An Intelligent Model for Optimizing Sparsity Problem Toward Movie Recommendation Paradigm Using Machine Learning

Manash Sarkar, Shiwangi Singh, V. L. Soundarya, Maroi Agrebi, and Ahmed Alkhayyat

1 Introduction

As we all know, the world is expanding at an unprecedented rate. Everyone is hurrying to achieve their ultimate objectives. This thirst leads to the growth of, practically, every industry. One of them is online business. We do not have time to go to the market, and this is not the end of it. We do not even have enough time to select an item from the collection. This sparked the birth of internet shopping, which has now grown into a massive tree with hundreds of branches. As the online market expands at an exponential rate, it is inevitable that competition will spread to other areas as well. Owners of separate websites must now entice their readers by giving appealing features. One of the features available to users is recommender engines. Recommender engines are the most well-known machine learning approach now in effect. We have all encountered services and websites that try to suggest books, movies, and articles based on our previous behavior. They attempt to infer tastes and

M. Sarkar (✉)
Department of Computer Science & Engineering, Atria Institute of Technology, Bengaluru, Karnataka, India

Expert Data Scientist, Limendo, Bolzano, Italy
e-mail: manash.sarkar@atria.edu; manash.sarkar@limendo.com

S. Singh · V. L. Soundarya
Department of Computer Science & Engineering, Atria Institute of Technology, Bengaluru, Karnataka, India

M. Agrebi
LAMIH UMR CNRS 8201, Department of Computer Science, Université Polytechnique Hauts-de-France, Valenciennes, France

A. Alkhayyat
College of technical engineering, The Islamic University, Najaf, Iraq
e-mail: ahmedalkhayyat85@iunajaf.edu.iq

preferences as well as recognize undiscovered interesting items. Recommendations are so commonplace in our environment that we sometimes fail to notice how effortlessly these systems created to improve our purchasing choices have been integrated into virtually every device and platform. Although we are accustomed to getting suggestions all the time, their integration into our everyday lives has a long and intriguing history that has been shaped by ups and downs, the ambition of a few pioneering firms, and the efforts of other dreamers. You may already be aware that a recommender system is a system that typically tries to forecast the value a user would attribute to a certain piece of material. These engines provide predictions by merging data models from several natural sources.

Sama et al. [1] have described in their research that using artificial intelligence technologies to create a movie recommendation system is difficult. Finding satisfying results is challenging because not all information obtained from the internet is trustworthy and useful. Experts advise using a recommendation system that gives pertinent information rather than continuously searching for the matching information. Collaborative filtering methods based on user information (such as gender, geography, or preference) are often effective. However, in today's world, everyone is concerned with advice or principles in order to make a choice. Our study focuses on developing a system for suggesting movies that may provide users with a list of movies that are comparable to the one they have chosen. Recommendation system, which falls under the category of model-based filtering approach, is explained in the research performed by Srikanth et al. [2]. It is built using cutting-edge deep autoencoders. NLP and computer vision applications employ deep autoencoders, an artificial neural network. Here, using an unsupervised learning technique, deep autoencoders are utilized to obtain the Top N movie recommendations.

Netflix's 3-month implicit value dataset with patterns discovered through collaborative filtering was utilized to assess the performance of the suggested recommender system. This chapter can successfully forecast the ratings that are absent from the sparse matrix that are first provided for input. The top model out of 20 created models is connected to the frontend.

By proposing movies that match the viewer's preferences, several online movie or video streaming companies may keep their users interested. Making more tailored suggestions is a major area of study for recommender systems. Two filtering strategies are covered in the study by Salmani et al. [3], content-based filtering and collaborative filtering. The former recommends items (movies) to users (viewers) based on their past viewing behavior and preferences, while the latter does so by considering the opinions and actions of other users (viewers) who have similar preferences (movies). Collaborative filtering has produced user-based, item-based, SVD++, and SVD algorithms, and their performance has been evaluated. A hybrid recommendation engine that combines both the content-based and SVD filtering models offers the greatest performance and improved movie selections for preserving active viewer engagement with the service. In order to suggest movies based on several criteria, the study done by Sunandana et al. [4] created a movie recommendation system. The project's main goal is to provide a system for movie

recommendation that would allow users to be prescribed movies. There are several algorithms that may be used to create a recommendation system. The content-based algorithm has been used to propose movies based on their similarities to other movies by looking at the movie's content. The degree of resemblance has been assessed using the cosine similarity method. The parameters in this case were taken from the TF-IDF vectorization output, and a linear kernel was used to get the cosine similarity. The most comparable movies are then recommended. The recommendation system uses things based on user preferences to help with decision-making. There are many different sectors where recommendation systems are employed, similar to social networking, advertising, and e-commerce. Collaborative filtering is a common approach for creating recommendation systems; however, it becomes less reliable if the amount of data is too little. As a result, using the weighted average technique can increase the precision of suggestions. Today, recommender system is also used for restaurant [5]. Restaurant recommendation is very useful for tourism industries.

The most common types of recommendation systems are as follows:

- *Content-based recommender systems*: To recommend things such as movies or music, content-based algorithms use meta data such as genre, producer, actor, and musician. A good example of such a recommendation would be recommending *Infinity War*, which stars Vin Diesel, because someone enjoyed *The Fate of the Furious*. Content-based systems are founded on the assumption that if you like one thing, you will probably like another similar thing.
- *Collaborative filtering recommender systems*: The behavior of a group of users is used to offer suggestions to other users in collaborative filtering. The preference of other users is used to provide recommendations.

There are two types of collaborative filtering systems:

- *Collaborative filtering based on users*: User-based collaborative filtering: In this paradigm, products are recommended to a user based on how popular they are among users who are similar between users.
- *Item-based collaborative filtering*: Related items are identified by these systems based on the users' earlier ratings. *Hybrid Recommender System*: Hybrid Recommender Systems merge any of the two interconnected systems in a customized sector. It is the most sought-after recommender system, as it merges the virtues of many recommender systems while also removing any potential flaws that might arise from using just one recommender system.
- *Weighted Hybrid Recommender*: The score of a recommended item is calculated using the results of all of the system's various recommendation strategies in this method. P-Tango, for example, combines content-based and collaborative recommendation systems, initially giving both equal weight but gradually changing the weighting as user ratings predictions are validated or disproved. Pazzani's hybrid does not use numerical scores, instead treating the output of each recommender as a series of votes that are then combined in a consensus-building procedure.

- *Switching Hybrid Recommender*: Switching Hybrid Recommender is a feature that allows you to switch between recommendation algorithms based on certain criteria. The switching hybrid recommender can deploy a content-based recommender system initially, and if that one fails, it will deploy a collaborative-based recommender system. This is possible if we combine content- and collaboration-based recommender systems.
- *Mixed Hybrid Recommender*: If a large number of recommendations can be made at the same time, mixed recommender systems should be used. Because recommendations from multiple techniques are given at the same time, the user can choose from a wide range of options. The majority of media and entertainment organizations use Smyth and Cotter's PTV system, which is mostly a recommended program, to recommend television viewing to customers.

Today, recommender system is the most powerful technique for intelligent business model to recommend the users' based on their choice and budget. There is an issue in recommender system called sparsity problem. In this problem, many values are not available in the dataset. This creates trouble at the time of analysis of historical data. In this research, the sparsity problem is optimized by proposing an intelligent model. An optimized decision to develop a collaborative filtering system that forecasts movie ratings for a user based on a large database of user ratings. A statistical analysis is carried out in order to create a pilot survey model for analyzing the real-time dataset. Ant Colony Optimization (ACO) is used to determine the group members' ratings for future recommendation. Finally, fourfold cross-validation is used to assess the performance of the proposed model.

The following sections comprise the remainder of this chapter: Section 2 reviews related works. Section 3 describes a fundamental of Big Data. Section 4 contains the proposed model, followed by Sect. 5's Results and Discussion. Section 6 contains the conclusion and future scope.

2   Similar Works Done

This section reviews earlier studies in the topic of recommendation systems, particularly those that use methods based on systems for recommending movies. The numerous studies are described in the section that follows.

According to this study undertaken by Omega et al. [6], user weighted average and movie weighted average have an impact on how users are recommended films. Additionally, it demonstrates that the weighted average recommendation system outperforms the collaborative recommendation system in terms of accuracy.

In their publication, Ajith et al. [7] go into great detail about the goal of their study, which is to provide a knowledge graph and particle filtering-based movie recommendation system. Every method of movie recommendation that is currently available makes use of clustering and machine learning technology. The groundbreaking movie recommendation system that is suggested in this chapter interacts

directly with the database, is very effective thanks to the expressive power of knowledge graphs, and offers the user the most pertinent suggestions. This movie recommendation tool uses particle filtering and a knowledge graph database to show the user the best movies based on their preferences for genres, directors, and other elements.

In contrast to previous systems, this recommendation system does not keep track of the browser history for forthcoming suggestions. The large amount of data makes it challenging for consumers to locate relevant movie resources fast; however, the movie recommendation model is a great way to deal with this problem. It is advised to apply GRU (Gate Recurrent Unit) to mine the rules and characteristics of their presence from the text data of movies in addition to deep neural networks (DNN) and factor decomposition machines FM to process the features of customers' wishes for various movie genres. By combining the link between users and movies, Xiong et al. [8] offer a movie recommendation model based on GRU, DNN, and FM in order to improve the accuracy of the suggested outcome. The Movielens-1m dataset is then used for a significant number of tests. The experimental findings demonstrate that the model outperforms CNN (Convolutional Neural Network) and LSTM (Long-term and Short-term Memory Network) recommendation models in terms of performance and recommendation accuracy.

Wang et al. [9] have discussed achieving lower marketing costs and higher revenues, traditional recommender models make use of technologies including collaborative filtering, matrix factorization, learning to rank, and deep learning models. Moviegoers, however, give the same film varying evaluations depending on the situation. The audience's mood, the setting, the weather, etc., are all significant movie-watching contexts. It is quite advantageous for recommender system builders to be able to use contextual data. However, widely used methods like tensor factorization require unreasonably large amounts of storage, which significantly affects their viability in actual environments. The MatMat framework is used in this study to create a context-aware movie recommendation system that is superior to conventional matrix factorization and similar in terms of fairness. By matrix fitting, MatMat factors matrices.

Due to their numerous uses, recommendation systems today play a significant role in society. Many businesses utilize it frequently in order to provide better customer experiences and accomplish growth by employing it properly. Information on the strategies and techniques utilized in recommendation systems is presented in this publication. Collaborative filtering approach, content-based approach, and hybrid approach are the three categories into which recommendation systems are divided. The study performed by Kukreja et al. [10] will outline the content-based method, its benefits and drawbacks, and how to use it to provide movie recommendations based on this concept. The report will also detail all other approaches, their methods, and their restrictions. Although we have chosen a content-based strategy, there may be various approaches depending on the business. Users can get their preferred movie programs from a wide range of films thanks to the user-oriented recommendation system. Qiu et al. [11] have discussed a type of heterogeneous network information called a knowledge graph that offers

a movie recommendation system a wealth of structural knowledge. It aids in resolving the cold start and data sparseness issues. We developed a knowledge graph-based movie recommendation model using this information. The knowledge graph is first integrated into the recommendation model, the vector representation is then updated using the engaging attention network represented by the user or object, rich local knowledge is then acquired through cross-interactive information dissemination, and predictions are finally produced. We experiment with open data sets, and the results show that our strategy is superior to the accepted practices. Lavanya et al. [12] have discussed recommendations that are so commonplace in our environment that we sometimes fail to notice how effortlessly these systems created to improve our purchasing choices have been integrated into virtually every device and platform. Although we are accustomed to getting suggestions all the time, their integration into our everyday lives has a long and intriguing history that has been shaped by ups and downs, the ambition of a few pioneering firms, and the efforts of other dreamers. You may already be aware that a recommender system is a system that typically tries to forecast the value a user would attribute to a certain piece of material. These engines provide predictions by merging data models from several natural sources. The dataset that was abandoned for the evaluation of qualities with CF includes nominal ratings for a wide range of films. A complex method of collaborative filtering is effective for the motivation of recommendations during evaluation and to link the performance of the arrangement with the inclusion of characteristics from user reviews and deprived of the attributes from user assessments. This method uses an enhanced methodology that is based solely on nominal ratings and movie characteristics like genres.

Agrawal et al. and Roy et al. [13, 14] have proposed a methodology that presents a hybrid approach that combines content-based filtering, collaborative filtering, SVM as a classifier, and genetic algorithm. Comparative results show that the proposed approach improves the accuracy, quality, and scalability of the movie recommendation system over pure approaches in three areas: accuracy, quality, and scalability. The benefits of both techniques are combined in a hybrid strategy that also seeks to minimize their drawbacks. This could happen when looking for things like books, music, films, and job listings. Therefore, it is necessary for a recommendation system to assist in recommending products to consumers that are more pertinent, accurate, and meet their wants and requirements. A movie recommendation system is used to make suggestions for films that are compatible with the user's interests and tastes. There are several ways to implement this system, including basic recommendation systems, collaborative recommendation systems, content-based recommendation systems, metadata-based recommendation systems, and demographic-based recommendation systems. In this study, they have employed a straightforward recommendation system, a content-based filtering strategy, and a collaborative filtering strategy. Python, the B/S framework, and the MySQL database have been used in the study performed by Xu et al. [15] to develop a system for recommending movies. The movie data for Douban is first crawled using crawler technology, after which the movie data is stored in MySQL. Users of the system can browse, look up, and collect movie-related data. The user-based collaborative

filtering algorithm is used to suggest movies that may be of interest to users based on their search history. The entire system of movie suggestion is made possible by the design of these useful components. In order to investigate user interests and offer the best products to users, collaborative filtering algorithms are frequently employed in e-commerce websites' recommendation systems. These systems are based on the analysis of a huge number of past behavior data from users. In this chapter, we concentrate on the development of a highly accurate and dependable algorithm for movie recommendation. It is important to note that the algorithm can be used in many different e-commerce domains and is not only restricted to movie recommendation. In the study undertaken by Zhou et al. [16], they have constructed a movie recommendation system in Ubuntu using Java language. The system can manage massive data sets thanks to the MapReduce framework and the item-based recommendation algorithm. Zhang et al. [17] have discueed these two problems in their study. First, a straightforward but very effective recommendation algorithm is suggested, which makes use of user profile information to divide users into various clusters. The size of the original user-item matrix is drastically reduced by creating a virtual opinion leader to represent each cluster as a whole. The weighted slope one-VU approach is then applied to the virtual opinion leader-item matrix to produce the recommendation results. When compared to traditional clustering-based CF recommendation schemes, our approach can significantly reduce the time complexity while still achieving comparable recommendation performance. By analyzing the user's profile to suggest the most relevant content, a recommendation system saves the user the time of looking for the information. For making suggestions, a number of strategies have been put forth by Hwang et al. [18] that included content-based, collaborative, and knowledge-based methodologies in their study. Recommendations for content like books, music, and videos are made using recommendation systems. They are also frequently utilized in online shopping. A collaborative filtering method based on genres, which is frequently used in movie recommendation systems, is used in particular by the South Korean film industry to make movie recommendations. This tactic might not be as effective when customers first discover movie suggestion services or have specific movie interests, like preferences for actors or directors.

One of the best and most appropriate techniques utilized in suggestion is collaborative filtering. Providing a prediction of the many goods that a user might be interested in based on their preferences is the core goal of the recommendation. Kharita et al. [19] have discussed in their study that when there is enough data, recommendation systems based on collaborative filtering approaches can produce predictions that are roughly correct. In the past, user-based collaborative filtering techniques have been quite effective at recommending products based on user preferences. However, there are also additional difficulties, such as data sparsity and scalability issues, which get worse as user and item counts rise. Finding the relevant information quickly on a big website is challenging. Selecting a movie that makes use of current technology to satisfy user needs has become more difficult as a result of the growth in the number of movies available. The growing usage of online services has led to a rise in the prevalence of recommendation systems. Today, the

goal of all recommendation systems is to use filtering and clustering algorithms to suggest content that people will find interesting. Users' preferences are initially discovered to allow them to evaluate movies of their choice by finding user profiles of people with similar interests.

To obtain precise recommendations based on customers' preferences, a hybrid recommendation model built on crowd search optimization is suggested. Collaborative and content-based filtering are combined to create the hybrid recommendation as described by Sarkar et al. [20].

The study undertaken by Joseph et al. [21] provides suggestions for a media commodity like movies to customers. After some usage, the recommender algorithm gets to know the user and suggests movies that are more likely to get good reviews. It is crucial for every area to implement recommender systems to give people advice on all technological elements. Different types of recommendation systems, including content-based, hybrid-based, and collaborative filtering-based systems, are available. Based on collaborative filtering, collaborative filtering for recommendations is separated into user-based and item-based filters. The study performed by Shrivastava et al. [22] aims to assess the collaborative filtering recommendation approach for a dataset of movies. The findings of the user-based and item-based recommendation methods compare how well each performed and which strategy produced the best outcomes.

3   Fundamentals of Big Data

Information that is present in large quantities and numbers is referred to as Big Data. Data refers to the various ways in which the computer has saved and recorded information as well as activities. This includes data that was provided by both people and machines. Big Data is a body of information that has grown significantly over time. Big Data is extremely large and complicated in terms of quantity and storage, making it impossible for conventional data management solutions to analyze and interpret it.

Big Data [23, 24] is collected in a number of methods that are critical. People's experiences and purchasing histories can be tracked through feedback and user reviews. This is also accomplished by conducting surveys that are presented to clients in the shape of a survey, which asks them different questions dependent on the data collecting goal. Every transaction a customer makes can earn them credits, and those points can then be redeemed for rewards based on specific standards. This makes it possible for the client to keep up a profile on the portal that details interests and behaviors; businesses use this kind of data.

There are some social media programs that monitor the information and activity on the concerned parties' profiles. In actuality, one of the most often used retargeting methods is social media. Other data collection techniques and resources exist as well, including cookies, email tracking, and satellite pictures, all of which operate similarly and are self-explanatory.

3.1 Properties of Big Data

Big Data, a large amount of data that is not handled by standard data storage or processing units makes up Big Data. Many multinational organizations utilize it to handle data and conduct business. The data flow would exceed 150 exabytes per day prior to replication. The properties of Big Data are explained by its 5 Vs (Fig. 1).

The 5 Vs of Big Data are as follows:

- Volume
- Veracity
- Variety
- Value
- Velocity
- *Volume*: Big Data is a term used to describe enormous amounts of data. Big Data is a term used to describe the enormous "volumes" of data that are produced every day from several sources, such as corporate processes, machines, human interactions, social media platforms, networks, and many more. About a billion messages, 4.5 billion "Like" button hits, and more than 350 million new postings are uploaded to Facebook every day. Big Data technology can manage large amounts of data.
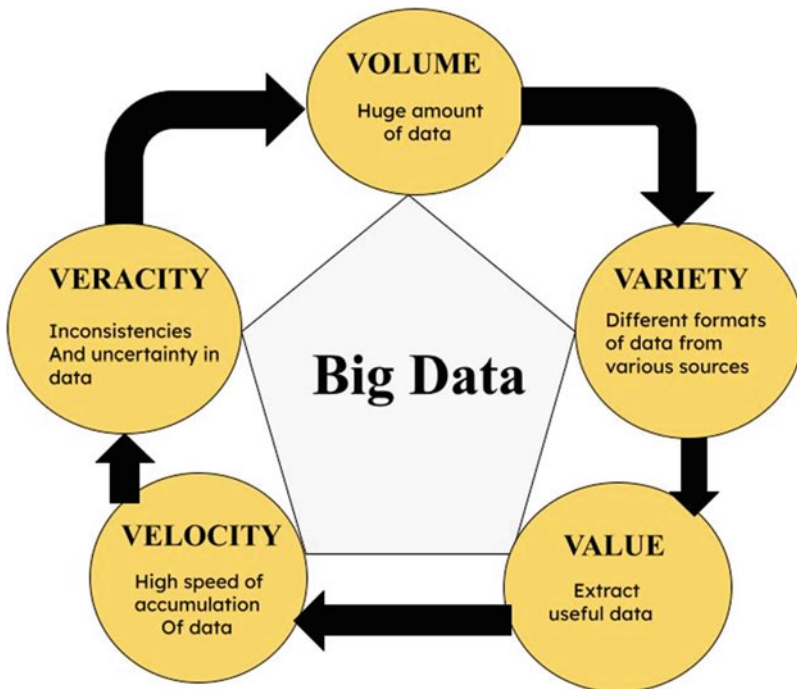

Fig. 1 Properties of Big Data

- *Variety*: Big Data can be gathered from a variety of sources and can be structured, semi-structured, and unstructured. Data was once only gathered through databases and sheets, but today it can also be found in PDF files, emails, audio files, social media posts, pictures, and videos. The data is arranged in the following way:
- *Structured data*: The structured schema contains each and every required column. It is arranged on a table. Structured data is kept in a relational database management system.

    – *Semi-structured Data*: Semi-structured data lacks a properly defined schema for technologies like XML, JSON, CSV, TSV, and email. The OLTP (Online Transaction Processing) systems handle semi-structured data. In order to store the data, tables or relations are used.
    – *Unstructured Data*: Unstructured data includes all unstructured files, including audio, log, and image files. Some businesses have a lot of data, but they are unsure how to value it because the data has not been processed.
    – *Quasi-structured Data*: The data format, which was created with effort and patience using a limited number of technologies, contains textual data with variations in data formats For example, a server creates and maintains a log file that contains a record of actions and is used by web servers (Fig. 2).

- *Veracity*: Veracity is a concept used to describe how reliable the data is. You may filter or manipulate data in a number of different ways. Veracity is the capacity to effectively handle and maintain data. Big Data is essential for business expansion as well, for instance, Facebook postings using hashtags.
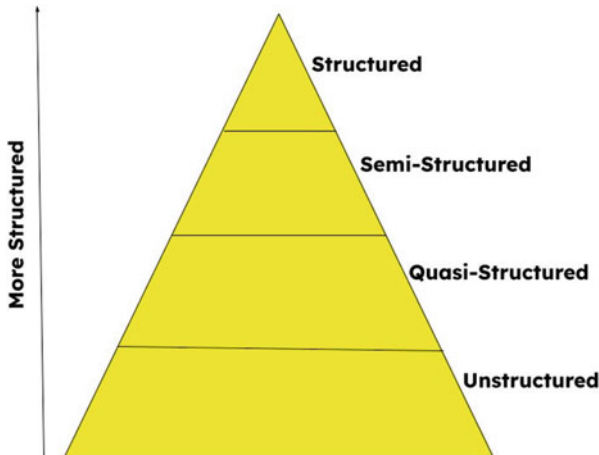

Fig. 2 Variety of data

- *Value*: The importance of value in large data cannot be overstated. The information we handle or store is not the issue. We store, process, and analyze useful and trustworthy information.
- *Velocity*: In contrast to other variables, velocity has a big impact. The rate at which data is produced in real time is known as velocity. It incorporates the incoming data sets' speed, rate of change, and activity bursts. Big Data's primary objective is to promptly supply highly demanded data. Big Data velocity is the rate at which data moves from sources including corporate processes, networks, social media sites, sensors, and mobile devices.

3.2 Big Data in Entertainment Industry

Nowadays, people are highly enthusiastic about discovering new material in terms of what they watch and choose to watch since media and entertainment have permeated every aspect of their life. The days of single-channel viewing, with no choice, integration, or consultation with viewers, are long gone. These factors are altering, though, since there are now millions of watching options to select from, and they can all be streamed across many platforms, making them much more accessible. Unquestionably, the use of Big Data, which is a mix of existing data, time series data, network data, and other types of data, has considerably assisted in the creation and successful application of such ideas and options in the media and entertainment sector. The only way to understand the behavior of consumers (who are the viewers here) is to analyze the different consumer data that is accessible throughout and as a product of multiple platforms. Big Data is aiding from a distance, but its influence is particularly obvious and felt in the sector. Let us examine Big Data's definition, applications in the media and entertainment industry, and sources. The recommendation system is very useful in entertainment industries [25]. Nowadays, OTT platform used the concept of recommender system to serve their customer smoothly.

3.2.1 Uses of Big Data in Media and Entertainment

The media and entertainment industry likewise integrates and gathers similar data from numerous sources in order to better understand viewer behavior and improve in a way that would allow them to thrive and be the viewers' favorite between each of them. The more you know your customer, the better you can predict their preferences and modify content, pricing, and user interface appropriately. This is a well-known marketing and financial reality (Fig. 3).

USES OF BIG DATA
IN
MEDIA AND ENTERTAINMENT

1 Predict Audience Interest	Optimization and Monetization **2**	**3** Understanding Audience Disengagement	Role of Advertisement **4**

Fig. 3 Uses of Big Data in entertainment. (Courtesy: www.google.com)

- *Predict Audience Interest*: Media firms and entertainment channels, particularly those that are focused on online streaming, can benefit from Big Data analysis. Because satisfying their customers' needs is so crucial to them, such channels want to be prepared with the content and categories necessary to accommodate practically any type of viewer and endeavor to ensure that content variety reigns supreme. Media and entertainment goliaths benefit from a plethora of classified data provided by Big Data. Search histories, genre ratings, followed social media trends, language, age, and other crucial pieces of this data can help businesses predict a particular viewer's interests, enabling them to not only customize the viewer experience but also come up with the most practical and popular program initiatives based on their audience's preferences. Some media organizations are also able to monitor how long viewers watch a video or movie for as well as their responses to it on other social media platforms. Additionally, there is a mechanism for tracking trending hashtags, and in a situation like this, user interaction makes it simpler to control and remove content, like a movie, from the board. They may occasionally provide the user a list of suggested readings to keep them interested in the site.
- *Optimization and Monetization*: Based on current market releases and trends, businesses may decide to add a specific movie to their list of available material just because it is popular and people would want to watch it. Because they often monetize such unconventional content, which keeps viewers interested and draws in new users seeking for the same content, this might bring in more money for the businesses than usual. Such discussion boards are more frequently used by them for amusement. There is also material when companies release a program or a film depending on how viewers respond to its trailers and elect to reserve a certain film or television program for a membership-only set of viewers, who are often charged requiring subscribers to pay a bundle to access these. These businesses further attempt to pique user interest by making only a small portion of a television show or movie available to viewers without a subscription. There are several types of material available to members and non-members of many entertainment firms, including Netflix, Hotstar, and Amazon Prime.

- *Understanding Audience Disengagement*: In any industry, losing a customer is still a frustrating problem for any company. The same is true for media and entertainment businesses; they provide memberships for access to certain material, which is a crucial step in the customer conversion procedure. Any company's membership duration in relation to its pricing is undoubtedly subjective, and it is continuously revised and updated. Big Data can also reveal information about returning clients and devoted fans. In certain cases, despite receiving several push-up reminders and calls to action, members choose to opt-out of the membership program and do not renew their memberships. This needs to be taken into consideration by the media and entertainment industries so that inaccurate information and stale, boring content may be seen. Big Data aids in acquiring up-to-date insights into customer behavior and might lead a company to modify content in response to the platform's widespread demand. As an illustration, the business may add multilingual material if the majority of customers choose not to renew their membership speech problems in the content introduced in the form of feedback.
- *Role of Advertisements*: The key factor that continues to determine a company's market worth and income is advertising. Inquiries from advertisers on a company's analysis of viewing behavior are essential. This may facilitate the timely and accurate distribution of relevant, tailored, and related adverts. The understanding of user behavior and what they are likely to buy as a consequence of targeted marketing is also aided by Big Data applications. Advertisements are common and seem to be an essential part of every entertainment sector. This gives companies the ability to act as retargeting agents, which means that if a customer is watching a show or a movie about one of their goods, an advertising will pop up more appropriately. The audience is more likely to purchase 3D glasses if they are marketed in the middle of a science fiction or technological film. An example of this is a fashion movie that mimics online shopping and encourages viewers to purchase items like clothing and freshly declared trends. Big Data enables businesses to construct successful advertising campaigns based on a variety of criteria, including weather, timing, and the use of second screens, in addition to enabling customers and businesses to generate content-related advertisements.

4 Proposed Model

An intelligent recommender model is proposed to minimize the sparsity problem. At the beginning, a statistical analysis using Brownian process is performed. A metaheuristic-based optimization technique called Ant Colony Optimization (ACO) is applied to optimize the sparsity problem.

4.1 Mathematical Background

At the beginning, a statistical analysis [26, 27] is performed to understand the nature of the data and its distribution. As per the real dataset, all the required features are not in scaled. Normalization is performed first to set the all features value with a certain range. After data visualization, it is found that some outlier data are there which are not ignorable. To find out the distance between the values, Brownian process [28] is deployed.

Assume that random distance time for approximating the sum of independent random values is considered.

Let $x_1 + x_2 + \ldots + x_n$ where x_i is a random variable $\forall i \geq 1$

As per the Brownian motion theory, it was a continuous time (t) where $0 \leq t \leq \infty$ and μ as a mean of the random data point $V(t)$ and

$$V(t) = 0 \tag{1}$$

$$P_\mu \{V(t) - V(s) \leq x\} = \phi \left[(x - \mu (t - s)) \Big/ (t - s)^{\frac{1}{2}} \right]$$

where

$$\forall \ 0 \leq s < t < \infty \tag{2}$$

and $0 \leq s_1 < t_1 \leq s_2 < t_2 \leq \ldots \cdots \leq s_n < t < \infty$ where $n = 2, 3 \ldots \ldots$ \quad (3)

The random data $V(t_i) - V(S_i)$ where $i = 1, 2, 3 \ldots \ldots n$ are stochastically independent. $Z(t)$ is a continuous function of t. The standard normal distribution is denoted as ϕ. If a random data x_i will be independent and distributed randomly with mean value μ and variance 1, then it will be as Eq. (4).

$$S_n = x_1 + x_2 + \ldots + x_n \text{ where } n = 0, 1, 2, \ldots \text{ and } V(n) = 0, 1, 2, \ldots \forall$$
$$\{ V(t), 0 \leq t < \infty \} \tag{4}$$

To implement the time in a discrete way, a random selection $S_n, n = 0, 1, 2, \ldots$ is considered. In this chapter, a random data selection is performed to avoid the biasness of the proposed model toward the particular class of data. To achieve this target, data sample with discrete time interval, the likelihood ratio of the $\{V(S), S \leq t\}$ under P_{μ_1} relative to P_{μ_2} will be as Eq. (5)

$$l(t, V(t); \mu_1, \mu_2) = \exp \left[(\mu_1 - \mu_2) \ V(t) - \tfrac{t}{2} \ (\mu_1^2 - \mu_2^2) \right]$$
$$\text{for any} -\infty < \mu_i < \infty \text{and} \ (i = 1, 2 \ldots) \text{ and } t > 0 \tag{5}$$

4.2   Ant Colony Optimization (ACO)

To achieve the desired output, a bio-inspired intelligent modll based on Ant Colony Optimization technique is deployed. At the beginning, a statistical analysis is performed to understand the data pattern and to visualize the data.

The Ant Colony Optimization metaheuristic's limited history is primarily one of experimental research. All early researchers were directed by trial and error, and most ongoing research efforts are still guided by trial and error. This is the normal scenario for almost all known metaheuristics: when experimental work has demonstrated the usefulness of the metaheuristic, researchers attempt to develop a theory to explain how it functions in addition to doing ever-more complex tests. Convergence is typically the first theoretically motivated topic to be addressed. If sufficient resources are available, will the metaheuristic discover the optimal solution? Numerous issues are frequently looked into, including the speed of convergence, ethical methods for choosing the metaheuristic's parameters, connections to preexisting strategies, identification of problem traits that increase the metaheuristic's likelihood of success, and comprehension of the significance of the various metaheuristic components. Here, we discuss the problems for which we currently have an answer. We focus on how ACO compares to other well-known techniques like stochastic gradient ascent as well as how various ACO algorithms converge to the best result.

4.2.1   Theoretical Considerations on ACO

The relatively broad concept of when attempting to demonstrate the theoretical merits of the ACO metaheuristic, ACO [29] presents the researcher with their first significant challenge. Generality is a desirable property that makes theoretical analysis much more difficult, if not impossible, because it enables ant-based algorithms to be applied to discrete optimization problems ranging from static problems like the traveling salesman problem to time-varying problems like routing in telecommunications networks. The convergence problem is the first theoretical aspect of ACO that we will look at in this chapter. Is it conceivable that the algorithm will eventually identify the ideal solution? ACO algorithms are stochastic search processes, and they may never find the best answer due of the bias provided by pheromone trails. This makes the topic intriguing. It is crucial to bear in mind that there are at least two different forms of convergence to take into account when analyzing a stochastic optimization algorithm: value convergence and solution convergence. Assuming indirectly that in the event of a problem with several optimum solutions, we are interested in assessing the likelihood that the algorithm will produce an ideal answer

at least once while investigating convergence in value. We want to determine the likelihood that the algorithm will provide an ideal result at least once. Convergence of values for ACO algorithms in ACObs; convergence of tmin and solutions for ACO algorithms in ACObs; small tmin parameter may vary during an algorithm run, as shown by tminy, and the iteration counter for the ACO algorithm is y. We next demonstrate that these findings hold even when common ACO components like search engine and heuristic information are taken into account. Finally, we discuss the consequences of these discoveries and demonstrate that the MMAS and ACS ACO algorithms, two of the most successful ACO algorithms now in use, directly benefit from the evidence of value convergence. Sadly, there are currently no findings on the rate of convergence of any ACO algorithm. Even if convergence may be demonstrated, there is presently no alternative to conducting large experimental procedures in order to assess algorithmic performance. The formal relationship between ACO and other techniques is another theoretical subject explored in this chapter. We embed ACO in the broader model-based search framework, to better comprehend the relationships between ACO, stochastic gradient ascent, and the more modern cross-entropy method [30, 31]. To find out the optimize solution of the users' rating, ACO, a metaheuristic algorithm have been applied**.** In this chapter, the users are considered ants and their corresponding rating are considered as pheromone. The users give their rating based on the choice. Similarity index is determine based on the rating value. The users too change their choice from movie I to movie j with probability.

$$P_{i,j} = \frac{\left(\tau_{i,j}^{\alpha}\right) \left(\eta_{i,j}^{\beta}\right)}{\sum \left(\tau_{i,j}^{\alpha}\right) \left(\eta_{i,j}^{\beta}\right)} \tag{6}$$

where

$\tau_{i,j}$➜ The amount of pheromone (Rating) on the edge i, j

α ➜ Parameter to control the influence of $\tau_{i,j}$

$\eta_{i,j}$ ➜ The desirability of edge i, j

β ➜ A parameter to control the influence of $\eta_{i,j}$

Algorithm
A well-known bio-inspired algorithm called Ant Colony Optimization is deployed for recommending movies

Algorithm: Ant Colony optimisation for movie recommendation
Begin
Input: Movielens dataset
Output: Optimistic decision making for movie recommendation
Initialization: A set of data from movie lens dataset is used for pilot survey.
Step 1: Randomly select the data from the dataset.
Step 2: User U with m numbers of features. $U = \{f_1, f_2, f_3,........, f_m\}$
Step 3: Consider the total number of users is n. $U = \{U_1, U_2, U_3,........, U_n\}$
Step 4: Determine the co-variance of the users' with their features.
Step 5: Pheromone update User rating "r," will rate the next rating "s" based on the following formula. $$s = \begin{cases} \max_{j \in N_i^k} \left\{ \tau_{ij}(t).\eta_{ij}^{\beta} \right\} & \text{if } q \prec Q_0 \\ \text{according to AS equation} & \text{otherwise} \end{cases}$$ N_i^k are the unrated movies of Kth user $\tau_{ij}(t)$ is rating in (i,j) edge during t turn and $\eta_{i,j}$ is the distance between user i to user j and "q" is the stochastic variable with uniform distribution. Q_0 is the threshold value in which initializing procedure begins.
End

4.3 Data Preparation

To validate the proposed model, a real-time dataset named "movies.csv" [32] is used as a test-bed.

5 Results and Discussion

A real-time dataset is used as a test-bed to validate the proposed model. In this research, 5000 data are considered. The users' ratings are visualized in Fig. 4.

The rating of the users' is varying based on their choice. Figure 5 shows the distribution of users' ratings. Some users' ratings are outliers. In this chapter, 5000 data is considered from Movielens and the outlier data is also considered for better results.

Figure 6 shows the difference in ratings between the user 20 and other users. It displays the distance between User Id 20 and other User Ids (Fig. 7).

5.1 Performance Analysis

Performance analysis based on various parameters of the confusion matrix is developed to assess the effectiveness of the proposed model. The fourfold cross-validation method is used on the provided data set in this chapter. Table 1 displays the various ratios that were chosen for the training and testing data sets over four
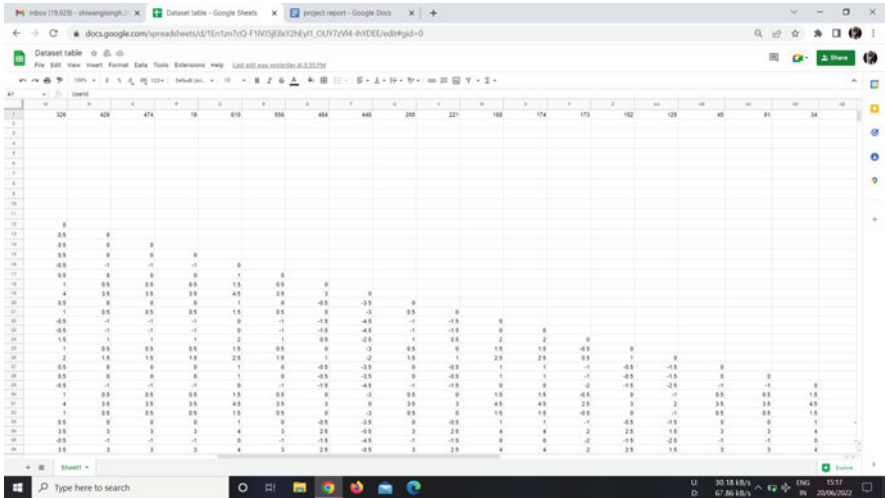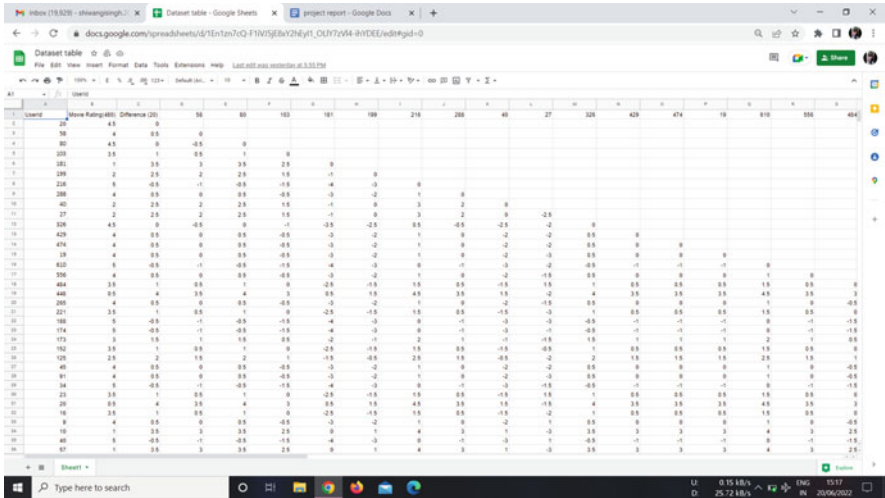
Fig. 4 Snapshot of difference in users and their ratings

iterations. The error is fixed after each iteration to make the model more accurate. The suggested model is used to provide users with the best recommendations based on their preferences.
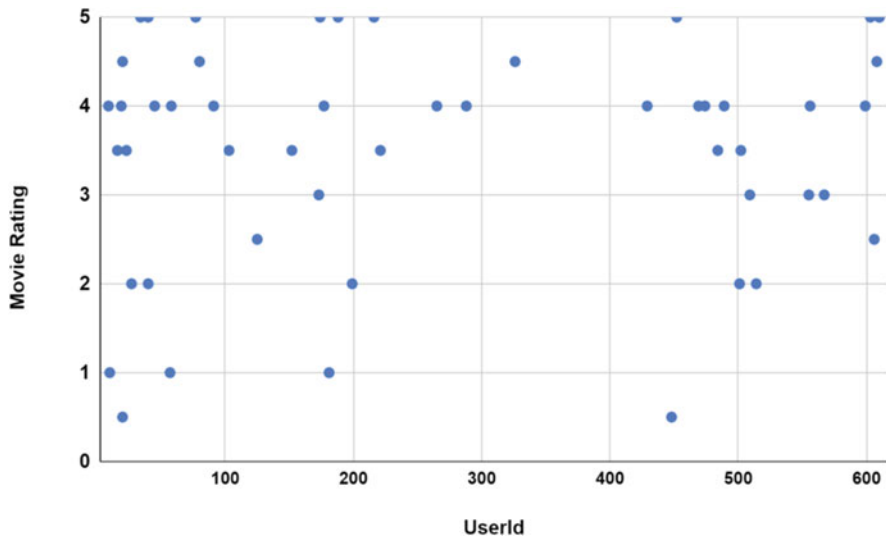
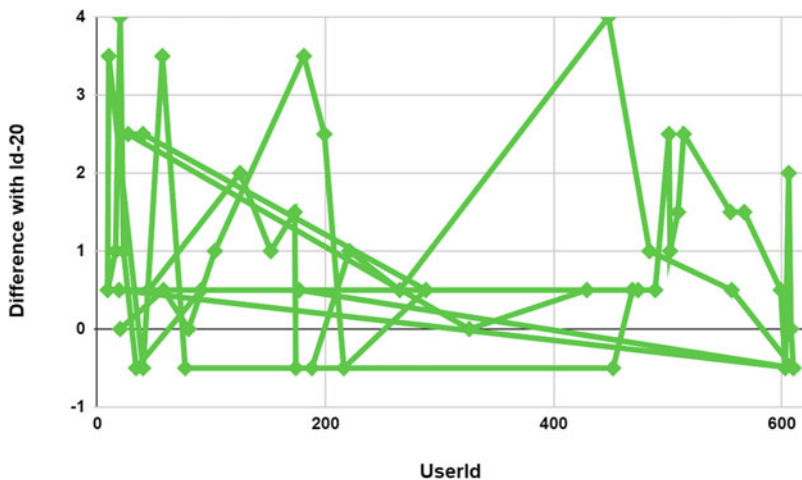**Fig. 5** Distribution of the users' rating

Fig. 6 Distance between user Id 20 and other users

5.1.1 Evaluation Metrics

Evaluation metrics are frequently used to gauge how accurate a recommendation system is. In areas like security, image processing, and information retrieval, the evaluation metrics are very helpful. In general, evaluation metrics make it simple to verify the effectiveness and efficiency of recommenders. Thus, fictitious frameworks could become actual applications. On the other hand, it identifies shortcomings,
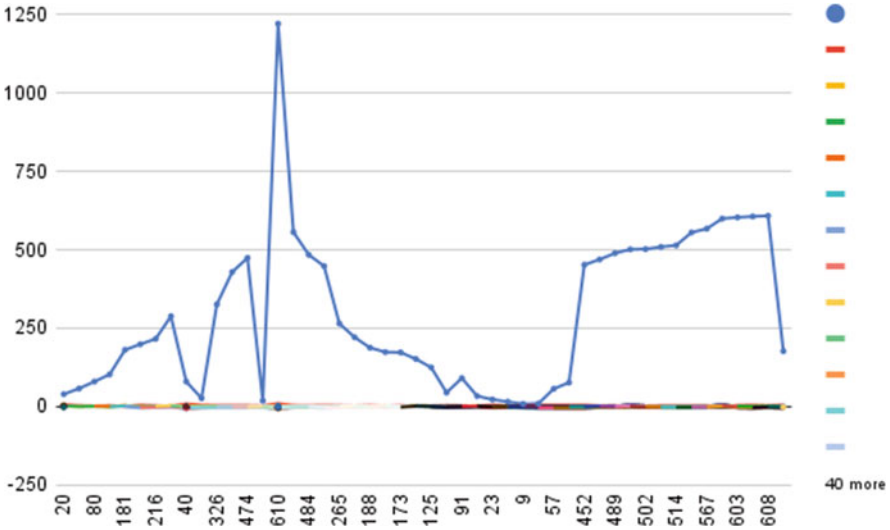
Fig. 7 Distribution of random user's ratings

Table 1 Performance at four different strategies

Training and testing strategy	Strategy 1 (50%–50%)	Strategy 2 (60%–40%)	Strategy 3 (70%–30%)	Strategy 4 (80%–20%)	Average
Accuracy	91.56	90.75	95.88	98.25	94.11
Recall	88.68	92.45	94.84	96.68	93.16
Precision	90.34	91.64	92.42	95.24	92.41

accuracy, and potential remedial actions. Furthermore, both explicit and implicit user activities are covered by these metrics. Four different strategies are applied for the dataset to validate the proposed model. The values of Accuracy, Recall, and Precision are shown in Table 1.

The average value of the Accuracy, Recall, and Precision, from Table 1, is shown in Fig. 8.

The average values of the Accuracy, Recall, and Precision are 94.11, 93.16, and 92.41, respectively.

6 Conclusion

The media sector has seen a significant amount of research work focused on various aspects of big data analysis, including the indexing, capture, storage, mining, and retrieval of multimedia big data. The entertainment industry now does a huge amount of business online in this digital age. During the pandemic, when social isolation is required, the idea of the entertainment through online become popular.
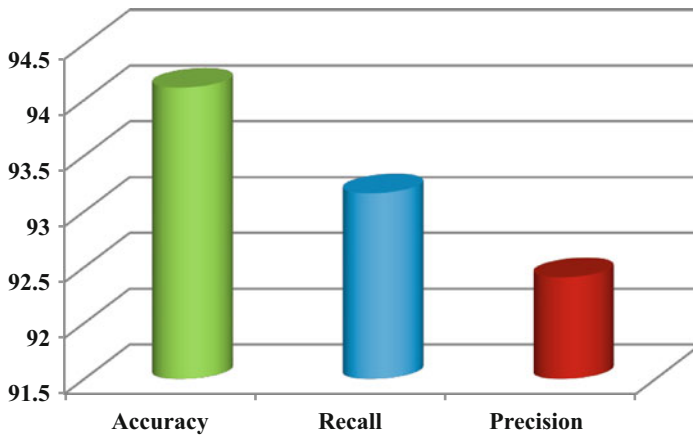
Fig. 8 Average value of the evaluation metrics

Industries use the idea of big data to represent customer feedback, comments, and choices in order to provide better service. In this chapter, an intelligent algorithm that makes an optimistic decision to develop a collaborative filtering system that predicts a user's movie rating based on a sizable user rating database. It recommends movies that are the best fit for users based on the genres of media they enjoy. The list of recommended movies is generated by the combined impact of user ratings and reviews. In order to create a pilot survey model to analyze the real-time dataset, a statistical analysis is carried out. Ant Colony Optimization (ACO) is used to rate the group members in order to make recommendations in the future. In a recommender system, sparsity problems will be optimized in this manner. As a result, the suggested model could be used to successfully train a recommender. The experimental findings show that there are very few false positive and false negative recommendations. Additionally, the abundance of data does not affect the model's accuracy. The results show that the proposed model's Accuracy, Recall, and Precision are 94.11, 93.16, and 92.41, respectively. Future research will focus on using deep learning-based intelligent models to reduce sparsity and enhance Precision and Recall values for recommendation systems in the context of big data.

References

1. Sama, L., Wang, H., & Makkar, A. (2021). Movie recommendation system using deep learning. In *9th International Conference on Orange Technology (ICOT)*, Tainan. https://ieeexplore.ieee.org/document/9680609
2. Purushothaman Srikanth, E. Ushitaasree, S. M. Bhargav Bhattaram, G., & Anand, P. (2021). Movie recommendation system using deep autoencoder, In *5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore. https://ieeexplore.ieee.org/document/9675960

3. Salmani, S., & Kulkarni, S. (2021). Hybrid movie recommendation system using machine learning. In *International Conference on Communication information and Computing Technology (ICCICT)*, Mumbai. https://ieeexplore.ieee.org/document/9510058

4. Sunandana, G., Reshma, M., Pratyusha, Y., Kommineni, M., & Gogulamudi, S. (2021). Movie recommendation system using enhanced content-based filtering algorithm based on user demographic data. In *6th International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore. https://ieeexplore.ieee.org/document/9489125

5. Roy, A., Banerjee, S., Sarkar, M., Darwish, A., Elhosen, M., & Hassanieen, A. E. (2018). Exploring New Vista of intelligent collaborative filtering: A restaurant recommendation paradigm. *Journal of Computational Science, Elsevier, 27*, 168–182.

6. Christ Zefanya Omega, H. (2021). Movie recommendation system using weighted average approach. In *2nd International Conference on Innovative and Creative Information Technology (ICITech)*, Salatiga. https://ieeexplore.ieee.org/document/9590147

7. Ajith, T. T., Ajay Krishnan, C. V., Nandakishore, J., Ananthu Subramanian, M. S., Siji Rani, S. (2021). Enhanced movie recommendation using knowledge graph and particle filtering. In *2nd International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy. https://ieeexplore.ieee.org/document/9591834

8. Xiong, W., & He, C. (2021). Personalized movie hybrid recommendation model based on GRU. In *4th International Conference on Robotics, Control and Automation Engineering (RCAE)*, Wuhan. https://ieeexplore.ieee.org/document/9638949

9. Wang, H. (2021). MovieMat: Context-aware movie recommendation with matrix factorization by matrix fitting. In *7th International Conference on Computer and Communications (ICCC)*, Chengdu. https://ieeexplore.ieee.org/document/9674549

10. Soni, N., Kumar, K., Sharma, A., Kukreja, S., & Yadav, A. (2021). Machine learning based movie recommendation system. In *IEEE International Conference on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking*, Dehradun. https://ieeexplore.ieee.org/document/9667602

11. Qiu, G., & Guo, Y. (2021). Movie big data intelligent recommendation system based on knowledge graph. In *IEEE International Conference on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking*, New York City. https://ieeexplore.ieee.org/document/9644835

12. Lavanya, R., & Bharathi, B. (2021). Systematic analysis of Movie Recommendation System through Sentiment Analysis. In *International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, Coimbatore. https://ieeexplore.ieee.org/document/9395854/authors#authors

13. Agrawal, S., & Jain, P. (2017). An improved approach for movie recommendation system. In *International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam. https://ieeexplore.ieee.org/document/8058367

14. Roy, S., Sharma, M., & Singh, S. K. (2019). Movie recommendation system using semi-supervised learning. In *Global Conference for Advancement in Technology (GCAT)*, Bangalore. https://ieeexplore.ieee.org/document/8978353

15. Xu, Q., & Han, J. (2021). The construction of movie recommendation system based on Python. In *IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, Chongqing. https://ieeexplore.ieee.org/document/9687872

16. Zhou, T., Chen, L., & Shen, J. (2017). Movie recommendation system employing the user-based CF in cloud computing. In *IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, Guangzhou. https://ieeexplore.ieee.org/document/8005971

17. Zhang, J., Wang, Y., & Yuan, Z. *Personalized real-time movie recommendation system: Practical prototype and evaluation.* Tsinghua Science and Technology. https://ieeexplore.ieee.org/document/8821512

18. Hwang, S., & Park, E. (2021). Movie recommendation systems using actor-based matrix computations in South Korea. In *IEEE Transactions on Computational Social Systems.* https://ieeexplore.ieee.org/document/9566476

19. Kharita, M. K., Kumar, A., & Singh, P. Item-based collaborative filtering in movie recommendation in real time. In *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, Jalandhar. https://ieeexplore.ieee.org/document/8703362

20. Sarkar, M., Roy, A., Agrebi, M., & AlQaheri, H. (2022). Exploring new vista of intelligent recommendation framework for tourism industries: An itinerary through big data paradigm. *Information, 13*(2), 70. https://doi.org/10.3390/info13020070

21. Anandkumar, R., Dinesh, K., Obaid, A. J., Malik, P., Sharma, R., Dumka, A., Singh, R., & Khatak, S. (2022). Securing e-Health application of cloud computing using hyperchaotic image encryption framework. *Computers & Electrical Engineering, 100*, 107860, ISSN 0045-7906. https://doi.org/10.1016/j.compeleceng.2022.107860

22. Sharma, R., Xin, Q., Siarry, P., & Hong, W.-C. (2022). Guest editorial: Deep learning-based intelligent communication systems: Using big data analytics. *IET Communications*. https://doi.org/10.1049/cmu2.12374

23. Sharma, R., & Arya, R. (2022). UAV based long range environment monitoring system with Industry 5.0 perspectives for smart city infrastructure, 108066, ISSN 0360-8352. *Computers & Industrial Engineering, 168*. https://doi.org/10.1016/j.cie.2022.108066

24. Rai, M., Maity, T., Sharma, R., et al. (2022). Early detection of foot ulceration in type II diabetic patient using registration method in infrared images and descriptive comparison with deep learning methods. *The Journal of Supercomputing*. https://doi.org/10.1007/s11227-022-04380-z

25. Sharma, R., Gupta, D., Maseleno, A., & Peng, S.-L. (2022). Introduction to the special issue on big data analytics with internet of things-oriented infrastructures for future smart cities. *Expert Systems, 39*, e12969. https://doi.org/10.1111/exsy.12969

26. Sharma, R., Gavalas, D., & Peng, S.-L. (2022). Smart and future applications of Internet of Multimedia Things (IoMT) using big data analytics. *Sensors, 22*, 4146. https://doi.org/10.3390/s22114146

27. Sharma, R., & Arya, R. (2022). Security threats and measures in the internet of things for smart city infrastructure: A state of art. *Transactions on Emerging Telecommunications Technologies*, e4571. https://doi.org/10.1002/ett.4571

28. Zheng, J., Wu, Z., Sharma, R., & Lv, H. (2022). Adaptive decision model of product team organization pattern for extracting new energy from agricultural waste, 102352, ISSN 2213-1388. *Sustainable Energy Technologies and Assessments, 53*(Part A). https://doi.org/10.1016/j.seta.2022.102352

29. Mou, J., Gao, K., Duan, P., Li, J., Garg, A., & Sharma, R. (2022). A machine learning approach for energy-efficient intelligent transportation scheduling problem in a real-world dynamic circumstances. In *IEEE Transactions on Intelligent Transportation Systems.* https://doi.org/10.1109/TITS.2022.3183215

30. Priyadarshini, I., Sharma, R., Bhatt, D., et al. (2022). Human activity recognition in cyber-physical systems using optimized machine learning techniques. *Cluster Computing*. https://doi.org/10.1007/s10586-022-03662-8

31. Priyadarshini, I., Alkhayyat, A., Obaid, A. J., & Sharma, R. (2022, 103970, ISSN 0264-2751). Water pollution reduction for sustainable urban development using machine learning techniques. *Cities, 130*. https://doi.org/10.1016/j.cities.2022.103970

32. Pandya, S., Gadekallu, T. R., Maddikunta, P. K. R., & Sharma, R. (2022). A study of the impacts of air pollution on the agricultural community and yield crops (Indian context). *Sustainability, 14*, 13098. https://doi.org/10.3390/su142013098