Recommender System(Movie/Web Show Recommendation)

A Project Work

Submitted in the partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE (BIG DATA ANALYTICS)

Submitted by:

SAKSHAM BHAMBOTA

ANJALI

TANISHKA TIWARI

University Roll Number

19BCS3820

19BCS3796

19BCS3819

Under the Supervision of:

MS. SEEMA

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
APEX INSTITUE OF TECHNOLOGY
CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,
PUNJAB

APRIL,2021

DECLARATION

We, **Anjali**, **Tanishka, Saksham** students of **'Bachelor of Engineering in Computer Science and Engineering'**, **session:2019-2023**, Department of Computer Science and Engineering, Apex Institute of Technology, Chandigarh University, Punjab, hereby declare that the work presented in this Project Work entitled '**Movie Recommendation System'** is the outcome of our own bona fide work and is correct to the best of our knowledge and this work has been undertaken taking care of Engineering Ethics. It contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

(Anjali, Tanishka, Saksham)

Candidate UID: 19BCS3796, 19BCS3819, 19BCS3820

Date: April 25,2021

Place:Chandigarh

ACKNOWLEDGEMENT

It gives me immense pleasure to express my deepest sense of gratitude and sincere thanks to my respected guide MS.SEEMA (Assistant Professor), CSE, CHANDIGARH UNIVERSITY.

I also wish to express my indebtedness to my parents as well as my family member whose blessings and support always helped me to face the challenges ahead.

ABSTRACT

A recommendation engine filters the data using different algorithms and recommends the most relevant items to users. It first captures the past behavior of a customer and based on that, recommends products which the users might be likely to buy. If a completely new user visits an e-commerce site, that site will not have any past history of that user. So how does the site go about recommending products to the user in such a scenario? One possible solution could be to recommend the best selling products, i.e. the products which are high in demand. Another possible solution could be to recommend the products which would bring the maximum profit to the business. Three main approaches are used for our recommender systems. One is Demographic Filtering i.e They offer generalized recommendations to every user, based on movie popularity and/or genre. The System recommends the same movies to users with similar demographic features. Since each user is different , this approach is considered to be too simple.

The basic idea behind this system is that movies that are more popular and critically acclaimed will have a higher probability of being liked by the average audience. Second is content-based filtering, where we try to profile the users interests using information collected, and recommend items based on that profile. The other is collaborative filtering, where we try to group similar users together and use information about the group to make recommendations to the user.

Table of Contents

INTODUCTION:-

A recommendation system is a type of information filtering system which attempts to predict the preferences of a user, and make suggests based on these preferences. There are a wide variety of applications for recommendation systems. These have become increasingly popular over the last few years and are now utilized in most online platforms that we use. The content of such platforms varies from movies, music, books and videos, to friends and stories on social media platforms, to products on e-commerce websites, to people on professional and dating websites, to search results returned on Google. Often, these systems are able to collect information about a users choices, and can use this information to improve their suggestions in the future. For example, Facebook can monitor your interaction with various stories on your feed in order to learn what types of stories appeal to you. Sometimes, the recommender systems can make improvements based on the activities of a large number of people. For example, if Amazon observes that a large number of customers who buy the latest Apple Macbook also buy a USB-C-to USB Adapter, they can recommend the Adapter to a new user who has just added a Macbook to his cart. Due to the advances in recommender systems, users constantly expect good recommendations. They have a low threshold for services that are not able to make appropriate suggestions. If a music streaming app is not able to predict and play music that the user likes, then the user will simply stop using it. This has led to a high emphasis by tech companies on improving their recommendation systems. However, the problem is more complex than it seems. Every user has different preferences and likes. In addition, even the taste of a single user can vary depending on a large number of factors, such as mood, season, or type of activity the user is doing. For example, the type of music one would like to hear while exercising differs greatly from the type of music he'd listen to when cooking dinner. Another issue that recommendation systems have to solve is the exploration vs exploitation problem. They must explore new domains to discover more about the user, while still making the most of what is already known about of the user. Three main approaches are used for our recommender systems. One is Demographic Filtering i.e They offer generalized recommendations to every user, based on movie popularity and/or genre. The System recommends the same movies to users with similar demographic features. Since each user is different , this approach is considered to be too simple. The basic idea behind this system is that movies that are more popular and critically acclaimed will have a higher probability of being liked by the average audience. Second is content-based filtering, where we try to profile the users interests using information collected, and recommend items based on that profile. The other is collaborative filtering, where we try to group similar users together and use information about the group to make recommendations to the user.

TECHNOLOGY USED

1. Python

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library.

2. PyCharm

PyCharm is an integrated development environment (IDE) used in computer programming, specifically for the Python language. It is developed by the Czech company JetBrains.[5] It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems (VCSes), and supports web development with Django as well as data science with Anaconda

LITEARATURE SURVEY

Content-based Filtering Systems (CBF based systems): In content-based filtering, items are recommended based on comparisons between item profile and user profile. A user profile is content that is found to be relevant to the user in form of keywords(or features). A user profile might be seen as a set of assigned keywords (terms, features) collected by algorithm from items found relevant (or interesting) by the user. A set of keywords (or features) of an item is the Item profile. For example, consider a scenario in which a person goes to buy his favorite cake ‘X’ to a pastry. Unfortunately, cake ‘X’ has been sold out and as a result of this the shopkeeper recommends the person to buy cake ‘Y’ which is made up of ingredients similar to cake ‘X’. This is an instance of content-based filtering. Advantages of content-based filtering are: • They capable of recommending unrated items • We can easily explain the working of recommender system by listing the Content features of an item. • Content-based recommender systems use need only the rating of the concerned user,and not any other user of the system. Disadvantages of content-based filtering are: • It does not work for a new user who has not rated any item yet as enough ratings are required contentbased recommender evaluates the user preferences and provides accurate recommendations. • No recommendation of serendipitous items. • Limited Content Analysis- The recommender does not work if the system fails to distinguish the items hat a user likes from the items that he does not like.

PROBLEM FORMULATION

- Why Recommendation systems?
- They help the user find items of their interest
- Helps the item provider to deliver their items to the right user
- To identify the most relevant products for each user
- Showcase personalized content to each user
- Suggest top offers and discounts to the right user
- Websites can improve user-engagement
- It increases revenues for business through increased consumption

OBJECTIVES

- This makes recommender systems essentially a central part of websites and e-commerce applications.
- The purpose of a recommendation system basically is to search for content that would be interesting to an individual.
- Recommendation systems are Artificial Intelligence based algorithms that skim through all possible options and create a customized list of items that are interesting and relevant to an individual.
- These results are based on their profile, search/browsing history, what other people with similar traits/demographics are watching, and how likely are you to watch those movies.

METHODOLOGY

- Content Based Recommender System recommends movies similar to the movie user likes and analyses the sentiments on the reviews given by the user for that movie.
- The details of the movies(title, genre, runtime, rating, poster, etc) are fetched using an API by TMDB, https://www.themoviedb.org/documentation/api, and using the IMDB id of the movie in the API, I did web scraping to get the reviews given by the user in the IMDB site using beautifulsoup4 and performed sentiment analysis on those reviews.

  Similarity Score :

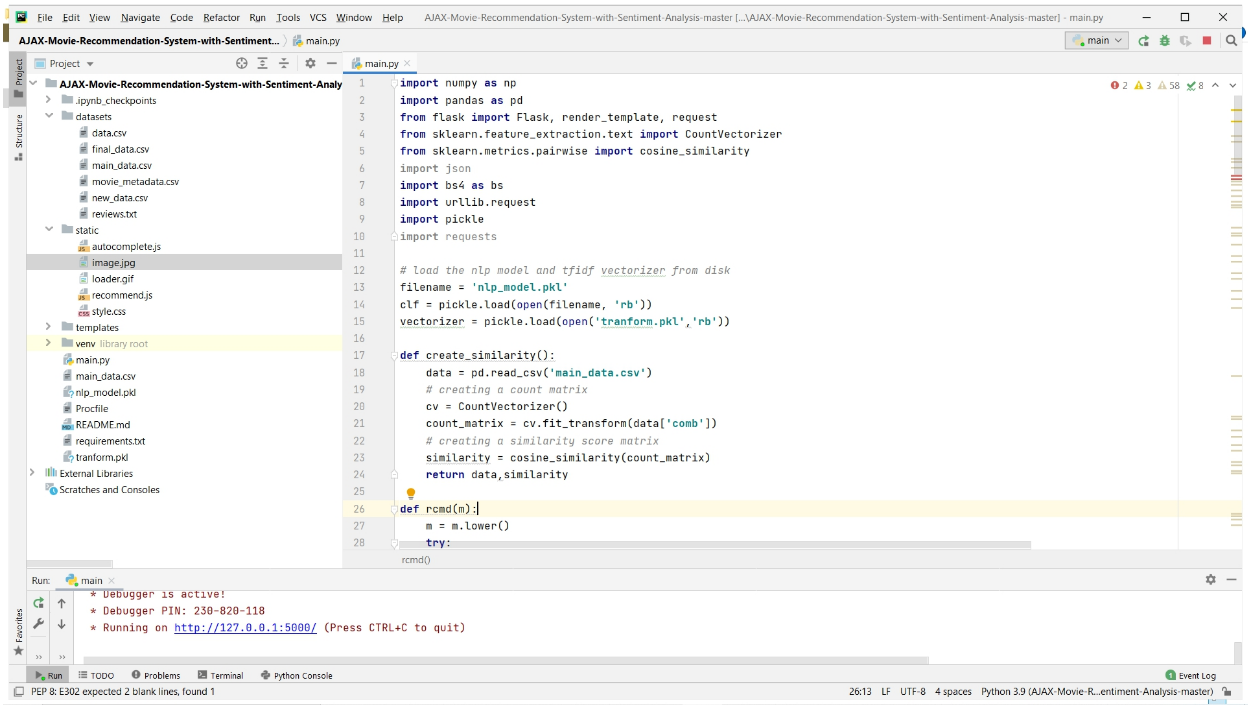
  - How does it decide which item is most similar to the item user likes? Here we use the similarity scores.

  - It is a numerical value ranges between zero to one which helps to determine how much two items are similar to each other on a scale of zero to one. This similarity score is obtained measuring the similarity between the text details of both of the items. So, similarity score is the measure of similarity between given text details of two items. This can be done by cosine-similarity.

CONCLUSION AND DISCUSSION

- we would conclude by stating that we hope you have got a basic idea of how this recommendation system works. We have discussed mainly one recommendation systems that was content based whereas there are several other systems that are used for recommendation purposes like Collaborative filtering, Hybrid models, also neural networks based approaches. Recommendation systems are very effective systems that are tremendous. People are trying to implement recommendations in various different applications.

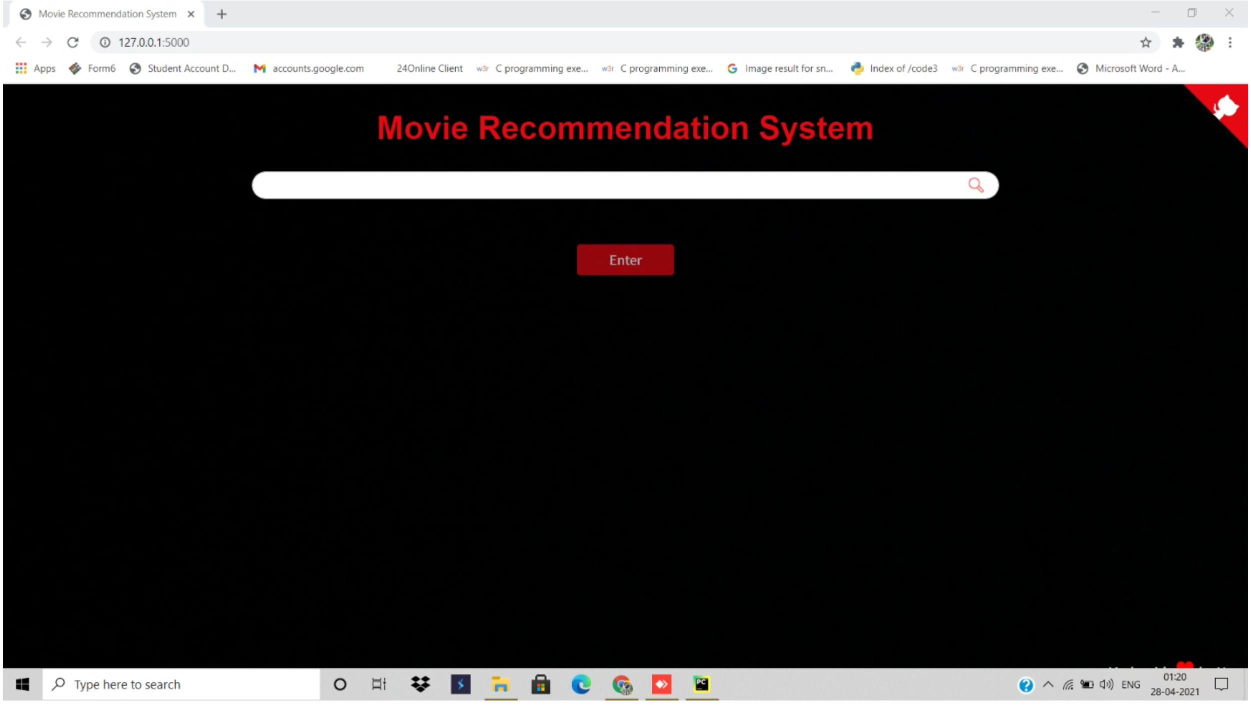
OUTPUTS

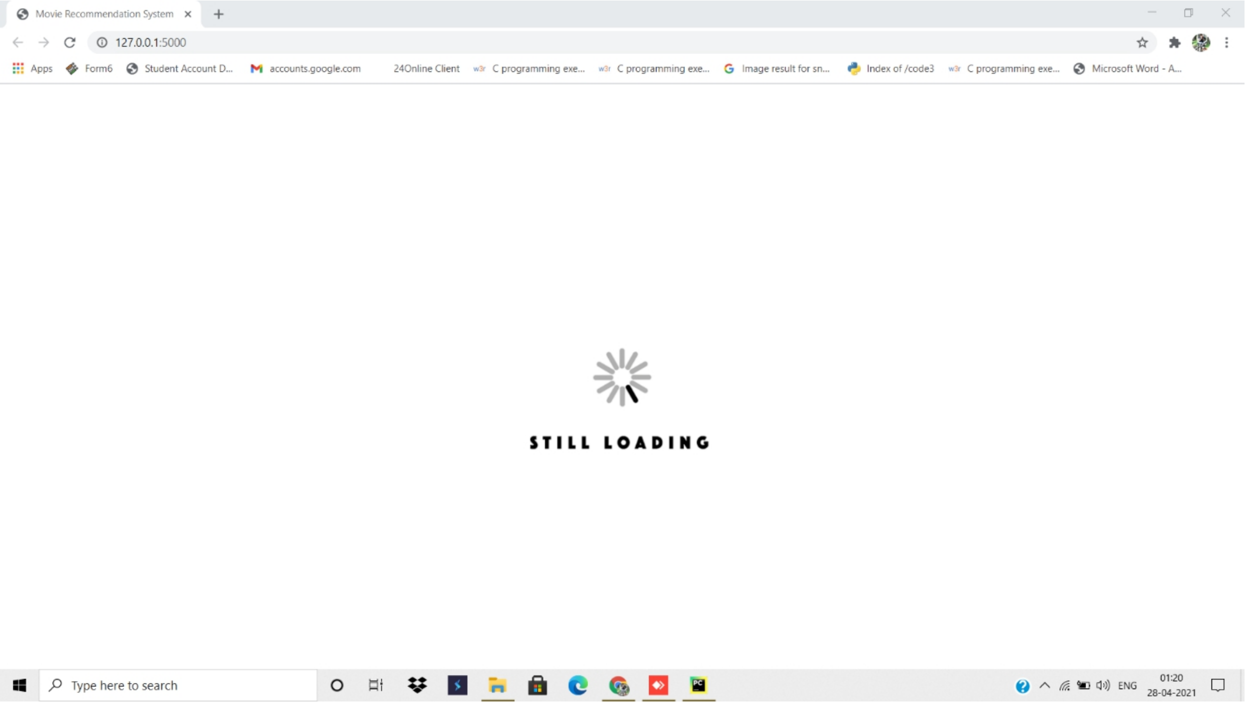

File Edit View Navigate Code Refactor Run Tools VCS Window Help
AJAX-Movie-Recommendation-System-with-Sentiment-Analysis-master [...\AJAX-Movie-Recommendation-System-with-Sentiment-Analysis-master] - main.py
AJAX-Movie-Recommendation-System-with-Sentiment...
main.py
Project
AJAX-Movie-Recommendation-System-with-Sentiment-Analy
.ipynb_checkpoints
datasets
data.csv
final_data.csv
main_data.csv
movie_metadata.csv
new_data.csv
reviews.txt
static
autocomplete.js
image.jpg
loader.gif
recommend.js
style.css
templates
venv library root
main.py
main_data.csv
nlp_model.pkl
Procfile
README.md
requirements.txt
tranform.pkl
External Libraries
Scratches and Consoles
rcmd()
Run:
main
Run
TODO
Problems
Terminal
Python Console
Event Log
26:13 LF UTF-8 4 spaces Python 3.9 (AJAX-Movie-R...entiment-Analysis-master)

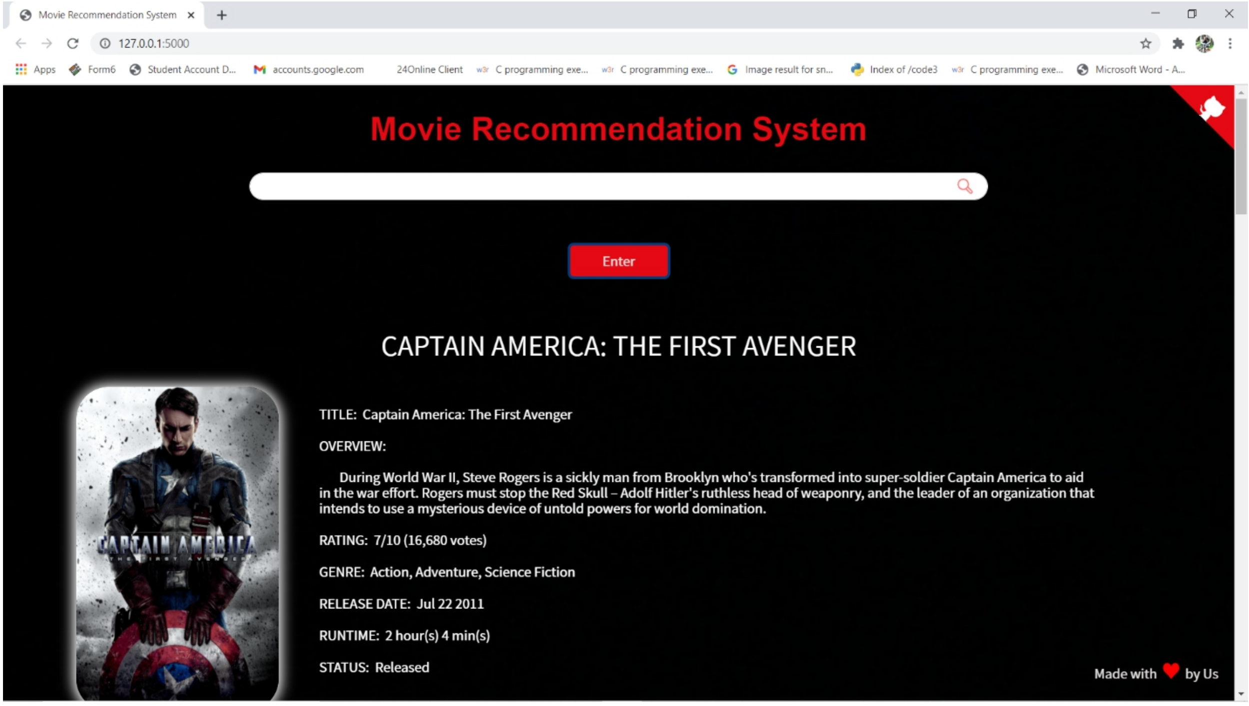
```python
import numpy as np
import pandas as pd
from flask import Flask, render_template, request
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics.pairwise import cosine_similarity
import json
import bs4 as bs
import urllib.request
import pickle
import requests

# load the nlp model and tfidf vectorizer from disk
filename = 'nlp_model.pkl'
clf = pickle.load(open(filename, 'rb'))
vectorizer = pickle.load(open('tranform.pkl','rb'))

def create_similarity():
    data = pd.read_csv('main_data.csv')
    # creating a count matrix
    cv = CountVectorizer()
    count_matrix = cv.fit_transform(data['comb'])
    # creating a similarity score matrix
    similarity = cosine_similarity(count_matrix)
    return data,similarity

def rcmd(m):
    m = m.lower()
    try:
```

```
 * Debugger is active!
 * Debugger PIN: 230-820-118
 * Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```

PEP 8: E302 expected 2 blank lines, found 1


Movie Recommendation System
127.0.0.1:5000
Movie Recommendation System
Enter
Type here to search
01:20
28-04-2021


Movie Recommendation System
127.0.0.1:5000
STILL LOADING
Type here to search
01:20
28-04-2021



Movie Recommendation System

CAPTAIN AMERICA: THE FIRST AVENGER

TITLE: Captain America: The First Avenger

OVERVIEW:

During World War II, Steve Rogers is a sickly man from Brooklyn who's transformed into super-soldier Captain America to aid in the war effort. Rogers must stop the Red Skull – Adolf Hitler's ruthless head of weaponry, and the leader of an organization that intends to use a mysterious device of untold powers for world domination.

RATING: 7/10 (16,680 votes)

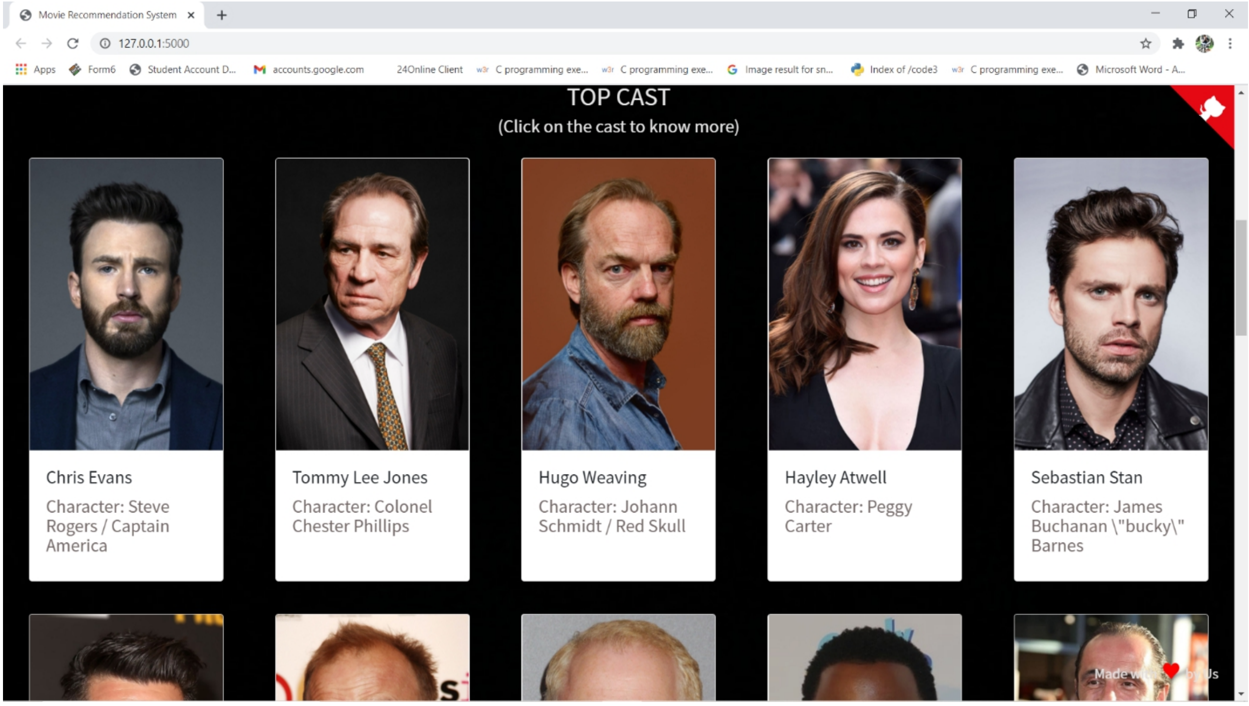
GENRE: Action, Adventure, Science Fiction

RELEASE DATE: Jul 22 2011

RUNTIME: 2 hour(s) 4 min(s)

STATUS: Released

Made with ♥ by Us



TOP CAST

(Click on the cast to know more)

Chris Evans

Character: Steve Rogers / Captain America

Tommy Lee Jones

Character: Colonel Chester Phillips

Hugo Weaving

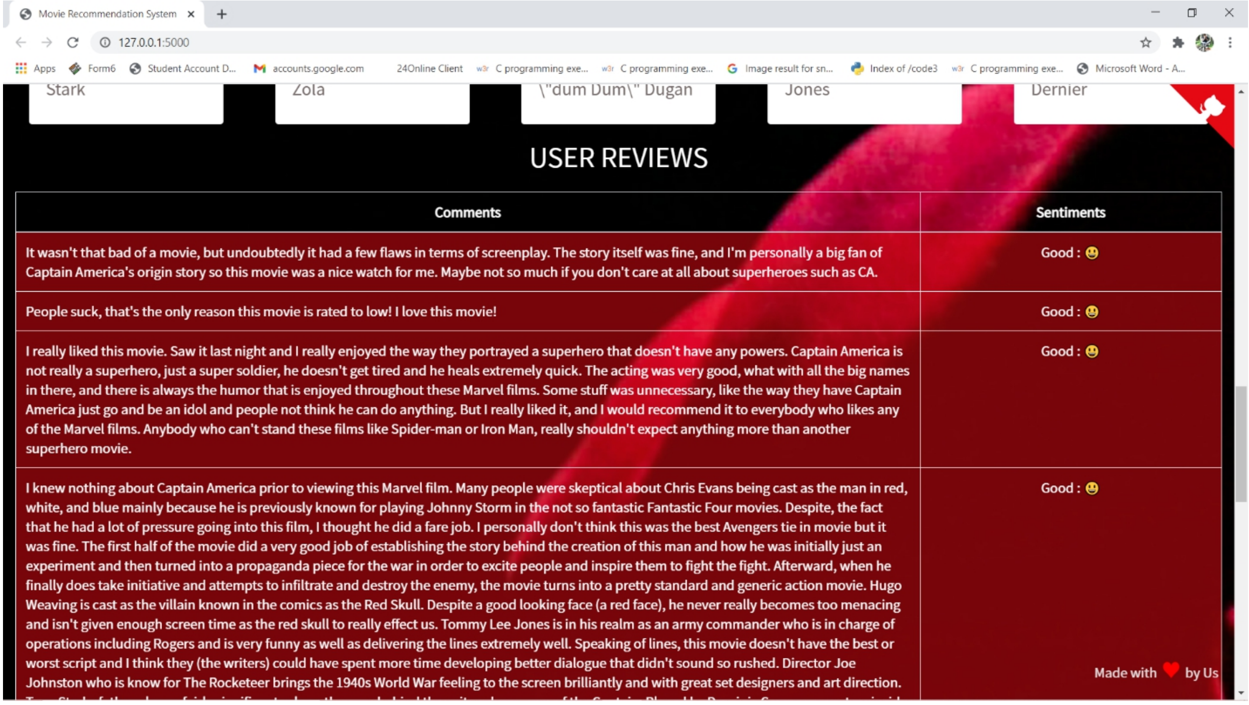
Character: Johann Schmidt / Red Skull

Hayley Atwell

Character: Peggy Carter

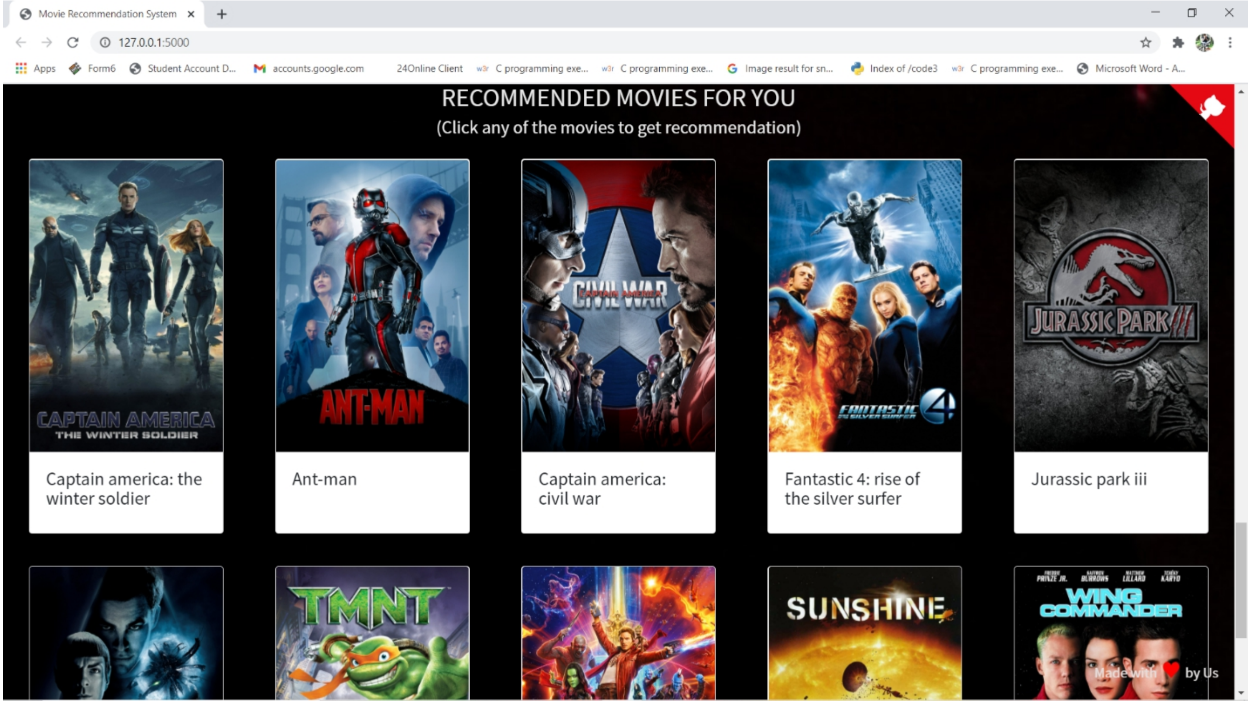
Sebastian Stan

Character: James Buchanan \"bucky\" Barnes



USER REVIEWS

Comments	Sentiments
It wasn't that bad of a movie, but undoubtedly it had a few flaws in terms of screenplay. The story itself was fine, and I'm personally a big fan of Captain America's origin story so this movie was a nice watch for me. Maybe not so much if you don't care at all about superheroes such as CA.	Good : 😃
People suck, that's the only reason this movie is rated to low! I love this movie!	Good : 😃
I really liked this movie. Saw it last night and I really enjoyed the way they portrayed a superhero that doesn't have any powers. Captain America is not really a superhero, just a super soldier, he doesn't get tired and he heals extremely quick. The acting was very good, what with all the big names in there, and there is always the humor that is enjoyed throughout these Marvel films. Some stuff was unnecessary, like the way they have Captain America just go and be an idol and people not think he can do anything. But I really liked it, and I would recommend it to everybody who likes any of the Marvel films. Anybody who can't stand these films like Spider-man or Iron Man, really shouldn't expect anything more than another superhero movie.	Good : 😃
I knew nothing about Captain America prior to viewing this Marvel film. Many people were skeptical about Chris Evans being cast as the man in red, white, and blue mainly because he is previously known for playing Johnny Storm in the not so fantastic Fantastic Four movies. Despite, the fact that he had a lot of pressure going into this film, I thought he did a fare job. I personally don't think this was the best Avengers tie in movie but it was fine. The first half of the movie did a very good job of establishing the story behind the creation of this man and how he was initially just an experiment and then turned into a propaganda piece for the war in order to excite people and inspire them to fight the fight. Afterward, when he finally does take initiative and attempts to infiltrate and destroy the enemy, the movie turns into a pretty standard and generic action movie. Hugo Weaving is cast as the villain known in the comics as the Red Skull. Despite a good looking face (a red face), he never really becomes too menacing and isn't given enough screen time as the red skull to really effect us. Tommy Lee Jones is in his realm as an army commander who is in charge of operations including Rogers and is very funny as well as delivering the lines extremely well. Speaking of lines, this movie doesn't have the best or worst script and I think they (the writers) could have spent more time developing better dialogue that didn't sound so rushed. Director Joe Johnston who is know for The Rocketeer brings the 1940s World War feeling to the screen brilliantly and with great set designers and art direction.	Good : 😃



RECOMMENDED MOVIES FOR YOU

(Click any of the movies to get recommendation)

Captain america: the winter soldier

Ant-man

Captain america: civil war

Fantastic 4: rise of the silver surfer

Jurassic park iii

1 REFERENCES

[1] D. Rattan, R. Bhatia, and M. Singh, “Software clone detection: A systematic review,” *Information and Software Technology*, vol. 55, no. 7, pp. 1165–1199, Jul. 2013.

[2] J. F. Islam, M. Mondal, and C. K. Roy, “Bug Replication in Code Clones: An Empirical Study,” in *2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER)*, 2016, pp. 68–78.

[3] M. R. Islam and M. F. Zibran, “A Comparative Study on Vulnerabilities in Categories of Clones and Non-cloned Code,” in *2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER)*, 2016, pp. 8–14.

[4] M. R. Islam, M. F. Zibran, and A. Nagpal, “Security Vulnerabilities in Categories of Clones and Non-Cloned Code: An Empirical Study,” in *2017 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 2017, pp. 20–29.

[5] C. K. Roy, M. F. Zibran, and R. Koschke, “The vision of software clone management: Past, present, and future (Keynote paper),” in *2014 Software Evolution Week - IEEE Conference on Software Maintenance, Reengineering, and Reverse Engineering (CSMR-WCRE)*, 2014, pp. 18–33.

[6] J. Krinke, “A Study of Consistent and Inconsistent Changes to Code Clones,” in *14th Working Conference on Reverse Engineering (WCRE 2007)*, 2007, pp. 170–178.

[7] D. Chatterji, J. C. Carver, N. A. Kraft, and J. Harder, “Effects of cloned code on software maintainability: A replicated developer study,” in *2013 20th Working Conference on Reverse Engineering (WCRE)*, 2013, pp. 112–121.

[8] D. Rattan, R. Bhatia, and M. Singh, “An Empirical Study of Clone Detection in MATLAB/ Simulink Models,” *International Journal of Information and Communication Technology*.

[9] D. Rattan, R. Bhatia, and M. Singh, “Detecting High Level Similarities in Source Code and Beyond,” *International Journal of Energy, Information and Communications*, vol. 6, no. 2, pp. 1–16, 2015.

[10] D. Rattan, R. Bhatia, and M. Singh, “Detection and Analysis of Clones in UML Class Models,” *International Journal of Software Engineering, IJSE*, vol. 8, no. 2, pp. 66–99, 2015.

[11] D. Rattan, R. Bhatia, and M. Singh, “Model clone detection based on tree comparison,” in *2012 Annual IEEE India Conference (INDICON)*, 2012, pp. 1041–1046.

[12] C. K. Roy and J. R. Cordy, “NICAD: Accurate Detection of Near-Miss Intentional Clones Using Flexible Pretty-Printing and Code Normalization,” in *2008 16th IEEE International Conference on Program Comprehension*, 2008, pp. 172–181.

[13] M. Mondal, C. K. Roy, and K. A. Schneider, “SPCP-Miner: A tool for mining code clones that are important for refactoring or tracking,” in *2015 IEEE 22nd International Conference on Software Analysis, Evolution, and Reengineering (SANER)*, 2015, pp. 484–488.