**RECOMMENDER SYSTEM DESIGN**

Project submitted to the

SRM University – AP, Andhra Pradesh

for the partial fulfillment of the requirements to award the degree of

**Bachelor of Technology**

In

**Computer Science and Engineering**  **School of Engineering and Sciences**

Submitted by

Sahithya Akula – AP21110011110



Under the Guidance of

**Dr. Mudassir Rafi**

**SRM University–AP**

**Neerukonda, Mangalagiri, Guntur**

**Andhra Pradesh – 522 240**

**Nov,2023**

# Certificate

Date: 28-Nov-23

This is to certify that the work present in this Project entitled “**Recommender System Design**” has been carried out by **Pradeep, Sudeepa, Sahitya, Mahendra** under my/our supervision. The work is genuine, original, and suitable for submission to the SRM University – AP for the award of Bachelor of Technology/Master of Technology in **School of Engineering and Sciences**.

## Supervisor

(Signature)

Dr. Mudassir Rafi

Designation, Affiliation.

## Co-supervisor

(Signature)

Prof. / Dr. [Name]

Designation, Affiliation.

# Acknowledgments

We are grateful to my UROP Advisor Dr. Mudassir Rafi who has been our mentor throughout the project and guided us accordingly like how exactly we have to do research.

Next, I express my deepest sense of gratitude to our college SRM University AP for giving us an opportunity to do research as a part of our curriculum which encouraged us to learn new things and made us think from different perspectives.

We have been interested in research lately and when we got this opportunity, We are very excited and thought of making use of this opportunity which will help us to bag a research publication under our name and helps us if we want to go abroad for a master's.

We are also grateful for this project as everyone contributed their part which made our project simpler and got completed in no time.

Our biggest concern was that maybe we will complete college with only a degree that contains the grades of only our academic subjects but when we got to know that our college is providing us with this research opportunity under supervision or guidance made us feel better as a part from the subjects, we will be doing research as a part of our curriculum and with the proper guidance, we are also hoping that the research paper we are going to submit at the final will publish and help us in many ways.

I would also like to express my sincere gratitude to Dr. Ajay Bhardwaj who is the UROP Organizer thank you so much sir for guiding us to approach or meet our faculty on a regular basis and showing them progress, and also giving deadlines to the activities of the UROP like monthly assessment, Report submission and many more which really made us think of contributing some of our time to the project and make progress accordingly.

Thank you so much for letting us express our gratitude toward everyone who has been part of UROP and guiding us accordingly.

# 

# Table of Contents

Certificate…………………………………………………………….i

Acknowledgements…………………………………………………ii

Table of Contents……………………………………………………iii

Abstract………………………………………………………………iv List of Publications………………………………………………….

Statement of Contributions…………………………………………

Abbreviations………………………………………………………..v

List of Tables…………………………………………………………vi List of Figures……………………………………………………….vii

List of Equations……………………………………………………viii

1. Introduction……………………………………………………..1
2. Methodology…………………………………………………….2
3. Discussion………………………………………………………..7
4. Concluding Remarks……………………………………………13
5. Future Work…………………………………………………….14

References……………………………………………………………15

# Abstract

As nowadays technology is increasing and people are getting a lot of selections or options for any product they want to purchase or they are confused also to decide what they want to do in their free time or thinking of going to a new place but, which place? because of these confusions, a technology is invented known as the Recommender system.

Nowadays people are showing interest in OTT (Over-The-Top) platforms like Amazon Prime Video, Netflix, Voot, and many more as they are cheap compared to the cost of a movie ticket for a whole family so even if it takes time people are preferring OTT platforms. But these OTT platforms contain a huge amount of data on movies of different genres and different languages. Because of this people also get in a state to decide which movie or series they want to watch in their free time.

The reason we have discussed OTT platforms is that we have designed a Recommender System for movies that are used in the OTT platforms to suggest to the respective user what they are interested to watch in their free time.

We will be using different techniques in designing this Recommender System.

Keywords: - selections, options, OTT platforms, confused, Recommender System, different genres, different languages, decide, interested, different techniques.

# Abbreviations

OTT Over – The – Top

CBF Content Based Filtering

CF Collaborative Filtering

MF Matrix Factorization

DCF Deep Collaborative Filtering

DL Deep Learning

LSA Latent Semantic Analysis

LDA Linear Discriminant Analysis

PR Predicted Rating

# List of Tables

Table 1 …………………………..………………………….. 8

Table 2 ……………………………………………………… 8

Table 3 ……………………………………………………… 10

Table 4 ……………………………………………………… 11

Table 5 ……………………………………………………… 12

Table 6 ……………………………………………………… 12

Table 7 ………………………………………………………. 13

# List of Figures

Figure 1 ……………………………...………………………. 9

Figure 2 ……………………………………………………… 9

Figure 3 ……………………………………………………… 10

Figure 4 ……………………………………………………… 11

# List of Equations

Equation 1 ………………………………………………………….... 3

Equation 2 …………………………………………………………… 3

Equation 3 ……………………………………………………………. 3

Equation 4 ……………………………………………………………. 4

Equation 5 ……………………………………………………………. 5

Equation 6 ……………………………………………………………. 6

**1.Introduction:**

Because of the increasing options in the everyday life of Human Beings, because of this people are not easily happy with options but they are confused to select the products which are available in the market at the same time thinking it is worth it or not.

This is also the same when it comes to Movies deciding or selecting this happens to many people whenever the people go through an online platform where it consists of all the movies in the world. Because of this huge number of options, it is not simple for people to select what they want to watch during their free time so that they can enjoy it.

The main advantage of selecting the Movies Database Is that the Movies are categorized into different genres, categories, and languages.

A recommender system is a system that provides personalized recommendations to users based on their interests, preferences, and behaviors. Recommender systems are becoming increasingly popular and are used in a wide range of applications from ecommerce to social media. Recommender systems are useful for making personalized selections, providing tailored recommendations, and improving user engagement. In this tutorial, we will discuss the design of a recommender system and how it can be implemented using various technologies. We will also explore some of the challenges associated with creating an effective and efficient recommender system.

The dataset contains information on different movies such as title, genre, release year, rating, and the number of ratings. We will use this data to create a recommender system that can suggest movies to users based on their preferences.

First, we need to pre-process the data. This includes cleaning the data to remove any invalid or missing values, and normalizing the ratings so that they are in the same range.

Next, we need to create a user profile. This is a collection of information about the user such as their favorite genres, the types of movies they like, and their preferred rating range.

These recommender systems have different techniques which are implemented and, in this paper, we will be discussing the different techniques which we used to build a particular recommender system.

The different techniques that we have thought of implementing are 1) Content-Based Filtering Technique, 2) Collaborative Based Filtering Technique, 3) Hybrid Technique, 4) Deep Collaborative Filtering Technique, and 5) Matrix Factorization Technique.

**2.Methodology:**

The methodology section basically focuses on the methods or techniques we have used to design this particular Movie Recommender System.

Content Based Filtering Technique: -

Content-based filtering is a type of recommendation system that uses the features of items in a dataset to recommend similar items to users. In the case of a movie dataset, content-based filtering can be used to recommend similar movies to a user based on the features of the movies they have already watched. This type of filtering can be used to determine which movies a user would be interested in watching based on their past viewing habits. The features used in content-based filtering can include genre, cast, director, plot, and other related information.

For example, a user who has watched a number of action movies might be recommended similar action movies based on the common features they have with the movies the user has already watched. Content-based filtering can also be used to recommend movies in different genres that the user might not have considered before. For example, a user who has only watched comedies might be recommended a romantic comedy or a thriller.

Create a data frame from the movie dataset. Calculate the similarity between movies using a distance metric such as Euclidean distance or Pearson correlation. Find the most similar movies for each movie. Create a recommendation system by recommending the most similar movies for each movie. Evaluate the performance of the system by measuring the accuracy of the recommendations.

Content-based filtering is an effective way to recommend movies to users based on their past viewing habits and preferences. By analyzing the features of the movies a user has watched, content-based filtering can accurately recommend similar movies that the user might enjoy.

Content-based filtering is a technique for recommending items to users based on their past history.

Formula:

1. Similarity Score = Σ (u1i \* u2i) / √ (Σu1i^2 \* Σu2i^2) ………. (1)
2. Prediction Rating = Σ (similarity scores \* rating of similar item)/ Σ (similarity scores)

………… (2)

1. Recommendation Score = Prediction Rating - Average Rating of Items ………(3)

Where u1i = User 1’s rating of item i u2i = User 2’s rating of item i

## 1.1 Collaborative Filtering Technique: -

Collaborative filtering is a type of recommendation system which uses the past user interactions to identify the potential items for the current user. It is a type of filtering technique which utilizes the data of user’s behaviors, activities, or preferences to make recommendations.

For movie dataset, the collaborative filtering technique can be used to recommend movies to a user based on the ratings of other users who have similar tastes. The technique uses the ratings of the users who have watched the same movies as the current user and compares them with the user’s ratings. Based on the comparison, the system suggests the movies with similar ratings to the current user. This helps the user to get personalized recommendations and explore movies that he might like.

For example, if a user has watched a set of romantic comedies and rated them highly, then the system can recommend other similar romantic comedies to the user. Similarly, if the user has watched a horror movie and rated it low, then the system can recommend other horror movies to the user with lower ratings.

Collaborative-based filtering is a type of recommendation engine that uses the opinion of the community to make predictions about a user's preferences. This technique works by collecting data from users on items they have interacted with, such as ratings, reviews, or purchases.

The formula for collaborative based filtering can be expressed as follows:

Predicted Rating (User, Item) = Σ (Similarity (User, OtherUser) x Rating

(OtherUser, Item)) / Σ Similarity (User, OtherUser) ………… (4)

Where:

User = The user for whom the predicted rating is being made

OtherUser = All other users who have interacted with the given item

Similarity (User, OtherUser) = A measure of how similar two users are in terms of their preferences

Rating (OtherUser, Item) = The rating given by the OtherUser for the given item

This formula can be used to predict ratings for items that the user has not yet interacted with, based on the ratings of similar users.

## 1.2 Hybrid Technique: -

Hybrid filtering techniques combine the best of collaborative filtering and contentbased filtering techniques to make recommendations. For example, a hybrid filtering system can use collaborative filtering to generate a list of similar users who have rated movies similarly in the past, and then use content-based filtering to narrow down the list of recommended movies by using the user's personal preferences.

For example, let's say a user has watched and rated several romantic comedies in the past. The hybrid filtering system would use collaborative filtering to identify a list of other users who have watched and rated similar movies in the past. Then, the system would use content-based filtering to narrow down the list of recommended movies by taking into account the user's personal preferences. The system could recommend movies that have a similar plot or genre, or even take into account the user's favorite actors or directors.

The hybrid filtering technique for a movie dataset would involve combining the properties of collaborative filtering and content-based filtering. First, collaborative filtering would look at user ratings and preferences to make recommendations. The content-based filtering would take into account attributes of the movies such as genres, actors, directors, etc. Both of these approaches would be combined to create a hybrid recommendation system that would take into account both user ratings and movie attributes to make recommendations.

PR = w1 \* Content-Based Rating + w2 \* Collaborative Filtering Rating ……… (5)

where w1 and w2 are weights to adjust the weight of content-based and collaborative filtering ratings respectively.

## 1.3 Deep Learning: -

1. Data Collection: The first step in developing a deep learning-based recommendation system using the movie dataset is to collect the necessary data. This data can be obtained from various sources such as movie websites, IMDb, Rotten Tomatoes, Netflix, etc. The data should include the movie titles, ratings, genres, release date, cast, etc.

1. Data Preprocessing: The next step is to preprocess the data. This involves cleaning the data, removing any irrelevant information, and formatting it properly. The data should also be split into training and testing sets.

1. Model Selection: After the data is preprocessed, the next step is to select a suitable model for the recommendation system. There are several models available such as collaborative filtering, content-based filtering, and deep learning models. The model should be selected based on the type of data and the desired output.

1. Model Training: Once the model is selected, it needs to be trained using the training data. This involves feeding the data into the model and adjusting the parameters to achieve the desired output.

1. Model Evaluation: After the model is trained, it needs to be evaluated to check the accuracy and performance of the model. This can be done by comparing the actual results with the predicted results.

1. Model Deployment: Once the model is evaluated, it can be deployed in the production environment. This involves deploying the model on the server and setting up the necessary infrastructure.

## 1.4 Matrix Factorization: -

Matrix factorization is a technique used in recommender systems to predict user ratings for items based on their past interactions. It is based on the assumption that there is a low-dimensional latent feature space that captures the similarity between items. By decomposing the user-item matrix into two smaller matrices, we can learn the latent features which characterize both users and items.

The goal of matrix factorization for the movie dataset is to learn the latent features of users and movies that are related to each other. The data is first split into a training and a test set. The training set is used to learn the latent features and the test set is used to evaluate the performance of the model.

The matrix factorization process involves two steps. First, the user-item matrix is decomposed into two matrices, one representing the user latent features and the other representing the item latent features. For each user and each item, a feature vector is learned which captures the user-item interaction.

Second, the user and item latent feature vectors are combined to predict the user ratings for items. The predicted ratings are then compared to the actual ratings in the test set to evaluate the model.

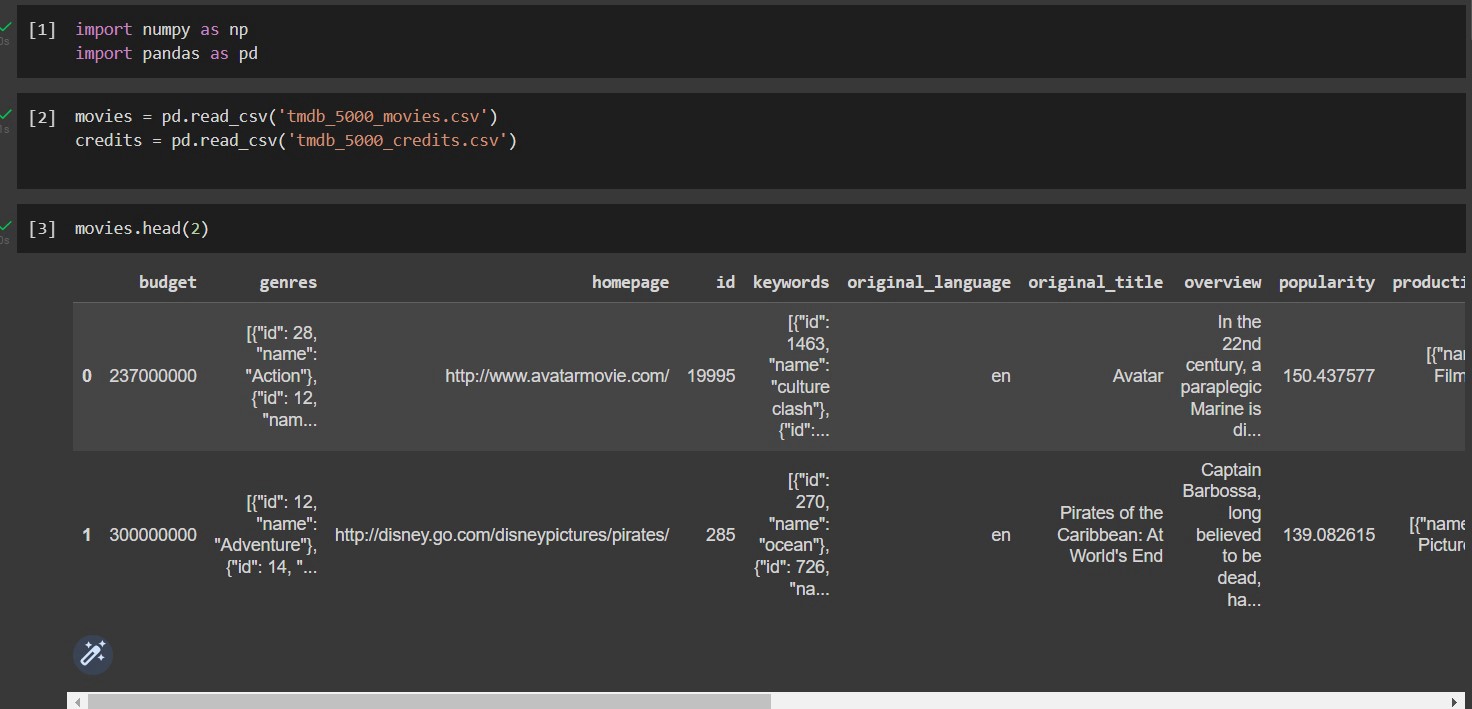
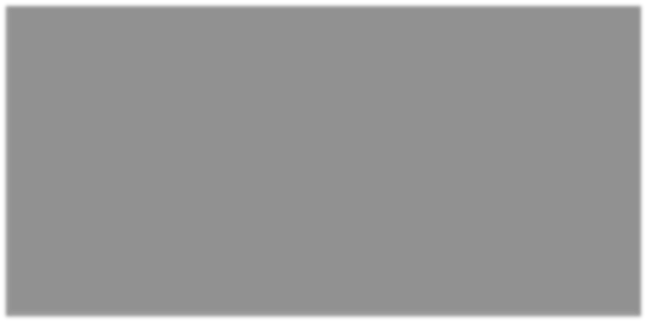
Matrix factorization can be used to improve the accuracy of recommender systems. It is a powerful technique that can help to accurately predict user ratings for items. In addition, it can be used to uncover hidden features which can be used to further improve recommender systems.

Predicted rating = Σ (user factor \* item factor) ………… (6) where user factor and item factor are matrices obtained from matrix factorization.

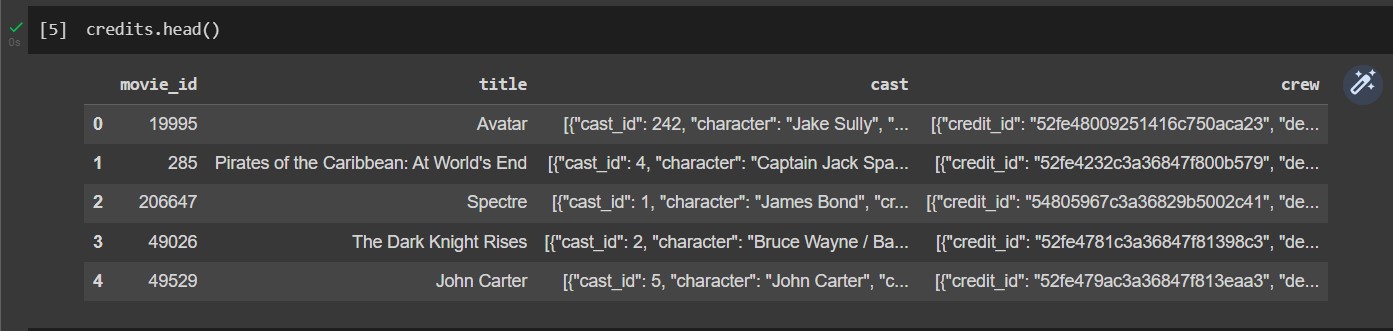
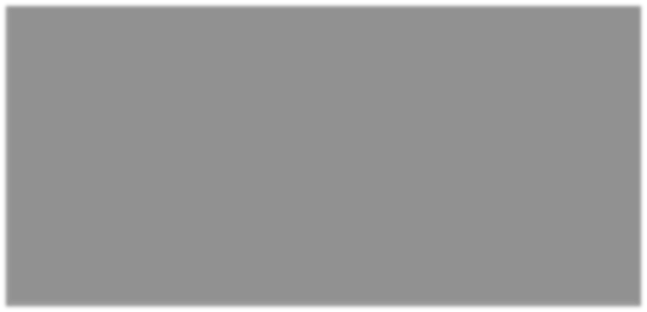
# 3.Discussion

**a Content Based Filtering Technique: -**

In this technique we will be taking a dataset containing different movies form online and we will take the data in the form .csv file and we will be accessing the data using code and we get the data in the following way.

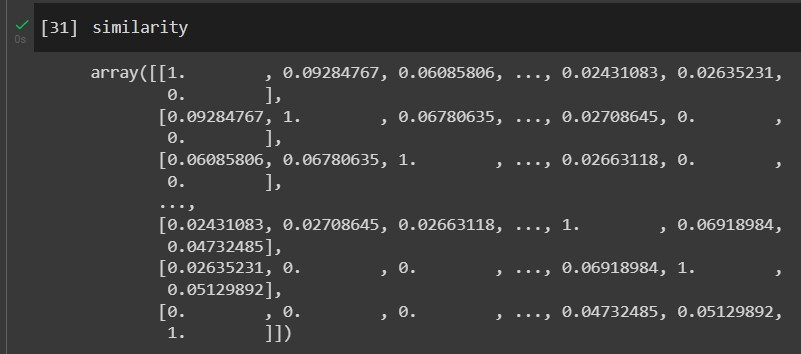
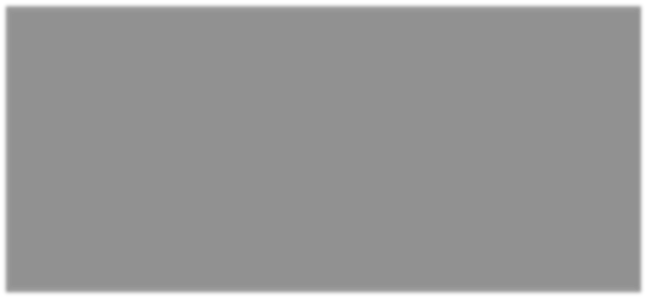


**Table 1. Accessing the movie data from Dataset using python**

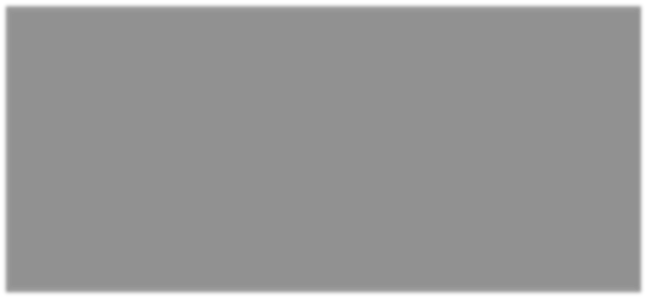


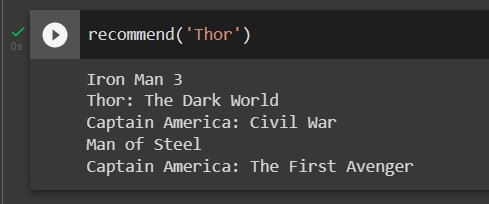
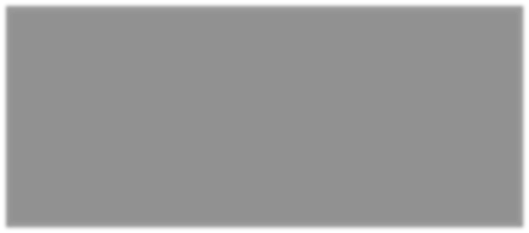
**Table 2. Accessing the credits data from Database using python**

Now we will be showing the matrix generated by the system.



**Figure 1. Generating the matrix for the data from Database using python**

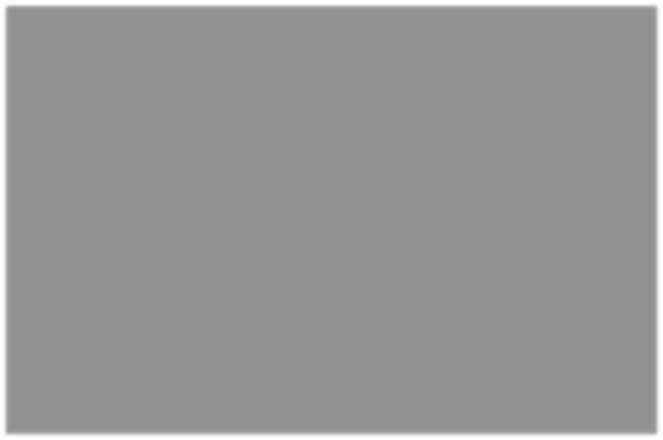
Result of Content Based Filtering Technique is:



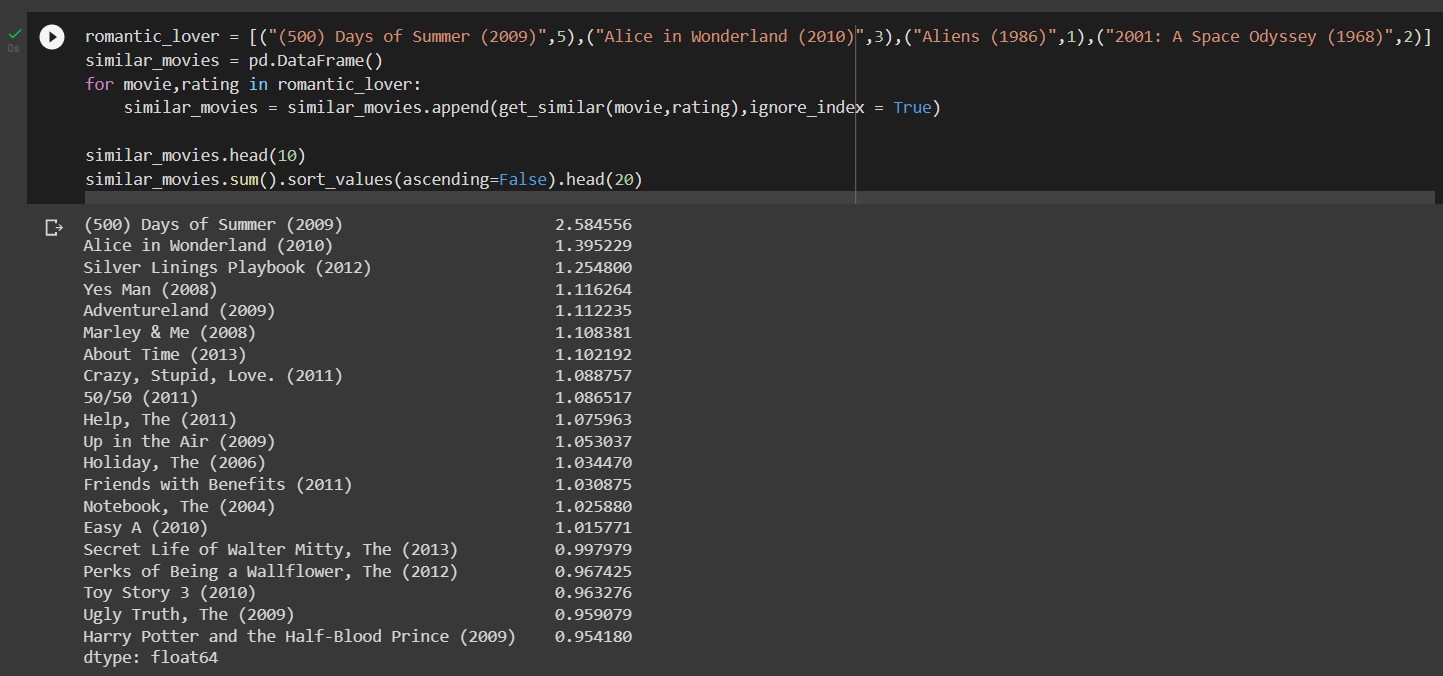
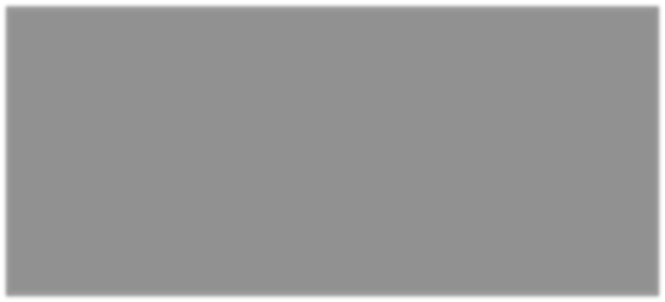
**Figure 2. Printing the recommended movies as output**

## b Collaborative Filtering Technique: -

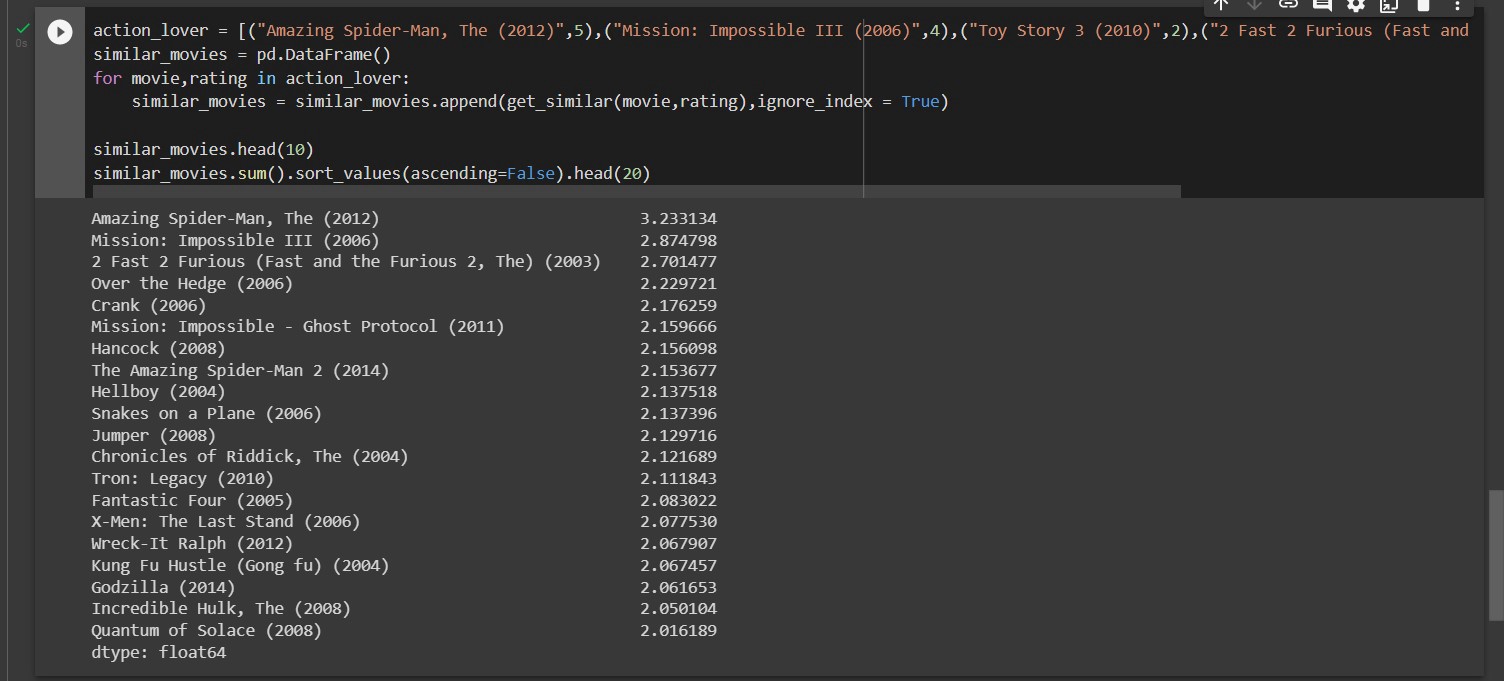
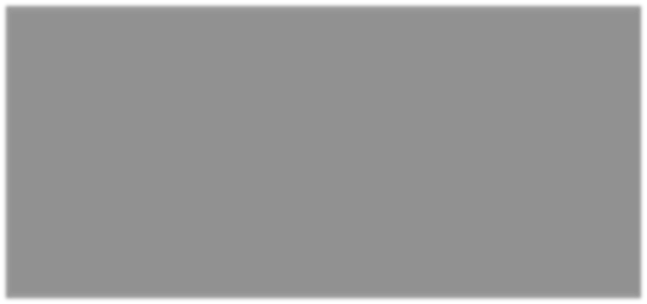
In this technique we will be taking a dataset containing different movies form online and we will take the data in the form .csv file and we will be accessing the data using code and we get the data in the following way. But in this technique, we will compare the ratings among different users.



**Table 3. Accessing the movies and ratings data from Dataset using python**



## Figure 3. List of recommended movies and their similarity index for USER 1

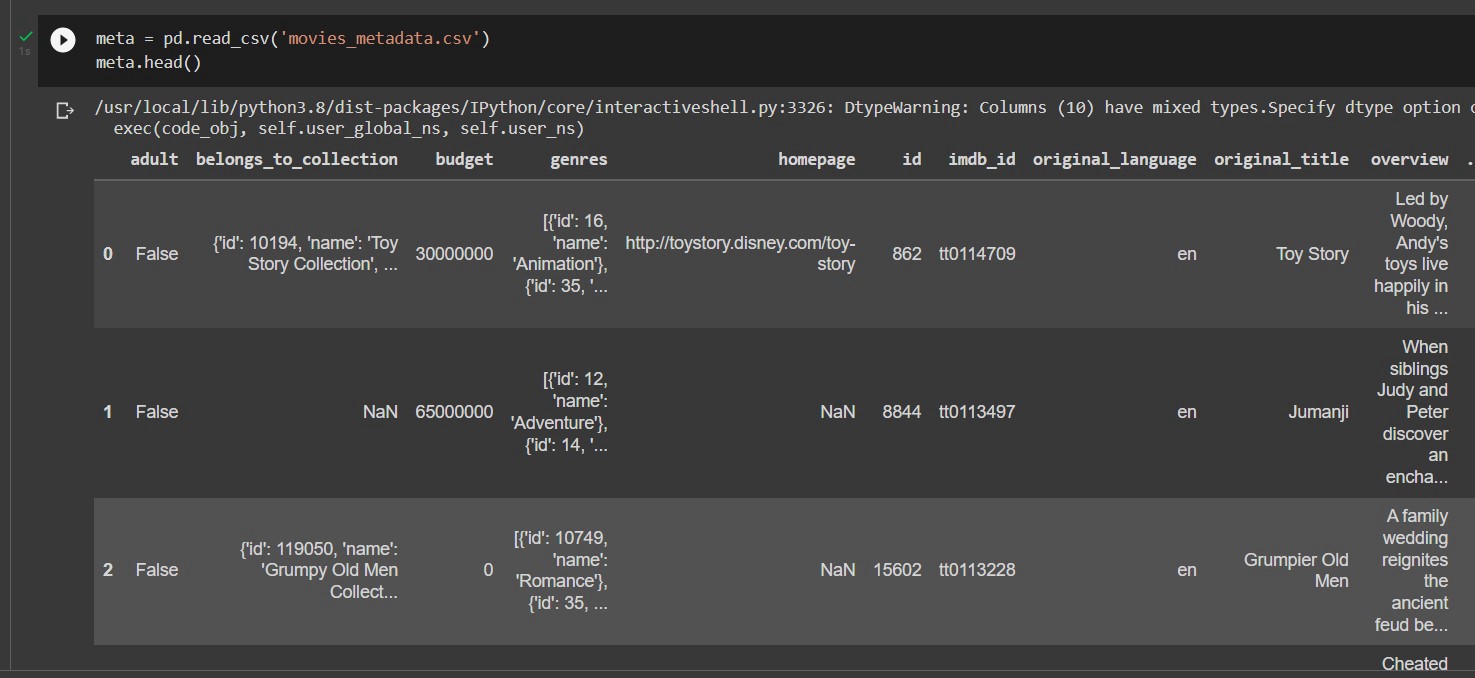
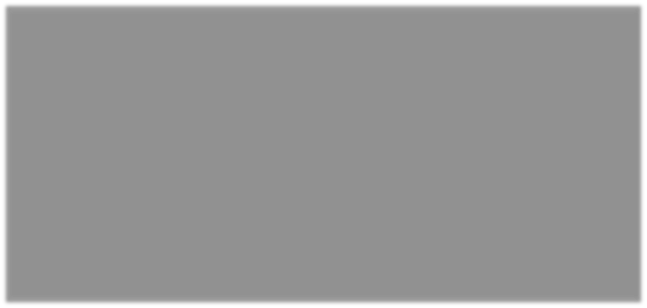
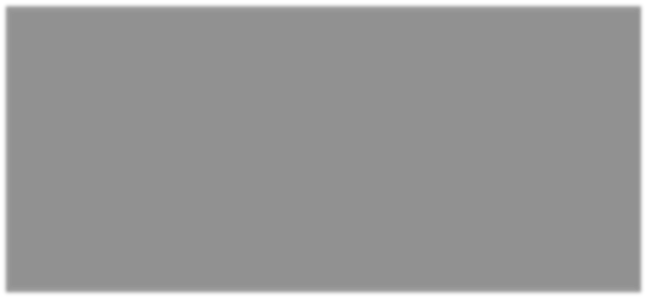


## Figure 4. List of recommended movies and their similarity index for USER 2

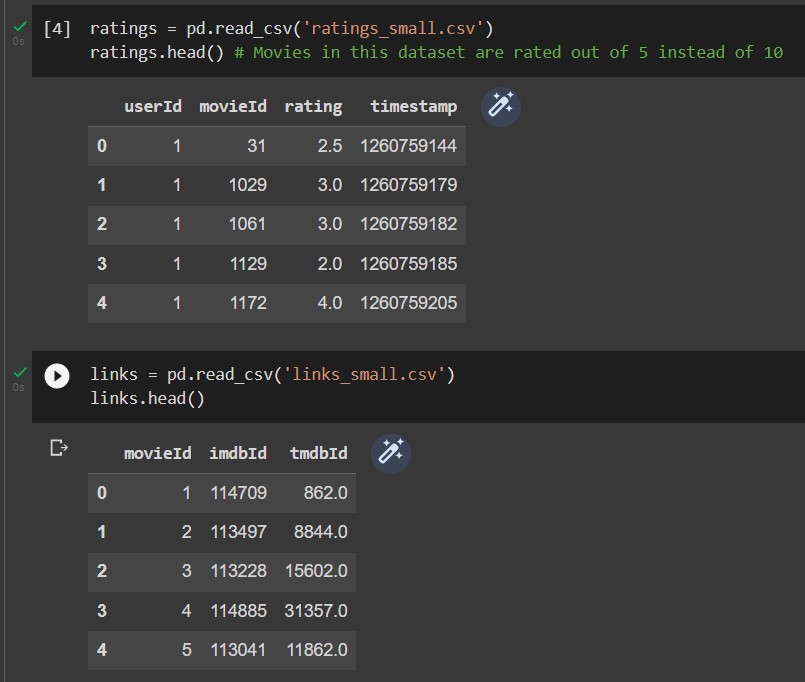
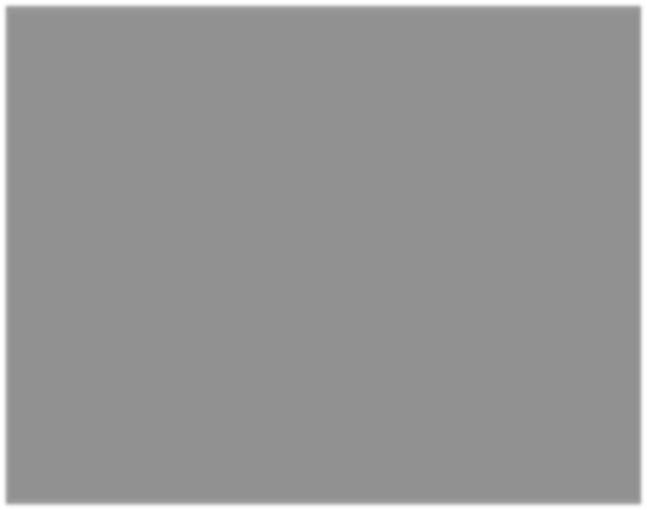
Here USER 1 likes Romantic and Comedy movies so he/she rates them accordingly and with the help of the logic in the code we get the recommendations of the movies based on the similarity index values. The same happens to USER 2 also, who likes Action Movies.

## c Hybrid Technique: -

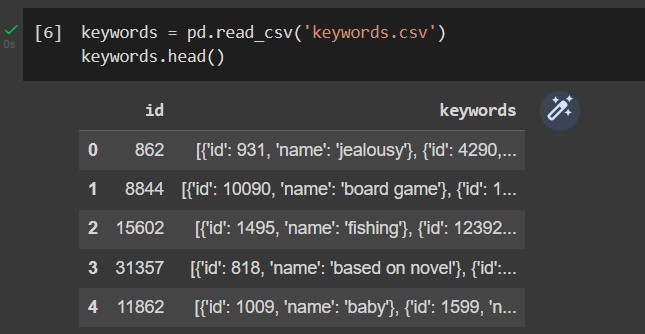
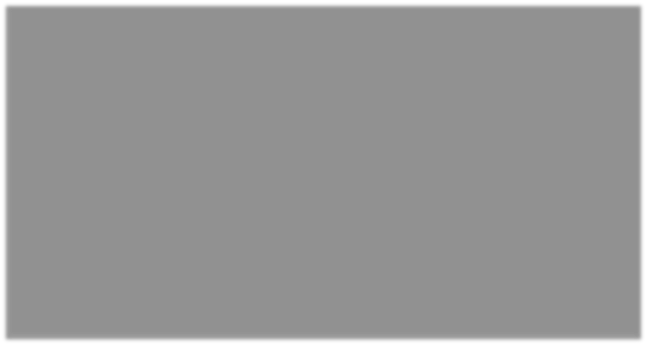
In this technique we will be taking a dataset containing different movies form online and we will take the data in the form .csv file and we will be accessing the data using code and we get the data in the following way. But in this technique, we will be combining both the Content based and Collaborative filtering techniques.



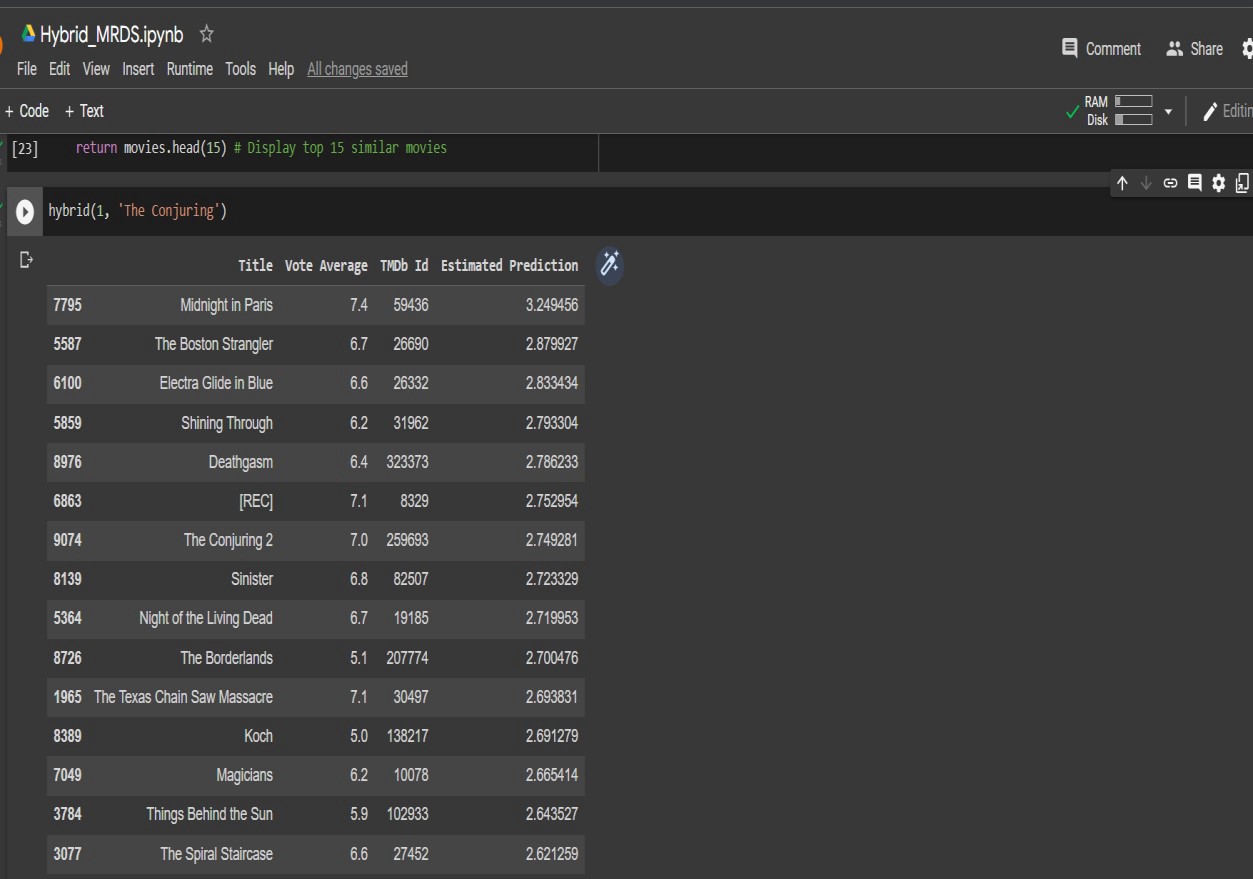
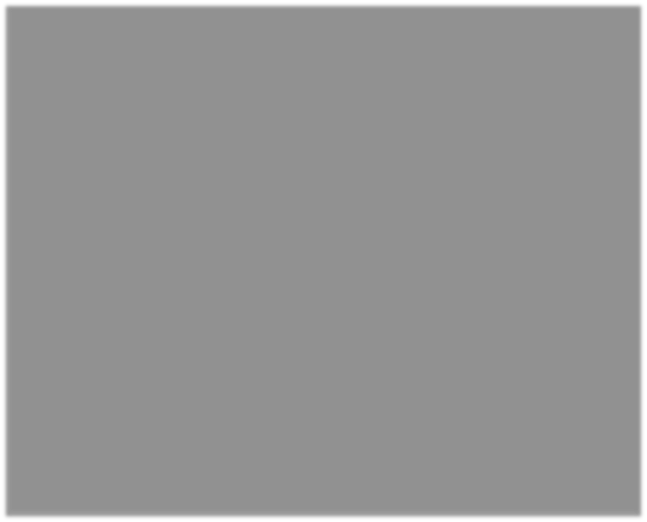
**Table 4. Accessing the movie meta data from Dataset using python**



**Table 5. Accessing the links\_small and ratings\_small from Database using python**



**Table 6. Accessing the keywords from Database using python**



**Table 7. The list has movies which are recommended using hybrid techniques for a specific user.**

# 4. Concluding Remarks

This report is just an introduction to the recommendation systems. there are still many more types of the recommendation systems.

This report mainly focusses on implementation of basic methods such as content based, collaborative filtering and hybrid recommendation systems.

Content based recommendation systems generates the recommendations based on user preferences whereas collaborative filtering predicts the interests of the user by identifying information from many users. Hybrid systems are combination of recommendation systems.

Recommender system design is a complex task that requires careful consideration of user preferences, business goals, and data sources. By combining AI and machine learning techniques with a well-crafted user interface, it is possible to create powerful and effective recommender systems that accurately match users with products, services, and content. With thoughtful design and implementation, businesses can greatly benefit from the power of recommenders to drive more effective marketing, increase customer satisfaction, and improve the overall user experience.

We have discussed the fundamentals of design for a recommender system, exploring the various components involved in creating a successful system. We have looked at the importance of data-driven approaches, leveraging algorithms to accurately capture user preferences, and discussed the challenges of balancing accuracy and scalability. We have also explored the challenges of implementing a recommender system in the real world, such as dealing with cold-start or limited data. Ultimately, designing a recommender system is an iterative process that requires thoughtful consideration of both the user experience and the data. With the right approach, a recommender system can be a powerful tool to help users find and discover new content.

# 5.Future Work

Future work in recommender system design should focus on developing more advanced hybrid approaches that combine traditional collaborative filtering methods with content-based approaches to increase the accuracy and effectiveness of the recommendations. Additionally, as the data available to recommender systems continues to grow, research should focus on developing ways to efficiently manage and analyze large-scale datasets such as those used in Netflix and Amazon. Finally, more research should be conducted on the impact of the user experience on the effectiveness of the recommendations, as well as on developing more personalized recommendations based on user preferences.

In future We are planning on creating a web page and later an application which asks user a particular set of questions about his/her interests and they have to choose their interests and when they choose all the things, the logic in my page runs accordingly and tries to give the best suitable or suggested movie to that particular user.

We will be using the same techniques as discussed in this report as they are also important for designing a recommender system.

# References

1. D. Ricci et al., “Recommender systems handbook,” 2011

1. C. C.Aggarwal, “Recommender systems: The textbook,” Springer International Publishing, 2016.

1. O. Nasraoui et al., “Recommender systems: Techniques, technologies, and trends,” IEEE Intelligent Systems, vol. 26, no. 2, pp. 24–36, 2011.

1. Design and Evaluation of a Recommender System” by R. B. Allen and P. Tuzhilin.

1. Recommender System Design: A survey” by G. Adomavicius and A. Tuzhilin.

1. L. Tang et al., “Social recommendation: A survey,” ACM Computing Surveys, vol. 52, no. 1, pp. 1–37, 2019.

1. Recommender System Design and Evaluation” by P. Brusilovsky and A. Kobsa.

1. X. Z. Gao et al., “Context-aware recommender systems,” ACM Computing Surveys, vol. 51, no. 1, pp. 1–44, 2018.

1. Pazzani et al., “Content-based recommendation systems,” in The adaptive web, pp. 325–341, Springer, 2007.

1. Design and Evaluation of a Recommender System” by M. J. Pazzani and D.

Billsus.

1. Y. Koren, “Collaborative filtering with temporal dynamics,” Communications of the ACM, vol. 53, no. 4, pp. 89–97, 2010.

1. Recommender System Design and Evaluation” by P. J. Melville and R. Mooney.