SUMMER TRAINING/INTERNSHIP

PROJECT REPORT

(Term June-July 2025)

MOVIE RECOMMENDATION SYSTEM

Submitted by

Banoth Srikanth
Registration Number :12303313
Course Code PETV719

Under the Guidance of

Prof.Mahipal Singh Papola

School of Computer Science and Engineering
LOVELY PROFESSIONAL UNIVERSITY

BONAFIDE CERTIFICATE

This is to certify that **Mr. BANOTH SRIKANTH**, a student pursuing Bachelor of Technology in Computer Science and Engineering (Data Science with Machine Learning) at Lovely Professional University, has successfully completed the project titled**: "MOVIE RECOMMENDATION SYSTEM"** under the guidance of **Mr. Mahipal Singh Papola**.

This certificate is issued as a recognition of the original and bonafide work carried out by the student during the project period, as part of his academic and professional development.

Date: 16-07-2025 Supervisor

Mr. Mahipal Singh Papola

ACKNOWLEDGEMENT

We would like to express our deepest appreciation to all those who provided us the possibility to complete this report. A special gratitude we give to our project supervisor, **Prof.Manipal Singh Popala** , in whose contribution stimulating suggestions and encouragement, helped us to coordinate our project especially in writing this report. Furthermore, we would also like to acknowledge with much appreciation the crucial role of the staff of Prime College, who gave the permission to use all required equipment and the necessary materials to complete the task. We are thankful and fortunate enough to get constant support from our seniors and every teaching staff of **BTech CSE department** which helped us successfully complete our project. We would also like to extend our regards to all the non-teaching staff of CSE department for their timely support. We have to appreciate the guidance given by other supervisor as well as the panels especially in our project presentation that has improved our presentation skills thanks to their comment and advices. Our appreciations also go to each and every one of our colleagues for their encouragement and support in developing the project. However, it would not have been possible without the kind support and help of many individuals. We would like to extend our sincere appreciation to all of them.

ABSTRACT

Here we have used an approach to a movie recommendation system using Cosine Similarity to recommend similar movies based on the one chosen by the user. The movies are converted into vectors using Count Vectorizer and Cosine Similarity is used to find the nearest vectors(movies) and recommend it to the user. Although the existing recommendation systems get the job done, it does not justify if the movie is worth spending time on. To enhance the user experience, this system performs sentiment analysis on the reviews of the movie chosen using Naive Bayes algorithm.

Keywords: Cosine similarity, Movie recommendation, Naive Bayes, Sentiment analysis

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1 Introduction

Movie Recommender is a movie recommendation system that provides personalized movie recommendations based on a user's movie preferences. It uses a content-based filtering approach, which recommends movies based on the user's preferences.

In today's digital world, there are countless options for movie streaming services, but finding the right movie to watch can be overwhelming. Movie Recommender is designed to solve this problem by providing movie recommendations, making it easier for users to find movies they will enjoy. Movie recommender has sentiment analysis built into it which classifies the user reviews which helps user in choosing right movies for them.

1.2 Problem Statement

The problem with traditional Movie Recommender: Content Based Movie Recommendation System with Sentiment Analysis is that they are often not personalized enough to meet the unique preferences and tastes of individual users. Many users find it difficult to discover new movies that they will enjoy watching, and this can result in reduced user engagement and satisfaction.

The Movie Recommender System aims to address this problem by providing personalized movie recommendations to users based on their viewing preferences. The system can accurately recommend movies that users are likely to enjoy watching, thereby increasing user engagement and satisfaction

1.3 Objectives

The project aims to meet the following objectives:

1. To provide personalized movie recommendations to users based on their viewing preferences.
2. To improve user engagement and satisfaction by accurately recommending movies that users are likely to enjoy watching.
3. To incorporate advanced algorithms and data processing techniques to accurately recommend movies to users.
4. To provide an easy-to-use interface for users to interact with the system, making it accessible and user-friendly.
5. Provide sentiment analysis to the movie reviews to help users decide properly

1.4 Scope and Limitation

1.4.1 Scope

This project has many scopes. They are

1. Personalization: A Movie Recommender: Content Based Movie Recommendation System with Sentiment Analysis should be able to provide personalized recommendations based on the user's preferences.
2. Content database: A Movie Recommender: Content Based Movie Recommendation System with Sentiment Analysis should have a comprehensive content database that includes a wide range of movies and their associated metadata, such as genre, director, actors, and ratings.
3. Recommendation engine: A Movie Recommender: Content Based Movie Recommendation System with Sentiment Analysis should have a robust recommendation engine that uses advanced algorithms and techniques to analyze user data and generate high-quality recommendations.
4. User interface: A Movie Recommender: Content Based Movie Recommendation System with Sentiment Analysis should have a user-friendly interface that allows users to easily search, browse, and rate movies, as well as provide feedback and receive personalized recommendations.

1.4.2 Limitation

1. The system relies heavily on user behavior data. If users provide inaccurate data, the recommendations may not be accurate.
2. The system is limited by the availability and quality of data on movies and user preferences. If there is insufficient data, the recommendations may not be relevant or useful.
3. The system may not be able to accurately recommend movies for users with niche or unique tastes that are not well-represented in the available data.
4. There is a possibility of bias in the recommendations, especially if the available data is biased towards certain genres, languages, or regions
5. Sentiment analysis can't be properly done in satirical movie reviews

1.5 Development Methodology

The development methodology used for this proj ect is incremental delivery, which is a kind of iterative software development process. The project is divided into a series of incremental iterations. Each function or features are being added upon the previous one to gradually deliver a full-fledged functional system.

The incremental delivery approach is particularly suited for projects which involves a high level of complexity, or where the requirements are not clearly-defined or may change over the period of time. Breaking the project down into smaller iterations, allows for a more manageable and flexible development process. Also, it allows for feedback and corrections to be made throughout the process rather than waiting until the end of the project to evaluate the final product

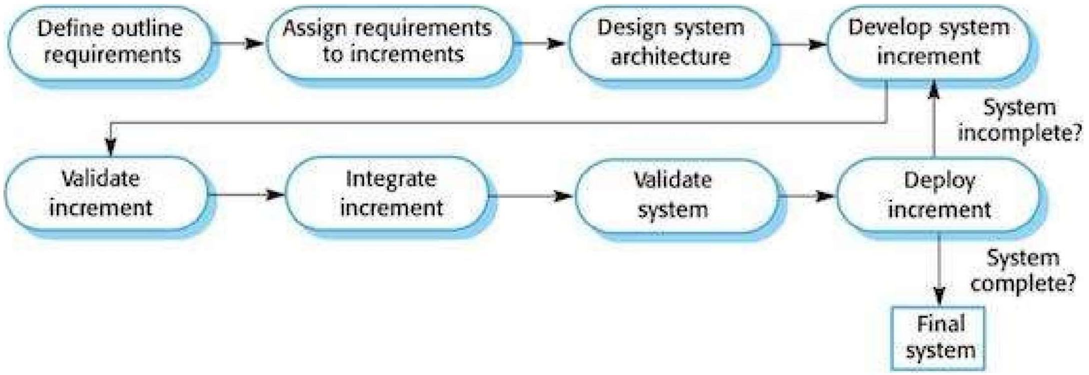


Figure 1. 1: Incremental Delivery

1.6 Report Organization

The report on "Movie Recommender: Content Based Movie Recommendation System with Sentiment Analysis" is based on six chapters. Each chapter follows the constructive building of this project. Chapter I gives an overview idea of our project. It anticipates and combines the main points to be described later in the chapters followingly. Similarly, chapter 2 usually contains the theoretical literature review. It gives an insight to distinguish the possible hypothesis, strategies and shortfalls in the current research. Chapter 3 studies the system such that information can be analyzed, modeled and developed. It also gives enough information to replicate the study. It addresses the problems from chapter 1 and explains the objects of each experiment. Chapter 4 contains an insight of the system design and algorithm being used while developing the system. Chapter 5 contains system testing, it discusses the execution of a program or system with the intent of finding errors. It also includes the examination of code as well as execution of that code in various environments and conditions. Chapter 6 describes the significance of "Movie Recommender System", moreover, discusses the future recommendations applicable to enhance the project.

CHAPTER 2
BACKGROUND STUDY AND LITERATURE REVIEW

2.1 Background Study

In recent years, the entertainment industry has been undergoing a significant shift towards digitalization, with an increasing number of consumers opting for online streaming services to watch movies. The vast selection of movies available online can often lead to confusion and decision paralysis, with consumers struggling to choose a movie that aligns with their preferences. This problem can be addressed through the development of a Movie Recommender system, which can provide personalized recommendations to consumers based on their viewing history and preferences.

Movie Recommender systems have been gaining popularity in recent years due to their ability to provide tailored recommendations to users, thus improving the user experience. These systems utilize machine learning algorithms and data analysis techniques to analyze user behavior and preferences, and suggest movies that are most likely to be of interest to the user. These algorithms consider various factors such as movie genre, actors, ratings, and user feedback to generate personalized recommendations.

The development of a Movie Recommender system requires expertise in various areas such as machine learning, data analysis, and software development. The system will need to be able to collect and analyze user data in real-time, and generate personalized recommendations efficiently. Additionally, the system will need to be scalable and capable of handling large amounts of data, as well as being user-friendly and easy to navigate.

The benefits of a Movie Recommender system are numerous, both for consumers and movie streaming services. Consumers will be able to easily find movies that align with their preferences, leading to a more satisfying viewing experience. On the other hand, streaming services will be able to provide personalized recommendations to users, leading to increased user engagement and retention.

In conclusion, the development of a Movie Recommender system has the potential to revolutionize the entertainment industry by providing personalized recommendations to consumers, thus improving the overall user experience. With the increasing demand for online streaming services, the development of such a system has become a necessity for movie streaming services to remain competitive in the market.

2.2 Literature Review

The entertainment industry has undergone a massive transformation in recent years due to the advancements in technology. As a result, various movie streaming services have emerged, offering users a vast library of movies to choose from. However, with so many options, it can be challenging to find a movie that suits one's preferences. To address this challenge, movie recommender systems have been developed. According to a study by Partho Pratim Pal and Sukanta Das [l], movie recommender systems use various techniques such as collaborative filtering, content-based filtering, and hybrid methods to recommend movies to users. Collaborative filtering techniques are based on user ratings and their similarity with other users, while content-based filtering methods analyze the movie's attributes to recommend similar movies. Hybrid methods combine both techniques to provide more accurate recommendations. Data security is a critical concern for any system that uses personal user data. As such, movie recommender systems must ensure that user data is secure. In their study, Junchao Zheng, et al. [2] propose a privacy-preserving movie recommendation method that employs homomorphic encryption to protect user data. The method ensures that the user's data is kept confidential while still providing accurate recommendations. User experience is essential in any application, and the movie recommender system is no exception. According to a study by Xiaoyan Wu and Michael Mandel [3], the user experience of a movie recommender system can be enhanced through personalization and interactivity. Personalization involves tailoring the recommendations to the user's preferences, while interactivity enables the user to provide feedback and improve the recommendation algorithm. One of the major challenges of movie recommender systems is dealing with the cold-start problem, where new users or movies have insufficient data to make accurate recommendations. In their study, Piyush K. Shukla and Mukesh Saraswat [4] propose a hybrid approach that combines content-based filtering and collaborative filtering techniques to overcome the cold-start problem. In conclusion, movie recommender systems have gained immense popularity in recent years due to the vast library of movies available on streaming services. To ensure the system's success, developers must address data security concerns, enhance the user experience, and overcome the cold-start problem.

CHAPTER 3
SYSTEM ANALYSIS

3.1 System Analysis

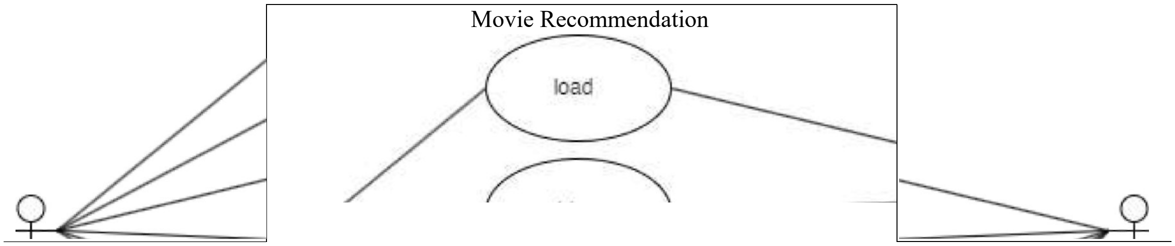
Systems analysis is a process of studying a system or organization in order to understand its components, how they interact and how they can be improved. It is a holistic approach that looks at the system as a whole and identifies the relationships between its parts. The goal of systems analysis is to identify problems and inefficiencies in the current system and to propose solutions for improvement.

3.1.1 Requirement Analysis

Requirement Analysis We've analyzed and validated the requirements, recorded and monitored the implementation throughout the project

3.1.1.1 Functional Requirements

1. Movie database: The system should maintain a database of movies with relevant metadata such as title, genre, rating, and cast.
2. Movie search: The system should allow users to search for movies.
3. Personalized recommendations: The system should provide personalized recommendations to each user based on their preferences.



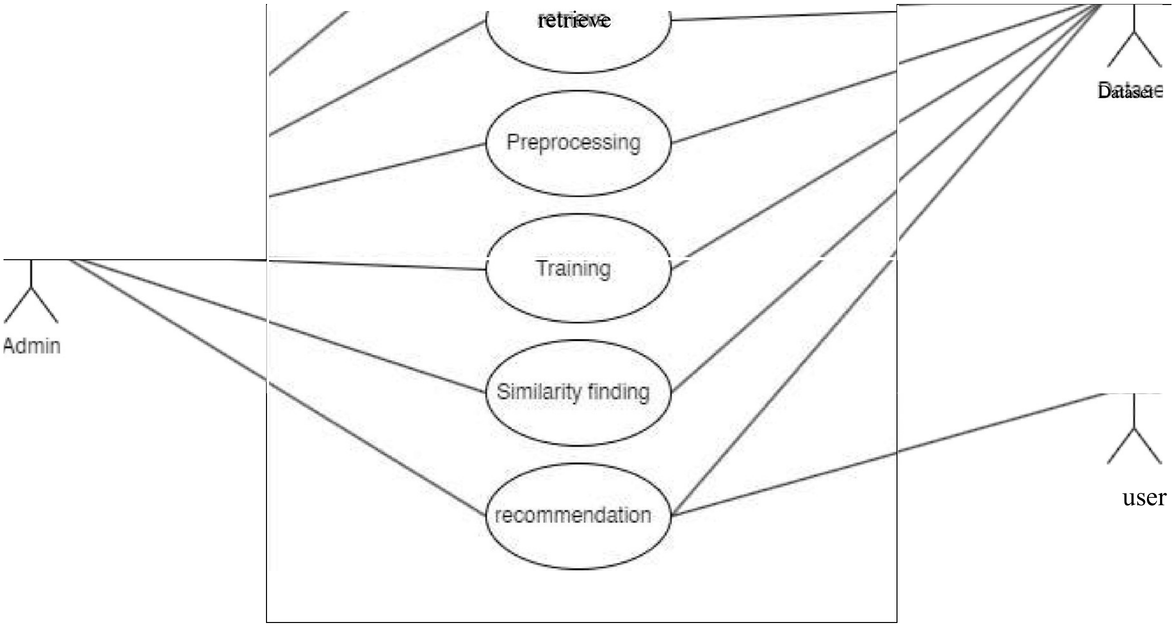


Figure 3. 1: Use case diagram of Movie Recommender

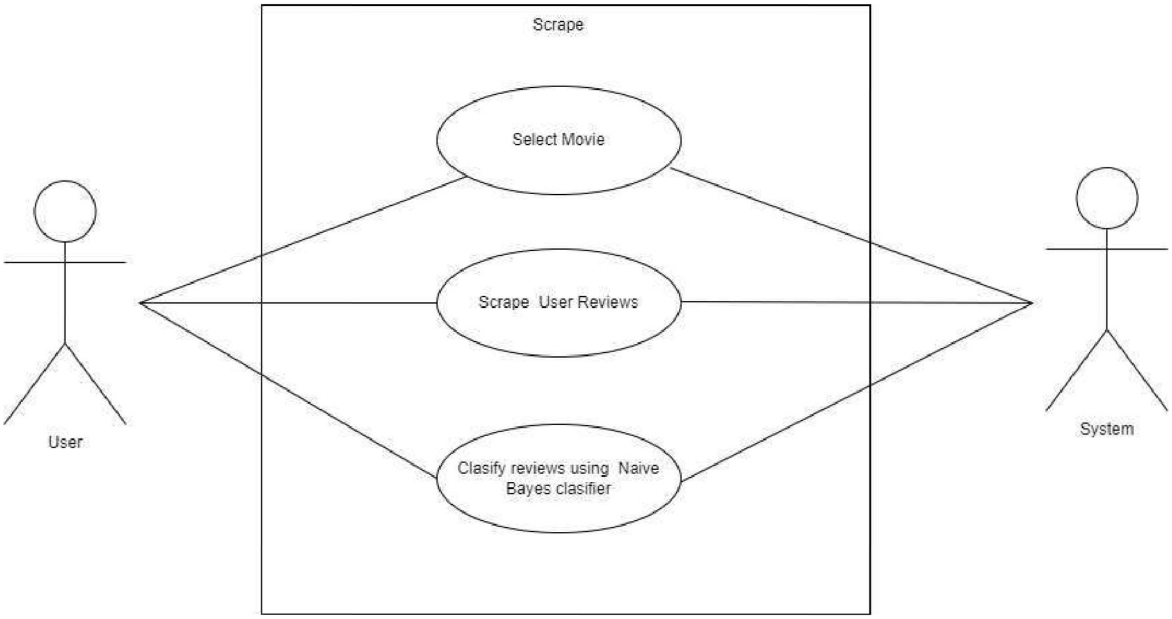


Figure 3. 2: Use Case diagram of sentiment analysis

3.1.1.2 Non-Functional Requirements

The points below focus on the non-functional requirement of the system proposed.

User friendly

User friendly generally means easy to read, use and communicate. The system is not complex and self-explanatory. Our system is well-organized, making it easy to locate different tools and options.

Reliability

The system is reliable. The system takes data from many trusted sources and organization.

Easy access

Our project is a web-based application. Considering this our platform can be accessed by anyone, anywhere where there is internet connection.

3.1.2 Feasibility Analysis

Feasibility studies aim to objectively and rationally uncover the strengths and weakness of an existing or proposed system, opportunities and threats as presented by the environment, the resources required to carry through, and ultimately the prospects for the success.

3.1.2.1 Technical Feasibility

The technical feasibility of the project is high, as the required hardware and software resources are widely available and accessible. Requirements of our system can be categorized as:

Hardware Requirements:

- A computer with a minimum of 2 GB of RAM
- Processor with 1.7-2.4 GHz speed ● Minimum of 2GB Graphic card

Software Requirements:

- Operating System: Windows 10, Linux, or MacOS
- Text Editor (VS-code)
- Anaconda distribution package
- Python libraries

- Anaconda distribution: Anaconda is a free and open-source distribution of the Python programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management system and deployment. Package versions are managed by the package management system conda. The anaconda distribution includes data-science packages suitable for Windows, Linux and MacOS.3
- Python libraries: For the computation and analysis we need certain python libraries which are used to perform analytics. Packages such as SKIearn, Numpy, pandas, Matplotlib, Flask framework, etc are needed.
- SKIearn: It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy. NumPy: NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python.
- Pandas: Pandas is one of the most widely used python libraries in data science. It provides high-performance, easy to use structures and data analysis tools. Unlike NumPy library which provides objects for multi-dimensional arrays, Pandas provides in-memory 2d table object called Data frame.
- Flask: It is a lightweight WSGI web application framework. It is designed to make getting started quick and easy, with the ability to scale up to complex applications. It began as a simple wrapper around Werkzeug.

3.1.2.2 Operational Feasibility

Operational feasibility measures how well a proposed system can solve the defined problem, and takes advantage of the opportunities identified during scope definition and how it satisfies the requirements identified in the requirements analysis phase. The system can be developed to be reliable, maintainable, usable, sustainable and affordable. So, this system is operationally feasible.

3.1.2.3 Economic Feasibility

Economic feasibility analyses the project's costs and revenue in an effort to determine whether it is possible to complete or not. There will not be any necessary equipment to be bought. However, the project will require domain, hosting and probably API which can be bought and configured with a suitable plan. Even if some features were to be added, it will be cost free as no extra equipment will be necessary. As the team already has everything needed, this system is economically feasible. Extensive databases can be maintained when the number of users of the app starts increasing.

3.1.2.4 Schedule Feasibility

It is the most important for the completion of the project on time. The project that we are proposing will too be completed within time constraints.

Activities	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12
Requirement Gathering and Analysis												
Design												
Implementation and Coding												
Testing												
Deployment												
Report Preparation												

Figure 3. 3: Gantt Chart

3.1.3 Analysis

3.1.3.1 Class Diagram

Class diagram is a static diagram. It represents the static view of an application. Class diagram is not only used for visualizing, describing, and documenting different aspects of a system but also for constructing executable code of the software application. Class diagram describes the attributes and operations of a class and also the constraints imposed on the system. The purpose of class diagram is to model the static view of an application. Class diagrams are the only diagrams which can be directly mapped with object-oriented languages and thus widely used at the time of construction. UML diagrams like activity diagram, sequence diagram can only give the sequence flow of the application, however class diagram is a bit different. It is the most popular UML diagram in the coder community.

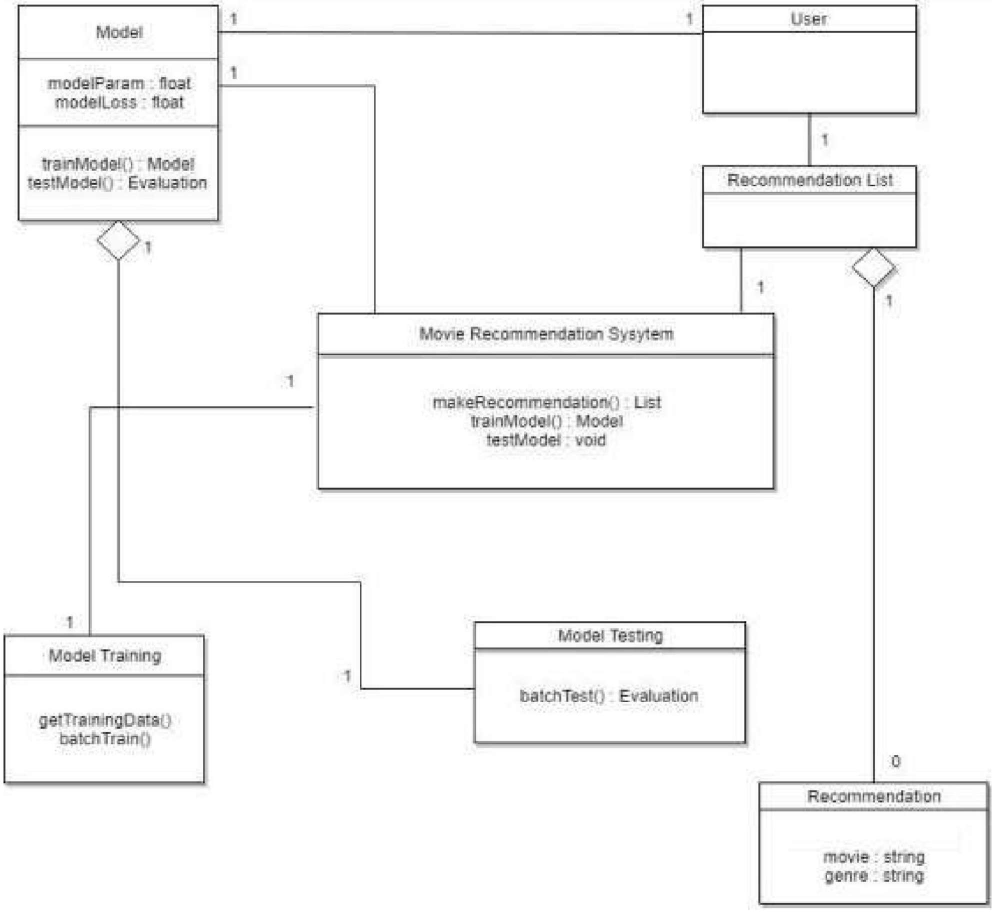


Figure 3. 4: Class Diagram of Movie Recommender System

Ago-check

- entity: string -

teature]   string

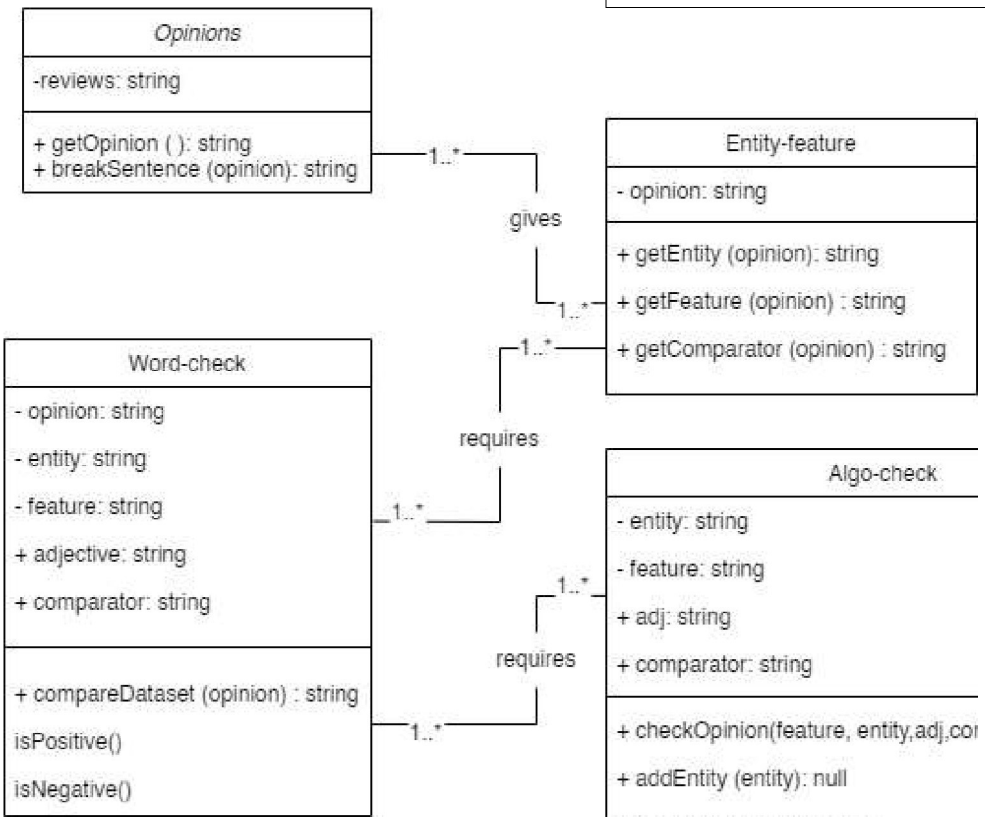
+ ad}: string

+ comparator: string

+ checkOpinion(feature, entityradj:comparator): string

+ addEntity (entity)' null

+ addFeature (feature}: null



gives

Figure 3. 5: Class Diagram of sentiment analysis

3.1.3.2 Sequence Diagram

Sequence diagrams can be useful reference diagrams for businesses and other organizations. We draw a sequence diagram to:

- Represent the details of a UML use case.
- Model the logic of a sophisticated procedure, function, or operation.
- See how tasks are moved between objects or components of a process.
- Plan and understand the detailed functionality of an existing or future scenario.

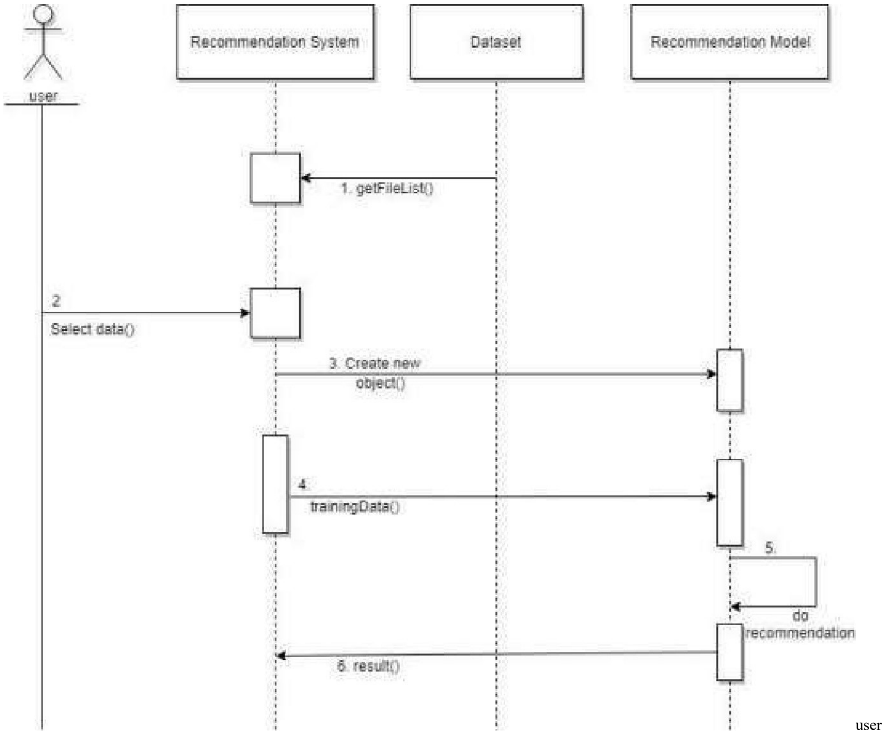


Figure 3. 6: Sequence diagram of movie recommender system

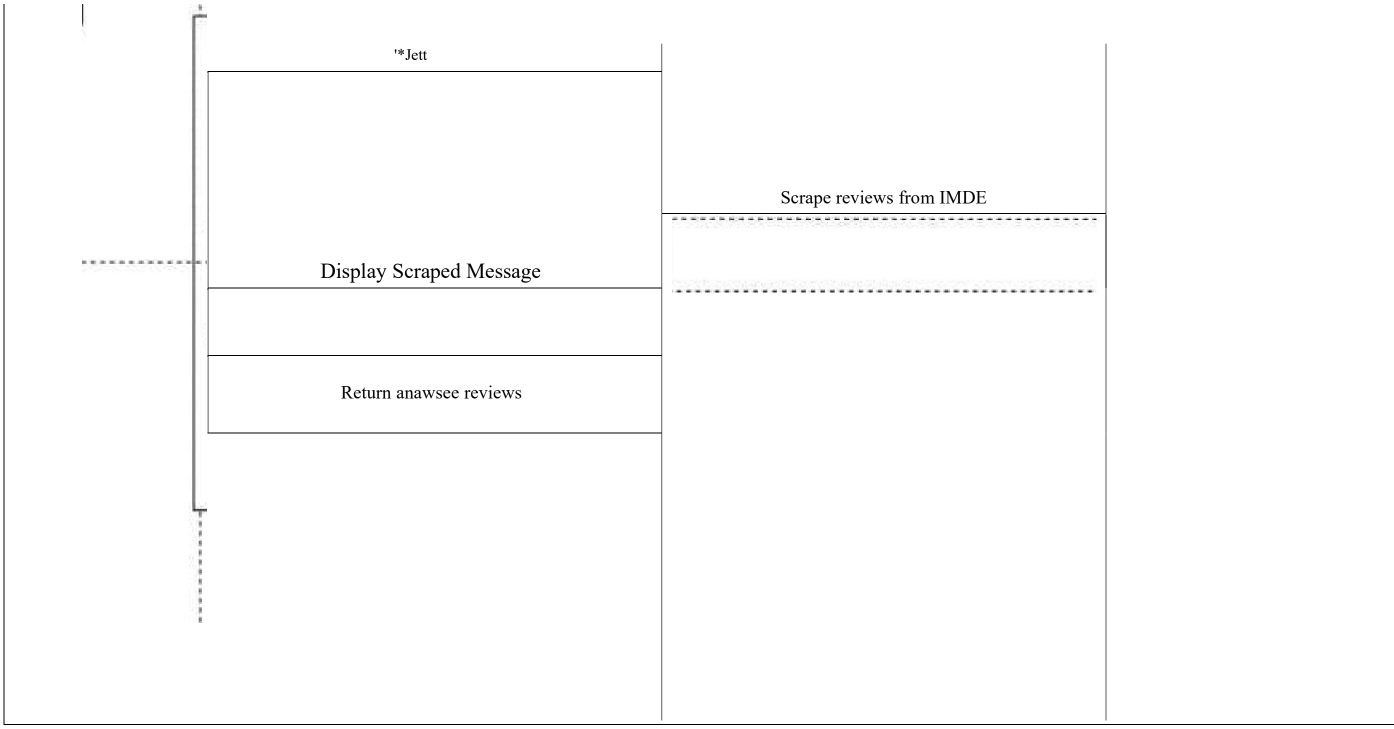


Figure 3. 7 Sequence diagram of sentiment analysis

3.1.3.3 Activity Diagram

Activity diagram is another important diagram in UML to describe the dynamic aspects of the system. Activity diagram is basically a flowchart to represent the flow from one activity to another activity. The activity can be described as an operation of the system. The control flow is drawn from one operation to another. This flow can be sequential, branched, or concurrent. Activity diagrams deal with all type of flow control by using different elements such as fork, join, etc. The basic purposes of activity diagrams is to capture the dynamic behavior of the system.

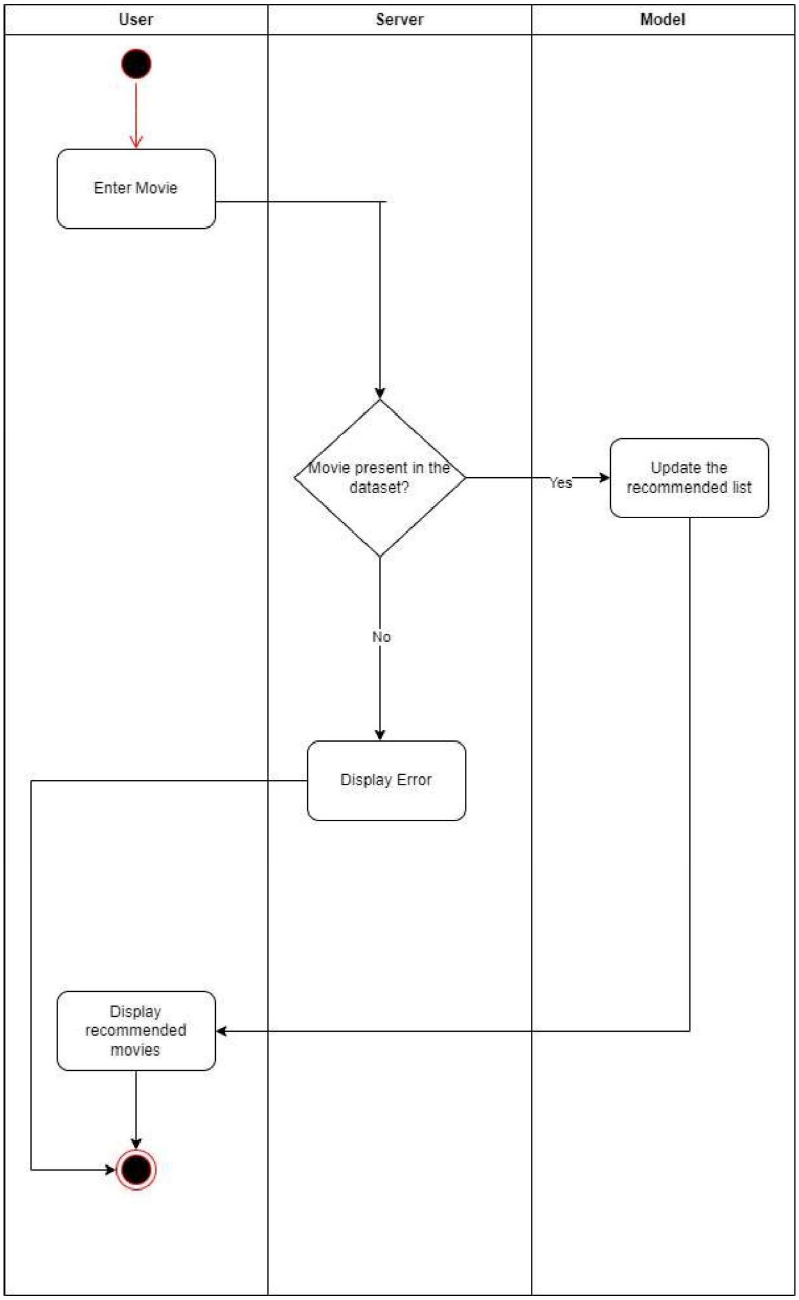
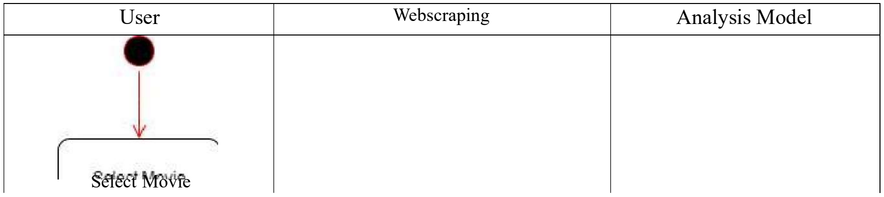

User
Server
Model
Enter Movie
Movie present in the dataset?
Yes
Update the recommended list
No
Display Error
Display recommended movies

Figure 3. 8: Activity Diagram of Movie Recommender System



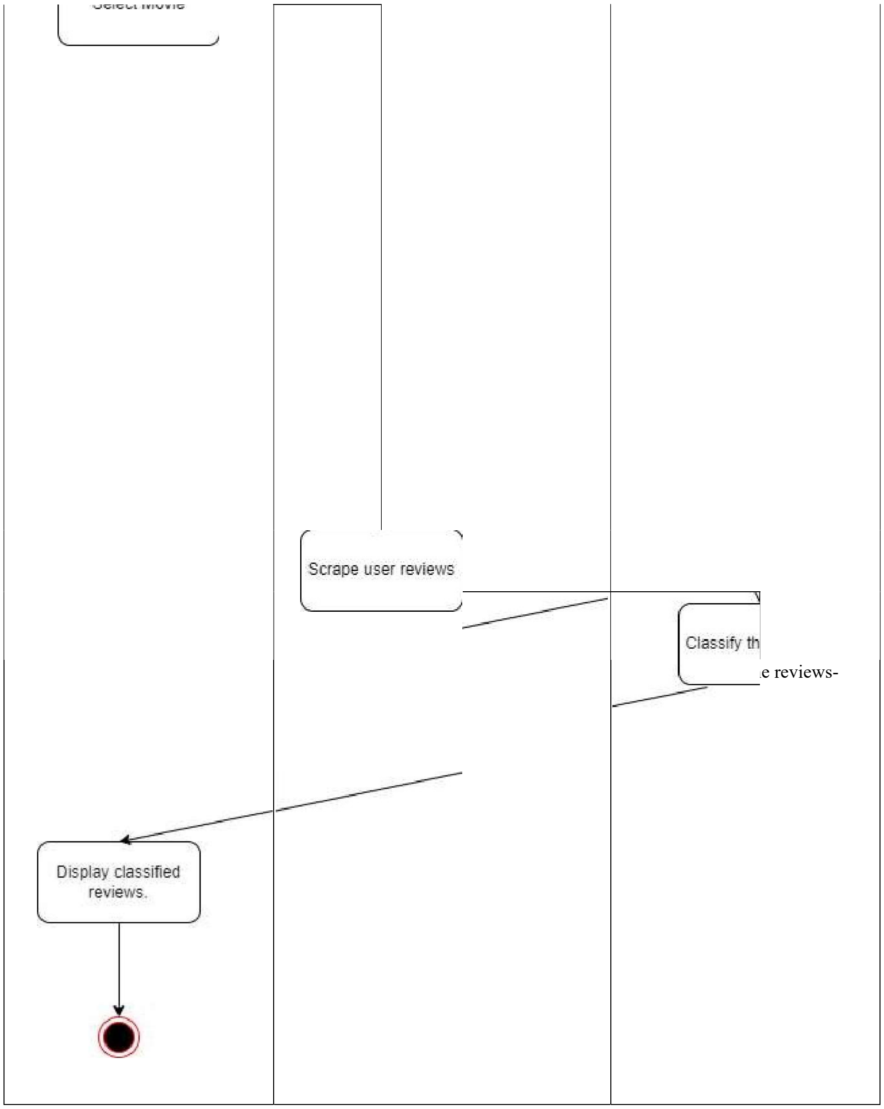


Figure 3 . 9.• Activity diagram of sentiment analysis

CHAPTER 4

SYSTEM DESIGN

4.1 Design

System design is the process of representing architecture, interfaces, components that are included in the system. i.e., system design can be seen as the application of system theory to product development.

4.1.1 Refinement of Class Diagram

The UML diagrams are now refined to shows more detail description of the system component which makes it easier to understand the overall working of the system. Refined UML diagrams include class diagram, sequence diagram and activity diagram of different system modules.

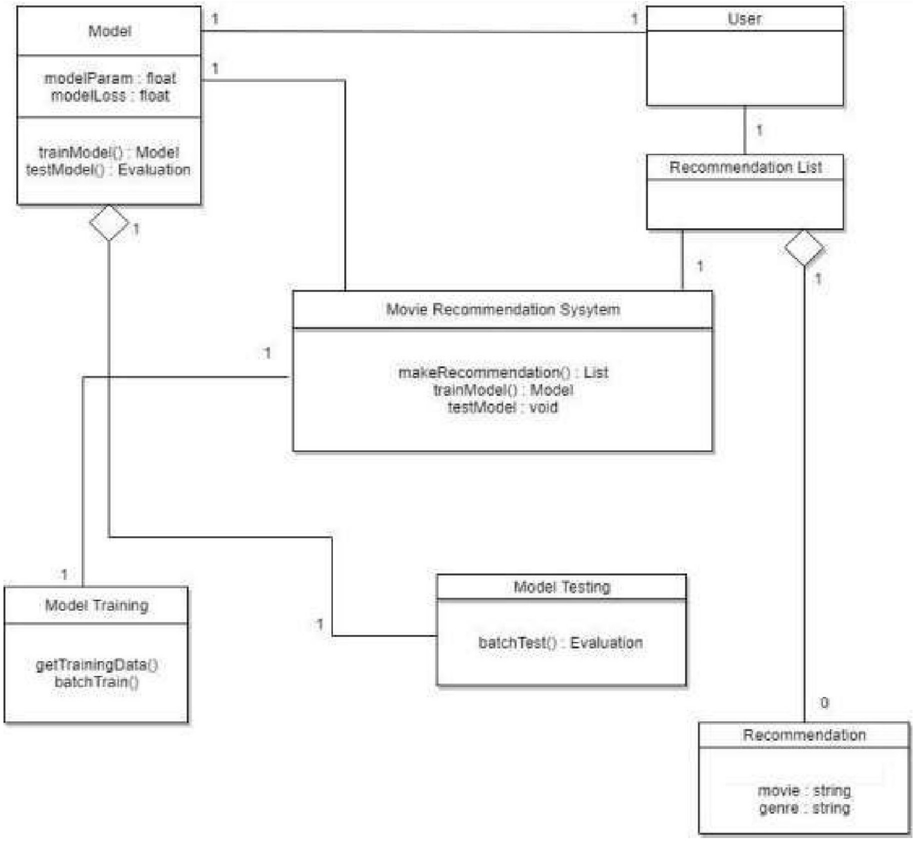


Figure 4. 1: Refined Class Diagram

4.1.2 Refinement of Sequence Diagram

A refined sequence diagram is a type of UML (Unified Modeling Language) diagram that provides a detailed view of the interactions between objects in a system or process. It shows the sequence of messages exchanged between objects or components of a system and the order in which these interactions occur.

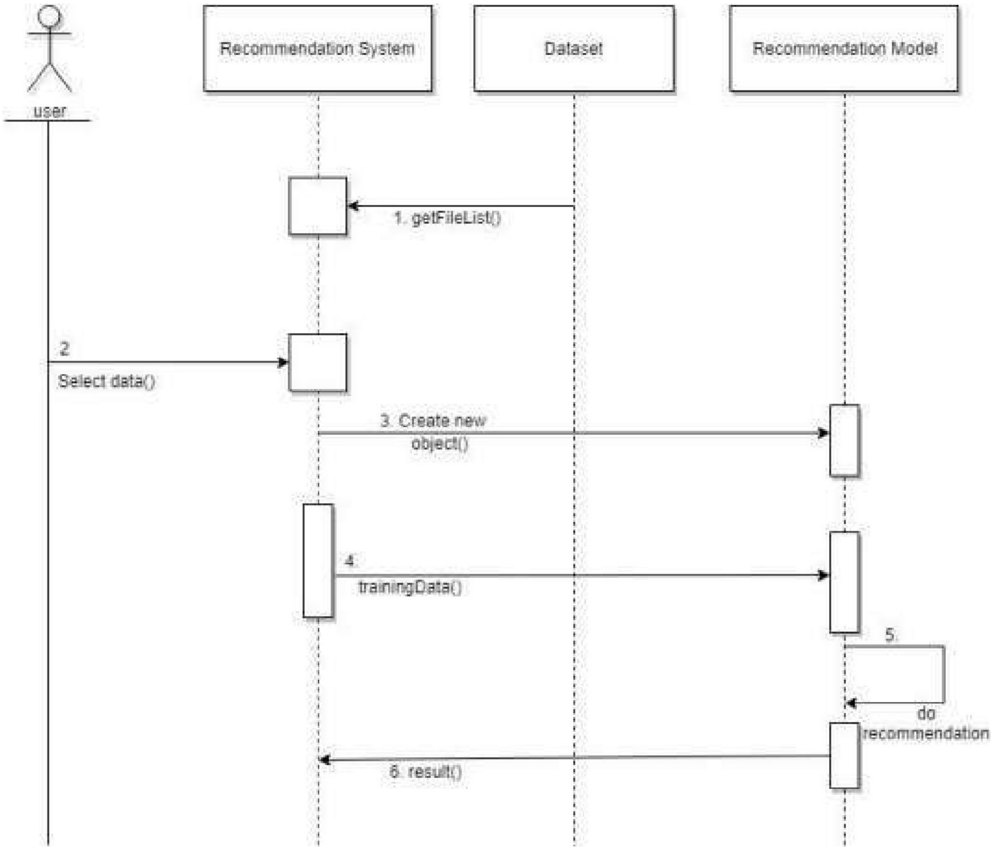


Figure 4. 2 Refined Sequence Diagram for Movie Recommender

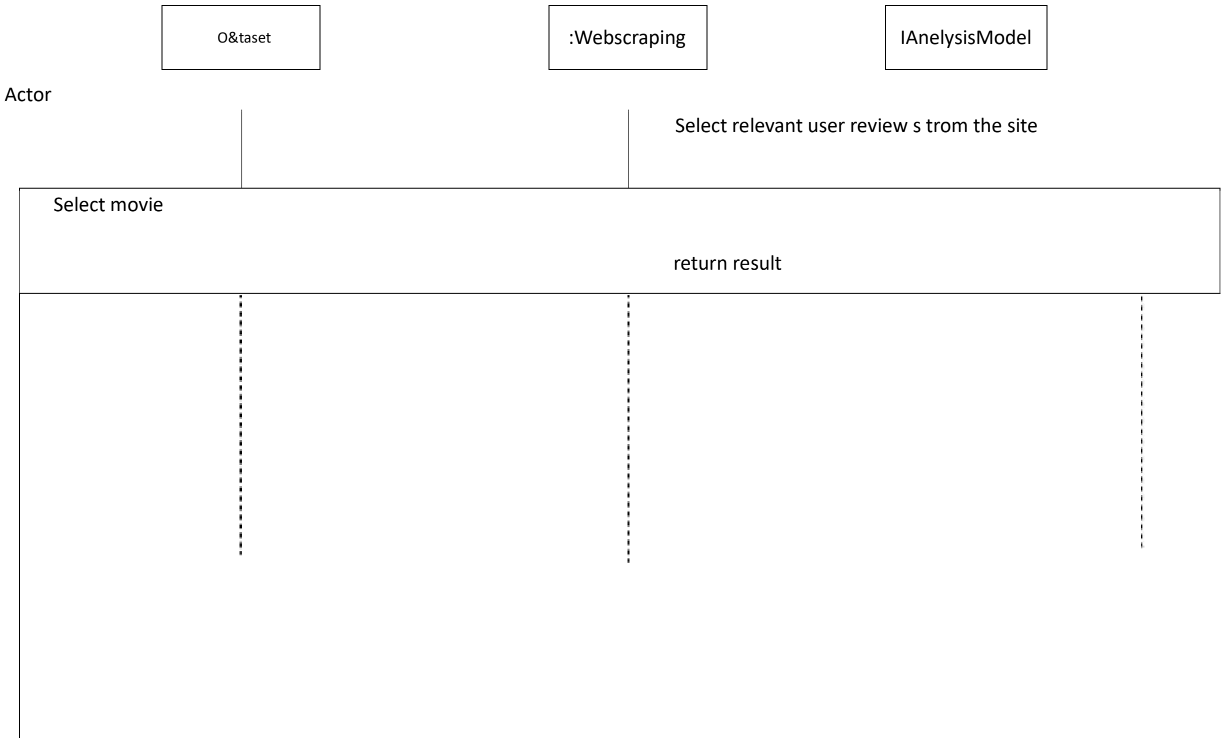


Figure 4. 3: Refined Sequence Diagram for sentiment analysis

4.1.3 Refinement of Activity Diagram

A refined activity diagram is a type of activity diagram that provides a more detailed representation of the activities and actions involved in a particular process or workflow. It builds upon the basic activity diagram by providing additional details and steps, allowing for a more thorough understanding of the process.

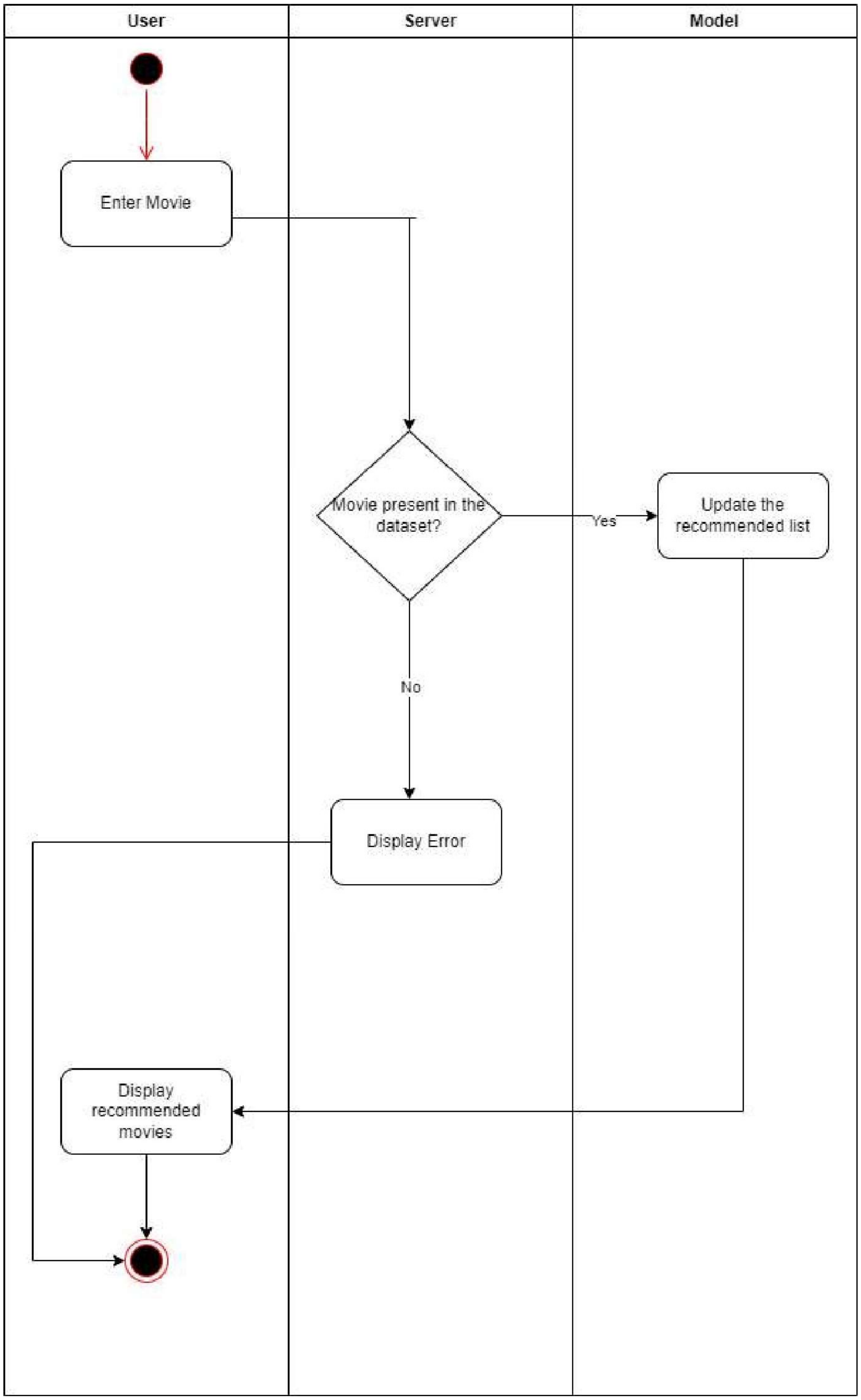

User
Server
Model
Enter Movie
Movie present in the dataset?
Yes
Update the recommended list
No
Display Error
Display recommended movies

Figure 4. 4: Refined Activity Diagram

4.1.4 Component Diagram

A component diagram is a type of UML diagram that shows the structural relationships and dependencies between the components of a software system. It illustrates how software components are connected and interact with each other within a system. Components can represent individual modules, libraries, executables, or other parts of a system.

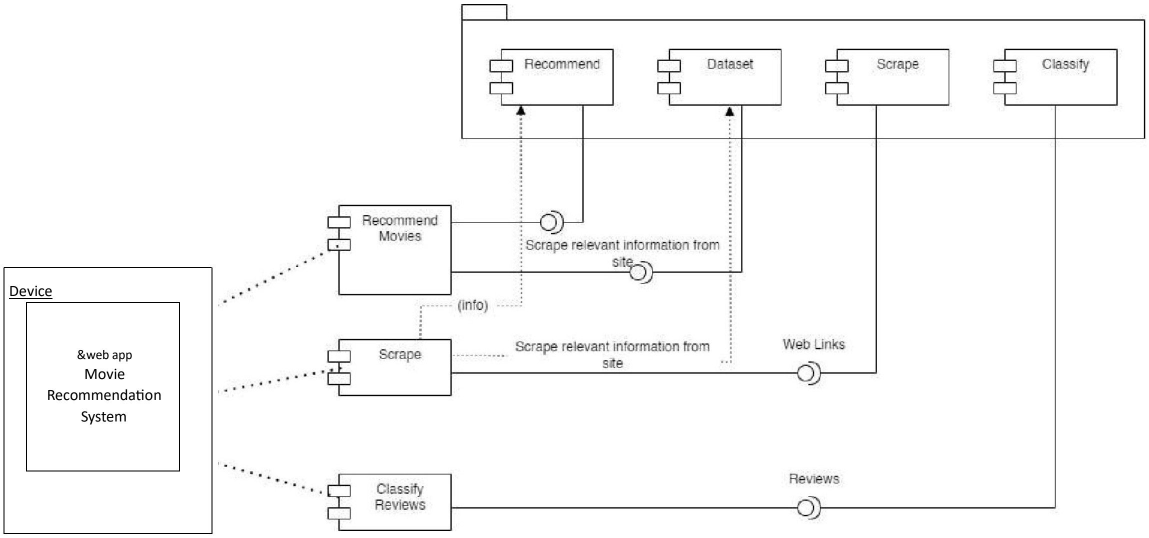
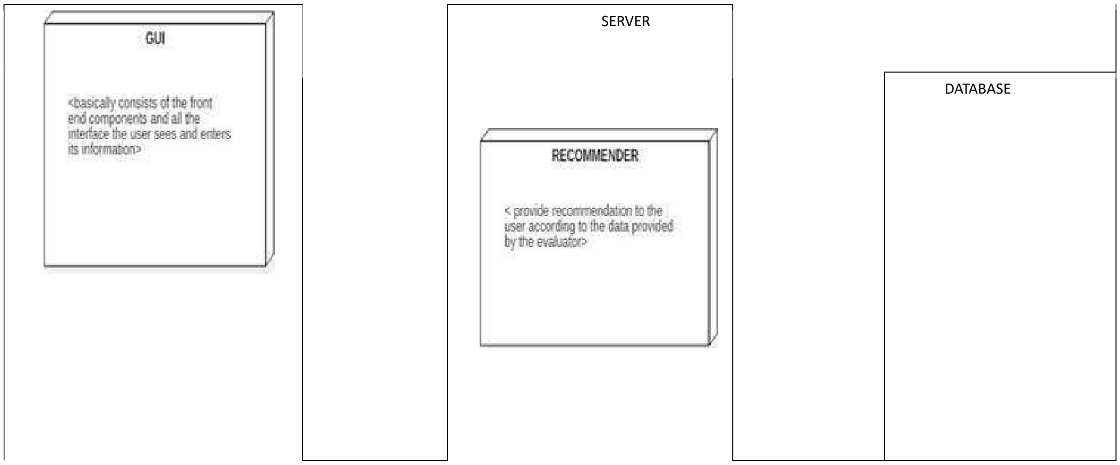


Figure 4. 5: Component Diagram

4.1.5 Deployment Diagram

Deployment diagram are UML structural diagrams that shows the relationships between the hardware and software components in the system and the physical distribution of the processing i.e., Deployment diagram are used to visualize the topology of the physical components of the system where software components are deployed.


GUI
<basically consists of the front end components and all the interface the user sees and enters its information>
SERVER
RECOMMENDER
< provide recommendation to the user according to the data provided by the evaluator>
DATABASE

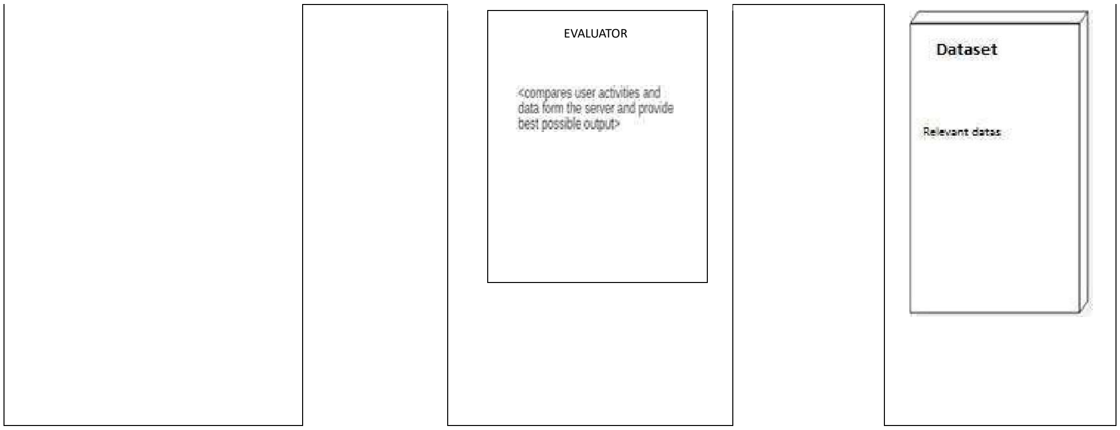

EVALUATOR
<compares user activities and data form the server and provide best possible output>
Dataset
Relevant datas

Figure 4. 6: Deployment Diagram

4.2 Algorithm Description

Count Vectorizer:

Count Vectorizer is a feature extraction technique that converts text documents into a matrix of token counts. It creates a vocabulary of all the unique words in the corpus and then counts the frequency of each word in each document. The resulting matrix represents the number of times each word appears in each document in the corpus, and this matrix can be used as input for various machine learning algorithms.

The project implements the following algorithms:

- Cosine Similarity:

Using the sklearn library in Python, the Cosine Similarity algorithm is used.

After the user is prompted to enter a movie, the algorithm provides 21 other movies like the one used as an input by the user.

In cosine similarity, vectors are taken as the data objects in data sets, when defined in a product space, the similarity is figured out. The smaller this distance, the higher the similarity, but the larger the distance, the lower the similarity. Cosine similarity is a measure that helps to find out how similar data objects are, regardless of size. Mathematically, it is the cosine of the angle between two vectors projected in a multi-dimensional space

$$\text{Cose —} \frac{\vec{a}.b}{\|\vec{a}\|\|\vec{b}\|} = \frac{\sum_1^n a_i b_i}{\sqrt{\sum_1^n a_i^2}\sqrt{\sum_1^n b_i^2}}$$

Where, d. b = aibi = aibi + a2b2 + + anbn is the dot product ofthe two vectors.

The angle between two vectors determines its direction and is measured in '0'. This angle can be calculated by using above equation.

When 0 = 0 ^{0} , the 'x' and' y' vectors overlap and prove to be similar.

When 0 = 90 ^{0} , the 'x' and' y' vectors are therefore dissimilar.

The steps involved are:

l . Firstly, the two datasets "tmdb 5000 credits.csv" and "tmdb 5000 movies.csv" are merged into one file and data pre-processing is done to remove the unnecessary tags and necessary tags like Director Name, Main Character Name, etc. are kept.

2. Now all the tags of 5000 movies are created and that is used as a reference point for each movie.

3. Then we compare each movie with the reference point and vectors are created for each moue.

4. Now whenever a user selects a movie the nearest 21 vectors ie movies are shown and the distance is calculated by using cosine similarity.

- NB algorithm

The NB Algorithm uses conditional probability to classify the given data set. Bayes theorem is used for the computation and used class levels represented as feature values or vectors of predictors for classification

$$\text{P(AIB) =} \frac{P(B|A)P(A)}{P(B)}$$

The above equation calculates the conditional probability of event A such that B has already occurred and this is used to for calculation in the NB classifier.

Types of models under NB algorithm

- 1 . Gaussian,
- 2. Multinomial,
- 3. Bernoulli.

The proposed system uses the multinomial NB model, which predicts the badge of a text such as a piece of email or newspaper article. The probability of each badge is calculated for a given sample and then badge with the highest probability is given as output.

Since this algorithm is mainly used for natural language processing and text data analysis it was a perfect choice for sentiment analysis of movie reviews.

CHAPTER 5
IMPLEMENTATION AND TESTING

5.1 Implementation Overview

5.1.1 Programing Language Tools

For the Movie Recommender: Content Based Movie Recommendation System with Sentiment Analysis project, the primary language used for the backend is Python, and for the frontend, HTML/CSS, JavaScript is used for the web interface.

Python libraries: For the computation and analysis we need certain python libraries which are used to perform analytics. Packages such as SKIearn, Numpy, pandas, Matplotlib, Flask framework, etc are needed.

SKIearn: It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

NumPy: NumPy is a general-purpose array-processing package. It provides a highperformance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python.

Pandas: Pandas is one of the most widely used python libraries in data science. It provides high-performance, easy to use structures and data analysis tools. Unlike NumPy library which provides objects for multi-dimensional arrays, Pandas provides in-memory 2d table object called Data frame.

Flask: It is a lightweight WSGI web application framework. It is designed to make getting started quick and easy, with the ability to scale up to complex applications. It began as a simple wrapper around Werkzeug.

5.1.2 Modules Description

This project implements following modules:

1 . Recommendation Module: This module is used to recommend movies according to the user input.

2. Sentiment Analysis Module: This module is used to classify the user reviews.

3. Scrape Module: This module is used to scrape the user reviews and movie information off of the Wikipedia and IMDB site.

5.2 Testing

System testing is actually a series of different tests whose primary purpose is to fully exercise the computer-based system. Although each test has a different purpose, all work to verify that all the system elements have been properly integrated and perform allocated functions. The testing process is actually carried out to make sure that the product exactly does the same thing what is supposed to do. In the testing stage following goals are tried to achieve:

- To affirm the quality of the project.

e To find and eliminate any residual errors from previous stages.

e To validate the software as a solution to the original problem.

e To provide operational reliability of the system.

Testing Methodologies

There are many different types of testing methods or techniques used as part of the software testing methodology. Some of the important testing methodologies are:

5.2.1 Unit Testing

Unit testing is the first level of testing and is often performed by the developers themselves. It is the process of ensuring individual components of a piece of software at the code level are functional and work as they were designed to. Developers in a testdriven environment will typically write and run the tests prior to the software or feature being passed over to the test team. Unit testing can be conducted manually, but automating the process will speed up delivery cycles and expand test coverage. Unit testing will also make debugging easier because finding issues earlier means they take less time to fix than if they were discovered later in the testing process. Test Left is a tool that allows advanced testers and developers to shift left with the fastest test automation tool embedded in any IDE.

Table 5. 1: Test Cases for Interface

	Test Cases	Feature	Test Description	steps To Execute	Expected Results	Remarks
1.	TC-IV-I	User Interface	Check all the textboxes and buttons	Check Page	• UI should be perfect • Text boxes and button should be aligned	Pass

2.	TC-1V-2	Required Fields	Check the required fields by not filling any data.	Enter invalid movie name	Error message should be shown	Pass
3.	TC-1V-3	Autocomplete	Check whether the movies are shown only one word entered.	Enter first word of the movie	Movies containing the entered word should be shown.	Pass

Table 5. 2: Test Cases for Model Accuracy

	Test Cases	Feature	Test Description	steps To Execute	Expected Results	Remarks
1.	TC-MA-1	Sentiment analysis testing	Tested on Test Data of our dataset	Check accuracy	Good Accuracy	Pass
2	TC-MA2	Recommendation testing	Tested on Test Data of our dataset	Check accuracy	Good Accuracy	Pass

Table 5. 3: Test Cases for Performance

	Test Cases	Feature	Test Description	Steps To Execute	Expected Results	Remarks
1.	TC-P-I	Waiting period.	System performs efficiently under normal and peak loads	Check loading time.	Recommends movie in minimal time	Pass

2.	TC-P-2	Waiting period.	System performs efficiently under normal and peak loads	Check loading time.	Performs sentiment analysis in minimal time.	Pass.

5.2.2 System Testing

System testing is a black box testing method used to evaluate the completed and integrated system, as a whole, to ensure it meets specified requirements. The functionality of the software is tested from end-to-end and is typically conducted by a separate testing team than the development team before the product is pushed into production.

Table 5. 4: Test Cases for Web Scraping

	Test Cases	Feature	Test Description	Steps To Execute	Expected Results	Remarks
1.	TC-WS-1	Information scrapping	Extract the relevant information about movies from the site.	Check the contents	Scrapped the relevant and needed information from the site.	Pass
2.	TC-WS2	Reviews scrapping.	Extract the user reviews from the site.	Check reviews.	Scrapped the relevant reviews from the site.	Pass

5.3 Result Analysis

The system was tested through unit testing and proved to be effective in executing its intended functions. The results showed that the project was able to meet its goals, but there is still room for improvement in terms of expanding the system's capabilities and increasing community involvement.

5.3.1 Evaluating Accuracy

In machine learning, accuracy is a common metric used to evaluate the performance of a classifier model. Accuracy measures the proportion of correctly classified instances among all instances in the dataset. To calculate accuracy, the first step is to divide the dataset into two parts: a training set and a test set. The training set is used to train the model, while the test set is used to evaluate the model's performance.

In classifier model the most common measure to evaluate accuracy are:

- Precision: Precision is the fraction of true positives among all the positive predictions made by the model. It measures how accurate the model is when predicting positive instances. The formula for precision is:

  Precision = True Positives / (True Positives + False Positives)

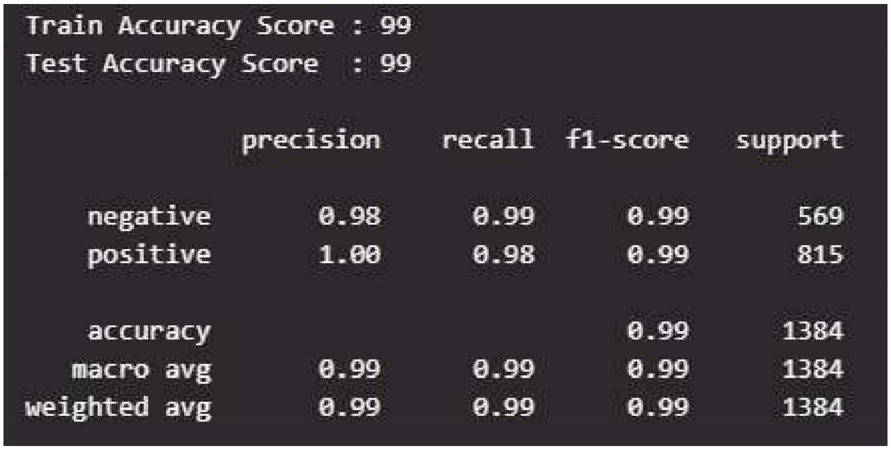
- Recall: Recall is the fraction of true positives among all the actual positive instances in the dataset. It measures how well the model is able to identify positive instances.

  The formula for recall is:

  Recall = True Positives / (True Positives + False Negatives)

- Fl score: The Fl score is the harmonic mean of precision and recall. It provides a single score that balances the tradeoff between precision and recall. The Fl score ranges from 0 to 1, where a score of 1 represents perfect precision and recall, and 0 represents the worst performance. The formula for Fl score is:

  Fl score = 2 * (Precision * Recall) / (Precision + Recall)



```
Train Accuracy Score : 99
Test Accuracy Score  : 99

              precision    recall  f1-score   support

    negative       0.98      0.99      0.99       569
    positive       1.00      0.98      0.99       815

    accuracy                           0.99      1384
   macro avg       0.99      0.99      0.99      1384
weighted avg       0.99      0.99      0.99      1384
```

Figure 5. 1: Accuracy of sentiment analysis

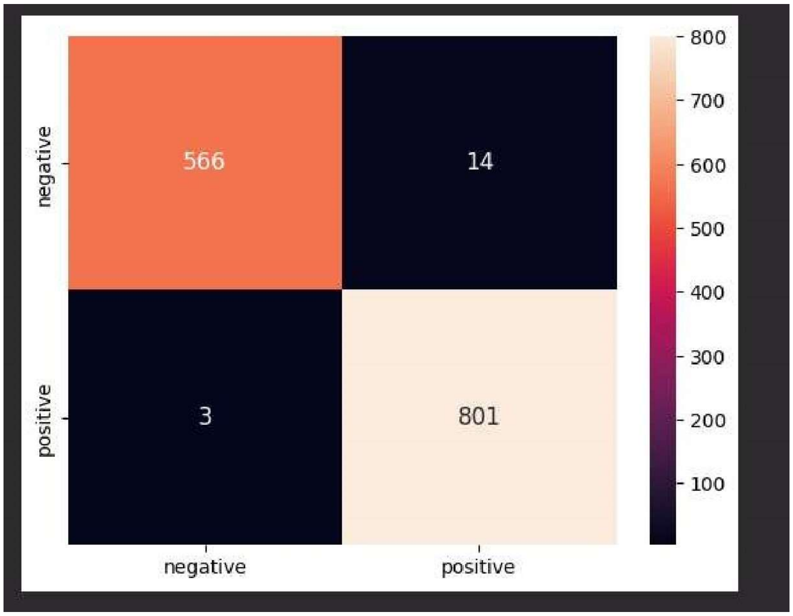


Figure 5. 2: Confusion Matrix

CHAPTER 6
CONCLUSION AND FUTURE RECOMMENDATIONS

6.1 Conclusion

Thus, the recommender system was successfully implemented. We found that ContentBased Collaborative Filtering was the best in our situation. For working on large dataset, it was an approach in implementing the algorithm and making it a Recommender System.

It was a challenge to implement a recommender system on this scale of huge data. Recommender systems have become ubiquitous. People use them to find books, music, news, smart phones, vacation trips, and romantic partners. Nearly every product, service, or type of information has recommenders to help people select from among the myriad alternatives the few they would most appreciate. Sustaining these commercial applications is a vibrant research community, with creative interaction ideas, powerful new algorithms, and careful experiments.

6.2 Future Recommendations

In the future, we would like to improve the suggestion process by:

1. Add registration feature:

   We can add user registration feature to store each users unique preferences and recommend them movies accordingly.

2. Introduce user dislike movie list. The user data is always useful in recommender systems. In the future we can collect more user data and add user dislike movie list. We will input dislike movie list into the recommender system as well and generate scores that will be added to previous result. By this way we can improve the result of recommender system.

3. Make the recommender system as an internal service. In the future, we can make it as an internal APIs for developers to invoke. Some movie lists in the website will be sorted by recommendation.

REFERENCES

[l] P. P. &. D. S. Pal, "A review of movie recommendation systems.," 2021.

[2] J. Z. Q. L. J. Y. Y. &. Y. H. Zheng, "Privacy-preserving movie recommendation based on homomorphic encryption.," IEEE Access, 7, 25916-25925., 2019.

[3] X. &. M. M. Wu, "User experience in movie recommendation systems: Personalization and interactivity.," International Journal of Human-Computer Interaction, 2019.

[4] P. K. &. S. M. Shukla, "A hybrid approach to solve cold start problem in movie recommendation system," International Conference on Information Systems and Computer Networks, 2018.

APPENDICES

(Screenshots)

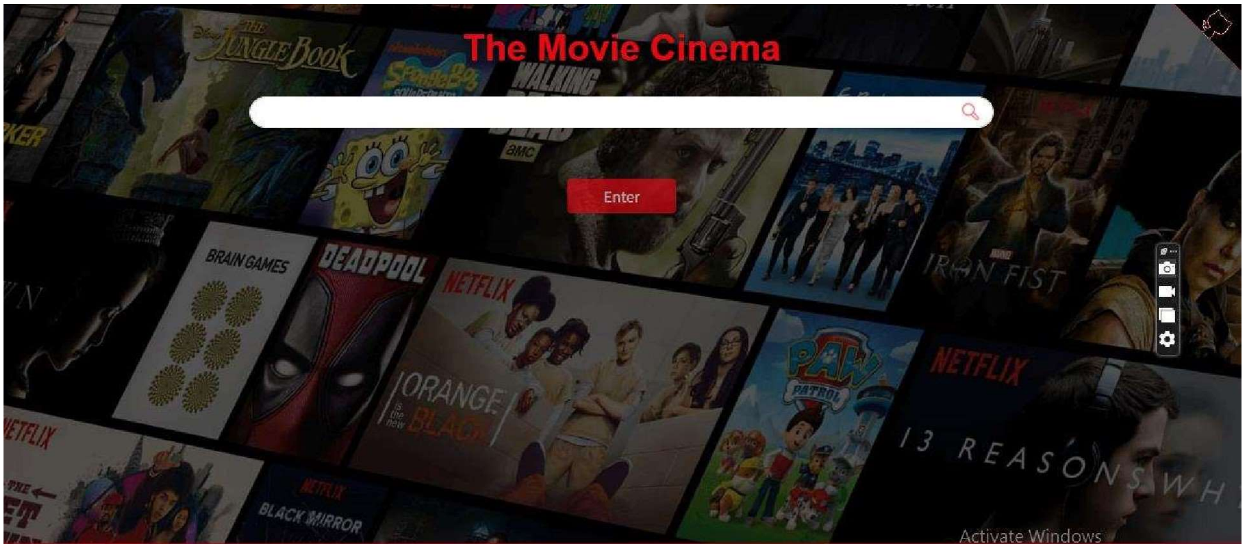


Fig: Homepage

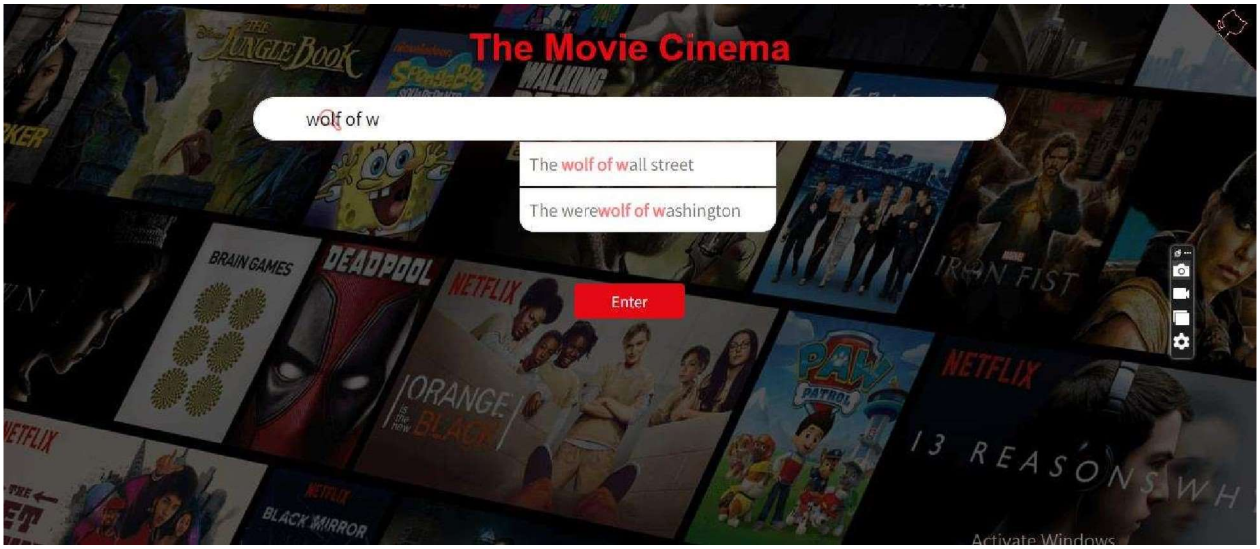
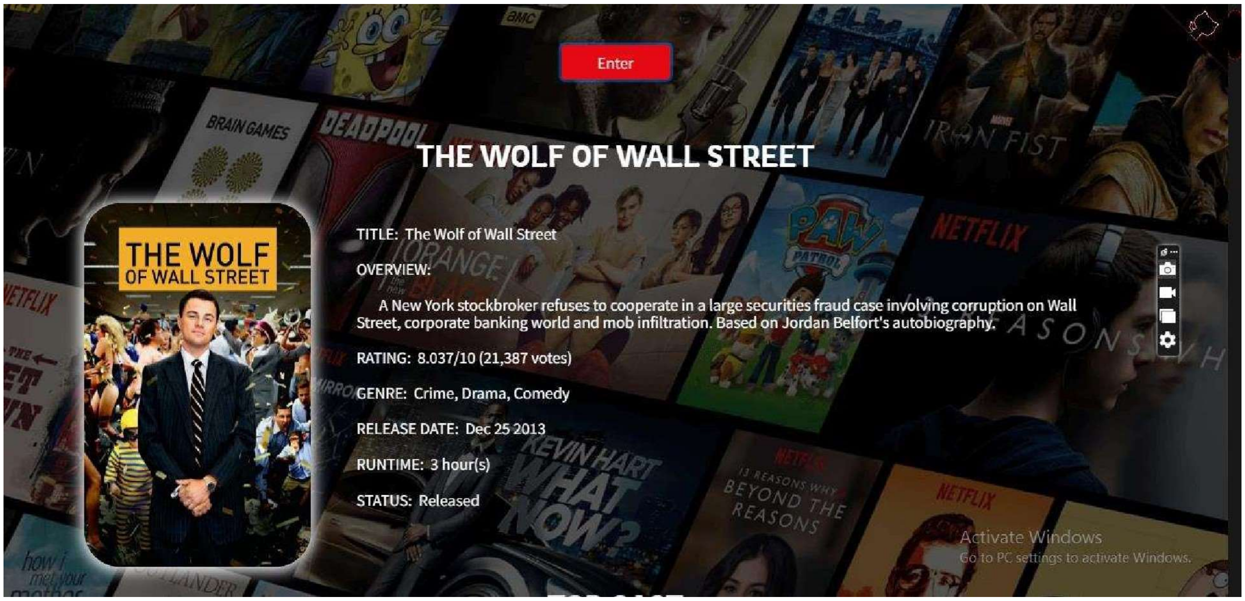


Fig: Autocomplete feature



THE WOLF OF WALL STREET

TITLE: The Wolf of Wall Street

OVERVIEW:

A New York stockbroker refuses to cooperate in a large securities fraud case involving corruption on Wall Street, corporate banking world and mob infiltration. Based on Jordan Belfort's autobiography.

RATING: 8.037/10 (21,387 votes)

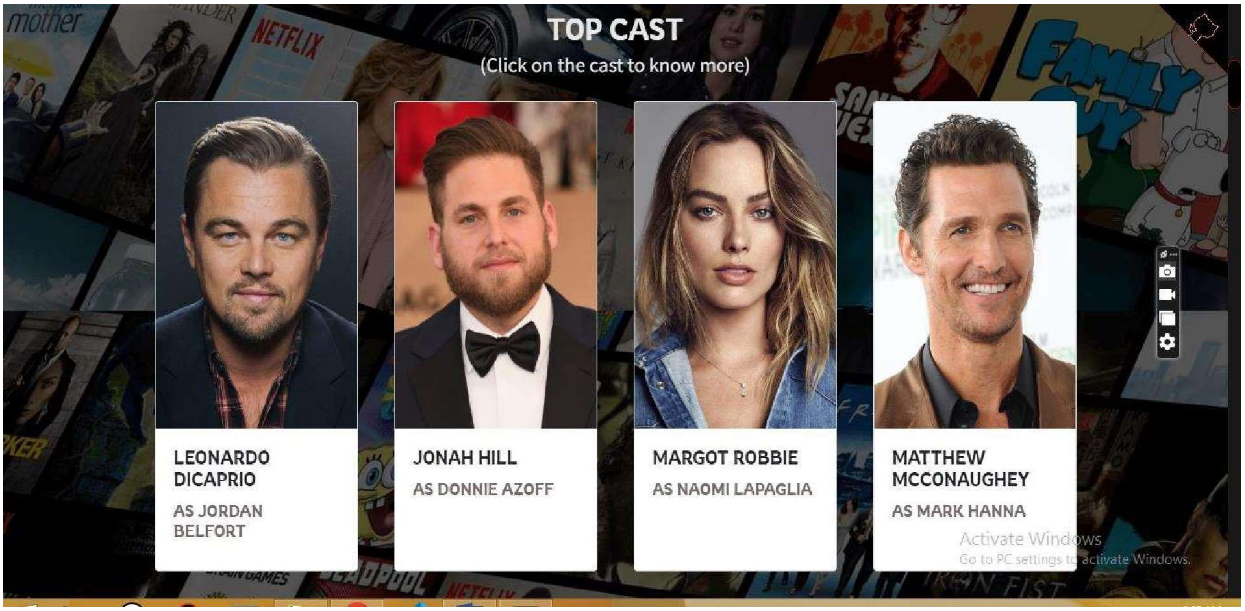
GENRE: Crime, Drama, Comedy

RELEASE DATE: Dec 25 2013

RUNTIME: 3 hour(s)

STATUS: Released

Fig: Information on selected movie


TOP CAST

(Click on the cast to know more)

LEONARDO DICAPRIO
AS JORDAN BELFORT

JONAH HILL
AS DONNIE AZOFF

MARGOT ROBBIE
AS NAOMI LAPAGLIA

MATTHEW MCCONAUGHEY
AS MARK HANNA

Fig: Cast of selected movie

Leonardo DiCaprio

Birthday: Nov 11 1974

Place of Birth: Los Angeles, California, LJSA

Biography,

Leonardo Wilhelm DiCaprio (born November 11, 1974) is an American actor and film producer. Known for his work in biopics and period films, DiCaprio is the recipient of numerous accolades, including an Academy Award, a British Academy Film Award, and three Golden Globe Awards. As of 2019, his films have grossed over $7.2 billion worldwide, and he has been placed eight times in annual rankings of the world's highest-paid actors, Born -in Los Angeles, DiCaprio began his career in the late 1980s by appearing in television commercials. In the early 1990S, he had recurring roles in various television shows, such as the sitcom Parenthood, and had his first major film part as author Tobias Wolff in This Boy's Life (1993). At age 19, he received critical acclaim and his first Academy Award and Golden Globe Award nominations for his performance as a developmentally disabled boy in What's Eating Gi lbert Grape (1993). He achieved internatiOnal stardom with the star-crossed romances Romeo +duliet (1996) and Titanic (1997). After the latter became the highest-grossing film at the time, he reduced his workload for few years. In an attempt to shed his image of a romantic hero, DiCaprio sought roles in other genres,

r rirna r'rarna

LEONA

DICAP

AS JOR

BELFOR

GREY

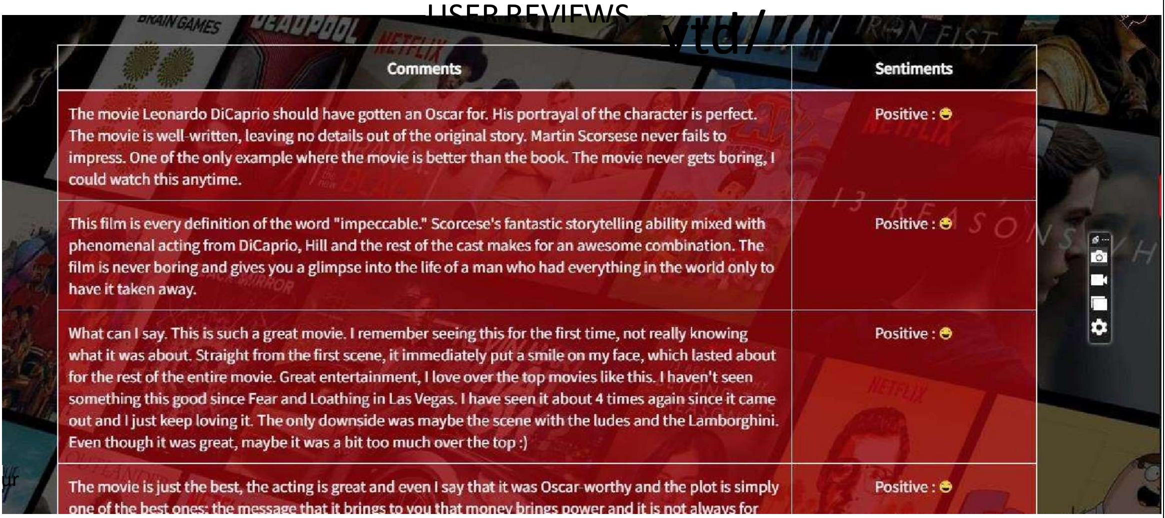
NNA

ctivate Wind ws

to ;'C, activate Windows.

Fig: Detailed information about the cast member when clicked upon

USER REVIEWS ytd/



Comments	Sentiments
The movie Leonardo DiCaprio should have gotten an Oscar for. His portrayal of the character is perfect. The movie is well-written, leaving no details out of the original story. Martin Scorsese never fails to impress. One of the only example where the movie is better than the book. The movie never gets boring, I could watch this anytime.	Positive : 😀
This film is every definition of the word "impeccable." Scorcese's fantastic storytelling ability mixed with phenomenal acting from DiCaprio, Hill and the rest of the cast makes for an awesome combination. The film is never boring and gives you a glimpse into the life of a man who had everything in the world only to have it taken away.	Positive : 😀
What can I say. This is such a great movie. I remember seeing this for the first time, not really knowing what it was about. Straight from the first scene, it immediately put a smile on my face, which lasted about for the rest of the entire movie. Great entertainment, I love over the top movies like this. I haven't seen something this good since Fear and Loathing in Las Vegas. I have seen it about 4 times again since it came out and I just keep loving it. The only downside was maybe the scene with the ludes and the Lamborghini. Even though it was great, maybe it was a bit too much over the top :)	Positive : 😀
The movie is just the best, the acting is great and even I say that it was Oscar-worthy and the plot is simply one of the best ones; the message that it brings to you that money brings power and it is not always for	Positive : 😀

metpur

movie Leonardo DiCaprio should have gotten an Oscar movie is well-written, leaving no details out of the original story. Martin Scorsese neverfails to impress. One ofthe only example where the movie is better than the book The movie never gets boring, I

l of the best ones; the message that it brings to you that money brings power and it is not always for est because in this movie we see the downfall of a man that will learn that actually in life the money

be a drug and will simply make you cause damage to yourself.

The movie isjust the best, the acting is great and even say that it was Oscar-worthy and the plot is simply one the b

Windows

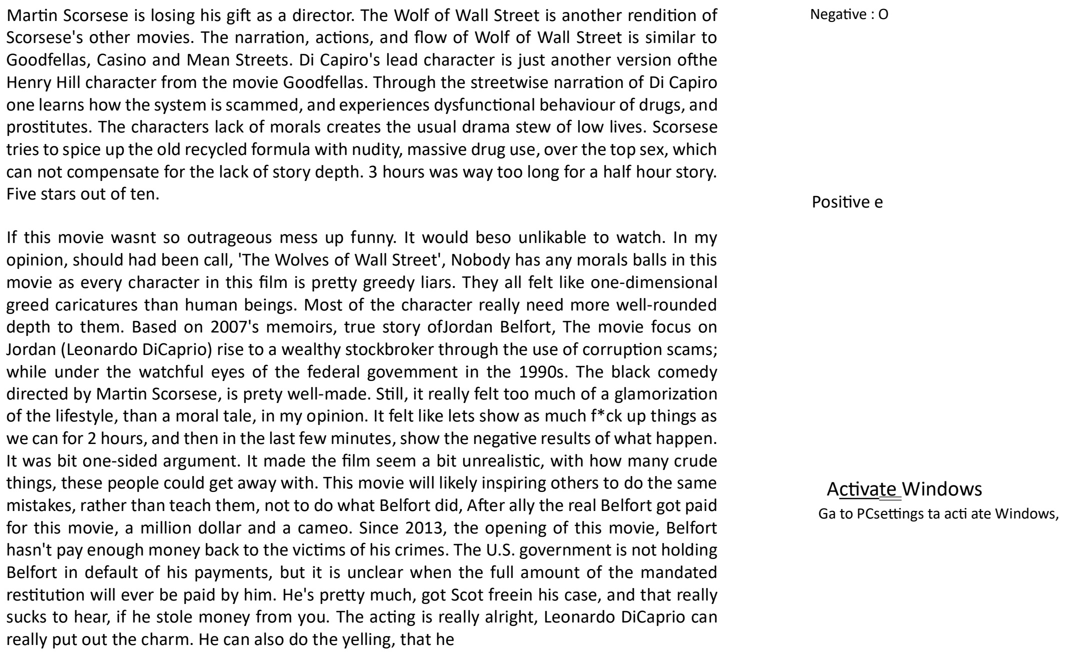
Positive : e

Activate

ate Wi

can Ga to PC settings ta acti

Fig: User Reviews being labelled "Positive" or "Negative"



Martin Scorsese is losing his gift as a director. The Wolf of Wall Street is another rendition of Scorsese's other movies. The narration, actions, and flow of Wolf of Wall Street is similar to Goodfellas, Casino and Mean Streets. Di Capiro's lead character is just another version ofthe Henry Hill character from the movie Goodfellas. Through the streetwise narration of Di Capiro one learns how the system is scammed, and experiences dysfunctional behaviour of drugs, and prostitutes. The characters lack of morals creates the usual drama stew of low lives. Scorsese tries to spice up the old recycled formula with nudity, massive drug use, over the top sex, which can not compensate for the lack of story depth. 3 hours was way too long for a half hour story. Five stars out of ten.

Negative : O

If this movie wasnt so outrageous mess up funny. It would beso unlikable to watch. In my opinion, should had been call, 'The Wolves of Wall Street', Nobody has any morals balls in this movie as every character in this film is pretty greedy liars. They all felt like one-dimensional greed caricatures than human beings. Most of the character really need more well-rounded depth to them. Based on 2007's memoirs, true story ofJordan Belfort, The movie focus on Jordan (Leonardo DiCaprio) rise to a wealthy stockbroker through the use of corruption scams; while under the watchful eyes of the federal govemment in the 1990s. The black comedy directed by Martin Scorsese, is prety well-made. Still, it really felt too much of a glamorization of the lifestyle, than a moral tale, in my opinion. It felt like lets show as much f*ck up things as we can for 2 hours, and then in the last few minutes, show the negative results of what happen. It was bit one-sided argument. It made the film seem a bit unrealistic, with how many crude things, these people could get away with. This movie will likely inspiring others to do the same mistakes, rather than teach them, not to do what Belfort did, After ally the real Belfort got paid for this movie, a million dollar and a cameo. Since 2013, the opening of this movie, Belfort hasn't pay enough money back to the victims of his crimes. The U.S. government is not holding Belfort in default of his payments, but it is unclear when the full amount of the mandated restitution will ever be paid by him. He's pretty much, got Scot freein his case, and that really sucks to hear, if he stole money from you. The acting is really alright, Leonardo DiCaprio can really put out the charm. He can also do the yelling, that he

Positive e

Activate Windows

Ga to PCsettings ta acti ate Windows,

Fig: User Reviews being labelled "Positive" or "Negative"

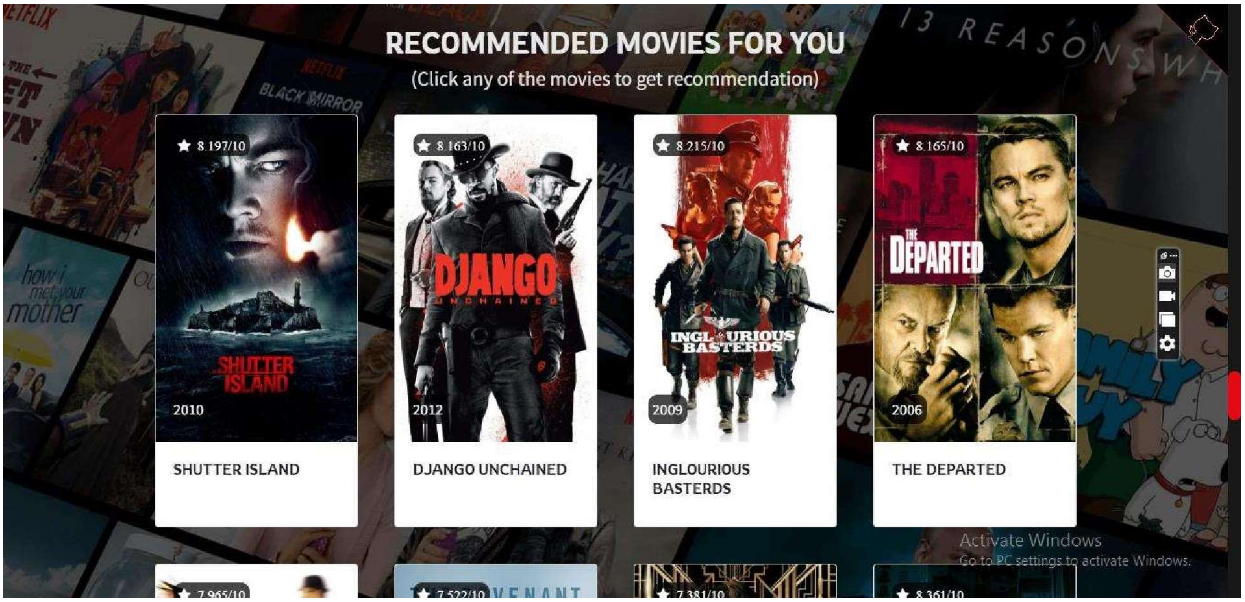


Fig: Recommended Movies

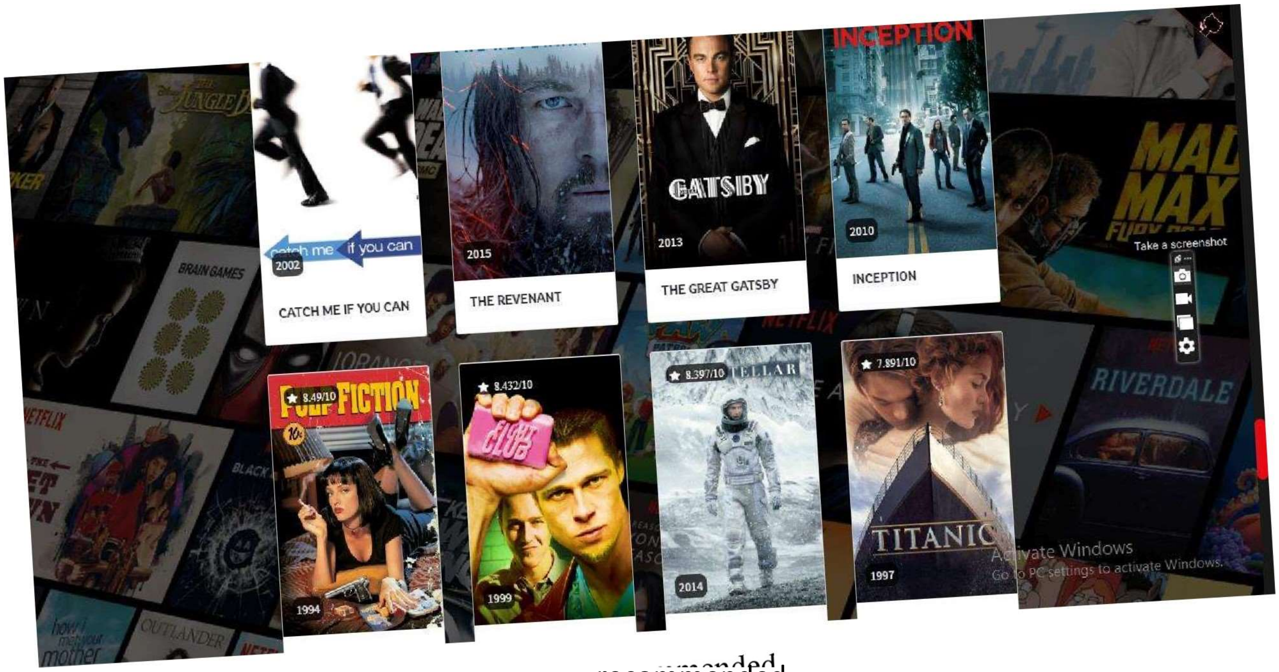


Fig: More recommended