



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 10    Issue: VI    Month of publication: June 2022**

**DOI: <https://doi.org/10.22214/ijraset.2022.44154>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# Movie Recommendation System Using Content-Based Filtering with Heroku Deployment

L. Malathi<sup>1</sup>, Md. Habeeb Ur Rahman<sup>2</sup>, Dr. Sundaragiri Dheeraj<sup>3</sup>, Mrs. B Vasundara Devi<sup>4</sup>

<sup>1</sup>B. Tech 4<sup>th</sup> year, CSE Department, Sreenidhi Institute of Science and Technology Yamnampet, Hyderabad

<sup>2</sup>B. Tech 4<sup>th</sup> year, ECM Department, Sreenidhi Institute of Science and Technology Yamnampet, Hyderabad

<sup>3,4</sup> Assistant Professor of CSE Department, Sreenidhi Institute of Science and Technology Yamnampet, Hyderabad

**Abstract:** *Now days peoples watching movies to make their busy life calm , watching movie is not that easy as it takes a lot of time to select a movie due to large data set of movies exist in the world and it is very difficult and time consuming process to people to select a movie due to large data set but we can achieve this with the help of movie recommendation system using content based filtering ,where this recommendation system recommend a movie to the user based on the content he/she is watching. In this paper I am building a recommendation system using content- based filtering with Heroku deployment. Generally, recommendation systems are build using content -based filtering approach or collaborative filtering approach here in this paper I developed recommendation system using content-based filtering approach with Heroku Deployment. Heroku Deployment enables easy deployment of application.*

**Keyword:** *Heroku Deployment, Content-based Filtering, Stream lit, Sklearn, Pickle, Pandas, NumPy, Cosine distance, Bag of Words, AST, Count Vectorizer, Cosine Similarity.*

## I. INTRODUCTION

Now days Recommendation system is playing a vital role in YouTube, Netflix, amazon prime, Instagram and Facebook etc., where system recommend an item to the user based on the ratings or preferences given to an item, the system predicts the ratings or preferences given to an item and recommend an item to user. This recommendation system helps you by recommending the things that you like when you search for something on the website. Assume that you have searched for some mobile phone in Amazon or Flipkart today, the next time when you open your amazon/flip kart app the system is going to show you some mobile phones that where related to your previous search, similarly it happens in amazon prime and Netflix where the system recommends a movie to the user based on his previous history and ratings. Anything you do in these websites there is a system which is watching your history and that will help you in suggesting the movies that you like to watch, and this recommendation is basically done by filtering techniques as we know there are three different types of filtering techniques such as content-based filtering, collaborative filtering and hybrid filtering. In this project I have used content-based filtering technique and I have taken the dataset from TMDB movie dataset this is the famous dataset and I used streamlit for website development along with I also used some python libraries such as pickle, sklearn, NumPy, pandas, requests and AST. The main idea of this project is to build a model which helps the user in easy recommendation of movies by the system based on their behavior.

## II. LITERATURE REVIEW

Sang-Min Choi et al. [1] discussed the shortcomings of the collaborative filtering approach, such as the sparsity problem and the cold-start problem. To avoid this problem, the authors proposed a solution based on category information. The authors proposed a film recommendation system based on genre correlations. According to the authors, the category information is present in the newly created content. Thus, even if the new content does not have enough ratings or views, it can still appear in the recommendations list thanks to category or genre information. The proposed solution is unbiased in terms of highly rated most watched content and new content that is not widely consumed. As a result, even a new film can be recommended by [2] suggested a hybrid approach for movie recommendation. According to the authors, both content-based filtering and collaborative filtering have drawbacks but can be useful in certain situations. As a result, the authors developed a hybrid approach that considers both content-based and contextual factors. Filtering and collaborative filtering the solution is implemented in the film 'More. System of recommendations Pearson correlation coefficient has not been used for pure collaborative filtering. Used. A new formula was used instead. However, this formula has a 'divide by zero' error. This mistake occurs.

When users have given the same rating to the films as a result, the authors ignored such users. In the case of p [4] proposed a 'Weighted KM-Slope-VU' collaborative filtering approach for movie recommendation. To divide the users into groups of similar users, the authors used K-means clustering. [5] discussed Simple Recommender System, Content-based Recommender System, and Personalized Recommender System. Finally, a Hybrid Recommender System based on Collaborative Filtering was proposed as a solution. System of Recommendation [6] proposed a solution that makes use of the K-means clustering algorithm. The authors used clusters to separate similar users. Later, for each cluster, the authors created a neural network for recommendation purposes. The proposed system includes steps such as data preprocessing, principal component analysis, clustering, and classification. Data Preprocessing for Neural Networks and Neural Network Construction User evaluation, user preference, and user feedback the consumption ratio has been considered.

Following the clustering phase, in order to predict the authors used a neural network to simulate the ratings that users might give unwatched movies. Finally, predicted high ratings are used to make recommendations. Gaurav Arora et al. [7] proposed a movie recommendation solution based on user similarity. The scope of the research paper is broad. is based on a hybrid approach that employs context-based filtering and collaborative filtering, but neither the parameters used, nor the internal working details have been disclosed. Subramanya swamy et al. [8] proposed a personalized movie recommendation solution that makes use of Technique for collaborative filtering in order to find the most similar user, the Euclidean distance metric was used. The user with the smallest Euclidean distance value is identified. Finally, movie recommendations are based on A specific user has given the highest rating.

The authors have even claimed that the recommendations change over time. That the system performs better as the user's tastes change over time Harper et al. [9] discussed the Movie Lens in detail. [10], recommendation systems can help with the problem of information overload.

The authors discussed data sparsity, cold start problems, scalability, and other issues. The authors conducted a review of nearly 15 research papers on movie recommendation systems. After reviewing all these papers, they discovered that many of the authors used collaborative filtering rather than Filtering based on content Furthermore, the authors discovered that many authors used a hybrid-based approach.

Even though although much research has been conducted on recommendation systems, there is always room for improvement in order to solve problems. The existing disadvantages. Niharika Immanent and others[11] proposed a hybrid recommendation technique that considers both content-based filtering and collaborative filtering approaches in a hierarchical manner to provide users with personalized movie recommendations. The authors have made movie recommendations using a proper sequence of images that describe the movie story plot, which is the most unique aspect of this research work. This contributes to better visuals. The author also described graph-based recommendation systems, content-based approaches, hybrid recommender systems, collaborative filtering systems, gene correlations-based recommender systems, and other approaches. The proposed algorithm is divided into four major phases. Initially, social networking websites such as Facebook are used to learn about the user's interests. Following that, the movie reviews must be analyzed, and recommendations made.

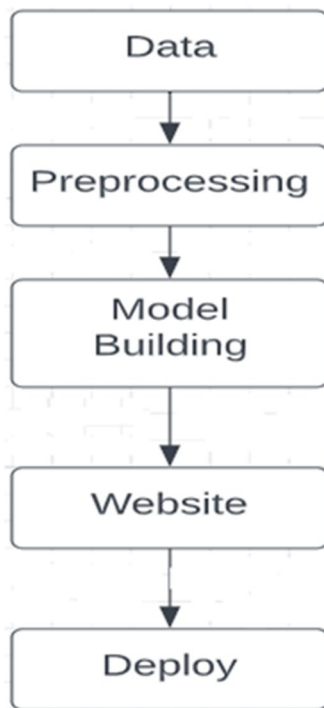
### III. PROPOSED METHODOLOGY

The dataset must be preprocessed, and the relevant features must be combined into a single feature. We'll need to convert the text from that feature into vectors later.

Later, we must determine the similarity of the vectors. Finally, obtain recommendations based on the system architecture described below. Firstly, I collected the dataset from TMDB movie dataset, where I have collected two datasets that is movies and credits dataset and displayed the data which is present in the movies and credits dataset by loading the dataset on to the console that I am working on , In the next step I have done data preprocessing to identify useful data and unnecessary data ,during this I faced problems like missing data values and data imbalance conditions and gave an explanation on how to solve this problem and then I visualized the dataset to know which features helpful for recommendation ,later on I have converted the movies data into vectors using count vectorizer tool which is present in scikit-learn library, this tool will convert the given text to vectors on the basis of count of each word that occur in the entire text without including stop words And cosine-similarity is used to identify the angle differences between two vectors to form similarity matrix to identify the similarity among the movies and a recommend method is built to recommend a movie to the user this is how the model is build using count vectorizer and cosine-similarity

#### IV. ARCHITECTURE DIAGRAM

The below architecture diagram shows the step-by-step process of model building of movie recommendation system.



##### A. Dataset, Exploratory, Data Analysis & Preprocessing

In this study, the 'TMDB 5000 Movie Dataset' is used to make movie recommendations. This dataset can be found at [kaggle.com](https://www.kaggle.com). The dataset is made up of two CSV files: 'tmdb 5000 movies.csv' and 'tmdb 5000 credits.csv'.

The attributes of the 'tmdb 5000 movies.csv' dataset are as follows:

- 1) 'budget': This term refers to the film's budget.
- 2) 'genres': It denotes the film's genres, such as Action, Documentary, and so on.
- 3) A film can have several genres.
- 4) 'homepage': It denotes the movie's homepage. It's essentially a website link.
- 5) 'id': This is the movie ID.
- 6) 'keywords': It denotes the film's keywords. Aside from the title, keywords provide quick information about the film.
- 7) 'Original language': Indicates whether the film was created in English or another language.
- 8) "Original title": Movie title
- 9) 'overview': Description of the movie
- 10) 'Popularity': It indicates popularity
- 11) 'Production companies': It consists of the names of companies which has produced the movie.
- 12) 'Production countries': It contains the names of the countries where the film was made.
- 13) 'Release date': It contains the movie's release date. The format is yyyy-mm-dd, where 'yyyy' represents the year of release, 'mm' represents the month of release, and 'dd' represents the day of release.,
- 14) 'Revenue': refers to the amount of money earned by the film.
- 15) 'runtime': It denotes the length of a movie. The term "runtime" refers to the length of the film.
- 16) 'Spoken languages': It contains the languages spoken in the film.
- 17) 'status': It denotes the movie's current status. A movie, for example, can be released or not released.
- 18) essentially indicates the status of that film
- 19) 'tagline': It is made up of the movie's tagline.
- 20) 'title': It is made up of the

budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	production_companies
37000000	[[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]]	http://www.avatarmovie.com/	19995	[[{"id": 1463, "name": "culture clash"}, {"id": 1463, "name": "culture clash"}]]	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	[[{"name": "Ingenious Film Partners", "id": 289...}]]
100000000	[[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Action"}]]	http://disney.go.com/disneypictures/pirates/	285	[[{"id": 270, "name": "ocean"}, {"id": 726, "name": "nautical"}]]	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	139.082615	[[{"name": "Walt Disney Pictures", "id": 2}, {"name": "Disney Studios", "id": 3}]]
45000000	[[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]]	http://www.sonypictures.com/movies/spectre/	206647	[[{"id": 470, "name": "spy"}, {"id": 818, "name": "action"}]]	en	Spectre	A cryptic message from Bond's past sends him o...	107.376788	[[{"name": "Columbia Pictures", "id": 5}, {"name": "Sony Pictures Entertainment", "id": 6}]]
50000000	[[{"id": 28, "name": "Action"}, {"id": 80, "name": "Drama"}]]	http://www.thedarkknightrises.com/	49026	[[{"id": 849, "name": "dc comics"}, {"id": 853, "name": "action"}]]	en	The Dark Knight Rises	Following the death of District Attorney Harvey...	112.312950	[[{"name": "Legendary Pictures", "id": 923}, {"name": "Warner Bros. Entertainment", "id": 924}]]
	[[{"id": 28, "name": "Action"}]]			[[{"id": 818, "name": "action"}]]		John Carter	John Carter is a		

B. The attributes of the 'tmdb 5000 credits.csv' dataset are as follows

- 1) 'Movie id': This is the movie ID.
- 2) 'title': This indicates the film's title.
- 3) 'cast': It is made up of the movie's cast. The cast refers to the actors and actresses who appear in the film.
- 4) 'Crew': It is made up of people who are involved in the film's production.

```
credits.head()
```

	movie_id	title	cast	crew
0	19995	Avatar	[[{"cast_id": 242, "character": "Jake Sully", "...}]]	[[{"credit_id": "52fe48009251416c750aca23", "de...}]]
1	285	Pirates of the Caribbean: At World's End	[[{"cast_id": 4, "character": "Captain Jack Spa...}]]	[[{"credit_id": "52fe4232c3a36847f800b579", "de...}]]
2	206647	Spectre	[[{"cast_id": 1, "character": "James Bond", "cr...}]]	[[{"credit_id": "54805967c3a36829b5002c41", "de...}]]
3	49026	The Dark Knight Rises	[[{"cast_id": 2, "character": "Bruce Wayne / Ba...}]]	[[{"credit_id": "52fe4781c3a36847f81398c3", "de...}]]
4	49529	John Carter	[[{"cast_id": 5, "character": "John Carter", "c...}]]	[[{"credit_id": "52fe479ac3a36847f81398c3", "de...}]]

#### IV. RESULTS

```
: recommend('Batman Begins')
```

```
The Dark Knight  
The Dark Knight Rises  
Batman  
Batman & Robin  
Batman
```

```
: recommend('Avatar')
```

```
Titan A.E.  
Small Soldiers  
Ender's Game  
Aliens vs Predator: Requiem  
Independence Day
```

### Movie Recommender System

Type or select a movie from the dropdown

Avatar

Show Recommendation

Titan A.E.



Small Soldiers



Ender's Game



Aliens vs Predator: Requiem



Independence Day



#### V. CONCLUSION

I cleaned up the data that I got from TMDB. We performed exploratory data analysis on the data, such as extracting valuable insights from it, removing missing values, and preprocessing the data so that we could train our model on it. When a user visits our website and types the title of a movie into the search bar, he or she will see some autosuggestions related to the movie.

#### REFERENCES

- [1] Sang-Min Choi, Sang-Ki Ko, and Yo-Sub Han. "A genre correlation-based movie recommendation algorithm." Expert Systems and Applications, 39.9 (2012), pp. 8079-8085.
- [2] George Lekakos and Petros Caravelas, "A Hybrid Approach to Movie Recommendation." Tools for multimedia and applications 36.1 (2008), pp. 55-70 Das, Debashis, Laxman Sahoo, and Sujoy Datta are the names of three people. "A recommendation system survey." International Magazine 160.7 of Computer Applications (2017).
- [3] Jiang, Zhang, and colleagues. "Personalized real-time movie recommendation system: Practical prototype and evaluation." 180-191 in Tsinghua Science and Technology 25.2 (2019).
- [4] S. Rajarajeswari and colleagues. "Movie Recommendation System." Computing, Information, and Emerging Research Applications and communication 329-340, Springer Singapore, 2019.
- [5] Ahmed, Muyeed, Mir Tahsin Imtiaz, and Raiyan Khan are the members "Clustering and recommendation system for movies a network for pattern recognition "IEEE 8th Annual Meeting 2018
- [6] M. A. Hossain and M. N. Uddin, "A Neural Engine for Movie Recommendation System," in 2018 4th International Conference on Electrical Engineering and Information and Communication Technology (ICEEICT), pp. 443-448, doi: 10.1109/CEEICT.2018.8628128.



- [7] Dr. Yogesh Kumar Sharma, Monika D.Rokade (2020). Detection of Malicious Network Packet Activity utilizing Deep Learning 29(9s), 2324 - 2331, International Journal of Advanced Science and Technology.
- [8] M. A. Hossain and M. N. Uddin, "A Neural Engine for Movie Recommendation System," in 2018 4th International Conference on Electrical Engineering and Information and Communication Technology (iCEEICT), pp. 443-448, doi: 10.1109/CEEICT.2018.8628128.
- [9] Monika D.Rokade and Dr. Yogesh kumar Sharma, "Deep and machine learning approaches for anomaly-based classification." "Detection of intrusions in imbalanced network traffic," IOSR Journal of Engineering (IOSR JEN), ISSN (e): 2250-3021, ISSN (p): 2278-8719
- [10] Dr. Yogesh Kumar Sharma, Monika D.Rokade "MLIDS: A Machine Learning-Based Intrusion Detection Approach" 2021 International Conference on Emerging Smart Computing and Datasets for Real-Time Networks IEEE Informatics (ESCI)
- [11] Dr. Yogesh Kumar Sharma, Monika D.Rokade (2020). Detection of Malicious Network Packet Activity utilizing Deep Learning 29(9s), 2324 - 2331, International Journal of Advanced Science and Technology.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)