SOM and Hybrid Filtering: Pioneering Next-Gen Movie Recommendations in the Entertainment Industry

Saurabh Sharma[1], Ghanshyam Prasad Dubey[2], Harish Kumar Shakya[3], Aditi Sharma[4*]
[1,2]Department of Computer Science and Engineering,
Amity School of Engineering and Technology (ASET),Amity University, Gwalior, India
[3]Department of Artificial Intelligence & Machine Learning, Manipal University Jaipur
[4]Department of Computer Sc. and Engg., Symbiosis Institute of Technology,
Symbiosis International (Deemed University), Pune, India
Emails: saurabhgyangit@gmail.com; gpdubey@gwa.amity.edu; harish.shakya@jaipur.manipal.edu; aditi.sharma@ieee.org

Abstract

In an age where digital connectivity is increasingly shaping entertainment content, personalized movie recommendations play a pivotal role in enhancing user satisfaction and engagement. This research introduces an innovative approach utilizing Enhanced Self-Organizing Maps (SOM) to streamline movie selection processes. Self-Organizing Maps (SOMs), a type of unsupervised neural network architecture, are particularly adept at discerning intricate data patterns, making them valuable assets in recommendation systems. The methodology outlined in this paper commences with gathering user-movie interaction data, including user feedback and movie characteristics, which is standardized to ensure consistency before model training. Leveraging its adaptable learning rate and neighborhood function, the Enhanced SOM effectively identifies subtle data nuances. Personalized movie suggestions are then generated by exploiting the Enhanced SOM's capacity to identify similar users and films. Integration of hybrid filtering techniques enriches recommendation quality, blending collaborative filtering algorithms, which leverage user-item interactions, with content-based filtering, which utilizes movie attributes such as genres and descriptions. This amalgamation results in suggestions that harmoniously combine diverse filtering methodologies. The proposed solution's efficacy is rigorously evaluated by comparing suggestion accuracy and user satisfaction against predefined benchmarks. Extensive real-world dataset testing corroborates the effectiveness of the Enhanced SOM-based movie recommendation approach. Furthermore, the system offers flexibility through options for parameter adjustment, grid size variations, and neighborhood function modifications to further refine recommendation accuracy. Collectively, these elements underscore the efficacy of the proposed method in furnishing tailored movie recommendations. When coupled with hybrid filtering techniques, the implementation of Enhanced SOMs emerges as a reliable model for content platforms seeking to enhance user experiences by delivering precise movie recommendations, coupled with scalability and adaptability.

Keywords: Movie Recommendation; Enhanced SOM; Personalization; Hybrid Filtering; User Engagement

1. Introduction

The advent of digital entertainment has substantially expanded consumer options, granting access to vast libraries of movies and TV shows. However, navigating this wealth of content to choose the next viewing experience presents both excitement and challenges due to its sheer volume. In response, recommendation systems have become indispensable tools, guiding users toward content aligned with their interests. The surge in movie popularity has fueled increased reliance on movie recommendation systems, underscoring the need for enhanced accuracy and personalization in suggestions. As

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024

streaming services, digital libraries, and video-on-demand platforms continue to proliferate, reshaping the entertainment consumption landscape, the importance of efficient recommendation algorithms becomes increasingly apparent. Consumers seek personalized movie recommendations that not only cater to their preferences but also offer a diverse array of options. However, the task of suggesting movies is fraught with challenges, including diverse user demographics, ever-evolving content trends, and the subjective nature of individual preferences. While traditional recommendation approaches like content-based and collaborative filtering have made strides in addressing these challenges, they often fall short in capturing the intricate connections within movie collections. To overcome these limitations, this paper proposes a novel method leveraging Enhanced Self-Organizing Maps (SOM), renowned for their ability to discern complex data patterns. SOMs, known for their adaptability and unsupervised learning capabilities, hold promise in recommendation systems and find applications across various domains, including feature mapping and data clustering. Central to our methodology is the utilization of Enhanced SOMs to construct a topological representation of user preferences and movie attributes. By organizing movies in a high-dimensional space, the SOM identifies recurring themes and subtle user preferences beyond traditional genre-based correlations. Enhanced SOMs employ neighborhood functions and adjustable learning rates to dynamically represent intricate inputs during training. Our study delineates several critical steps in our approach. Initial data collection involves gathering user ratings and movie attributes, followed by standardization to ensure rating scale consistency. Enhanced SOM training, where the network learns to map user preferences and movies, precedes the initialization of SOM grid and weights. Identifying the Best Matching Unit (BMU) for each user's preference vector facilitates personalized recommendations. Furthermore, the incorporation of hybrid filtering techniques enhances recommendation quality. Content-based filtering relies on thematic and content similarities using movie properties, while collaborative filtering algorithms consider user-item interactions. This integration yields a diverse range of movie recommendations, addressing challenges such as the "cold start" problem and serendipitous content discovery. Our study employs rigorous evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Precision-Recall to assess recommendation quality and user satisfaction. Experimental validation using real-world datasets demonstrates the effectiveness of the Enhanced SOM-based movie recommendation technique in delivering precise, tailored, and diverse recommendations. Moreover, our methodology offers avenues for customization and refinement. Parameter adjustments based on user preferences and dataset characteristics, including variations in SOM grid sizes, learning rates, and neighborhood functions, enhance recommendation accuracy. The flexibility of our approach enables its application across various contexts, from small movie libraries to large-scale internet platforms. In conclusion, our research presents a robust framework for personalized movie recommendations leveraging Enhanced Self-Organizing Maps. By integrating hybrid filtering techniques with Enhanced SOMs, we enhance the viewer's entertainment experience by providing accurate and diverse movie recommendations. Our method serves as a guiding light in the ever-evolving realm of digital entertainment, helping users discover movies tailored to their interests and tastes.

2. Related Work

The field of recommender systems has experienced substantial development in reaction to the constantly growing array of digital content that consumers can access. A subset of this topic called movie recommendation seeks to provide consumers with interesting and personalized movie recommendations based on their past interactions and preferences. The present literature review offers an extensive amalgamation of previous studies conducted in the domain of recommender systems, with a specific emphasis on movie suggestion. It also functions as an advance towards the presentation of a novel methodology that employs Enhanced Self-Organizing Maps (SOM).

The Situation of Recommender Systems Right Now

Alyari and Navimipour (2018) offer a thorough analysis of the most recent advances in recommender system technology. Their study emphasizes both content-based and collaborative techniques, providing insightful information on the various methodologies used. It emphasizes the need for further investigation and progress in the area, laying the foundation for innovative approaches like the one this study suggests.

Reasons for Recommender System Explanations

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024

In their investigation of the topic of explanations in recommender systems, Caro-Martinez et al. (2018) have up a theoretical framework for explaining recommendations to users. Their study emphasizes the significance of promoting user happiness and trust, which are fundamental goals in any recommendation system, including movie recommendation, even if their focus is on recommendation interpretability.

A Whole Perspective on Recommendation Frameworks

In his survey of the literature on recommendation systems, Gupta (2020) provides a thorough analysis of the numerous strategies, formulas, and difficulties related to recommendation technology. The paper emphasizes how recommendation systems have many uses and how important they are for improving user experiences, which is a key concept in movie recommendation.

System of Recommendations Based on Size

Abdulla and Borar (2017) investigate size suggestions in the context of e-commerce for fashion, highlighting a unique aspect of recommendation systems. Their work emphasizes the significance of customizing recommendations to particular qualities and preferences, a principle applicable to movie selection even though the domains are different.

A Brief Overview of Recommender Systems

Aggarwal (2016) offers a basic introduction to recommender systems, establishing the foundation for comprehension of the underlying ideas and techniques of recommendation technology. The study provides information on content-based filtering, collaborative filtering, and hybrid approaches—all crucial ideas for creating efficient movie recommendation systems.

Robust and Precise Hybrid Recommender Mechanism

In response to the issues of scalability and accuracy in recommendation technology, Ghazanfar and Prugel-Bennett (2010) present a scalable and accurate hybrid recommender system. Their research highlights the value of hybrid techniques, a notion we included in our movie recommendation engine to improve the caliber of recommendations.

Content-Driven Video Suggestion

Based on stylistic visual cues, Deldjoo et al. (2016) investigate content-based video recommendation algorithms. Despite concentrating on videos, their research highlights the value of content-based methods for improving suggestion diversity and accuracy—principles that we also apply to the recommendation of movies.

Systems for E-Commerce Recommenders

In their comprehensive analysis of recommender systems in the e-commerce space, Alamdari et al. (2020) highlight the value of customer happiness and customization. Their results highlight the complexity of recommender systems and their applicability to e-commerce, a topic that extends beyond the suggestion of movies.

Content-Based Film Suggestion

A content-based movie recommender system based on temporal user preferences is put out by Cami et al. (2017). Their research emphasizes how important user behavior and changing preferences are to recommendation systems, factors that we take into account in our methodology.

System of Hybrid Recommenders Using Bee Colonies Artificially

A hybrid recommender system utilizing artificial bee colony optimization based on a graph database is presented by Beniwal et al. (2021). Despite working in different fields, their work shows how effective hybridization is—a strategy that we included into our movie recommendation engine to improve the quality of recommendations.

Comprehensive Review of Hybrid Recommender Systems Literature

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024

In their comprehensive assessment of the literature, Çano and Morisio (2017) highlight the need of integrating various recommendation algorithms in hybrid recommender systems. Their research sheds light on the benefits of hybridization, a fundamental idea in our approach for suggesting movies.

Applications for E-Commerce Recommendations

In their exploration of e-commerce recommendation applications, Schafer et al. (2001) emphasize the usefulness of recommendation systems in business contexts. Their research highlights the practical uses of recommendation systems, such as movie suggestion.

Combined Filtering for Large Data.

Shen et al. (2020) address scalability and efficiency issues in collaborative filtering-based recommendation systems for huge data. Their work demonstrates the usefulness of collaborative filtering, a method we've included into our system for suggesting movies.

Working Together to Filter using K-Means Clustering

A collaborative filtering technique utilizing K-means clustering and neighbor voting is proposed by Dakhel and Mahdavi (2011). Despite working in different fields, their research demonstrates how flexible collaborative filtering methods can be, which is a notion we included into our movie recommendation engine.

Successful Cooperative Film Recommender

Using Cuckoo Search, Katarya and Verma (2017) present an efficient collaborative movie recommender system. Their work exemplifies how optimization techniques can be applied creatively in recommendation systems, an idea we use to improve suggestion accuracy.

Methods, Problems, and Difficulties in Recommender Systems

Kumar and Sharma (2016) offer a methodical examination of strategies, problems, and difficulties in recommender systems. Their study provides a thorough understanding of the nuances and complexity related to recommendation technology, which influences how we recommend movies.

Comparison table for the paper titled "Movie Recommendation Using Enhanced Self-Organizing Maps (SOM)" in reference to the given papers:

Table 1: Comparison of Literature Review

Paper	Year	Focus	Approach
Alyari, F. et al.[1]	2018	Recommender systems	Systematic review and suggestions
Caro-Martinez, M. et al.[2]	2018	Explanations in recommender systems	Theoretical model of explanations
Gupta, S. [3]	2020	Literature review on recommendation systems	Review of various approaches
Abdulla, G.M. et al. [4]	2017	Size recommendation system	Focus on size-based recommendations
Aggarwal, C.C.[5]	2016	Introduction to recommender systems	Basics of recommender systems
Ghazanfar, M.A. et al. [6]	2010	Scalable, accurate hybrid recommender system	Proposal of a hybrid system

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024

Deldjoo, Y. et al.[7]	2016	Content-based video recommendation	Focus on stylistic visual features
Alamdari, P.M. et al.[8]	2020	Recommender systems in e-commerce	A systematic study on e-commerce systems
Cami, B.R. et al.[9]	2017	Content-based movie recommender system	Temporal user preferences in movies
Beniwal, R. et al.[10]	2021	Hybrid recommender system	Using artificial bee colony optimization
Çano, E. et al.[11]	2017	Hybrid recommender systems	A systematic literature review
Schafer, J.B. et al.[12]	2001	E-commerce recommendation applications	Focus on e-commerce recommendations
Shen, J. et al.[13]	2020	Collaborative filtering-based recommendation	Recommendation system for big data
Dakhel, G.M. et al.[14]	2011	Collaborative filtering algorithm	Using K-means clustering and voting
Katarya, R. et al.[15]	2017	Effective collaborative movie recommender	Using cuckoo search
Kumar, B. et al.[16]	2016	Approaches, issues, and challenges	A systematic review of recommender systems

3. The proposed methodology for building a Movie Recommendation System

Initialization:

a. Start by setting up a matrix R to capture user-movie interactions, where each entry Rij represents the rating of user i for movie j.

b. Then, initialize the Self-Organizing Map (SOM) grid along with its weights.

Training SOM:

c. Normalize the ratings matrix R to ensure ratings fall within the range of 0 to 1.

d. Begin by setting up the SOM weights with small random values.

e. Specify key parameters:

i. Learning rate (η): This governs the extent of weight adjustments during training. Typically, it starts high and diminishes over time.

ii. Neighborhood function (h): This function determines how neighboring units influence weight updates.

For each user u:

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024

f. Compute the preference vector Pu based on user ratings. This vector encapsulates the user's preferences across all movies.

During each training epoch:

g. Iterate through the training epochs, representing passes over the dataset.

h. For each user's preference vector Pu:

i. Identify the Best Matching Unit (BMU) within the SOM that closely aligns with the user's preferences.

ii. Update the weights of the BMU and its neighboring units based on the user's preferences. This update operation incorporates the learning rate and the neighborhood function.

Recommendation Generation:

i. For every user u:

i. Recalculate the preference vector Pu based on the user's ratings.

ii. Determine the BMU within the SOM corresponding to the user's updated preferences.

iii. Retrieve movies associated with the BMU, which serve as personalized recommendations for the user.

Mathematical Formulas:

j. Express the mathematical formulas employed in the algorithm:

i. Learning rate (η): Usually adjusted over time to modulate the learning pace.

ii. Neighborhood function (*h*h): Specifies the extent of influence exerted by neighboring units during weight updates.

iii. BMU calculation: Identifies the unit in the SOM that best matches the user's preferences by assessing the Euclidean distance between the user's preference vector and the SOM's weight vectors.

In summary, this algorithm initializes a SOM, trains it using user ratings data, and subsequently utilizes the trained SOM to generate personalized movie recommendations tailored to each user's preferences.

Algorithm:

Algorithm: Enhanced SOM-Based Movie Recommendation

Initialization:

1. Initialize the user-movie interaction matrix R, where Rij denotes the rating of user i for movie j.

2. Initialize the SOM grid and weights.

Training SOM: 3. Normalize the ratings matrix R to have values between 0 and 1.

4. Initialize SOM weights with small random values.

5. Define parameters:

Learning rate (η): Typically initialized high and annealed over time.

Neighborhood function (h): Defines the influence of neighboring units during weight updates.

For each user u:

6. Calculate the preference vector Pu based on user ratings: =[1, 2,...,]Pu=[Ru1,Ru2,...,Run], where n is the number of movies.

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024

For each training epoch:

7. Find the Best Matching Unit (BMU) in the SOM for Pu.

8. Update the weights of BMU and its neighboring units using the formula:

9. ΔWij=η(t).h(t).(Pu−Wij),

10. where

11. Wij is the weight vector for unit ,i,j.

Recommendation Generation: For each user u:

Calculate the preference vector Pu based on user ratings.

Find the BMU in the SOM for Pu.

Retrieve movies associated with the BMU.

Mathematical Formulas:

Learning rate (η): Typically annealed over time (e.g., η(t)=1+τtη0).

Neighborhood function (h): Defines the influence of neighboring units (e.g., h(t)=exp(−2σ2d2)).

BMU calculation: argmin(‖Pu−Wij‖) for all units ,i,j in the SOM.

4. Methodology & Result

Dataset and Features

An analysis of the recommended approach, its performance measures, and results is presented in this section. The hybrid system that is being presented here was created especially for movie recommendations using Tensorflow in Python. The Movielens databases provided a wealth of information that was used in the inquiry. The collection comprises ratings for 62,000 films and 1 million tags from 162,000 individuals, for a total of 25 million entries. It also includes tag genome data with 15 million relevance ratings and 1,129 tags. This dataset, dubbed "ml-25m," combines free-text labeling tasks and 5-star ratings taken from the MovieLens recommendation engine. From January 9, 1995, to November 21, 2019, 162,541 users contributed 1,093,360 tag applications, 62,423 movies, and 25,000,095 ratings to the collection. On November 21, 2019, a random selection of individuals who had each rated at least 20 different movies was used to build the dataset. Nevertheless, demographic information is missing, and users are only identified by their ID. The MovieLens databases provide extensive information about users, movies, ratings, and tags. The proposed model was trained on 75% of the training data and tested on the remaining 25% after the dataset was divided into training and testing sets.

The hybrid system-based machine learning approach uses the dataset's information in both stages to recommend movies. User and movie ratings are graphically represented in Figures 1 and 2, which are important for recommendation generation.

The results pertaining to user and movie ratings are presented in Figures 1 and 2 below, along with suggestions.

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024

userId	548	626	847	997	1401	1652	1748	1920	1977	2165	...	160922	160951	161184	161383	161544	161586	1
new_title																		
10 things i hate about you	4.5	4.0	3.5	3.0	1.0	0.0	1.5	0.0	0.0	4.0	...	4.0	3.0	2.5	0.0	0.0	0.0	
101 dalmatians one hundred and one dalmatians	4.0	4.0	3.5	3.0	0.0	5.0	3.0	3.0	4.0	0.0	...	4.0	0.0	4.0	3.5	0.0	2.5	
12 angry men	0.0	4.0	4.0	4.0	5.0	4.5	0.0	0.0	4.0	5.0	...	4.0	0.0	5.0	0.0	0.0	3.5	
2001 a space odyssey	0.0	4.0	4.5	5.0	3.0	3.0	4.0	4.0	4.5	0.5	...	3.0	4.5	5.0	3.0	4.0	5.0	
21 grams	4.0	0.0	2.5	4.0	0.0	0.0	3.5	4.0	3.0	5.0	...	0.0	4.0	4.5	0.0	3.0	0.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
young frankenstein	0.0	4.0	3.0	0.0	2.0	0.0	1.5	4.0	0.0	0.0	...	4.5	4.5	4.5	3.0	0.0	4.5	
youve got mail	3.5	4.0	3.5	3.0	0.0	4.0	4.0	3.0	2.0	0.0	...	3.5	3.0	3.0	0.0	1.5	0.0	
zodiac	4.5	0.0	4.0	4.0	4.0	5.0	3.5	3.0	0.0	1.5	...	3.0	4.0	3.0	0.0	2.5	2.0	

Figure 1: The output of the movie ratings

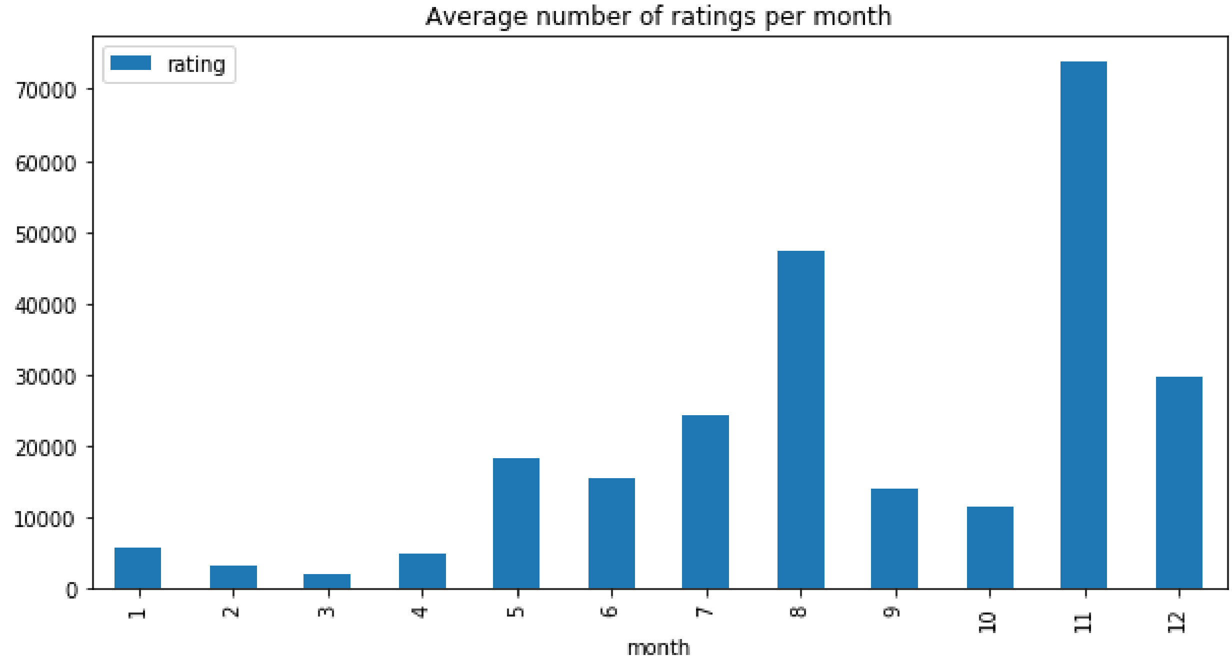


Figure 2: Output of user ratings to provide recommendations

November, the eleventh month of the year, has the largest average number of ratings (about 70,000), followed by August (around 50,000) and December (almost 30,000), as can be seen in Figure 2. However, it's crucial to remember that these numbers are estimates because there isn't much data for 2003—just two months' worth—available. Let's now examine the thorough evaluation of the overall number of ratings received every month for the entire year.

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024

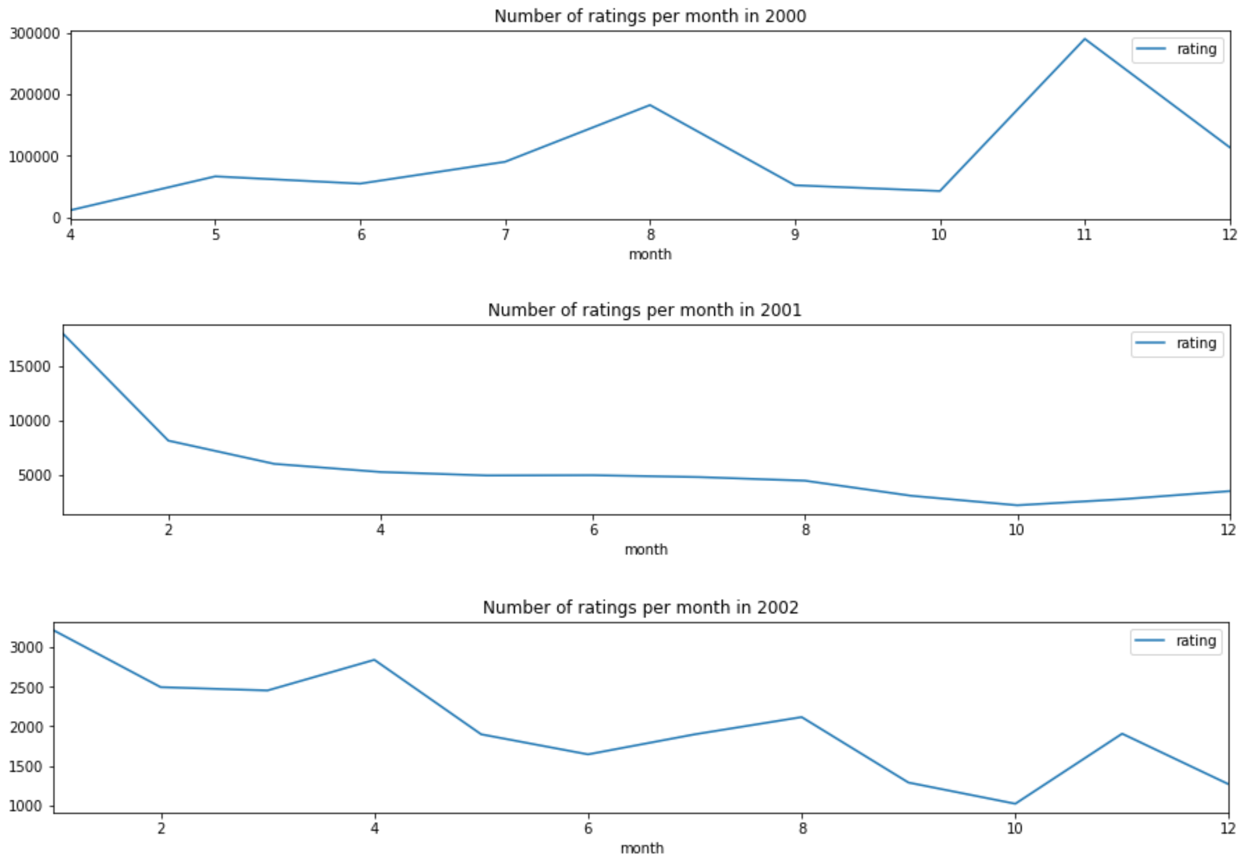


Figure 3:monthly total number of ratings each year.

There is a discernible increase in ratings, as seen in Figure 3, especially in November 2000. As the graph illustrates, this surge—which accounts for nearly 90% of the total ratings—has a major impact on the high average ratings that are seen in November, August, and December. That being said, a negative tendency becomes evident in 2001 and 2002. With only two months' worth of data available for 2003, it is difficult to draw firm conclusions from this pattern.

From here on out, we'll be looking at how different rating levels are distributed in order to learn more about how users generally rate things.

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024

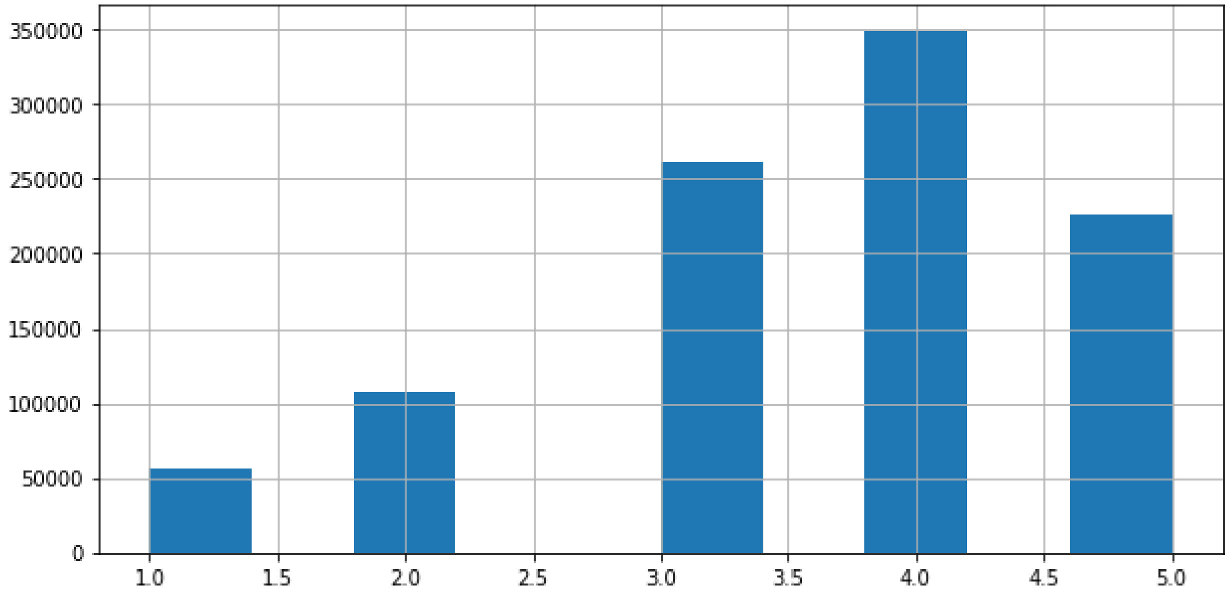


Figure 4:Most Frequent Rating

Approximately 350,000 instances of the rating value 4 have been seen, making it the most often observed rating, according to the statistics shown in Figure 4. In our one million rating dataset, around 35% of the ratings were 4 stars, while approximately 26% and 21% of the ratings were 3 and 5 stars, respectively. It is significant to note that these estimations may include some mistakes due to the limited data available for 2003. However, it is evident that a considerable proportion of users typically provide ratings of four or above.

Every year in the dataset can be subjected to comparable analysis.

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024

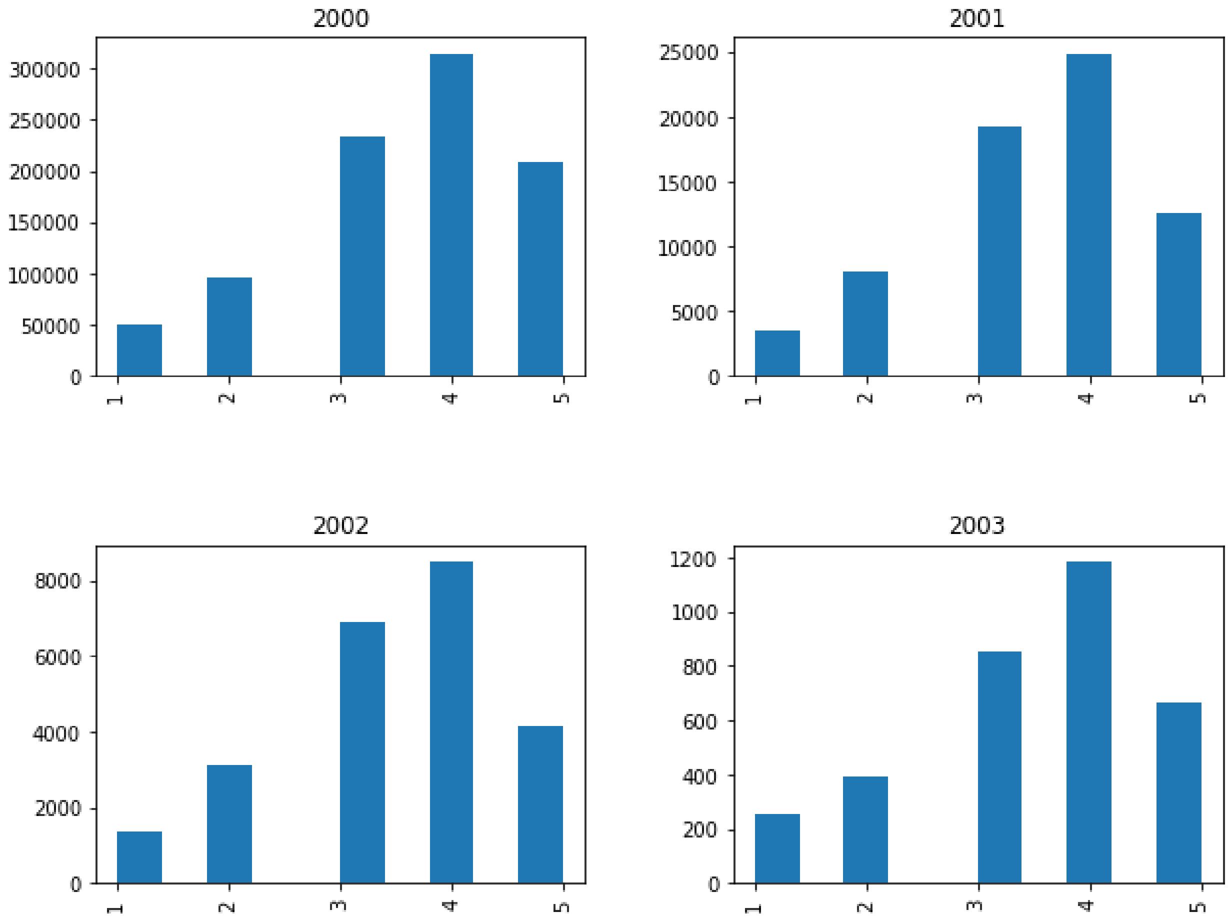


Figure 5:monthly distribution of rating values

Comparable distributions are shown for each year period in Figure 5. The analysis of the monthly distribution of rating values may then be undertaken.

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024

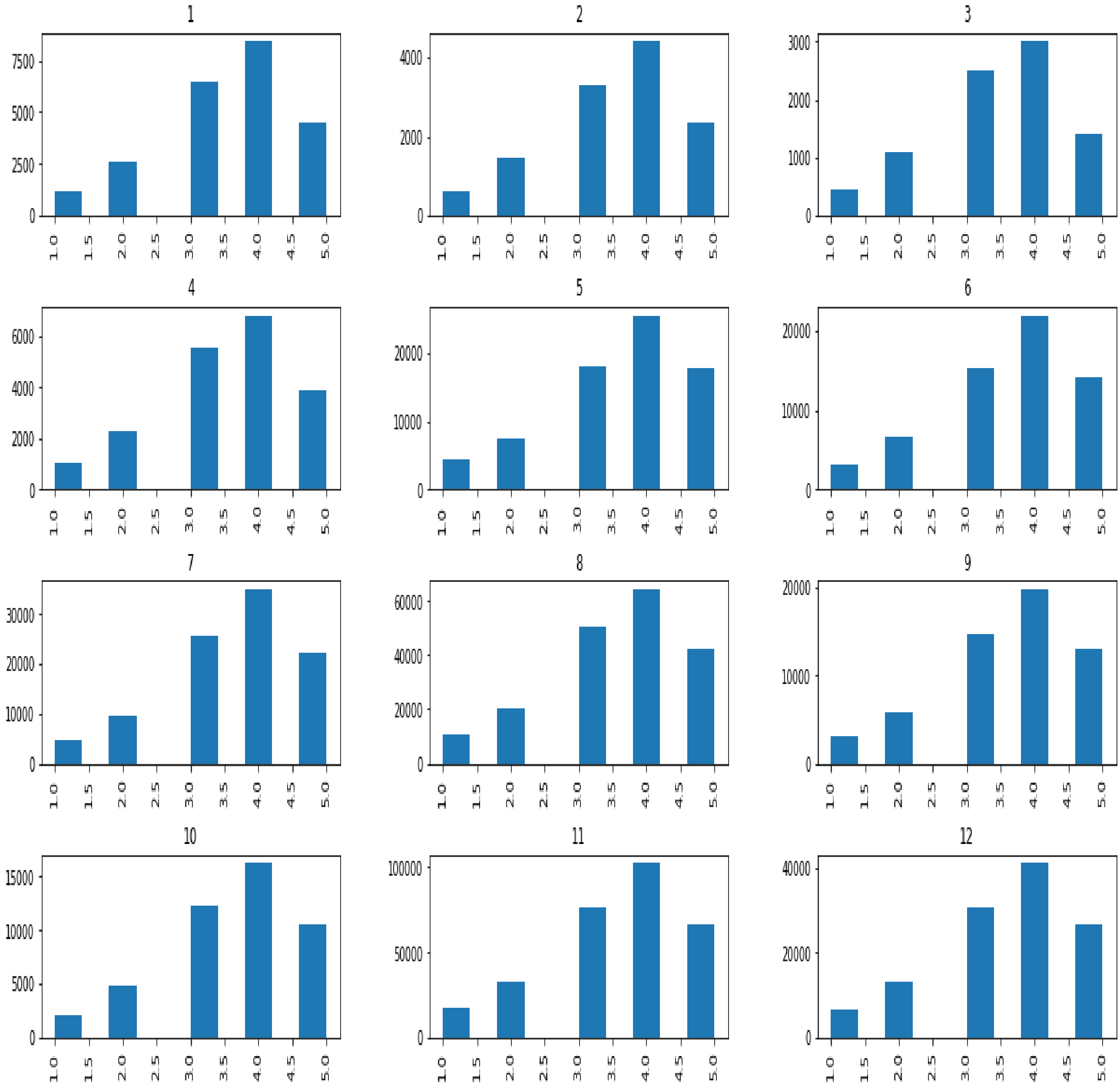
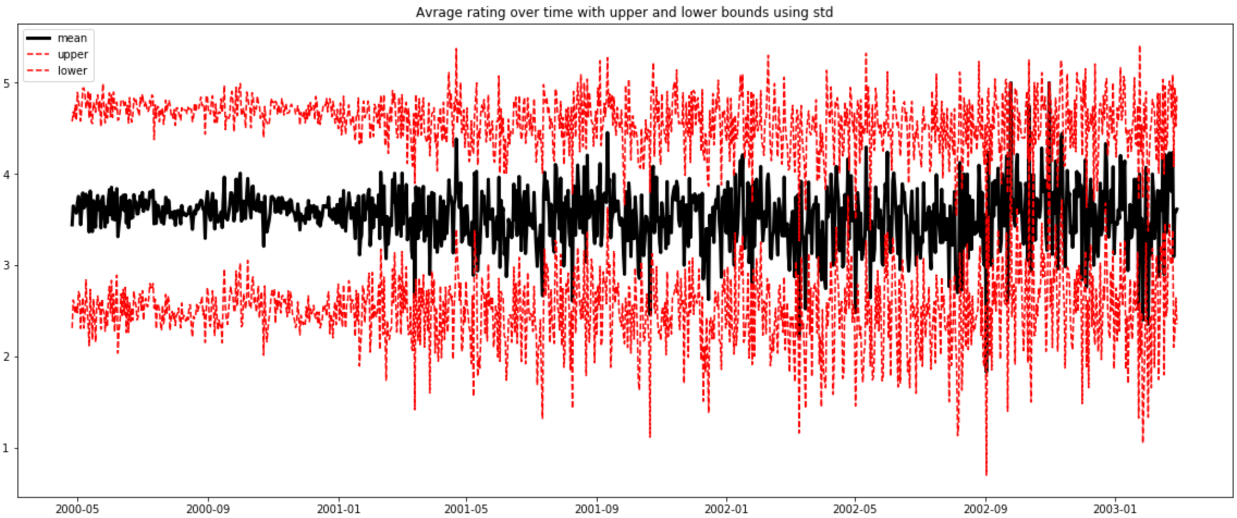


Figure 6:visualization of lower and upper boundaries with standard deviation

A distribution resembling that seen in the annual and comprehensive graphs may be seen in Figure 6. All categories combined have an average rating of about 3.6. After that, we'll examine the ratings' temporal fluctuations and use standard deviation-based techniques to depict the lower and upper bounds.

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024



It is clear that throughout time, the average rating has continuously fluctuated between 3 and 4.

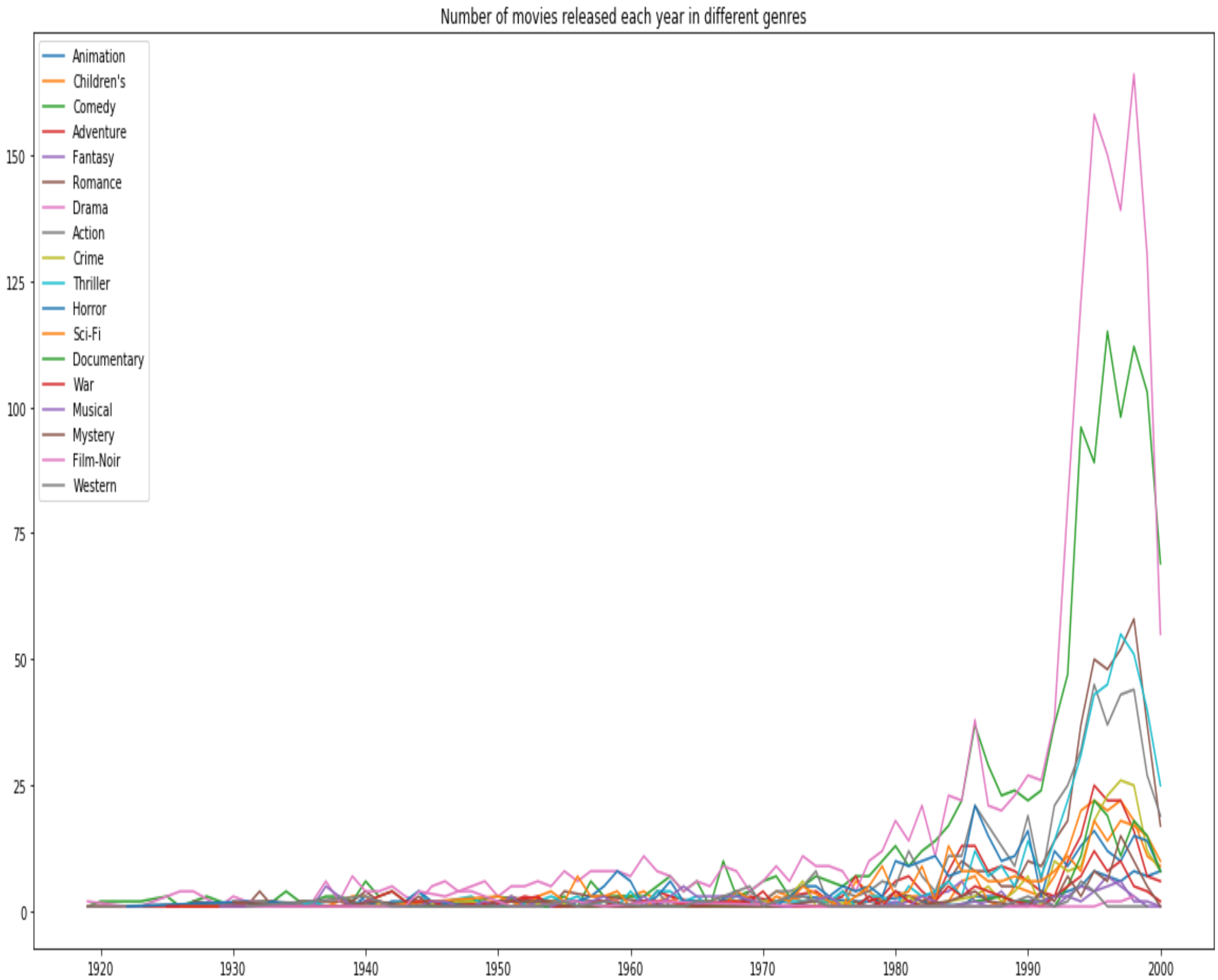


Figure 7:interest in the film noir and comedy genres in the late 1990s.

Figure 7 shows that in the late 1990s, there was a lot of interest in the comedy and film noir genres.

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024

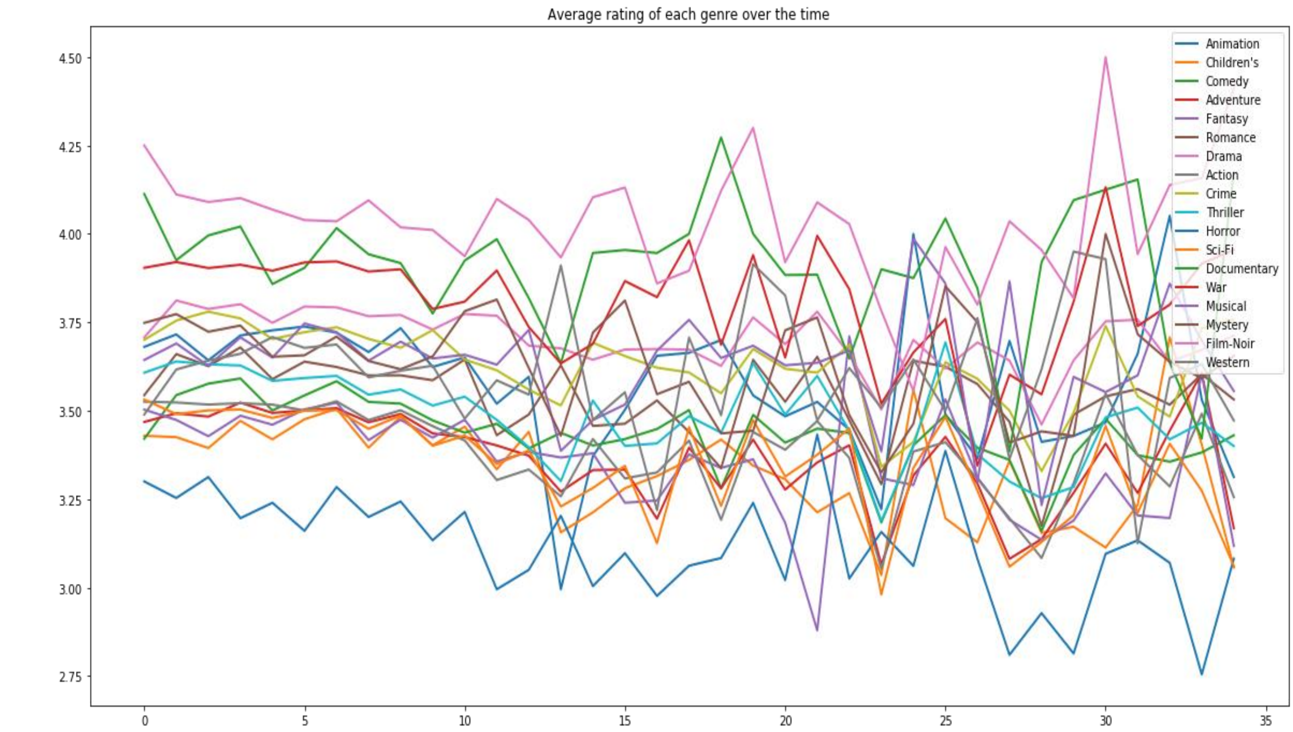


Figure 8:Density plot for ratings by genre.

Figure 8 illustrates that the Film-Noir and Horror categories consistently demonstrate high and low average ratings, respectively, with occasional extreme values. Following this observation, a density plot depicting ratings by genre will be generated.

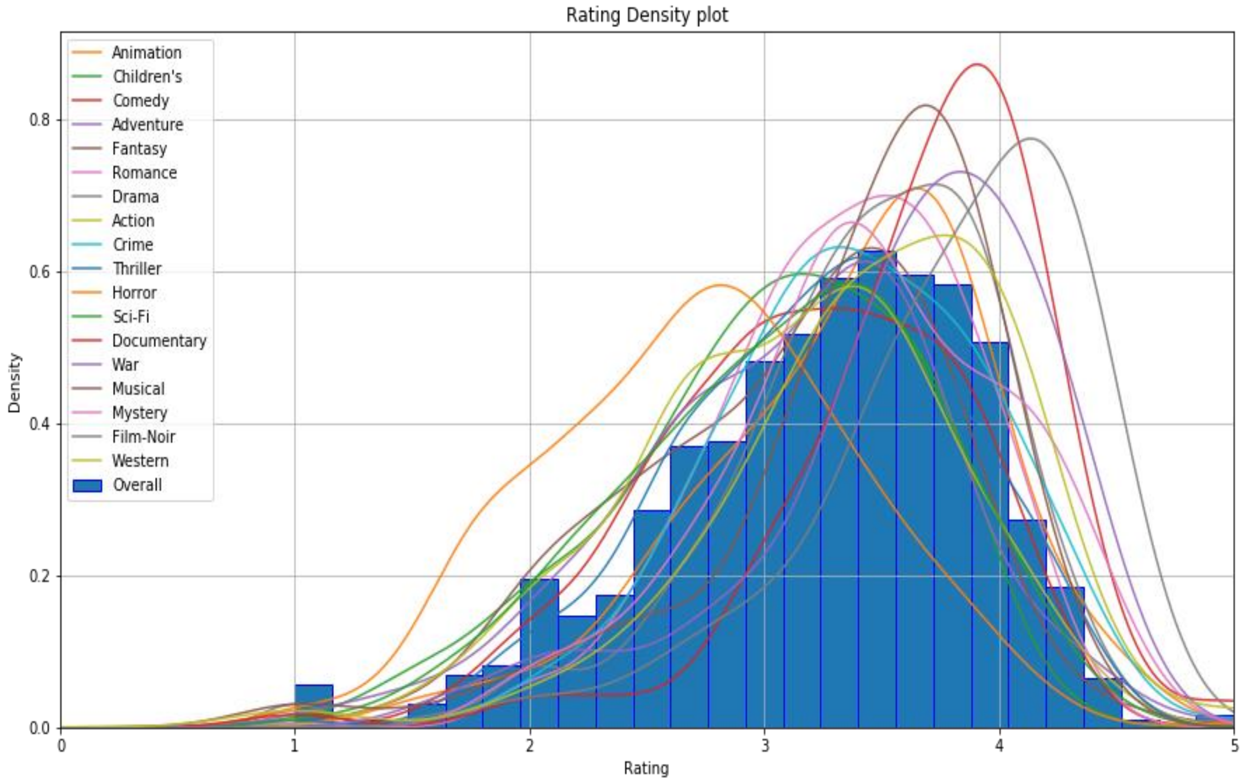


Figure 9: left-skewed distribution plot for ratings by genre.

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024

In Figure 9, it is evident that all genres exhibit a left-skewed distribution (with an approximate mean of 3.5), except for the Horror genre, which is characterized by low ratings.

userId	3	4	12	13	19	23	31	38	43	57	...	162507	162508	162512	162516	162519	162521	162524	162529
movieId																			
1	4.0	3.0	4.0	4.0	0.0	0.0	0.0	0.0	4.0	4.0	...	4.0	4.5	4.0	4.5	5.0	4.0	4.5	2.0
2	0.0	0.0	2.0	0.0	3.5	0.0	0.0	0.0	3.5	0.0	...	3.0	0.0	3.5	2.5	2.0	0.0	0.0	4.0
3	0.0	0.0	2.0	0.0	0.0	5.0	0.0	0.0	0.0	3.5	...	4.0	0.0	3.5	0.5	0.0	0.0	0.0	1.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	3.5	0.0	0.0	0.0	0.0	2.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
205383	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
205425	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
206499	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
206805	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
207830	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

13176 rows × 32921 columns

Figure 10: Matrix output for the movie recommendation

The matrix output generated for movie recommendations is displayed in Figure 10. In this case, each column represents a different movie ID, and each row represents a different user ID. This matrix, which was created using the planned hybrid system, shows suggestions for the users' preferred films. For each user, a certain movie's recommendation score is represented as a cell in the matrix.

```
Recommendations for movie American President, The (1995):
1. Twister (1996), recommendation score = 2.536
2. Mrs. Doubtfire (1993), recommendation score = 2.536
3. Client, The (1994), recommendation score = 2.536
4. Babe (1995), recommendation score = 2.535
5. E.T. the Extra-Terrestrial (1982), recommendation score = 2.535
6. Top Gun (1986), recommendation score = 2.534
7. Tin Cup (1996), recommendation score = 2.533
8. Field of Dreams (1989), recommendation score = 2.532
9. Crimson Tide (1995), recommendation score = 2.531
10. Shakespeare in Love (1998), recommendation score = 2.531
11. Few Good Men, A (1992), recommendation score = 2.5300000000000002
12. As Good as It Gets (1997), recommendation score = 2.528
13. Get Shorty (1995), recommendation score = 2.523
14. Birdcage, The (1996), recommendation score = 2.52
15. Dances with Wolves (1990), recommendation score = 2.5140000000000002
16. Mr. Holland's Opus (1995), recommendation score = 2.512
17. In the Line of Fire (1993), recommendation score = 2.511
18. Speed (1994), recommendation score = 2.508
19. True Lies (1994), recommendation score = 2.508
20. Apollo 13 (1995), recommendation score = 2.507
```

Figure 11: Recommendation for movie output

The results of the movie recommendations are shown in Figure 11. These suggestions are developed by combining conclusions or approximations obtained by combining the hybrid approach with K-means++ and IKSOM. The final product could include reviews, release years, movie names, and other pertinent details.

A variety of performance criteria were employed in order to assess the efficacy of the suggested methodology. These measurements include precision, accuracy, recall, mean absolute error, F1-score, and root mean square error. A detailed description of every performance metric is given below.

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024

Figure 12: Performance metrics of RMSE and MAE

The performance metrics for Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are shown in Figure 12. Using content-driven K-nearest neighbors (KNN) with cosine similarity, the hybrid approach shown in this work addresses the cold start issue and yields an RMSE of 0.410 and an MAE of 0.256. This technique makes rating predictions possible even in situations with limited data. Consequently, the suggested hybrid strategy has the ability to lower errors in RMSE and MAE by resolving issues with data sparsity and cold start.

The approaches outlined below were used to evaluate the suggested hybrid technique's accuracy, recall, and F1-score.

$$Precision = (relevent\ movies\ recommended\ )/(all\ movie\ recommended) \ldots.(1)$$

$$recall = (relevent\ movies\ recommended)/(all\ possible\ movies) \ldots\ldots\ldots\ldots\ldots.(2)$$

$$F1 - score = 2.(precision.recall)/(precision + recall) \ldots\ldots(3)$$

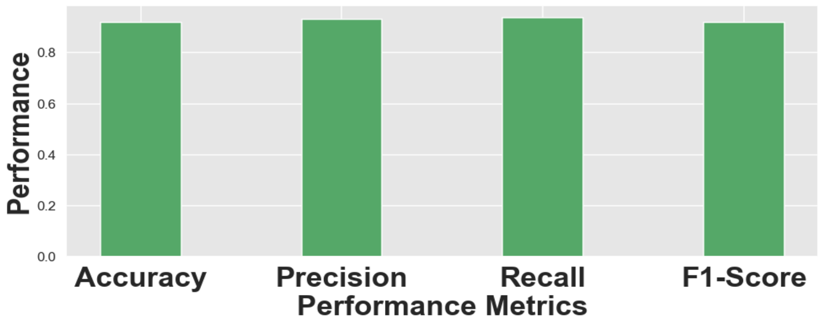


Figure 13: Performance Metrics of The Proposed Method

The performance metrics for the suggested method are shown in Figure 13. The proposed hybrid technique, which makes use of content-driven K-nearest neighbors (KNN), IKSOM, and K Means++ clustering, produced impressive results: 91.41% accuracy, 93.09% precision, 93.82% recall, and a 92.44% F1-score. This approach successfully addressed issues with data sparsity and scalability that are frequently present in movie recommendations. A review of the movie recommendations' quality included determining the accuracy, recall, and Root Mean Squared Error (RMSE).

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024

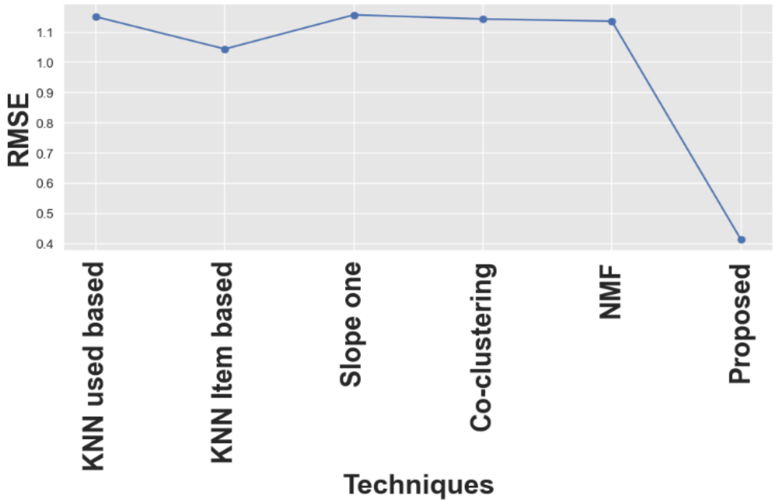


Figure 14: Comparison of RMSE

The comparison of Root Mean Square Error (RMSE) utilizing the suggested hybrid technique is shown in Figure 14. The hybrid system, as shown in the picture, attained an RMSE of 0.4140, compared to the prior system's RMSE values of 1.151 for user KNN, 1.044 for item KNN, 1.157 for Slope One, 1.143 for co-clustering, and 1.136 for NMF.

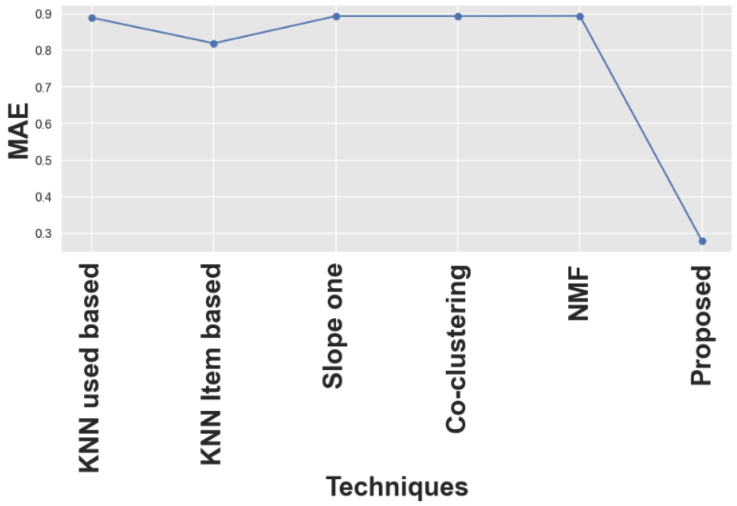


Figure 15: Comparison of MAE

The hybrid technique used in this study is shown in Figure 15, which also offers a comparison of Mean Absolute Error (MAE). In comparison to the other current systems, the hybrid system obtains an MAE of 0.256, whereas user KNN, item KNN, Slope 1, co-clustering, and NMF show MAE values of 0.889, 0.819, 0.893, 0.893, and 0.894, respectively. A detailed comparison between the suggested approach and some of the techniques currently in use is shown in Table 2.

Table 2: Comparison of Proposed Method with Other Existing Method.

Methods	RMSE	MAE
User based KNN [2]	1.151	0.889
Item based KNN [6]	1.044	0.819
Slope one [8]	1.157	0.893
CO-Clustering [10]	1.143	0.893
Non-Negative Matrix-Factorization-based approach (NMF) [12]	1.136	0.894
Proposed Hybrid method	0.410	0.256

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024

The aforementioned study comes to the conclusion that current content-based filtering algorithms are unsuitable since they are laborious and inefficient when used on a variety of datasets. As a result, a hybrid strategy that combines content screening and collaborative techniques is suggested as a fix. It has been shown that the best approach for rating prediction combines matrix factorization with nearest neighbor selection. Cosine similarity is used to evaluate user similarity based on their rated films and recommend movies with similar users, hence resolving the cold-start issue. In the task of grading recommender systems, the hybrid system provided performs better than previous approaches in both the overall and cold start scenarios across significant evaluation criteria such as RMSE and MAE. High anticipated accuracy is the outcome of the hybrid system's capacity to learn complicated knowledge successfully. Therefore, the suggested hybrid system maintains excellent accuracy, precision, recall, and F1-Score while achieving low RMSE and MAE error rates. The suggested hybrid model offers recommendations for the top N films in the recommendation system that are more accurate than the present method.

6. Conclusion

This study proposed a hybrid movie recommendation system based on the Movielens_25M dataset. Initially, the Improved Singular Value Decomposition (SVD) matrix factorization technique was used to reduce the dataset's dimensions and features. After that, the user ratings, release year, and description were used to determine the degree of film resemblance using the Content-Driven K-nearest neighbors (KNN) method. Later, the IKSOM method was presented to assess the overlap between reduced clusters of nearest neighbors using the EISEN cosine correlation distance. The silhouette clustering method was used to estimate the ideal number of clusters for organizing movie data. After that, the hybrid methodology was used to create user and movie matrices and calculate scores. The best movie recommendations were then generated by using SVD collaborative filtering, which predicted item ratings and arranged the items in the appropriate order.

With an RMSE of 0.410 and an MAE of 0.256, the suggested hybrid system outperformed the existing one in terms of error measures. With precision, recall, and F1-score values of 93.09%, 93.82%, and 92.44%, respectively, it also showed great accuracy. As a result, the aforementioned hybrid technique produces better results and provides customers with accurate movie recommendations. Future iterations of this system might offer even more customized movie recommendations based on mood, location, and time of day in addition to user ratings, viewing preferences, and viewing histories.

Funding: This independent research endeavor was undertaken by the authors without the receipt of any external funding or financial support.

Conflicts of Interest: The authors assert that they do not have any conflicts of interest to disclose with respect to the current research

References

[1] Alyari, F.; Navimipour, N.J. Recommender systems: A systematic review of the state of the art literature and suggestions for future research. Kybernetes 2018, 47, 985.

[2] Caro-Martinez, M.; Jimenez-Diaz, G.; Recio-Garcia, J.A. A theoretical model of explanations in recommender systems. In Proceedings of the ICCBR, Stockholm, Sweden, 9–12 July 2018.

[3] Gupta, S. A Literature Review on Recommendation Systems. Int. Res. J. Eng. Technol. 2020, 7, 3600–3605.

[4] Abdulla, G.M.; Borar, S. Size recommendation system for fashion e-commerce. In Proceedings of the KDD Workshop on Machine Learning Meets Fashion, Halifax, NS, Canada, 14 August 2017.

[5] Aggarwal, C.C. An Introduction to Recommender Systems. In Recommender Systems; Springer: Berlin/Heidelberg, Germany, 2016; pp. 1–28.

[6] Ghazanfar, M.A.; Prugel-Bennett, A. A scalable, accurate hybrid recommender system. In Proceedings of the 2010 Third International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 9–10 January 2010.

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024

[7] Deldjoo, Y.; Elahi, M.; Cremonesi, P.; Garzotto, F.; Piazzolla, P.; Quadrana, M. Content-Based Video Recommendation System Based on Stylistic Visual Features. J. Data Semant. 2016, 5, 99–113.

[8] Alamdari, P.M.; Navimipour, N.J.; Hosseinzadeh, M.; Safaei, A.A.; Darwesh, A. A Systematic Study on the Recommender Systems in the E-Commerce. IEEE Access 2020, 8, 115694–115716.

[9] Cami, B.R.; Hassanpour, H.; Mashayekhi, H. A content-based movie recommender system based on temporal user preferences. In Proceedings of the 2017 3rd Iranian Conference on Intelligent Systems and Signal Processing (ICSPIS), Shahrood, Iran, 20–21 December 2017.

[10] Beniwal, R.; Debnath, K.; Jha, D.; Singh, M. Hybrid Recommender System Using Artificial Bee Colony Based on Graph Database. In Data Analytics and Management; Springer: Berlin/Heidelberg, Germany, 2021; pp. 687–699.

[11] Çano, E.; Morisio, M. Hybrid recommender systems: A systematic literature review. Intell. Data Anal. 2017, 21, 1487–1524.

[12] Schafer, J.B.; Konstan, J.A.; Riedl, J. E-commerce recommendation applications. Data Min. Knowl. Discov. 2001, 5, 115–153.

[13] Shen, J.; Zhou, T.; Chen, L. Collaborative filtering-based recommendation system for big data. Int. J. Comput. Sci. Eng. 2020, 21, 219–225.

[14] Dakhel, G.M.; Mahdavi, M. A new collaborative filtering algorithm using K-means clustering and neighbors' voting. In Proceedings of the 11th International Conference on Hybrid Intelligent Systems (HIS), Malacca, Malaysia, 5–8 December 2011; pp. 179–184

[15] Katarya, R.; Verma, O.P. An effective collaborative movie recommender system with cuckoo search. Egypt. Inform. J. 2017, 18, 105–112.

[16] Kumar, B.; Sharma, N. Approaches, Issues and Challenges in Recommender Systems: A Systematic Review. Indian J. Sci. Technol. 2016, 9, 1–12

[17] Phani, S.. , Sandeep, Kotte. , Raghavendra, N.. , Sharma, Aditi. , Pandey, Jitendra. , Chouhan, Vikas. Outlier Management and its Impact on Diabetes Prediction: A Voting Ensemble Study. Journal of Journal of Intelligent Systems and Internet of Things, vol. 12, no. 1, 2024, pp. 08-19. DOI: https://doi.org/10.54216/JISIoT.120101

[18] Sharma S, Shakya HK. Hybrid Real-Time Implicit Feedback SOM-Based Movie Recommendation Systems. InInternational Conference on Computing, Communications, and Cyber-Security 2022 Oct 21 (pp. 371-388). Singapore: Springer Nature Singapore.

[19] Sharma S, Shakya HK. Recommendation System for Movies Using Improved version of SOM with Hybrid Filtering Methods. In2023 6th International Conference on Information Systems and Computer Networks (ISCON) 2023 Mar 3 (pp. 1-7). IEEE.

[20] S., Malathi. , S., Hemamalini. , M., Ashwin. , Benny, Rijo. Knowledge Navigator: Revolutionizing Education through LLMs in Generative AI. Journal of Fusion: Practice and Applications, vol. 16, no. 1, 2024, pp. 209-222. DOI: https://doi.org/10.54216/FPA.160114

[21] Sharma S, Shakya HK. Recommendation Systems for a Group of Users Which Recommend Recent Attention: Using Hybrid Recommendation Model. InInternational Conference on Advanced Communication and Intelligent Systems 2022 Oct 21 (pp. 659-672). Cham: Springer Nature Switzerland.

[22] Mohamed, A. H., E. Omar, A. (2024). Improving Arabic Spam classification in social media using hyperparameters tuning and Particle Swarm Optimization. Journal of Fusion: Practice and Applications, 16(1), 08-22. DOI: https://doi.org/10.54216/FPA.160101

[23] Shanker, Jata. , K.Gupta, N.. , Sharma, Aditi. Enhanced Heart Disease Prediction Using Machine Learning Techniques. Journal of Journal of Intelligent Systems and Internet of Things, vol. 12, no. 2, 2024, pp. 19-33. DOI: https://doi.org/10.54216/JISIoT.120202

[24] Sharma, S., Shakya, H.K. and Marriboyina, V., 2021. A location based novel recommender framework of user interest through data categorization. Materials Today: Proceedings, 47, pp.7155-7161.

[25] Dash, R.K., Nguyen, T.N., Cengiz, K. et al. Fine-tuned support vector regression model for stock predictions. Neural Comput & Applic 35, 23295–23309 (2023). https://doi.org/10.1007/s00521-021-05842-w

[26] S. Phani Praveen, Kotte Sandeep, N. Raghavendra Sai, Aditi Sharma, Jitendra Pandey, Vikas Chouhan. (2024). Outlier Management and its Impact on Diabetes Prediction: A

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024

Voting Ensemble Study. Journal of Journal of Intelligent Systems and Internet of Things, 12 (1), 08-19 (Doi : https://doi.org/10.54216/JISIoT.120101)

[27] Sharma, S., Shakya, H.K. and Mishra, A., 2022. 11 Study and Estimation of Existing Software Quality. Multi-Criteria Decision Models in Software Reliability: Methods and Applications, p.235.

[28] Sharma S, Mishra A, Shakya HK. A Time-Variant Software Stability Model for Error Detection. InMulti-Criteria Decision Models in Software Reliability 2022 Nov 30 (pp. 277-290). CRC Press.

[29] Sharma S, Shakya HK. Hybrid Real-Time Implicit Feedback SOM-Based Movie Recommendation Systems. In International Conference on Computing, Communications, and Cyber-Security 2022 Oct 21 (pp. 371-388). Singapore: Springer Nature Singapore.

Doi: https://doi.org/10.54216/FPA.160204
Received: April 08, 2024 Revised: May 26, 2024 Accepted: June 30, 2024