn Data science is the future and how powerful it is, if used properly. The 12-week internship provided me an exposure of industry standard work, and helped me enhance my skills. It was an exhilarating and challenging experience to deal with the coding, mathematics of data science. But my peers helped me realize, resolve and work on project efficiently.

A movie recommendation system is a complex data science project that involves several phases, including data collection and preprocessing, exploratory data analysis, feature engineering, model selection and training, and deployment and integration. The project requires a variety of tools and technologies, including Python, pandas, scikit-learn, streamlit, Kaggle, and cloud computing platforms. By implementing a movie recommendation system, you can gain valuable skills and knowledge in data analysis, machine learning, and software development, as well as learn about the challenges and opportunities of working with real-world datasets. With proper planning and execution, a movie recommendation system can provide a powerful tool for users to discover new movies and improve their viewing experience, while also showcasing your skills and abilities as a data scientist.

### **8.2 DATES OF CONTINUOUS EVALUATION**

First evaluation: 11/03/2023

Second Evaluation: 01/04/2023

Third Evaluation: 20/04/2023

### 8.3 PROBLEMS ENCOUNTERED AND POSSIBLE SOLUTIONS

- **Limited data:** One challenge with a content-based system is that it relies heavily on the availability of data on movie features, such as genre, actors, and plot. If the database does not have comprehensive data on these features, the system's recommendations may be inaccurate or incomplete. A possible solution is to supplement the existing data with data from other sources, such as social media or user-generated content.
- **Overfitting:** Another potential issue is overfitting, where the system becomes too specialized to the data it has been trained on and fails to generalize to new data. A possible solution is to use regularization techniques to prevent overfitting and ensure that the system can generate accurate recommendations for a range of users and movies.
- **Cold start problem:** A common problem with content-based systems is the cold start problem, where the system has insufficient data to generate accurate recommendations for new users or movies. A possible solution is to use hybrid systems that combine content-based and collaborative filtering techniques, or to use user profile data to generate initial recommendations for new users.
- **Bias:** Finally, content-based systems can be susceptible to bias if certain features are over-represented in the database, leading to skewed recommendations. A possible solution is to use algorithms that adjust for bias and ensure that recommendations are based on a balanced representation of all available features.

Overall, the development of a content-based movie recommendation system requires careful attention to data quality, model accuracy, and algorithmic bias to ensure that the system can generate accurate and fair recommendations for all users.

## 8.4 SUMMARY OF INTERNSHIP/PROJECT WORK

The movie recommendation project was done in my 12-weeks internship that involved developing a content-based filtering system for generating movie recommendations. The project involved several stages of development, including identifying key movie features, analysing data to assign weights to each feature, designing a rating system, developing an input/output interface, and testing the system to ensure its accuracy and reliability. As an intern working on the project, I had the opportunity to gain hands-on experience in data analysis, software development, and testing. I was able to work with a team of experienced developers and learn from their expertise, as well as contribute my own ideas and insights to the project. Throughout the internship, I was able to grow my skills in several areas, including data analysis, programming, and project management. I learned how to work collaboratively with a team, communicate effectively, and solve complex problems using a variety of tools and techniques.

Overall, the internship was a valuable learning experience that helped me develop new skills and gain real-world experience in a dynamic and challenging environment. It gave me the opportunity to work on a meaningful project that had the potential to impact people's lives, and I am proud to have been a part of it.

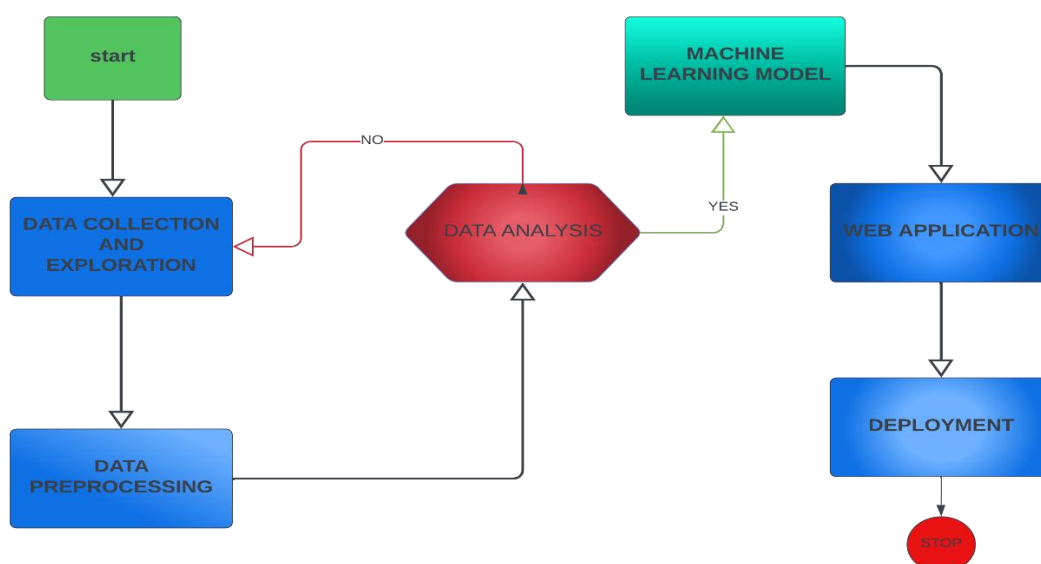


Figure 10 Project Work Flow



## 8.5 LIMITATIONS AND FUTURE ENHANCEMENT

The main limitation of this system is lack of proper data and module from which the posters are fetched which using and API. If the TMDB website goes under maintenance then the system will show traceback error of the posters fetching engine in the backend.

Limitations:

1. Limited scope - The system is limited to recommending movies based on the user's historical data and does not consider the user's social or demographic characteristics.
2. Cold start problem - The system may not be able to provide recommendations for new users with no historical data.
3. Limited diversity - The system may provide recommendations for similar movies and may not explore new genres or categories.
4. Lack of transparency - The system's decision-making process may be difficult to explain to users and may lead to low user trust and acceptance.

Future Enhancements:

1. Hybrid recommendation system - Developing a hybrid system that combines both content-based and collaborative filtering techniques could enhance the system's accuracy and diversity.
2. Deep learning techniques - Incorporating deep learning techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) could improve the system's feature extraction and decision-making capabilities.
3. Natural language processing - Using natural language processing (NLP) techniques to analyse user reviews and comments could provide more accurate and personalized recommendations.
4. Improved user interface - Developing an improved user interface that enhances the system's usability and transparency could increase user trust and satisfaction.

## REFERENCES

During the internship and project work I took some help from below:

- <https://streamlit.io/>.
- <https://www.python.org/>.
- <https://pandas.pydata.org/>.
- <https://towardsdatascience.com>
- <https://www.geeksforgeeks.org>
- <https://www.javatpoint.com/data-science>
- <https://www.w3schools.com/datascience>
- <https://neptune.ai>
- <https://www.ibm.com/in-en/data-science>

