Recommendation Systems: Overview, Future Scope, and Effectiveness of Modelling Based on Mind Maps

Shivam Patel,
Department of Computer Science Engineering, Chandigarh University, Gharaun-140413, Punjab, India
Email: shivampatel2765@gmail.com

Kavita,
Uttaranchal University, Dehradun Email: kavita@ieee.org

Sonali Dash,
Department of Computer Science Engineering,
Chandigarh University, Gharaun-140413, Punjab, India

Rajendra Prasad
Uttaranchal University, Dehradun Email: rajendraprasad@uumail.in

Abstract—Recommendation systems/engine do the filtrations the data for that it uses different algorithms and recom- mend/predict the most suitable items for users. It starts capturing which users may purchase. Three main methods are used in our recommendation programs. Another Demographic Filtering i.e. Basic premise behind this system is that such films are they are very popular and highly respected they will have highs opportunities to appeal to the average audience. One is cooperation filtering, when we try to bring the same person together to form the group to make recommendations for the user. Hybrid Recommendation System for movie use a combination of filtering in conjunction with content of complimentary programs. One of the best ways to get the recommendation is using the sites like IMDB, rotten Tomato's etc. because they give the ratings to movies based on user reviews and critics comments and various factors included. Hence, we can also use ratings as a base to make our model to predict the recommended movies based on some kind or numerical rating or value. Hence, Higher the values of rating or number the more the movies are recommended to the users.

INTRODUCTION

Recommendation system, a form of information a sugges- tion system that attempts to predict favorites user, and make suggestions-based preferences. They may have a less services to make good suggestions. If the video streaming applications can predict and play user-friendly video and movies, the people will just stop to use it. This will lead to more pressures by tech organizations in developing their recommendation plans. This problem is tough than it looks. Always the user has different preferences and preferences. For example, the type of video one can makes while cooking is very different from the type of the video we watch when we are travelling.

Another issue that recommendation plans to resolve by testing vs problem of exploitation [1]. They have to check new domains to find out extra knowledge about the person/user, while standing to make the most from what we know till now about the user. Three main methods are used for us recommendation systems.

A compliment system that creates a certain type of in- formation filter trying to present things with interest user defined [2], [13]. Most users are there who are here for e- commerce customer applications, that help peoples in taking decisions by giving personal info [3]. Aside the accuracy of the compliment, the efficiency of the computer is most considerate in the field of Information Technology [4], [22]. Usually, Recommendation had to manage users whose numbers are in millions, making computer measurement instantaneous. Below the limits of time and space the use of many predictive algo- rithms is accessible quickly their volume limit of controllable data potential. To manage large databases, some upgrades are on information

representation and recommendations modeling should be done [5], [14].

II. LITERATURE REVIEW

There is a recommendation system named MOVREC that is presented by D.K. Yadav that is based on shared filtering method. There is a technique called collaborative filtering that uses information about user. That information is processed and the recommendation system give result of about the movie that has high degree or which is most rated.

Also, the Luis M. Capos researched about two techniques of recommendation algorithms based on content and collab- oration, which was result to a failure but with that failure he proposed a new algorithm which was the combination of both the algorithms (Bayesi Network and Collaboration method). Also, a combined system of made by Harpreet kaur. That program is made with the combination of content and collaboration method [23].

There are many factors included while a movie is considered as a recommendation to any user like what is the content of movie like, what is the rating is movie by censor board and also what are the reviews of critics and other viewers on the movies are also considered. Utkarsh Gupta compiles such crucial information using a caterpillar [6], [15]. This is an effective method based on hierarchical compilation of a recommendation system. Predicting the voting system level is used. The proposed System has a low error and has a better combination of similar features demonstrated by Urszula Kuzelewska et al., He proposed the integration as a way to deal with the recommendation program [7], [16].

III. PROPOSED SYSTEM

System we are proposing will be having a website with multiple web-series or movies and shows. Registration process can also be added but for now we are skipping it as we are focused on algorithms and refining our model at best. What our model is focused on is feedback every user provide to certain movie and shows and the processing the feedback and categorizing the results and give some ratings calculated and most rated movie gets recommended. We are also doing real time analysis to ensure that the system is working fine and is based on people's behavior [3], [17]. The below [Figure 1] is shown the Flowchart of System Architecture.

The Workflow of model is like first we preprocess the data and create the data frame from that and then with features will do different types of functions like features dimensionality reductions and also apply multiple ML Algorithm to get final results and at the end we have used hybrid approach by combining two different algorithms to get final output as recommended movie or show[21-35]. Below is the proper flow is given [Figure 2].

This is [Figure 2] shows how we start with the data, preprocess it and do multiple things to make data modeling ready and then applying ML Algorithms for getting the perfect results and get most perfect recommendations.

IV. METHODOLOGY

Some of the different Approaches we have used are given as:

A. Demographic Filter- provides users with the same people background with the same popular and well-rated movies regardless of genre or other features [8], [18]. [Figure 3 and Figure 4] Therefore, since it does not take into account individual tastes, it provides a simple but easy-to-use effect [9].

Before getting started with this –

- Need of scoring metric to rate movies
- Calculating the scores for each different movie
- Sorting the scores and recommending the best results. Used WR [Eq 1] derived below in: -

Weighted Rating (WR) = ((v/v+m). R) + ((m/v+m).C)

(1)

where, V: Number of votes; M: Minimum Votes; R: Average Rating; C : Mean votes across the data

B. Content-based Filtering Systems:

Main benefit of using this is that there is no need of remembering the matrices. Here what we do is from the ratings of existing users make ML

models that can predict the ratings effectively and hence we will get the ratings of all movies and thus can perform recommendations [10]. Using different types of differentiating factors to recommend like the predictions can be based on user's previous watch or likes, this can be translated as feedback but from external sources [11], [19].

For showcasing [Figure 4] the work flow of this algorithm we can use an example of some features extracted from Apple App store. Here in the diagram each function has some category (let's say action, thriller or something else). I t can be simplified as a matrix of binary where 1 represents that application contains that feature [12], [20].

We will use cosine function to calculate the similarity [Eq.2] of the matches for movies [2], [21].

Derived as follows:

$$\text{Similarity} = \cos(\theta) = (A.B)/(||A||\ ||B||) \qquad (2)$$

V. DESIGN:

The Above [Figure.5] shows of Activity of Recommending movies starts and what is does after till the End where we get recommended movies or shows.

Above Figure [Figure 6] is explained in below explanations.

The Above Diagram shows that the User has the following authorities: -

.

He can create and edit his/her account He can view profiles

He can search for various movies

He can rate and edit movie rating as per his choice The admin has the following authorities: -

He can edit account only if the user permits him to do so He can view profiles

He can search for movies

He can edit movie description

The Movie Maker has the following authorities

- He can edit movie description
- He can check for success prediction

VI. CONCLUSION:

A hybrid approach is taken to get the final results because depending upon any single algorithm creates a biasness and dependency towards that algorithm. Also, by using hybrid approach we have increased the accuracy of our model because it uses the advantages of both the algorithms and reduces disadvantages of both also. We also used techniques such as Similarity, Classification and Clustering to get the better results and hence ultimately reducing the MAE and helping in improving accuracy as well as precision. In the future we can work on others domains also like songs, news, books etc. using the same approach.

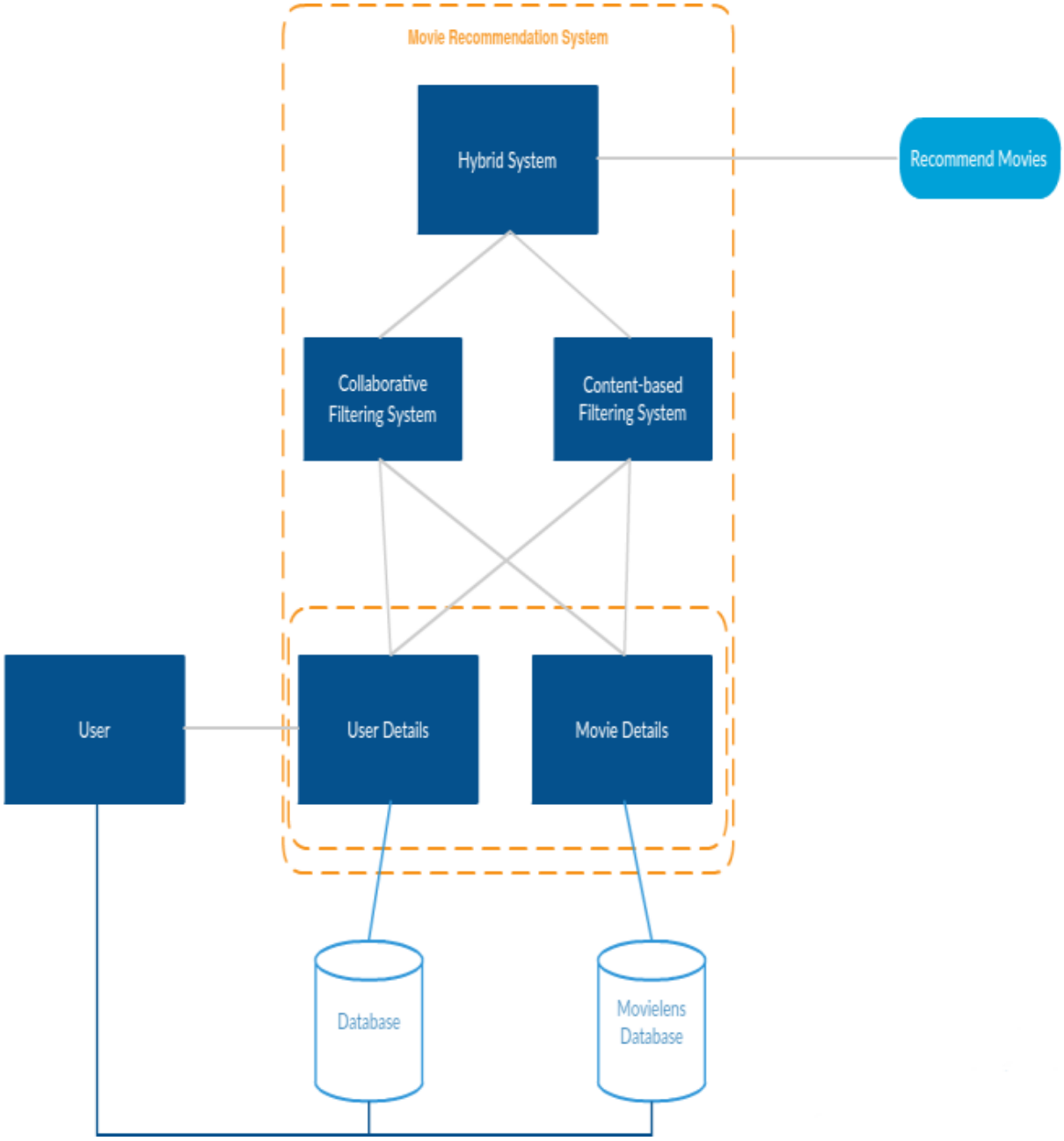

Fig. 1. 1. Flowchart of System Architecture

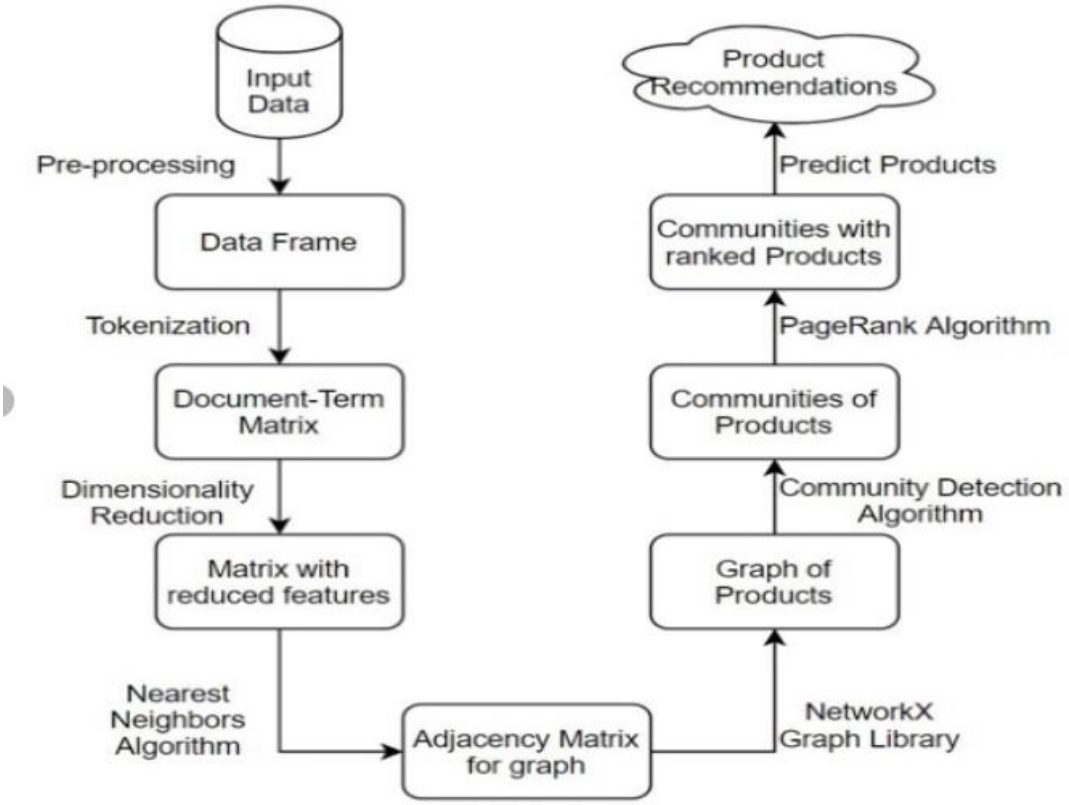

Fig. 2. Workflow of the System

```
#Sort movies based on score calculated above
q_movies = q_movies.sort_values('score', ascending=False)

#Print the top 15 movies
q_movies[['title', 'vote_count', 'vote_average', 'score']].head(10)
```

	title	vote_count	vote_average	score
1881	The Shawshank Redemption	8205	8.5	8.059258
662	Fight Club	9413	8.3	7.939256
65	The Dark Knight	12002	8.2	7.920020
3232	Pulp Fiction	8428	8.3	7.904645
96	Inception	13752	8.1	7.863239
3337	The Godfather	5893	8.4	7.851236
95	Interstellar	10867	8.1	7.809479
809	Forrest Gump	7927	8.2	7.803188
329	The Lord of the Rings: The Return of the King	8064	8.1	7.727243
1990	The Empire Strikes Back	5879	8.2	7.697884

Fig. 3. Demographic Filtering

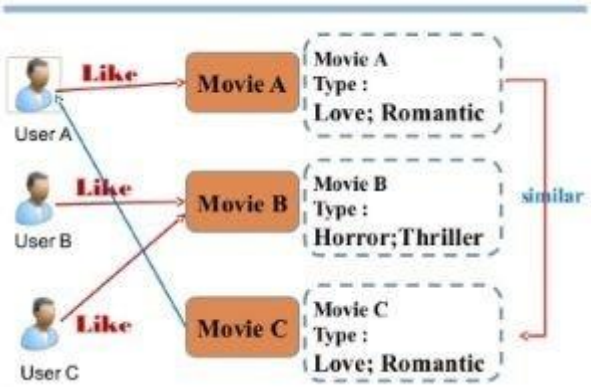

Fig. 4. How Content Based Filtering works

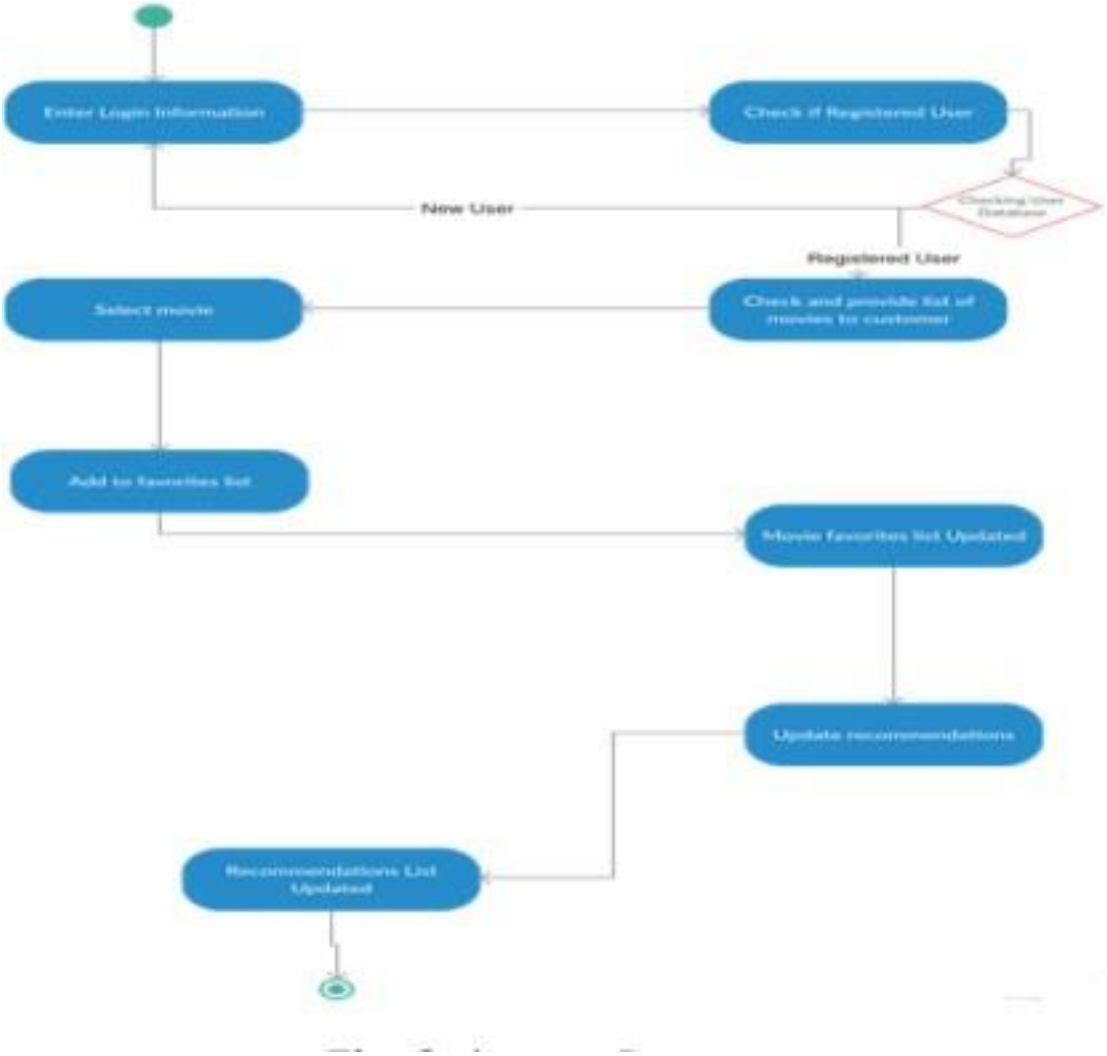

Fig. 5. Flowchart of Activity

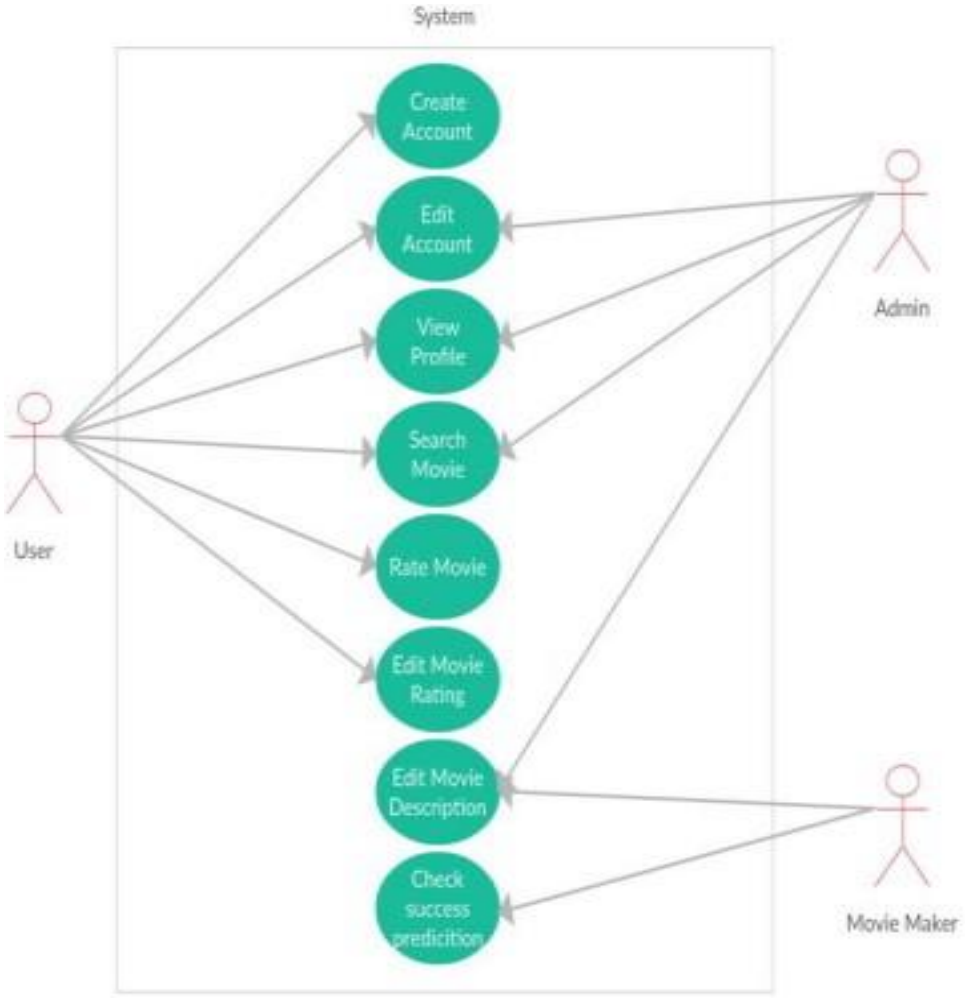

Fig. 6. Flowchart of Use Case

```
In [16]: get_recommendations('The Dark Knight Rises')

Out[16]: 65                                The Dark Knight
         299                                Batman Forever
         428                                Batman Returns
         1359                                       Batman
         3854    Batman: The Dark Knight Returns, Part 2
         119                                 Batman Begins
         2507                                    Slow Burn
         9              Batman v Superman: Dawn of Justice
         1181                                         JFK
         210                                Batman & Robin
         Name: title, dtype: object

In [17]: get_recommendations('The Avengers')

Out[17]: 7                   Avengers: Age of Ultron
         3144                                Plastic
         1715                                Timecop
         4124                     This Thing of Ours
         3311                  Thank You for Smoking
         3033                        The Corruptor
         588      Wall Street: Money Never Sleeps
         2136          Team America: World Police
         1468                           The Fountain
         1286                            Snowpiercer
         Name: title, dtype: object
```

Fig. 7.

```
pop= df2.sort_values('popularity', ascending=False)
import matplotlib.pyplot as plt
plt.figure(figsize=(12,4))

plt.barh(pop['title'].head(6),pop['popularity'].head(6), align='center',
        color='skyblue')
plt.gca().invert_yaxis()
plt.xlabel("Popularity")
plt.title("Popular Movies")
```

```
Text(0.5,1,'Popular Movies')
```

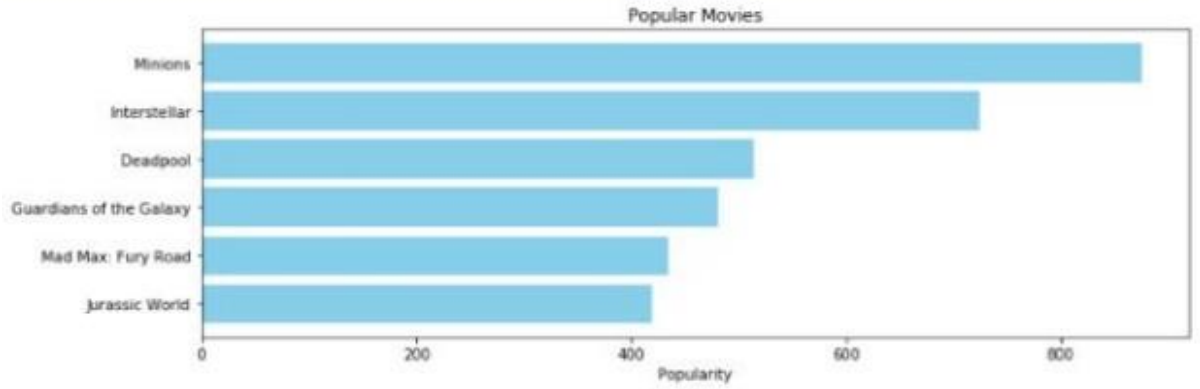

Fig. 8.

```
reader = Reader()
ratings = pd.read_csv('the-movies-dataset/ratings_small.csv')
ratings.head()
```

Out[32]:

	userId	movieId	rating	timestamp
0	1	31	2.5	1260759144
1	1	1029	3.0	1260759179
2	1	1061	3.0	1260759182
3	1	1129	2.0	1260759185
4	1	1172	4.0	1260759205

```
Note that in this dataset movies are rated on a scale of 5 unlike the earlier one.
```

In [34]:
```
data = Dataset.load_from_df(ratings[['userId', 'movieId', 'rating']], reader)
#data.split(n_folds=5)
```

In [36]:
```
svd = SVD()
#evaluate(svd, data, measures=['RMSE', 'MAE'])
cross_validate(svd, data, measures=['RMSE', 'MAE'], cv=5, verbose=True)
```

```
Evaluating RMSE, MAE of algorithm SVD on 5 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Mean    Std
RMSE (testset)    0.8995  0.8951  0.8919  0.8934  0.9028  0.8966  0.0040
MAE (testset)     0.6926  0.6907  0.6876  0.6879  0.6922  0.6902  0.0021
Fit time          10.85   15.22   12.99   12.19   15.55   13.36   1.79
Test time         0.26    0.46    0.14    0.48    0.40    0.35    0.13
```

Fig. 9.

```
In [29]: get_recommendations('The Dark Knight Rises', cosine_sim2)

Out[29]: 65                The Dark Knight
         119                 Batman Begins
         4638    Amidst the Devil's Wings
         1196                The Prestige
         3073           Romeo Is Bleeding
         3326              Black November
         1503                      Takers
         1986                      Faster
         303                     Catwoman
         747                Gangster Squad
         Name: title, dtype: object

In [30]: get_recommendations('The Godfather', cosine_sim2)

Out[30]: 867       The Godfather: Part III
         2731       The Godfather: Part II
         4638    Amidst the Devil's Wings
         2649           The Son of No One
         1525              Apocalypse Now
         1018           The Cotton Club
         1170     The Talented Mr. Ripley
         1209               The Rainmaker
         1394               Donnie Brasco
         1850                     Scarface
         Name: title, dtype: object
```

Fig. 10.

```
In [38]: ratings[ratings['userId'] == 1]
```

Out[38]:

	userId	movieId	rating	timestamp
0	1	31	2.5	1260759144
1	1	1029	3.0	1260759179
2	1	1061	3.0	1260759182
3	1	1129	2.0	1260759185
4	1	1172	4.0	1260759205
5	1	1263	2.0	1260759151
6	1	1287	2.0	1260759187
7	1	1293	2.0	1260759148
8	1	1339	3.5	1260759125
9	1	1343	2.0	1260759131
10	1	1371	2.5	1260759135
11	1	1405	1.0	1260759203
12	1	1953	4.0	1260759191
13	1	2105	4.0	1260759139
14	1	2150	3.0	1260759194
15	1	2193	2.0	1260759198
16	1	2294	2.0	1260759108
17	1	2455	2.5	1260759113
18	1	2968	1.0	1260759200
19	1	3671	3.0	1260759117

```
In [39]: svd.predict(1, 302, 3)

Out[39]: Prediction(uid=1, iid=302, r_ui=3, est=2.8519641281630195, details={'was_impossible': False})
```

Fig. 11. Result for Collaborative based Filtering Method

REFERENCES

[1] Bahuguna, A., Ashraf, A., Kavita, Verma, S., Negi, P. (2023). Brain Tumor Classification from MRI Scans. In: Hassanien, A.E., Castillo, O., Anand, S., Jaiswal, A. (eds) International Conference on Innovative Computing and Communications. ICICC 2023. Lecture Notes in Networks and Systems, vol 537. Springer, Singapore. https://doi.org/10.1007/978-981-99-3010-4_57

[2] Gupta, H., Kaur, A., Kavita, Verma, S., Rawat, P. (2023). Recognition of Handwritten Digits Using Convolutional Neural Network in Python and Comparison of Performance for Various Hidden Layers. In: Hassanien, A.E., Castillo, O., Anand, S., Jaiswal, A. (eds) International Conference on Innovative Computing and Communications. ICICC 2023. Lecture Notes in Networks and Systems, vol 537. Springer, Singapore. https://doi.org/10.1007/978-981-99-3010-4_58

[3] Dhetarwal, M., Ashraf, A., Verma, S., Kavita, Rawat, B. (2023). Employee Turnover Prediction Using Machine Learning. In: Hassanien, A.E., Castillo, O., Anand, S., Jaiswal, A. (eds) International Conference on Innovative Computing and Communications. ICICC 2023. Lecture Notes in Networks and Systems, vol 537. Springer, Singapore. https://doi.org/10.1007/978-981-99-3010-4_55

[4] Thind, R., Divya, K., Verma, S., Kavita, Kaur, N., Uniyal, V. (2023). Voice Email for the Visually Disabled. In: Hassanien, A.E., Castillo, O., Anand, S., Jaiswal, A. (eds) International Conference on Innovative Computing and Communications. ICICC 2023. Lecture Notes in Networks and Systems, vol 537. Springer, Singapore. https://doi.org/10.1007/978-981-99-3010-4_60

[5] Jain, Shikha, Navneet Kaur, Sahil Verma, Kavita, A. S. M. Sanwar Hosen, and Satbir S Sehgal. 2022. "Use of Machine Learning in Air Pollution Research: A Bibliographic Perspective" Electronics 11, no. 21: 3621. https://doi.org/10.3390/electronics11213621H.

[6] K. Hade, M. Maurya, A. Magar, D. B. Rao, and A. Professor, “Social Distancing Violation Alert System,” International Research Journal of Engineering and Technology, 2021.

[7] M. Robakowska, “The use of drones during mass events,” Disaster Emerg. Med. J, vol. 2, no. 3, pp. 129–134, 2017.

[8] J. Liu et al., "A Comprehensive Privacy-Preserving Federated Learning Scheme with Secure Authentication and Aggregation for Internet of Medical Things," in IEEE Journal of Biomedical and Health Informatics, doi: 10.1109/JBHI.2023.3304361.

[9] Ramisetty, Sowjanya; Anand, Divya; Kavita; Verma, Sahil; Jhanjhi, N. Z.; Masud, Mehedi; Baz, Mohammed, Computer Systems Science and Engineering 2023, 45(2), 1971-1983. https://doi.org/10.32604/csse.2022.021924

[10] Kaur, Ramanpreet, Divya Anand, Upinder Kaur, Sahil Verma, Kavita, Seok-Wook Park, A. S. M. Sanwar Hosen, and In-Ho Ra. 2023. "An Advanced Job Scheduling Algorithmic Architecture to Reduce Energy Consumption and CO2 Emissions in Multi-Cloud" Electronics 12, no. 8: 1810. https://doi.org/10.3390/electronics12081810

[11] Upadhyay, Shrikant, Mohit Kumar, Aditi Upadhyay, Sahil Verma, Kavita, A. S. M. Sanwar Hosen, In-Ho Ra, Maninder Kaur, and Satnam Singh. 2023. "Digital Image Identification and Verification Using Maximum and Preliminary Score Approach with Watermarking for Security and Validation Enhancement" Electronics 12, no. 7: 1609. https://doi.org/10.3390/electronics12071609J.

[12] Adda, “Economic activity and the spread of viral diseases: Evidence from high frequency data,” The Journal of Economics, vol. 131, no. 2, pp. 891–941.

[13] S. Rebelo, M. S. Trabandt, and Eichenbaum.

[14] X. Wang, “Deep Learning in Object Recognition, Detection, and Seg- mentation,” Foundations and Trends in Signal Processing, 2016.

[15] Y. C. Hou, M. Z. Baharuddin, S. Yussof, and S. Dzulkifly, "Social Dis- tancing Detection with Deep Learning Model," 2020 8th International Conference on Information Technology and Multimedia, vol. 2020, pp. 334–338, 2020.

[16] J. Redmon, S. Divvala, R. Kirchick, and A. Farhadi. [Online].

Available: http://pjreddie.com/yolo/

[17] R. Kirchick, T. Darrell, J. Donahue, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580–587, 2014.

[18] J. Redmon, R. Kirchick, A. Farhadi, and S. Divvala, "YOLO: Unified, real-time object detection," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788, 2016.

[19] I. A. Shah, Q. Sial, N. Z. Jhanjhi, and L. Gaur, "The Role of the IoT and Digital Twin in the Healthcare Digitalization Process: IoT and Digital Twin in the Healthcare Digitalization Process," Digital Twins and Healthcare: Trends, Techniques, and Challenges, pp. 20–34, 2023.

[20] N. Z. Jhanjhi, S. N. Brohi, N. A. Malik, and M. Humayun, "Proposing a hybrid rpl protocol for rank and wormhole attack mitigation using machine learning," 2020 2nd International Conference on Computer and Information Sciences (ICCIS), pp. 1–6, 2020.

[21] K. Hussain, S. J. Hussain, N. Jhanjhi, and M. Humayun, "SYN Flood Attack Detection based on Bayes Estimator (SFADBE) For MANET," in 2019 International Conference on Computer and Information Sciences (ICCIS), 2019, pp. 1–4.

[22] I. A. Shah, Q. Sial, N. Z. Jhanjhi, and L. Gaur, "Use Cases for Digital Twin," Digital Twins and Healthcare: Trends, Techniques, and Challenges, pp. 102–118, 2023.

[23] N. Ravi, "Securing VANET Using Blockchain Technology," 2021 4th International Conference on Signal Processing and Information Security (ICSPIS), vol. 23, pp. 80–83, 2021.

[24] N. S. Vishnu, PDF Malware Classifiers - A Survey, Future Directions and Recommended Methodology" in Security Handbook. USA: CRC Press.

[25] K. Srinivasan, L. Garg, D. Datta, A. A. Alaboudi, N. Z. Jhanjhi,

R. Agarwal, and A. G. Thomas, "Performance comparison of deep cnn models for detecting driver's distraction," Materials & Continua, vol. 68, no. 3, pp. 4109–4124, 2021.

[26] A. Almusaylim, Z. Jhanjhi, N. Z. Alhumam, and A, "Detection and mitigation of RPL rank and version number attacks in the internet of things: SRPL-RP," Sensors, vol. 20, no. 21, pp. 5997–5997, 2020.

[27] Abhishek P. Patil , Neelika Chakrabarti, A review into the evolution of HIPAA in response to evolving technological environments, Journal of Cybersecurity and Information Management, Vol. 4 , No. 2 : Special No.-RIDAPPH , (2020) : 5- 15 (Doi : https://doi.org/10.54216/JCIM.040201)

[28] Mukesh Soni , YashKumar Barot , S. Gomathi, A review on Privacy-Preserving Data Preprocessing, Journal of Cybersecurity and Information Management, Vol. 4 , No. 2 : Special No.-RIDAPPH , (2020) : 16-30 (Doi : https://doi.org/10.54216/JCIM.040202)

[29] Mustafa Tanriverdi, A Systematic Review of Privacy Preserving Healthcare Data Sharing on Blockchain, Journal of Cybersecurity and Information Management, Vol. 4 , No. 2 : Special No.-RIDAPPH , (2020) : 31-37 (Doi : https://doi.org/10.54216/JCIM.040203)

[30] Parth Rustagi , Rohit Sroa , Priyanshu Sinha , Ashish Sharma4 , Sandeep Tayal, HomeTec Software for Security Aspects of Smart Home Devices Based on IoT, Journal of Cybersecurity and Information Management, Vol. 5 , No. 1 , (2021) : 5-16 (Doi : https://doi.org/10.54216/JCIM.050101)

[31] Aldin Justin sundararaj , K. Martin Sagayam , Ahmed A. Elngar , A N Subash , B C Pillai, Design, development and performance estimation of 110 kW kinetic heating simulation facilities for material studies–Phase I, Journal of Cybersecurity and Information Management, Vol.

5 , No. 1 , (2021) : 17-28 (Doi : https://doi.org/10.54216/JCIM.050102)

[32] Kanika Sharma , Achyut Shankar , Prabhishek Singh, Information Security Assessment in Big Data Environment using Fuzzy Logic, Journal of Cybersecurity and Information Management, Vol. 5 , No. 1 , (2021) : 29-42 (Doi : https://doi.org/10.54216/JCIM.050103).

[33] Sennan, S., Somula, R., Luhach, A. K., Deverajan, G. G., Alnumay, W., Jhanjhi, N. Z., ... & Sharma, P. (2021). Energy efficient optimal parent selection based routing protocol for Internet of Things using firefly optimization algorithm. Transactions on Emerging Telecommunications Technologies, 32(8), e4171.

[34] Adeyemo, V. E., Abdullah, A., JhanJhi, N. Z., Supramaniam, M., & Balogun, A. O. (2019). Ensemble and deep-learning methods for two-class and multi-attack anomaly intrusion detection: an empirical study. International Journal of Advanced Computer Science and Applications, 10(9).

[35] Shafiq, M., Ashraf, H., Ullah, A., Masud, M., Azeem, M., Jhanjhi, N. Z., & Humayun, M. (2021). Robust Cluster-Based Routing Protocol for IoT-Assisted Smart Devices in WSN. Computers, Materials & Continua, 67(3).