

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

**Project Report
Semester-IV(Batch-2022)**

**Title of the project:
“Movie Mate:Your Family Companion”**



Submitted To:
Dr.Jatin Arora

Submitted by:
Kashish Mahana(2210990497)
Khushi Bansal (2210990511)

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
CHITKARA UNIVERSITY INSTITUTE OF ENGINEERING AND TECHNOLOGY,
CHITKARA UNIVERSITY PUNJAB**

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 Naïve Bayes algorithm.

Keywords: *Cosine similarity, Movie recommendation, Naïve Bayes, Sentiment analysis*

TABLE OF CONTENTS

ACKNOWLEDGEMENT.....	i
-----------------------------	----------

ABSTRACT	ii
-----------------------	-----------

TABLE OF CONTENTS	iii
------------------------------------	------------

CHAPTER 1

INTRODUCTION	1
---------------------------	----------

• 1.1 Introduction.....	1
• 1.2 Objectives	1
• 1.3 Background Study.....	2

CHAPTER 2 PROBLEM DEFINITION AND REQUIREMENTS.....

3

2.1Problem Statement.....	3
---------------------------	---

2.2 Software Analysis.....	3-4
----------------------------	-----

2.3 Hardware Requirements.....	4
--------------------------------	---

CHAPTER 3 PROPOSED DESIGN/ METHODOLOGY.....

5

3.1 Analysis.....	5
-------------------	---

3.1.1 Class Diagram.....	5-6
3.1.2 Sequence Diagram.....	6-7
3.1.3 Activity Diagram.....	8-9
3.2 Algorithms.....	10-13

CHAPTER 4 Result.....13

4.1 Result Analysis.....	13
4.2 Evaluating Analysis.....	13-14
4.2 Screenshots.....	14-18

REFERENCES.....19

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 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.

1.3 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.

CHAPTER 2

PROBLEM DEFINITION AND REQUIREMENTS

2.1 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

2.2 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 SKlearn, Numpy, pandas, Matplotlib, Flask framework, etc are needed.
- **SKlearn:** 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.

2.3 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

CHAPTER 3

PROPOSED DESIGN/METHODOLOGY

3.1 Analysis

3.1.1 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 application, however class diagram is a bit different. It is the most popular UML diagram in the coder community.

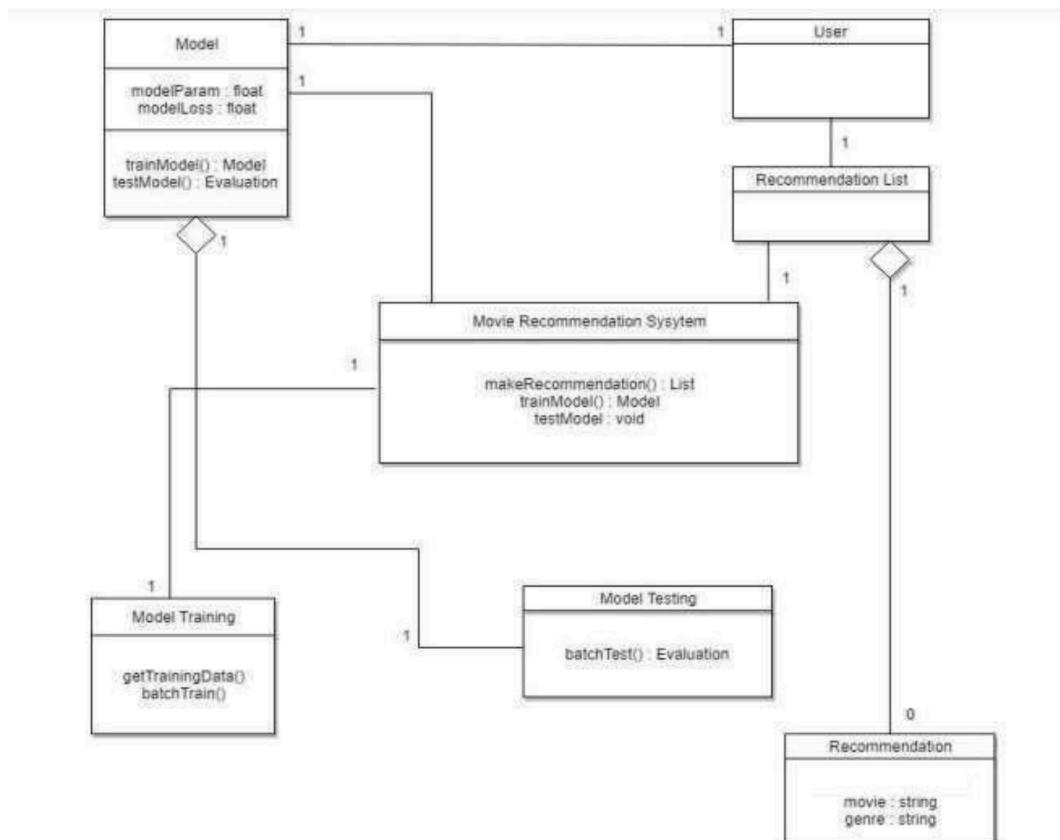


Figure 3. 4: Class Diagram of Movie Recommender System

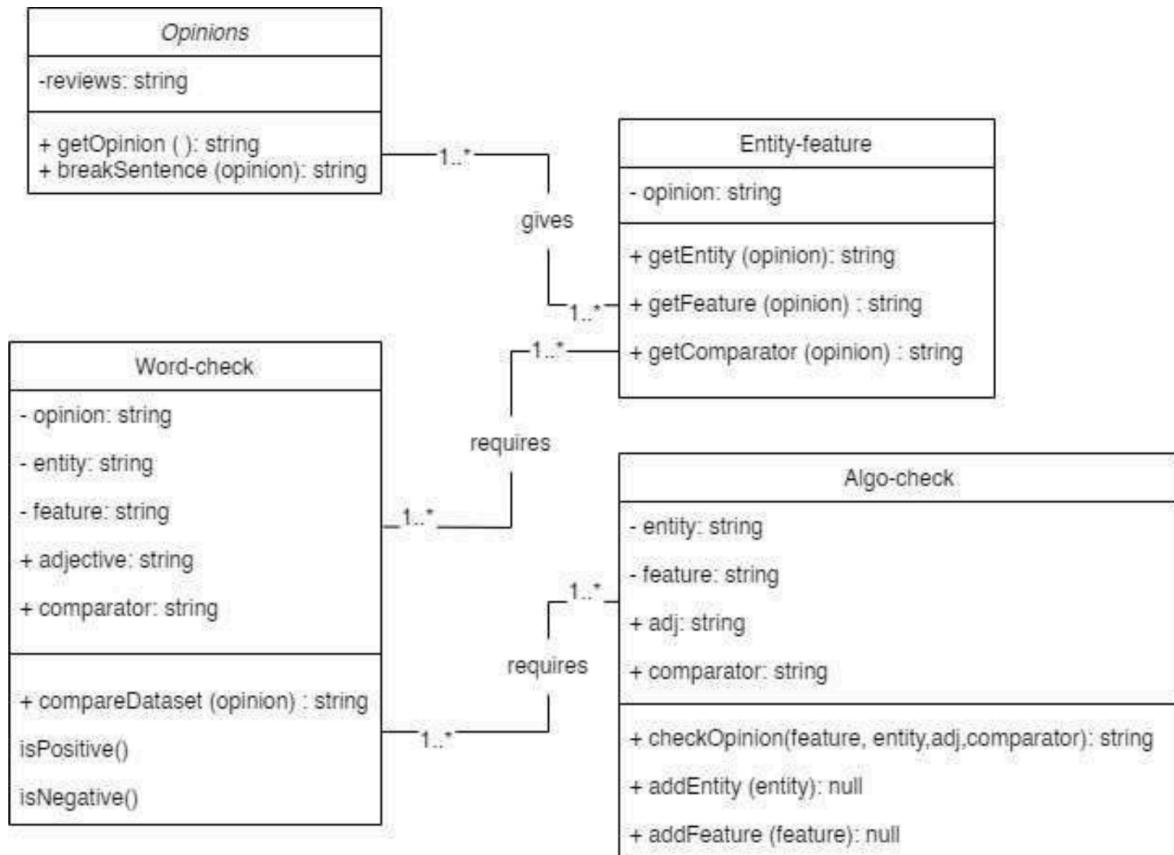


Figure 3. 5: Class Diagram of sentiment analysis

3.1.2 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.

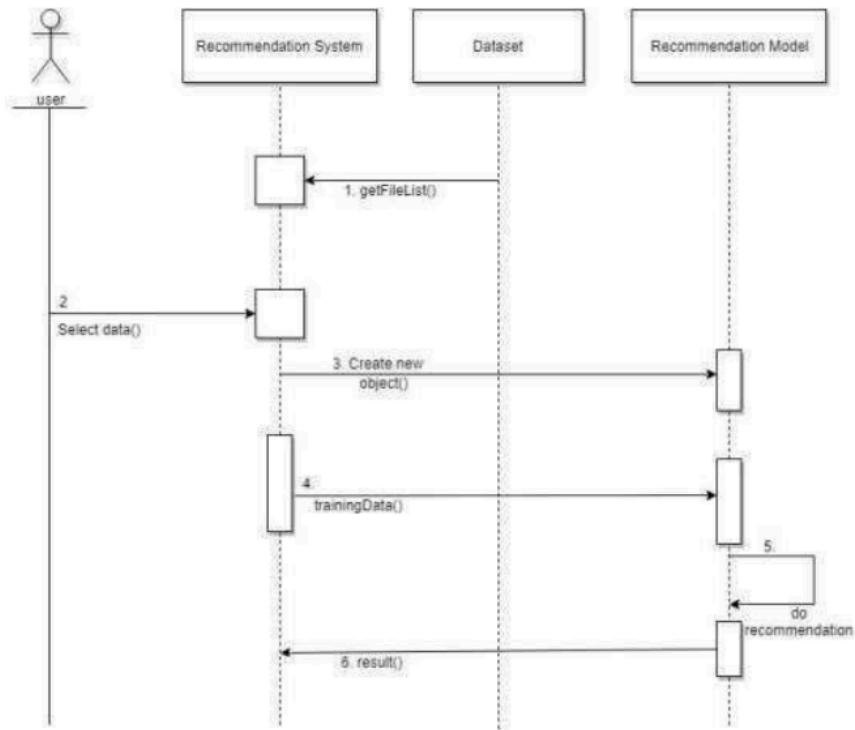


Figure 3. 6: Sequence diagram of movie recommender system

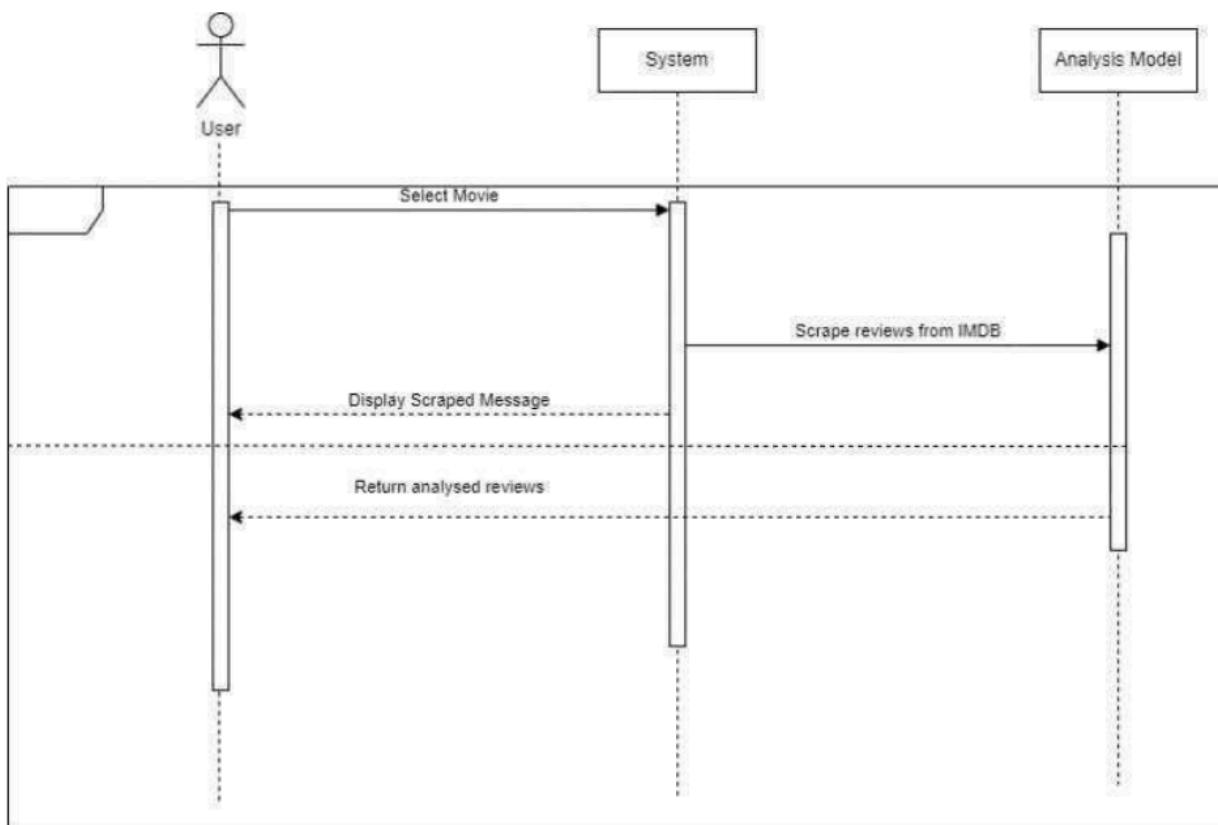


Figure 3. 7 Sequence diagram of sentiment analysis

3.1.3Activity 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.

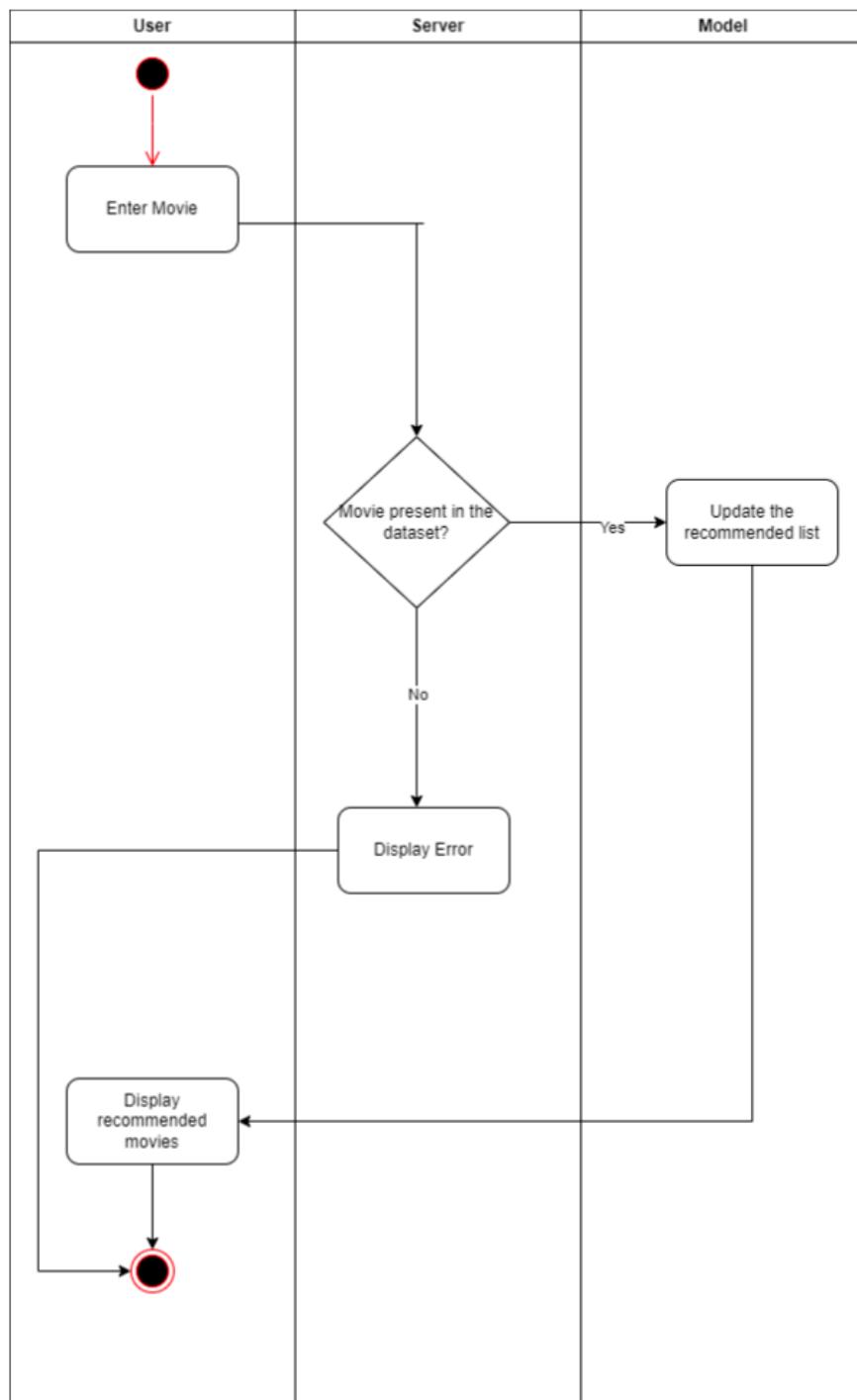


Figure 3. 8: Activity Diagram of Movie Recommender System

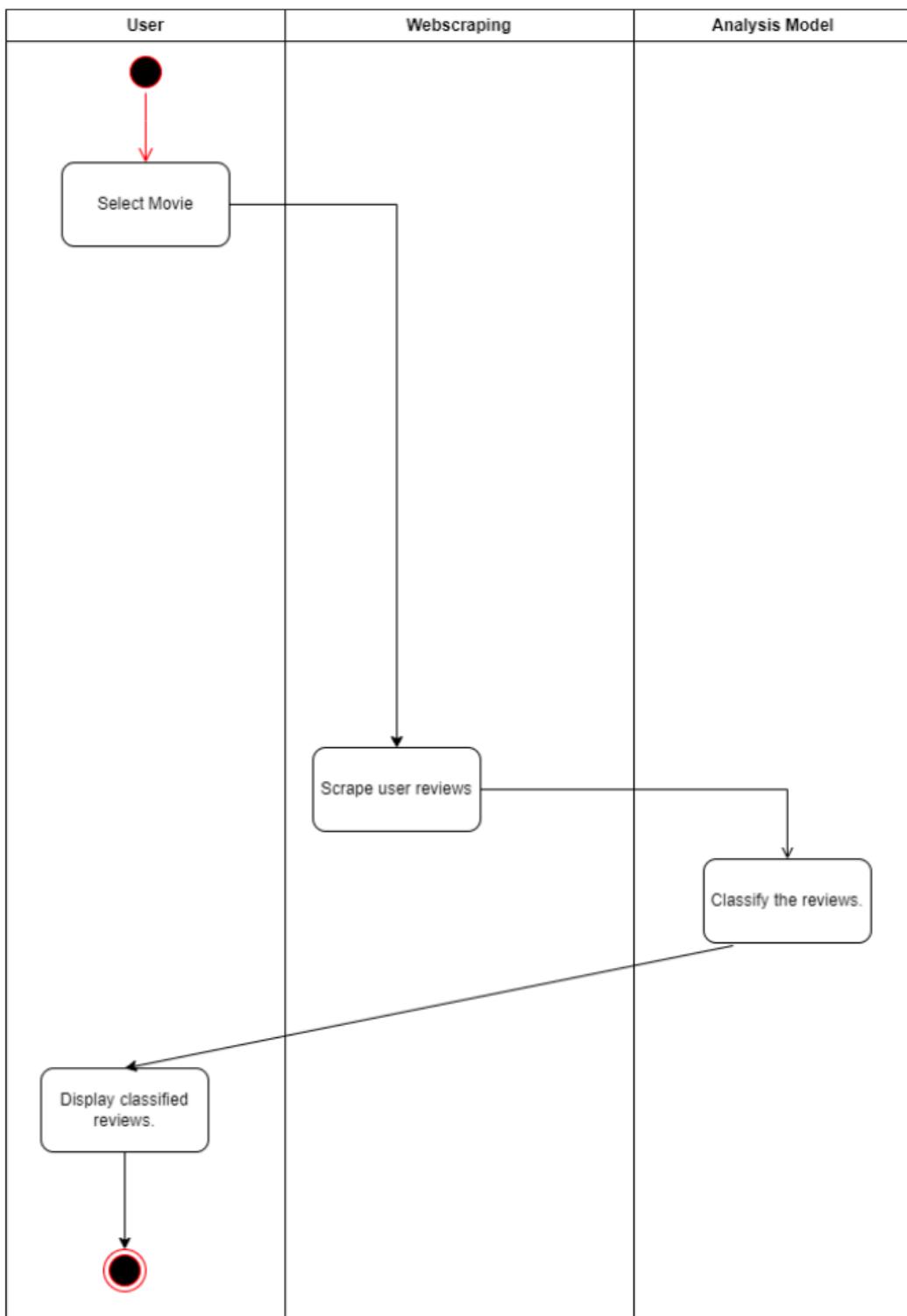


Figure 3. 9: Activity diagram of sentiment analysis

3.2 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

$$\sum^n aibi$$

$$\sqrt{\sum^n a^2} \sqrt{\sum^n b^2}$$

Where, $\vec{a} \cdot \vec{b} = \sum^n aibi = a1b1 + a2b2 + \dots + anbn$ is the dot product of the two vectors.

The angle between two vectors determines its direction and is measured in ‘θ’. This

angle can be calculated by using above equation.

When $\theta = 0^\circ$, the 'x' and 'y' vectors overlap and prove to be similar. When $\theta = 90^\circ$, the 'x' and 'y' vectors are therefore dissimilar.

The steps involved are:

1. 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 movie.
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

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

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 4

RESULT

4.1 RESULT ANALYSIS

Thus, the recommender system was successfully implemented. We found that Content- Based 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.

4.2 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:

$$\text{Precision} = \text{True Positives} / (\text{True Positives} + \text{False Positives})$$

- 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:

$$\text{Recall} = \text{True Positives} / (\text{True Positives} + \text{False Negatives})$$

- F1 score: The F1 score is the harmonic mean of precision and recall. It provides a single score that balances the tradeoff between precision and recall. The F1 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 F1 score is:

$$\text{F1 score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

- | 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

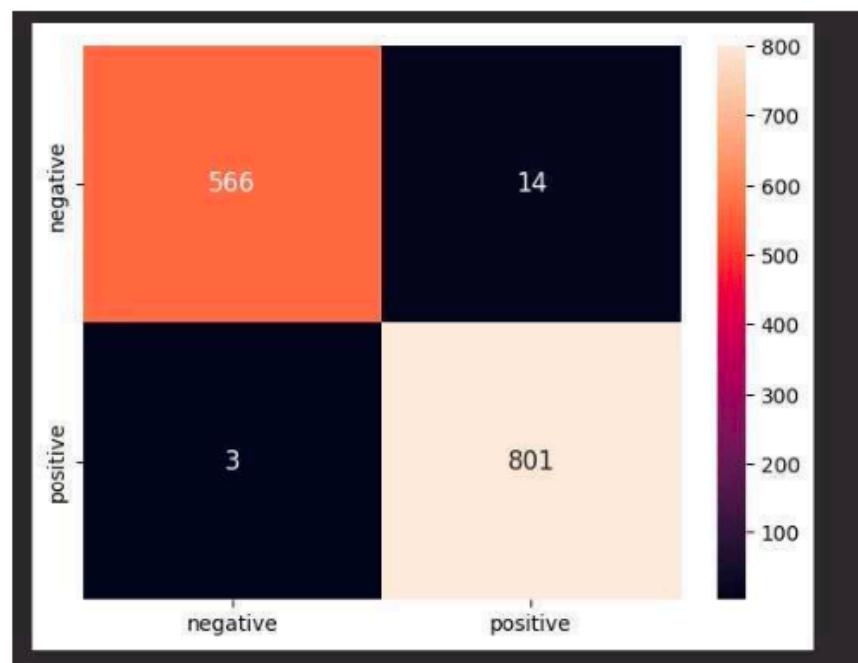


Figure 5. 2: Confusion Matrix

4.3 Screenshots

APPENDICES

(Screenshots)

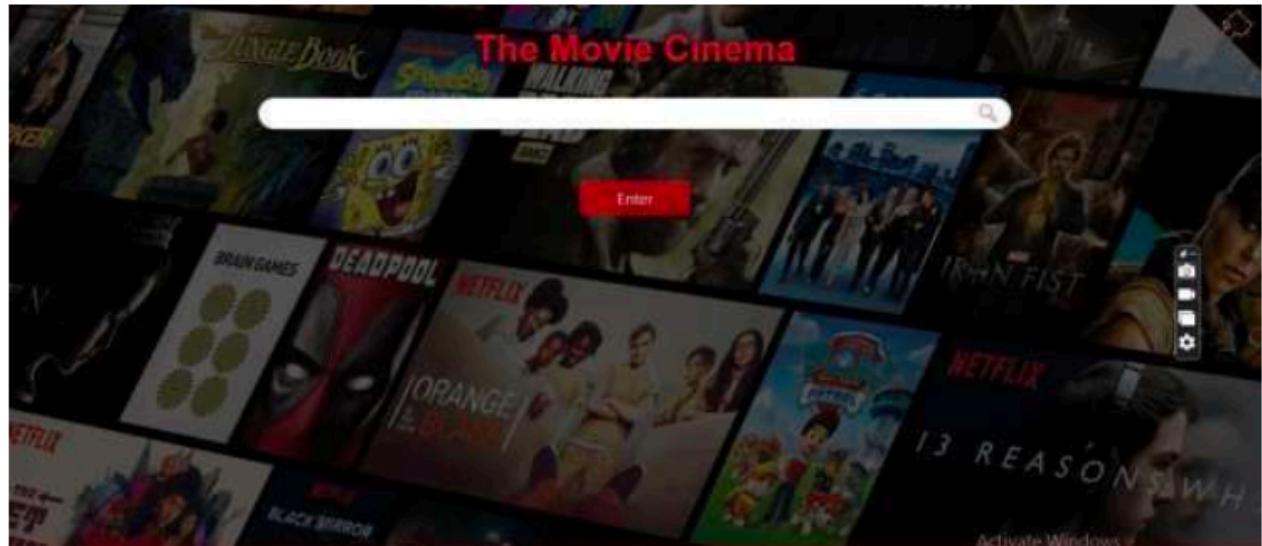


Fig: Homepage

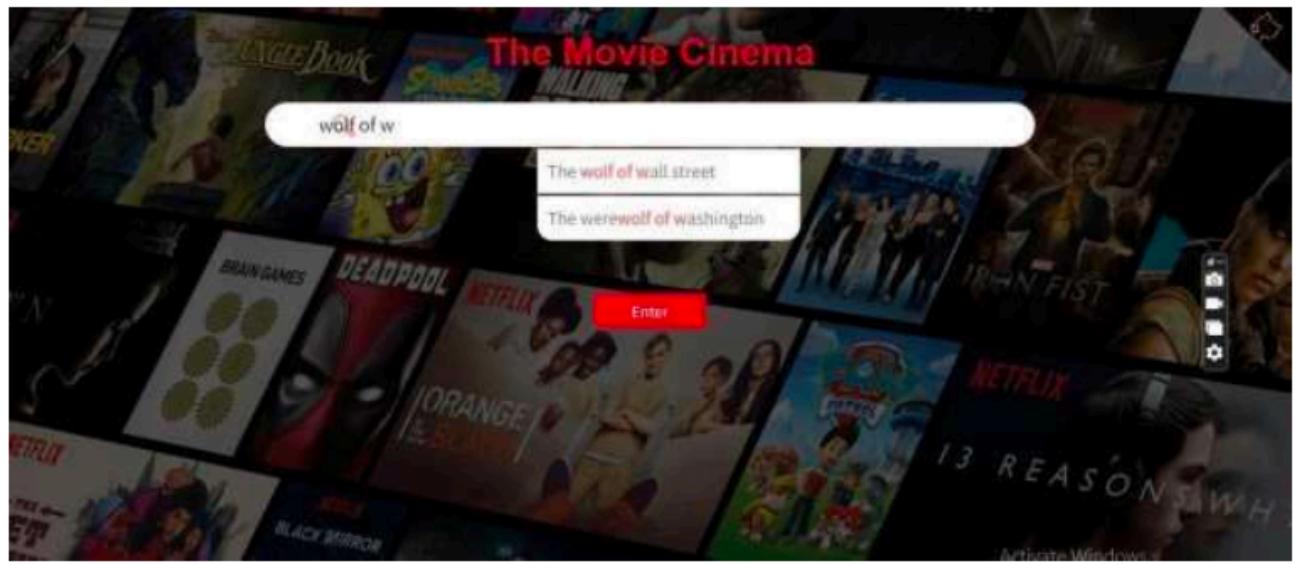


Fig: Autocomplete feature

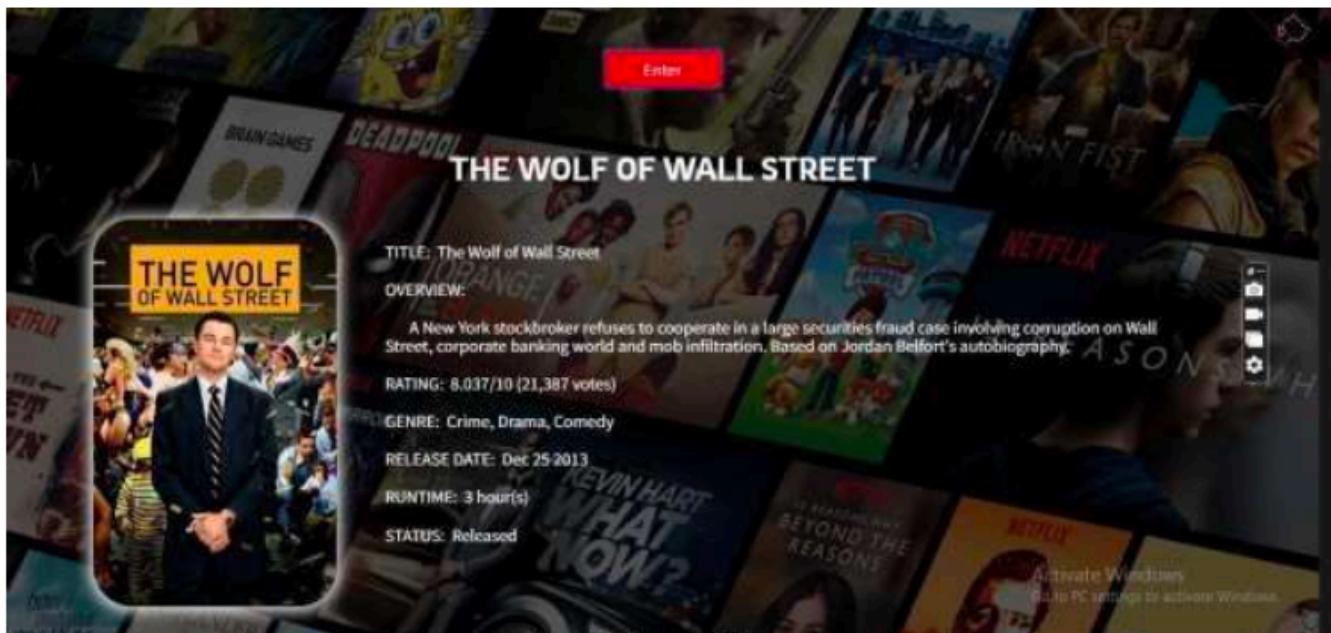


Fig: Information on selected movie



Fig: Cast of selected movie



Fig: Detailed information about the cast member when clicked upon

USER REVIEWS	
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 : 8
This film is every definition of the word "impeccable." Scorsese'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 : 8
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 Jutes and the Lamborghini. Even though it was great, maybe it was a bit too much over the top :)	Positive : 8
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 the best because in this movie we see the downfall of a man that will learn that actually in life the money can be a drug and will simply make you cause damage to yourself.	Positive : 8

Fig: User Reviews being labelled “Positive” or “Negative”



Fig: User Reviews being labelled “Positive” or “Negative”

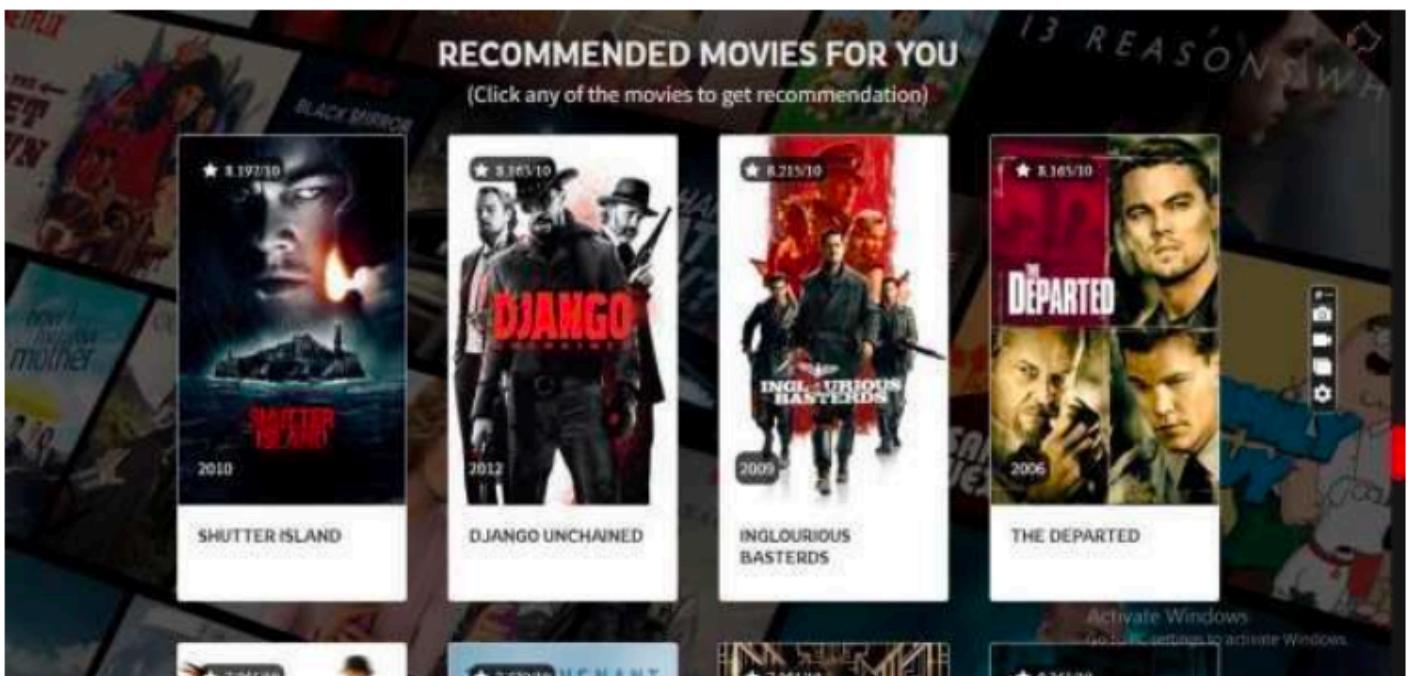


Fig: Recommended Movies



Fig: More recommended

REFERENCES

- [1] 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.