Recommender System

Project report submitted to the Amrita Vishwa Vidyapeetham in partial fulfillment of the requirement for the Degree of

BACHELOR of TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

Submitted by

Devata Dinesh Vamsi Durga Bhaskar AM.EN.U4CSE18314
Madabhushi Aditya AM.EN.U4CSE18333
N Chaithanya AM.EN.U4CSE18138
Dande Dharani AM.EN.U4CSE18364

AMRITA SCHOOL OF ENGINEERING

AMRITA VISHWA VIDYAPEETHAM

(Estd. U/S 3 of the UGC Act 1956)
AMRITAPURI CAMPUS

KOLLAM -690525

December 2021

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

AMRITA VISHWA VIDYAPEETHAM

(Estd. U/S 3 of the UGC Act 1956)

Amritapuri Campus

Kollam -690525

BONAFIDE CERTIFICATE

This is to certify that the project report entitled **"Recommender System"** submit- ted by **Devata Dinesh Vamsi Durga Bhaskar (AM.EN.U4CSE18314), Madabhushi Aditya (AM.EN.U4CSE18333), N Chaitanya (AM.EN.U4CSE18138) and Dande Dharani (AM.EN.U4CSE18364)**, in partial fulfillment of the requirements for the award of Degree of Bache- lor of Technology in Computer Science and Engineering from Amrita Vishwa Vidyapeetham, is a bonafide record of the work carried out by them under my guidance and supervision at Amrita School of Engineering, Amritapuri during Semester 8 of the academic year 2021-2022.

Your Guides Name
Project Guide
Greeshma Sarath Ma'am.

Coordinator name
Project Coordinator
Lekha S Ma'am

Dr. Prema Nedungadi
Chairperson
Dept. of Computer Science & Engineering

Reviewer

Place : Amritapuri
Date :
21 December 2021

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

AMRITA VISHWA VIDYAPEETHAM

(Estd. U/S 3 of the UGC Act 1956)

Amritapuri Campus

Kollam -690525

DECLARATION

We, **Devata Dinesh Vamsi Durga Bhaskar(AM.EN.U4CSE18314), Madabhushi Aditya(AM.EN.U4CSE18333), N Chaithanya (AM.EN.U4CSE18138) and Dande Dharani(AM.EN.U4CSE18364)** hereby declare that this project entitled **"Recommender System"** is a record of the original work done by us under the guidance of **Greeshma Sarath Ma'am**, Dept. of Computer Science and Engineering, Amrita Vishwa Vidyapeetham, that this work has not formed the basis for any degree/diploma/associationship/fellowship or similar awards to any candidate in any university to the best of our knowledge.

Place : Amritapuri

Date :

21December 2021

Signature of the student

Signature of the Project Guide

Acknowledgements

We as a team would like to express our very great appreciation towards our project guide **Greeshma Sarath** and project mentor **Lekha S** for their valuable and constructive suggestions throughout the planning and development of our project, which is very supportive to us to come up with this successful outcome. Their willingness to give their time for our improvement and corrections was very much appreciated.

Also, each and every member of our team has an equal part of work in this project. Each person has contributed their best efforts in each phase that could be esteemed very much. Without which our project would not yield a high output.

Regards,
Team 39BD

Abstract

The internet is widely used these days, due to the advancement of technology. Music and movies are significant aspects of human life that provide people with relaxation during difficult times. More than any other activity, people like to listen to music or watch movies. Millions of individuals across the world now have access to a vast amount of music and movie content from a variety of genres due to internet technology. The main challenge for users is choosing appropriate music or movies from such a large collection.

To overcome this problem a recommender system is developed. System which makes suggestions that usually filters the given data using multiple methods and suggests the relevant one to the customer's benefit. This recommender system will recommend music and movies to the user based on their mood along with an emoticon symbolizing their mood. The mood of the user will be derived by performing sentiment analysis.

The recommendation is done in three ways :

1. Suggestions based on Machine Learning approach (Content- Based Recommendation technique)
2. Suggestions based on User’s mood detection Using Face Detection technique
3. Suggestions based on User’s mood detection using the Conversation of the user with the Chatbot

Contents

Chapter 1

Introduction

1.1 Chatbot and Emotion:

Nowadays it is the era of Artificial Intelligence, i,e. Intelligence of a machine that could interpret a task and perform it, without much explicitly hand-coded by a man. With the advancements of Artificial Intelligence, Machine Learning and Deep Learning machines had started to impersonate as human. Such one beautiful example is Chatbot , a conversational software agent that is activated by natural language processing.

Emotion detection is an important process in this project that which requires the accuracy. And this can be done effectively with the help of facial expressions which is how each and everyone can understand and interpret an emotion. Some research shows that when a person's facial expressions are readable, it can actually vary your interpretation of the task in comparison of what is actually being spoken and it can also control how the conversation turns out to be. Humans are capable of perceiving emotions which are exceedingly important for a communication to be successful and effective.

1.2 Use of Chatbots

Major purpose of the chatbot is to support and scale business teams in their economic relations with the end-customers. It could be used vastly in message/chat applications like Instagram, Slack, Telegram, Facebook Messenger, Text Messages, etc... Chatbot applications lie in streamlining the interactions between people and their services, thus enhancing the customer experience.

An AI Chatbot can analyze the data and perform tasks much better than humans, which yields greater accuracy. In our project, we use Chatbot for identifying the mood of the user based on the conversation that the user is having with the chatbot. The Chatbot module of the application makes the use of AI techniques for its implementation. Our chatbot is rule based which is the AI methodology used to design a simple Chatbot. We have made use of a rule based chatbot as far as our application requires a simple chatbot.

Chapter 2

Problem Definition

In this project we built a Recommendation System that recommends movies and songs based on users emotion through chat with a chatbot or based on users emotion through face detection technique.

Also, here we implemented a chatbot that is used to identify the emotional tone of the user and provides songs accordingly to the mood of the user and an ML model that would recommend the movies to the users according to their preferences and finally a song recommender based on face recognition technique.

Chapter 3

Related Work

The Recommendation system topic is a live platform for researchers to work on all the time, and the recommendations a person requires include movies, books, web series, songs, venues, videos, and websites, among other things. Content filtering, memory-based filtering, and model-based filtering techniques are used in many recommendation systems.

User input, data collecting, ratings, preferences, similarities, and other criteria all play a major role in online system recommendations. The information is gathered based on the input and then used to display the recommendations.The recommendation is based on the ratings and popularity for those who have provided no input or previous choices.

A large group of people are interested in researching everything on the internet. In today's world, filtering algorithms that forecast user preferences are a must. These types of methodologies are used in a variety of fields, including movies, videos, books, news, and ecommerce purchases, where data provided by the user is analyzed and the best options are presented to the customer. Using K-means and collaborative approaches the above purpose is solved.

On a smaller dataset, John Dalton discovered that user-based collaborative filtering lowers mean square error. Racheal Carson presented a method based on a content-based approach that considers related objects, that is, the user will be recommended items with similar content. It uses cosine similarity algorithms to find similar content. Mohammad used a collaborative filtering approach. This will be based on user ratings, limiting one's ability to travel further.

Chapter 4

Requirements

The design of this project contains some sort of software and hardware adjustments to be taken care of. The specifications of the project are listed below:

4.1 Hardware

For running this project any device which could have a CPU and memory unit like a Standard PC, Desktop, Laptop with a minimum available configuration is good enough.

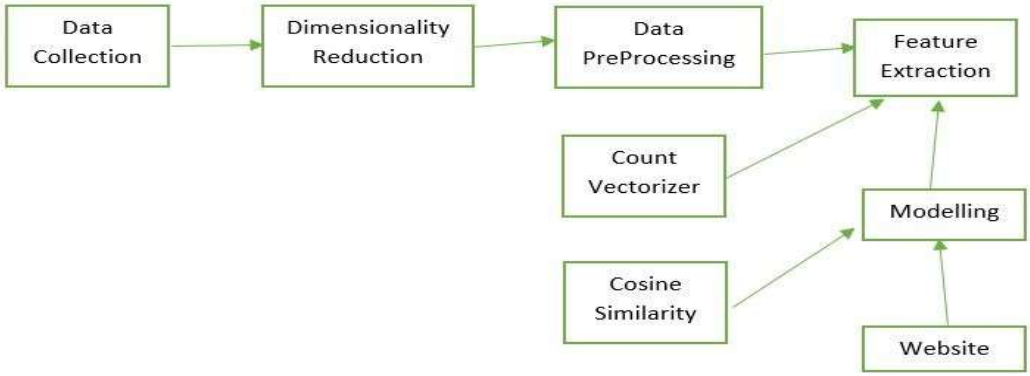
4.2 Software

a. Docker for chatbot application
b. Visual Studio
c. Windows 10/11
d. Python
e. Jupyter Notebook and Streamlit framework

Chapter 5

Proposed System

A. Suggestions based on Machine Learning approach. (Content- Based Recommendation technique)



Block diagram

Steps involved:

1. Data Collection
2. Dimensionality Reduction(Dropping Features)
3. Data Preprocessing
    a. Merging DataSets
    b. Removing Null values\tuples
    c. Conversion of features
    d. Removing Stopwords
    e. Lemmatization
4. Feature Extraction
    a. Count_Vectorizer
5. Building Model
    a. Cosine Similarity
6. Website Creation

1. Data Collection:

Data collection is an important step in the process of Recommendation System. Many datasets were available throughout the internet, but selecting the appropriate dataset for training the model is a crucial part. Here the chosen dataset for our Implementation 1 is from Kagale (Which is an openSource). Contains both movies and credits datasets for 5000 movies taken from TMDB website in separate files.

Movies.csv —--------> 5000 x 20
20 Features include- Budget, Genre, Homepage, ID, Keywords, Original Language, Overview, Popularity, Revenue, Production_comapnies, spoken_languages, Status, Tagline, Title, Vote_count,...

Credits.csv —--------> 5000 x 4
4 Features include- Movie_id, Title, Cast, Crew

2. Dropping Features:

Training the model with all the features may lead to the problem called overfitting of the model. In order to avoid that situation, we need to drop some features that are not necessarily required for the Recommendation purpose to the user. Like original_language, original_title, budget, production_companies, status,... need to be dropped for better suggestions because they have no significance in matching the recommendations.

3. Data PreProcessing:

a. Merging Datasets

For easy understanding of the model, Here we combined the both datasets into a single dataset on the basis of the common column ‘movie_id’. Now our dataset dimension is,

movies.csv —--------> 5000 x 7

b. Removing Null Values

Null Values in the dataset cause trouble in finding the similarities among the movies for the ML model. So before we start exploring into the model we need to remove the tuples/rows that are having Null values.

c. Conversion of Features

Almost all features in the Dataset are in the form of key-value pairs,i,e. Nested-Dictionary Format. So they need to be transformed into a simple Linear format. Hence using the 'literal_eval' function we can convert them into a linear data Structure format.

d. Lemmatization

Words that have different verb forms all have the same meaning. In order to have grammatical structure they are used. But in Learning for the model they are not required. So all these words need to be converted into a single original root form.

Example: do < —- {do, does, did, done}

This can be done using the PorterStemmer function from nltk Library. This process is called Stemming.

e. Removing Stop words

Words that are commonly repeated like a, an, the, and, of, for, … have no use in Recommendation. Hence needed to be removed.

4. Feature Extraction:

Extracting the features(Most commonly/frequently occurring words and important words) from the attributes is termed to be called 'tags'. And Convert each tuple into multidimensional space in terms of those Weighted words and finally change them into the form of Vectors, Using the CountVectorizer function. Here we extracted 5000 important words aka features.

5. Modeling:

Now we build our model with the help of the ‘Cosine Similarity function’, Which is used to find the Similarity between two movies based on the above extracted features.

$$Cosine\ similarity(u,p) = \frac{\sum_i u_i p_i}{\sqrt{\sum_i u_i^2}\sqrt{\sum_i p_i^2}}$$

After that Now we transform the dataset into a 5000 feature space matrix and will use those scores for similarity comparisons.

B. Suggestions based on User’s mood detection Using Face Detection technique.

Steps involved:

1. Selecting Languages
2. Choosing the Singers
3. Emotion Detection
4. Recommendation of songs according to the emotion

Selecting Language : The language of the user choice to be selected first. This is the first step of the recommendation process.

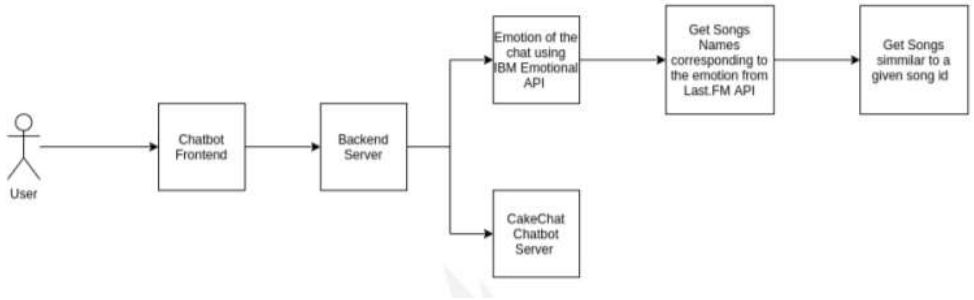
Choosing the Singers : Based on the language selected by the user, singers choice can be done.

Emotion Detection : Firstly, the face of the user will be detected. Emotion is detected using the h5 model. Different emotions like sad, happy, neutral, anger, fear, surprise will be displayed based on the user’s expression.

Recommendation of song : Finally, song will be recommended based on the user's emotion.

We use the h5 model to detect the emotion of the user. H5 is a file format to store structured data. It is elaborated as Hierarchical Data Format. This model can be saved by using model.save()

C.Suggestions based on User's mood detection using the Conversation of the user with the Chatbot.



Process flow diagram

- Chatbot is built on keras and tensorflow.

Model:

- Hierarchical Recurrent Encoder-Decoder (HRED) architecture for handling deep dialog context.
- Multilayer RNN with GRU cells.

Word embedding layer:

- Initialized using w2v model trained on corpus.
- Embedding layer is either fixed or fine-tuned along with other weights of the network.

- Reranking of the generated candidates is performed according to the log-likelihood.
- Lexical similarity between samples of the model and some fixed dataset. Lexical similarity is a cosine distance between the TF-IDF vector of responses generated by the model and tokens in the dataset.

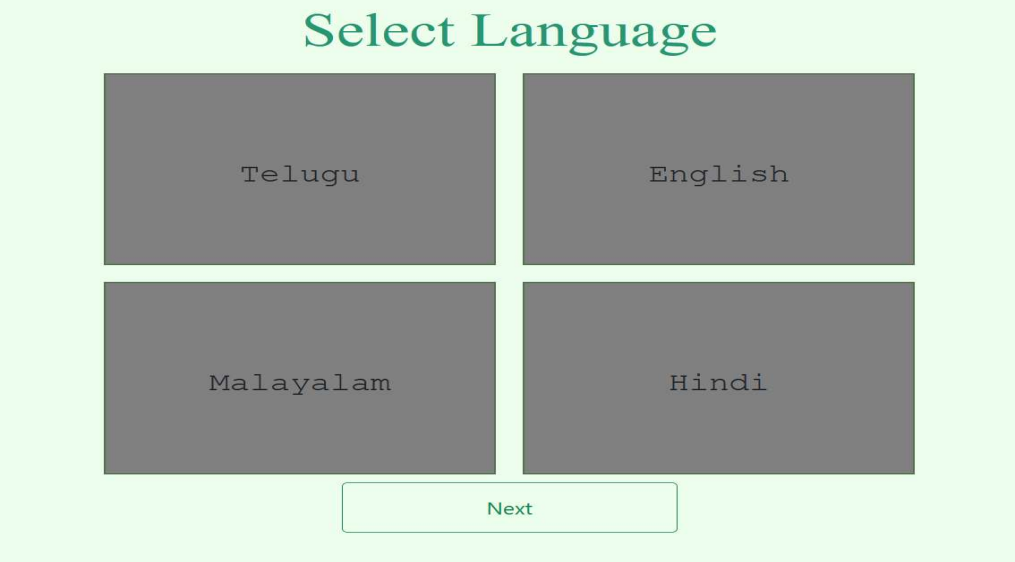
The model was trained on a preprocessed Twitter corpus with ~50 million dialogs (11Gb of text data). To clean up the corpus, we removed

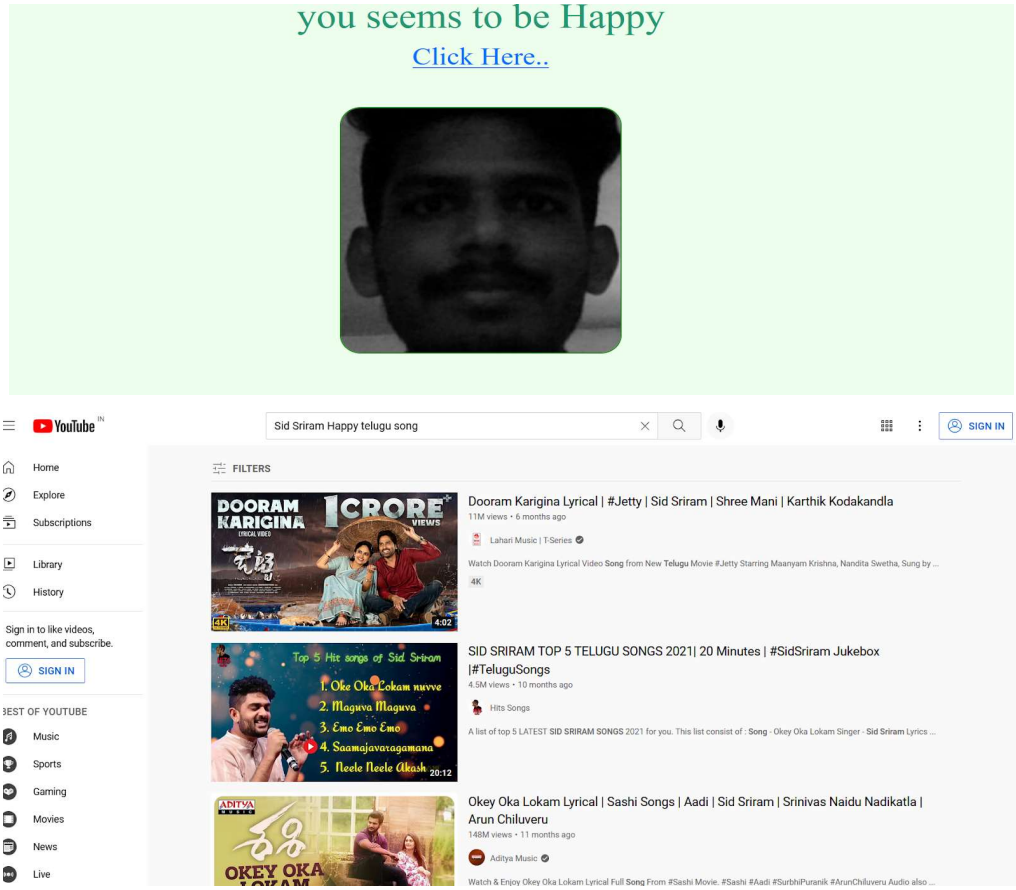
- URLs, retweets and citations;
- mentions and hashtags that are not preceded by regular words or punctuation marks;
- messages that contain more than 30 tokens.

Chapter 6

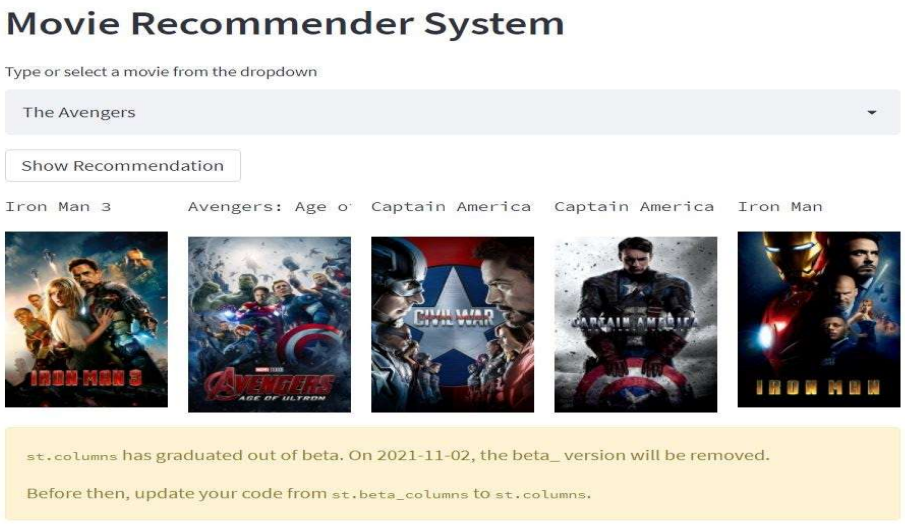
Result and Analysis

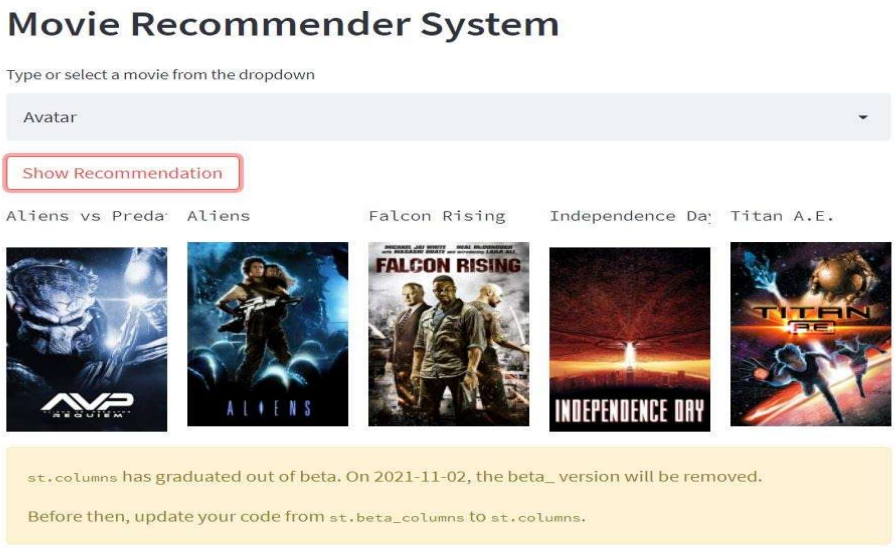
1. **Suggestions based on User's mood detection Using Face Detection technique.**

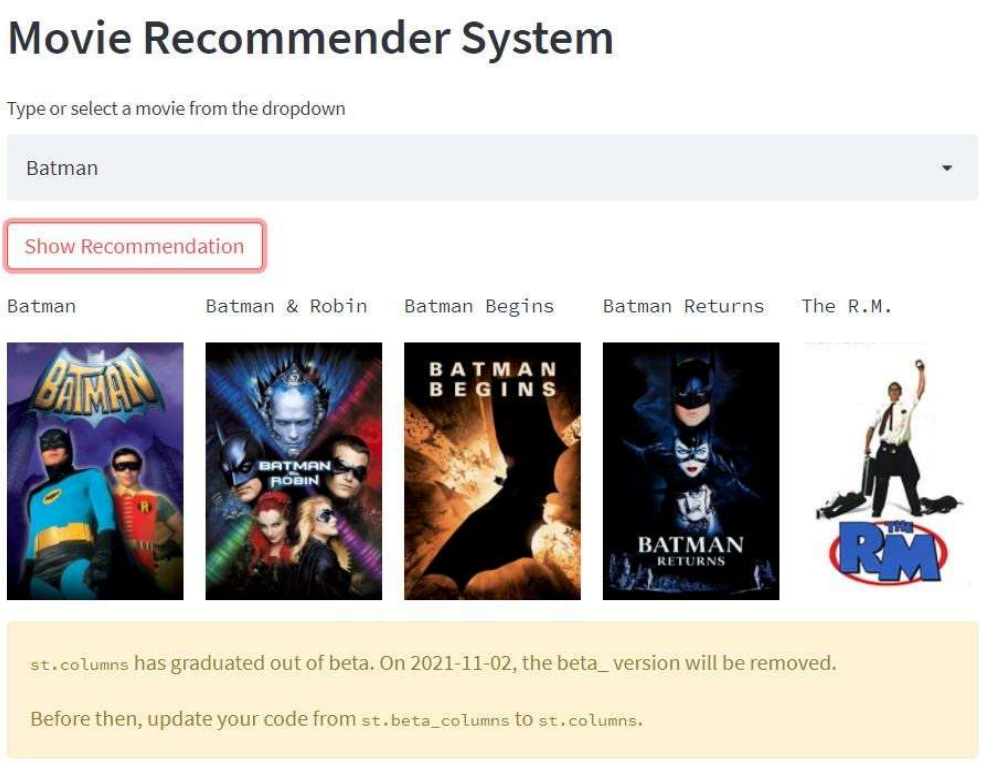




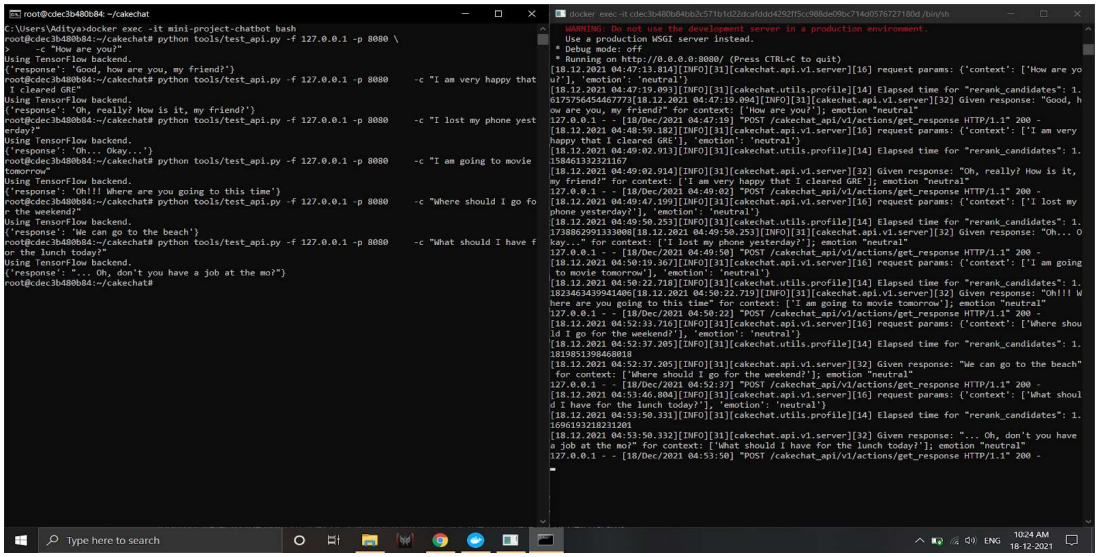
2. Suggestions based on Machine Learning approach. (Content- Based Recommendation technique)




Movie Recommender System
Type or select a movie from the dropdown
Avatar
Show Recommendation
Aliens vs Preda
Aliens
Falcon Rising
Independence Da
Titan A.E.
st.columns has graduated out of beta. On 2021-11-02, the beta_ version will be removed.
Before then, update your code from st.beta_columns to st.columns.


Movie Recommender System
Type or select a movie from the dropdown
Batman
Show Recommendation
Batman
Batman & Robin
Batman Begins
Batman Returns
The R.M.
st.columns has graduated out of beta. On 2021-11-02, the beta_ version will be removed.
Before then, update your code from st.beta_columns to st.columns.

3. Suggestions based on User's mood detection using the Conversation of the user with the Chatbot.



Chapter 7

Conclusion

By searching the vast media collection, which contains many songs and movies of many genres, the recommender system we built will reduce human effort, which is time demanding. The user will be given recommendations based on their mood. As a result, this system will deliver greater user satisfaction in less time and effort, as recommendations for music and movies will be made automatically based on their mood.

Ratings could be viewed as a way to improve the system's performance. This system can be expanded to suggest books, places, restaurants, and other items based on the user's requirements. The emotion category could be expanded to include more conflicting emotions like hostility, anxiety, jealousy, and enthusiasm, among others.

Chapter 8

Future Work

Initially, we thought of implementing a Complete end-to-end Chatbot Recommender System that could recommend the songs and music to the user based on their facial expression recognition, user preference and emotional tone analysis from the chat using IBM cloud (Any cloud).

As of now, we could only complete the Suggestions based on Facial expression recognition, Suggestions based on User's preference (Content-Based Recommendation Filtering) and a Chatot Implementation using Tensorflow.

As future work, We would like to continue to develop this project and build an Complete end-end-end ChatBot Recommender System which includes emotional tone analysis and detect emotion of the user. And thereby fetches the Songs from Last.fm Api and displays them to the user. Also, willing to include other categories of Recommendations into it.

References

[1] Mohammad Nuruzzaman Phd, Omar Khadeer Hussain: Survey on Chatbot Implementation in Customer Service Industry through Deep Neural Networks. https://www.researchgate.net/publication/330027049

[2] Samuvel, D. J., Perumal, B., & Elangovan, M : Music recommendation system based on facial emotion recognition March 2020.

https://doi.org/10.17993/3ctecno.2020.specialissue4.261-271

[3] Ashu Abdul, Jenhui Chen, Hua-Yuan Liao and Shun-Hao Chang : An Emotion-Aware Personalized Music Recommendation System Using a Convolutional Neural Networks Approach.
https://doi.org/10.3390/app8071103

[4] Shivani Shivanand, K S Pavan Kamini, Monika Bai M N, Ranjana Ramesh, Sumathi H R : Chatbot with Music and Movie Recommendation based on Mood.
https://www.ijert.org/chatbot-with-music-and-movie-recommendation-based-on-mood

Appendix A

Source code

Chatbot Code:
https://drive.google.com/file/d/1bmWEGrgj9cNfmasVJC7KmgBxFMDjGoLj/view?usp=sharing

Face Detection Code:
https://drive.google.com/drive/folders/1Cwf2lQKBxdCw9wz-8E3GEVt8p8dv3El2?usp=sharing

Movie Recommendation Code:
https://drive.google.com/file/d/1B4ReBTQFSSeNk0vhwBKq54MhB6PdDjJv/view?usp=sharing