



Development a Hybrid Recommender System Based-Classification Techniques in Data Mining Algorithms and Collaborative Filtering

Huda Rashid Shakir^{1*}, Sarah Shihab Ahmed², Sadiq Abdul Aziz Mehdi³

¹ Third Karkh Directorate, Ministry of Education-Iraq, Baghdad 10006, Iraq

² Department of Physiology, Collage of Medicine, Al Iraqia University, Baghdad 10006, Iraq

³ Department of Computer Science, Collage of Education, Mustansiriyah University, Baghdad 10006, Iraq

Corresponding Author Email: hudarashid@uomustansiriyah.edu.iq

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/jesa.580706>

ABSTRACT

Received: 20 February 2025

Revised: 30 June 2025

Accepted: 9 July 2025

Available online: 31 July 2025

Keywords:

Naïve-Bayes, recommender system, collaborative filtering, data mining

The volume of data has significantly expanded across every device, including those used by the cinema, television, and electronic industries. To improve user simplicity and knowledge, it is feasible to extract important knowledge and give consumers access to more relevant data. The manner that items are sought after has been altered by recommending algorithms. Predicting individual tastes is done using the filtering process. Based on user ratings, a list of suggested movies is categorized and assessed. Naive Bayes classifiers have shown their efficacy in a variety of applications, especially systems for recommending movies. In order to establish a rating of suggested movies according to user forecasts, this research study suggests a system for recommendation that uses a Naive Bayes classifier for offering personalized suggestions using the LDOSCOMODA dataset. Contextual information was used to increase the number of ratings while enabling the system to forecast unrated movies. The usefulness of the suggested recommender technique for producing precise and pertinent suggestions is demonstrated by the findings of the experiment. A proportion of 0.98 was attained by the forecast, 0.98 by precision, and 0.99 by recall. The suggested was contrasted with earlier efforts. The outcomes are better than those of the same, it was determined.

1. INTRODUCTION

Systems that provide suggestions are becoming more and more popular among consumers since they make selecting goods and services easier and help address the issue of knowledge saturation. In overall, systems for suggestions cater to user choices, allowing the user to both expressly and intuitively verify material [1]. Recommendation engines are programs that gather user preferences for a category of content, such as films, books, music, websites, and so on. Via the collection of feedback from users and the tracking of user activity through the viewing of movies, downloading of programs, or visiting of websites, it is feasible to easily gather the most comprehensive data and suggestions [2]. Recommendation engines take advantage of user demographics like age and gender as well as social media data like follows and posts. The likelihood that forecasting, endorsement, and balance work in regards to security, precision, and dispersal increases with recommendations devices [3]. Following a previous consideration of the suggestion strategy, systems for suggestions have been applied to an already-existing data set. Machine learning is then employed, and the data set is assessed against the data used for training and testing either once or repeatedly.

In order to categorize an item, knowledge must be obtained and acquired via the system of suggestions. For this reason,

data sets will be required, and these data will require context data—that is, data about the user at the time the item is applied or decision is made to assess the item [4]. Customized and unidentified techniques are the two types of methods Recommending Systems (RS) fall under. Algorithms that are not customized from their instances, all customers get the same suggestions (i.e., generic suggestions for the newest songs, movies, or hotels). Regarding the customized computations, customers are presented with multiple choices that are designed to offer the most suitable suggestions using non-custom techniques [5].

Due to the extraordinary advancements in computer technology, online users are increasingly using suggested movies for a variety of purposes, includes making purchases. When contrasted to the conventional manner of buying, suggested movies give customers more effective ways to learn about the products, such as their cost, accessibility, vendor, and even the production procedure. Internet users now find it challenging to make fast and correct decisions due to the abundance of knowledge available to them (such as details saturation) [6].

Therefore, the availability of a wide variety of items is one of the primary issues with online shopping. Customers find it tedious and challenging to review every single one, which might make it impossible for them to make a decision. Consequently, they discontinue purchasing from the website,

start visiting rival websites, and eventually, prospective buyers decide not to make any purchases there. Thus, one of among the most crucial elements of suggested films and web-based businesses is the customization of items and offerings [7].

However, as consumer loyalty in suggested movie environments especially those that involve internet retailers is far more nuanced and delicate than it is in conventional company settings, suggested film organizations must create techniques like person-to-person recommenders. Therefore, it is imperative that suggested filmmakers set up an individualized making choices support system. This is because, with the right knowledge at hand, users can make decisions more quickly and satisfactorily, which increases the likelihood of retaining customers, fostering or strengthening their loyalty, and generating a profit [8].

Customized systems for suggestions, or RS, are one type of computerized system that helps users make decisions online. Its primary goal is to filter knowledge to reduce the amount of data consumed and enhance user-system relationships by offering customized suggestions. Offering tailored offerings to each client in order to discern their unique tastes and satisfy their demands is known as customization, and it's one of the most successful marketing techniques. One technological tool that aids in the creation of customized characteristics is the recommendation system. Analyzing the qualities and prospective interests of clientele is a promising and novel area for RSs [9].

Recommending Systems analyze consumer purchase patterns and draw conclusions from them in an effort to suggest appropriate products, assisting the customer in selecting the best alternative. among the most well-known suggestion systems on Amazon.com, for example, uses the data page for every title to supply specifics regarding the text and buying knowledge in addition to two book suggestions that are frequently obtained by those who obtained the current edition and an introduction to the researchers whose works have been regularly picked up. Another recommendation system that allows users to rate the degree of contentment for buyers as well as sellers is Ebay.com [10].

The ability to assist users in directing vast volumes of knowledge is one of the objectives of recommendations systems. Recommendation systems boost movie ticket sales by turning visitors into purchasers, showcasing new products, building client loyalty, raising satisfaction among clients, and raising the likelihood that a happy consumer will return. Research indicates that recommendations are an essential component of electronic commerce, with websites utilizing individualized recommendation systems can boost sales by up to 38% via suggested products [11].

One of the most popular and effective recommendation systems amongst these is the collaborating filter (CF), which is regarded as one of the essential elements of effective suggested movie algorithms. Even though CF is widely used, it has a number of drawbacks, such as cold start, flexibility, sparseness of the customer-item matrix, and consumer taste changes with time [12].

Some of the most significant issues with CF is cold begin, where the system hasn't collected enough data to offer reliable guidance yet. The most frequent scenario is when the software is unable to accurately suggest any readily accessible things to an unfamiliar user (the novel user's difficulty) or suggest something new to current users (the newly product's issue) [13].

Scalability issues arise when a system must devote a lot of money and time extracting commonalities across every client, even though there is a large amount of customer-related information and a wide range of things available. This reduces the system's efficacy and effectiveness. The final result of CF in neighborhood-based CF methods is significantly influenced by the similarity between users. Furthermore, a limited user-item matrix may result from a broad range of items if transactions or achieving are insufficient to calculate parallels and anticipate suggested items. In this scenario, the algorithm may perform poorly when it comes to accurately recommending items or effectively identifying comparable users [14].

As a result, one of the primary issues with RSs is that sparsity makes it difficult for them to make exact suggestions. Considering the problem as a matter of time is crucial when dealing with all of these issues because it affects CF systems greatly because user preferences may alter over time. Recently made purchases by those who are the target audience are more indicative of their tastes and desires than previous purchases, according to a number of investigations, and these purchases also allow us to make more accurate forecasts about what they will buy in the future. These restrictions typically prohibit recommendation systems from achieving their main objective, which is to suggest suitable things to suitable people [15, 16].

Accurate forecasts alone are insufficient to assess achievement. Additionally, competitive bids ought to present a variety of beneficial goods that cater to the preferences and inclinations of many people. Regretfully, the constantly changing user profile combined with the recommendation algorithms' lack of diversification and stability have grown into a major obstacle. Although numerous research has been conducted to address these issues and enhance CF, no comprehensive framework exists at this time to address or mitigate every one of these issues. Thus, the primary goal of the current work is to propose a new CF recommendations system with increased accuracy in recommendations settings by using a set of data mining techniques. The K-nearest neighbor (KNN) technique, which is the most often used CF technique, is improved in this framework that used by CF [17, 18].

To do this, consumers will first be divided into sections according to the LRFM (the regency, rate, financial, and duration) parameters (at the stage of the good group). Following this, every one segment's cart contents will be analyzed using the association guidelines depending on the user-class the framework. The item classes anticipated for each target consumer as a result of categorization and association regulations will be used as the starting point for the CF. In order to minimize the matrix's size based on the outcomes of segmented and association regulations, the CF system modifies the matrix of user items in a different manner depending on the time variable, which includes the time of buying it and the time of providing the product on the web page [19].

Based on the aforementioned concerns, the primary objectives of this study are to: determine the limitations of conventional CF in terms of suggesting suitable products to suitable users; visible an extensive and novel recommendation engine to do away with the constraints of traditional CF; and evaluate the suggested product system using actual data in order to produce suggestions that are more precise and in line with customers' preferences.

Two main approaches can be used to categorize various types of tailored recommendations.

1. Content-driven Methods: The target user receives suggestions based on how similar the suggested components are to formerly favored components. Content-driven filtering, which depends on personas created with retrieved characteristics, has shown to be quite successful. This guarantees that suggestions can be rapidly altered in accordance with what the user wants [20].

2. Collective filtering methods: These methods recommend items that have been liked by people with similar preferences to the target user. A user-item matrix, or database of feedback provided by users for various items, is utilized in shared filtering. Next, recommendations for products that other users with similar tastes have favorably reviewed are sent to the users. There are two main kinds of filtering that are collaborative [21]:

A- Memory collaborative filtering of methods

Collective filtering methods, or memory-based techniques, use the complete set of previously evaluated items to produce recommendations according to feedback from consumers. The fact that these techniques save and use the initial rating information straight from memory while generating suggestions is how they got their moniker [22]. Memory-based techniques include the following steps:

1. The process of determining how comparable an intended consumer or object is to other users or products that are included in the evaluation matrix. The region is a set of related users or items that show common interests or traits. This area is determined by the method of calculation.

2. Choose a region that is comparable to the intended user or object. The K-nearest-neighborhood method, where K is the total number of closest neighbors, is used to accomplish this.

3. After the surrounding area has been determined, calculated averages of the evaluations from this particular group are used to provide forecasts.

4. The application of a Top-T recommendations algorithm comes last. The items that should be suggested are determined by combining the user-item matrix of the k most comparable users that were found and selected [23].

B- Model-based filtering

By understanding the fundamental elements of the approach, modeling-based strategies leverage techniques from machine learning and statistics to forecast a user's preferences for an item. This approach offers multiple benefits over memory-based approaches. Situational simulation, synthetic neural networks, trees of decisions, grouping, and Naive Bayesian are some of these methods [24].

When contrasted with memory-based techniques, model-based cooperative filtering recommendation systems provide a number of benefits [25], as mentioned in:

1. Adaptability: By creating models with fewer variables than the initial dataset, techniques based on models improve performance and decrease runtime complication. As such, they can be used to solve scaling and sparseness problems.

2. Rapider Forecast Speeds: Since it takes fewer seconds to search the predictive model than the complete dataset, modelling-based solutions are typically faster.

3. Model-based approaches are thought to be more accurate than memory-based structures in terms of precision. This is particularly valid in the case of hidden factor systems. Memory-based devices are easy to create, but their precision is frequently lacking.

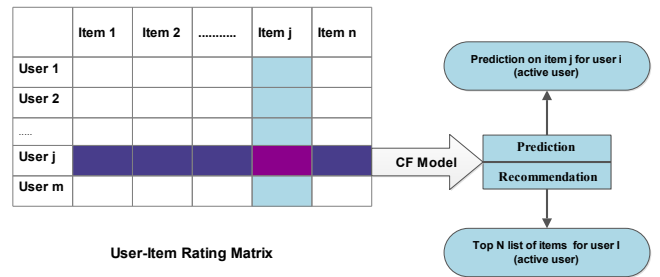


Figure 1. Shown collaborative filtering process [25]

In collective filtering, the two most used metrics for resemblance are cosine-based and correlation-based techniques. The amount of linear association among the two parameters is ascertained using the Pearson correlation factor. The relationship between users 'a' and 'u' in the user-based similarity assessment can be computed, as shown in Figure 1, which shows the teamwork filtering procedure.

Collective filtering is useful in areas where there is little content about items, especially beliefs, and where machine understanding is difficult. Even in the case where individual profiles are devoid of significant material, it can nonetheless propose items. Nevertheless, the cold-start challenge, which appears when consumers have blank accounts because they have not evaluated any things and the system does not know their tastes, can impair recommending system efficiency [26].

One possible solution to recommendations algorithms' cold-starting problem is to use a hybrid method. This entails combining collaborative filtering approaches with other approaches like knowledge-driven, centered around content, or demographics filtering. Methods of active learning can also be utilized to ask users to score or comment on particular items. By utilizing data from outside sources or areas connected to the target consumers and objects, transferring knowledge approaches also show value. Ensemble approaches use complementary models to further improve system durability and efficiency [27].

The organization of this manuscript is as the following. Section 2 discussed the relevant related work. In section 3 discuss the model-based recommendation algorithm using Naive Bayes. Section 4 describe the proposed Naive Bayes - based movie recommendation system design and deployment. In section 5 we present the performance evaluation. Section 6 describes the recommendation performance evaluation. Section 7 illustrates the conclusions.

2. RELATED WORK

The naïve Bayesian suggestion method was employed in a study by the study [28] to enhance movie suggestions by taking into account consumer tastes and the features of films in the Movie Lens dataset. The aim of the project was to precisely characterize the movie-watching interests of users and suggest films that suit their tastes. The naïve Bayesian system for suggestions was evaluated using arbitrarily selected user suggestions, and the results showed that it was highly accurate (0.8321), precise (0.9117), and recall (0.87436).

Developed a collaborative filter-based technique in the study [29] that improves object learning. By determining suitable and improper situations, this technology aims to improve the way recommendations are made by offering educational resources that take into account each user's unique

context. To make instructional material relevant to the specific fields of study, it is created, categorized, and revised. Together with metadata that may be used in future instances, these educational objects are kept in a library. The tool has a number of features that help writers assess the reliability of instructional materials. The main goal of this project is to apply collaborative filtering approaches to improve the process of generating, editing, and versions learning objects. Additionally, user behaviors are being observed to enhance the procedure for making recommendations.

CRecSys, a suggestion system, was created in reference [30]. In order to anticipate objects to unrated movies, this method uses contextual information to compute the semantic correspondence between films. Network-matching and collaboration filter-item-based algorithms are used to construct an RDF graph by extracting contextually object-based attributes from a dataset. We used a dataset of 1,400 movies from well-known movie databases including IMDB and the Rotten Tomatoes website. The film's data was divided into two sets for the purpose of the research. Experiments showed that the suggested approach successfully tackles concerns associated with beginnings that are cold and inadequate coverage in conventional systems for suggestion, as well as certain enduring difficulties. A number of test measures that regularly showed favorable results were used to validate the system's performance. Ten classic recommendation techniques and two more recent models were compared, and the results showed that the suggested approach performed better.

An effective context-aware recommending model that acts as an actor in the recommended model and ensures that the suggestions that are made have the least amount of mistake was recently presented in the study [31]. The investigators used crawling technologies to prepare the data, removing unnecessary terms from user evaluations that they had collected from web platforms. The recommendations model performed much better once they extracted contextual characteristics from these reviews. The Artificial Language Tool Kit was used on the Python framework for analyzing data. The reviews underwent preprocessing, and their text was merged. The density-based combination of the user's feelings, comprising favorable, neutral, and unfavorable was performed using this combined characteristic as input. The most preferred users from each emotion category were then determined using a Deeply Recurrent Neural Network system. The NYC Restaurants Rich dataset was used to establish and test the machine learning model's variables in order to put this strategy into practice. The model's efficacy was assessed through contrasting it to other models and concentrating on measures of trust and recall. Compared to the deep learning models that have been examined thus far, our suggested model has a precision level of 99.8%, indicating a comparatively greater degree of precision.

A system of suggestions that forecasts ratings by combining environmental information and deep learning algorithms was shown in the study [32]. Their method comprises using a sophisticated encoding and a neural network to effectively capture features for score forecasting. To reliably predict user preferences, the framework considers context, item, and personal data. For their investigation into this suggestion system, the investigators used three records: CARSKIT, Concordia Film, InCarMusic, and Restaurant Tijuana. As for the results, the investigators showed that in datasets containing several contextual elements, their method produced

predictions with a higher degree of reliability than existing systems for recommendation. In smaller databases, the accuracy ratings ranged from 0.03 to 0.09 and revealed distinct results for several categories, including films (0.01), restaurants (0.07), and songs (0.06). Since it was quicker than KNN but slower in contrast to other already available approaches, one drawback of this concept is that it responds more slowly than other systems for recommending things.

A deep learning system named TACDA was presented in the study [33] with the goal of enhancing the precision of Top-N suggestions by the integration of aware contextual states into an encoding system. The investigators successfully handle cold start concerns and sparseness by including trust data into their auto encoding module, which addresses issues like contextual noise mitigation and user scant preferences. To provide improved recommendations, converts each label set into a Multi-Label Classification (MLC) problem [34]. Experimental results show that the proposed ANN performed better than previous approaches on the Binary Resonance (BR) Instance-Based Classification algorithm, BR Decision Trees, and Multi-label SVM on the itinerary Adviser and LDOS-CoMoDa Dataset. Additionally, the recommended ANN operates better than cutting-edge methods, increasing accuracy by 1.1 to 6.1 percentage points.

In order to calculate similarity between users and on the features of the video suggestion system in an effective manner, Li et al. [35] developed the approach known as PCA for systems that recommend videos. Enhancing the efficacy of a context-aware recommendations model through collective filtering utilizing the analysis of principal components (PCA) is the main objective. The collection includes basic user data along with 14 contextual components. Only six environment parameters are employed since adding more will increase the system's complexity and processing time. Of the information, 25 percent is utilized to evaluate and 75 percent is used for training. Three distinct measures were used to assess the effectiveness of this framework: the median absolute error (MAE), mean squared error (MSE), root average square error (RMSE), and recall, precision, and the f measurement. For the suggested approach, precision is 94.74%, recall is 79.63%, and the F measurement is 88.67%.

In an effort to create an integrated representation of the similarity of users for CF, Deng et al. [36] have suggested a combination of models to evaluate the similarity among users in a thorough and impartial manner. A resemblance requirement for items determined by the divergence measure Kullback-Leibler (KL) has been developed in the proposed model, and it serves as an indicator of weight for the resultant rectification of a proximity-significance-singularity framework. Simultaneously, the suggested model takes into account an imbalance component and a user preference factor to differentiate the scoring priorities of various users and enhance the output algorithm's dependability. The suggested user model for dispersed data is suitable and enhances the caliber of suggestions and prediction precision in an efficient manner.

Iftikhar et al. [37] have investigated the function of clarification in RS in their studies and conclude that a rationale serves as a justification for the data pertaining to suggestions and is crucial in assisting users in assessing the suggestions they receive regarding online shopping, social media, search engine, and other website-related websites. Three primary topics of the study are addressed: the goals and assessment of explaining efficiency, the effect of explaining relationships on

consumers, and the increasing influence of reasons on social media platforms.

A fuzzy mixed technique has been suggested by the study [38] in an effort to broaden the range of top suggestions. according to the profiles of users, the hybrid fuzzy RS proposed in this study can produce a variety of suggestions. This approach uses a fuzzy-based CF in conjunction with the content-driven filtering technique, allowing users to be assigned to distinct groups due to fuzzy and ambiguous user profiles.

Alhijawi et al. [39] have carried out research with the goal of offering a useful suggestion based on consumer preferences for a precise product selection. This study makes the case that consumers subtly convey their opinions regarding particular product attributes. Customer preferences based on product characteristics can be combined in some favorite matrices that is shown. Furthermore, he has employed weighted associations to identify recurring trends in product purchases in order to enhance the caliber of suggestions. The study's findings demonstrate that the suggested approach minimizes the sparsity issue and outperforms the other algorithms.

Margaris et al. [40] made an effort to solve the scarce user-item matrices issue in order to enhance memory-based CF techniques. They populated the matrix by estimating similarities between different components based on their attitudes (song title, publishing data, vocalist, duration, and so forth). A banking RS that categorizes user preference histories while ignoring inter-user choice consistency. They provided suggestions to every user according to the defined classes. used a combination of CF techniques and user demographics segmentation to suggest films. They used a weighted additive function for combining the user grouping similarities and the CF similarities in order to forecast the choices of new customers.

Sallam et al. [41] developed a model wherein the consumer's rating matrices and the matrix derived from personal data provided by users are combined to improve tourist RS. An improved surrounding region for consumers is defined by a newly developed resemblance operation, which improves the quality of the suggestions. a recommendation engine for internet merchants selling computed digital goods. Using the RFM framework and comparable relative choices, they segmented the consumer base in this method (determining the FM parameters at the scale of product class). Then, they made it possible to suggest items to customers on two different levels of item categories by utilizing CF techniques and regulations for associations.

Logesh et al. [42] suggested a novel paradigm for RSs on an online book retailer. In this model, the technique known as K-means is initially used to categorize consumers based on their demographics. They then gave the users in that group suitable suggestions based on the relationship rules that were retrieved for each cluster. The tourist attractions in this framework were initially categorized into five groups. The user-item matrices of the CF method were then used to assign ratings to these activities, which allowed the user to identify which attractions were most attractive to them. Next, a list of destinations is suggested using the CBF approach, which is based on how similar the sites of interest are to one another.

Aljunid and Dh [43] enhanced CF frameworks for the travel and tourism sector by using users' location data. With the usage of this data, they gave the users two different kinds of suggestions: one kind was generic and was based on the area's well-known spots, while the other kind was personalized

according to on the user's hobbies and inclinations. Working together, we can improve the users' tailored suggestions by modeling their interests. The investigation's phases were:

1. Compiling input from users.
2. Simulating user preferences (film industry).
3. Using KNN to anticipate user interests.
4. Formulating suggestions for the user.

Using this method, they first used data mining to find relevant and helpful characteristics for the user, followed by they collaborated with users who were similar to them to develop the customized model.

Li and Ye [44] offered a film recommended system in which individuals with comparable tastes were found using an organizational approach, and each user received appropriate suggestions based on their cluster and location. They did this by concurrently segmenting their consumer base using the matrix of user items and the RFM algorithm's parameters. Appropriate suggestions were given to each segment's clients according to the rules that were retrieved from the sections. Based on the backdrop, numerous research projects have been conducted to improve the RS by addressing the issues of chilly start, dense matrix structures, flexibility, and shifting client interests. These can be evaluated from a number of angles:

This research conducted audience segmentation using demographics data; some of the investigations broadened and refined RSs in a variety of industries, including movies and films.

- Following the cluster identification process, a subset of these experiments retrieved the relationship rules associated with each cluster independently and made product recommendations to customers inside that cluster only using these guidelines. On the other hand, a different set of research improved the computation of client characteristics by utilizing the clusters derived by categorization, which in turn enhanced the efficiency of CF algorithms.

- The item's classification approach was also applied to studies such as movies. Based on the ratings these classes receive from the users, these studies complete the user-item matrix, lower the number of products, or ascertain the resemblance of clients.

- By merging CF and CBF structures, hybrid RS has been used in a few other research [45].

- Lastly, a few studies have taken into account the temporal dimension in relation to how users' interests evolve over time. These assessments show that although numerous research have been conducted to enhance CF suggestions, only one or two issues have been addressed in each study. Stated differently, no all-encompassing model has been developed to date for addressing or lessening the fundamental constraints of CF systems. Since online clothing sales are relatively new in Iran, establishing RS for this kind of online shopping is essential to raising consumer happiness and loyalty levels and, ultimately, driving up sales.

Thus, in order to greatly enhance the efficiency of CF structures, the current work attempts to address issues with cold begin, dense matrices, flexibility, and the evolution of consumer preferences over time. It does this by enhancing the technique of KNN via:

1. Segmenting customers using LRFM characteristics at the item classification to assess the Duration of the Relationship with the Client and RFM of that item's group sales.
2. Obtaining rules for association at each group by separating them from user-item categorization matrices.

3. Segmenting clients according to demographic characteristics.
4. Decreasing the length of the matrix of user items and altering it according to the time variable.
5. Using the weighted total of the segmented and CF technique findings to create a new resemblance value [46].

3. MODEL BASED RECOMMENDATION ALGORITHM USING NAIVE BAYES

The two types of collaboration approaches to recommendation that are most frequently distinguished are memory-based and model-based. The conventional method, referred known as user-based, is classified as storage-based since it generates suggestions by directly accessing the evaluation database that is kept in recollection. Model-based methods, on the other hand, treat the initial information offline prior to execution using methods like item-based filtration and decreasing dimensionality. Only the precomputed or acquired framework is required to make recommendations during runtime. Many model-based methods exist for item recommendation, such as Boltzmann Equipment, neural networks, Latent Semantic Evaluation, Latent Dirichlet Allocations, Greatest Entropy, also and Solitary Value Decomposition [47].

The difficulty of learning the corresponding functional from a set of training examples can be viewed as the assignment of classification. Informally, this role is frequently referred to as the categorization model. A traditional method that's frequently used in data mining uses Bayes classifiers. The results obtained from samples are regarded as being distinct and have the same probabilities for different random variable patterns. The Naïve-Bayes hypothesis uses an occurrence's recurrence rather than its chance of happening. Moreover, each sample feature has independent between many variables [48].

Using user data analysis, the subsequent probabilities $P(A|B)$ is computed. User data includes a variety of characteristics, including name, gender, age, profession, and favorite movie genre. The qualities of a user are represented via a vector formed by these properties. The sample set must take into account every conceivable arrangement that includes these properties in order to calculate the subsequent likelihood. All of the likelihood values are going to be zero if there is no characteristic connection between them. The study's attribute requirements are assumed to be separate from one another by the Naive Bayes categorization. Eq. (1) shows how to compute the posterior likelihood under this presumption, taking into account the fact that each person's values for attributes are independent of each other [49].

$$p(a \setminus b) = \frac{p(a)p(b \setminus c)}{p(b)} = \frac{p(a)}{p(b)} \prod p(b \setminus a) \quad (1)$$

Taking x as an identifiable variable and presuming that each of the defining characteristics of each specimen are uncorrelated, we calculate the chances of subcategory dependent on x and classify the grouping with the highest likelihood. Naive Bayes's categorization principle is based on its fundamental idea. It becomes essential to examine user activity data, such as application rate, in systems that offer suggestions. We can recommend films for a specific user by developing classifiers based on this data, which computes a

function of predictions that determines the likelihood of being appreciated by that user.

$$y = f(x) = \arg \max p(c) \prod p(x \setminus c) \quad (2)$$

A combination of many sample attributes is represented by $\mu p(x \setminus c)$. It is possible to regard the qualities and values as independent of one another. The largest value, represented by the c in $P(c)$ [50], must be determined.

1. Hybrid Side Knowledge: To get the optimum efficiency of the system and avoid the drawbacks and restrictions of systems that offer suggestions, this technique integrates a variety of suggestions. This system integrates multiple algorithms to provide predictions that are more precise and useful [51].

2. Context-Aware Recommendations Systems: An Overview These algorithms make advantage of extra data to give the user the best possible experience. "Relevant knowledge" refers to data that describes the environment in which human actions and system elements take place. To obtain setting, there are three approaches [52]:

- By giving precise answers to questions.
- Using the user's precise location as an instance of implicit contextual data GPS, which can be obtained from a smartphone.
- Integrating data mining and statistical methods.

4. THE PROPOSED NAIVE BAYES-BASED MOVIE RECOMMENDATION SYSTEM DESIGN AND DEPLOYMENT

Since there are a lot of movies on each website, making it hard to select the greatest ones already out there. We propose to grade movies with consideration for user circumstances by utilizing the Naïve-Bayes system. utilizing the assessment in the study to model the collective probability distribution of several random elements and develop a movie choice strategy based on the various user situations. As seen in Figure 2, the suggested model was divided into two stages: the training stage and the testing stage.

In this stage, a collection of data of LDOSCOMODA Film consumers that completed retraining was used to create the suggested model offline. There are thirty-two variables in the dataset, 14 of which are situational factors that characterize the environment in which a movie is watched as see Table 1. Furthermore, the 20 parameters offer description about the material itself (e.g., a director, region where film was created) and generic customer data (e.g., age, sex), position (town and region). The languages used in the film, the year it was published, the genre(s) it corresponds to (genre1, genre2, and genre3), and facts about the three principal performers and actresses engaged in the film's making are all included in this database. Additionally, every film has a budget parameter added.

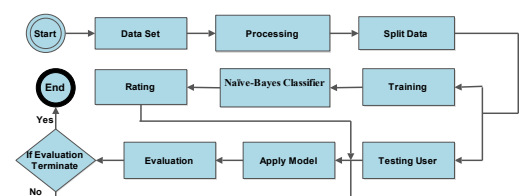


Figure 2. Flowchart of proposed model

Table 1. Context-related information

User ID	Item ID	Age	Sex	City	End Emo	Time	Day Type	Season	Location	Weather	Social	Domin Emo	Rating
32	11	34	1	30	2	3	2	2	1	1	1	3	5
34	7	25	1	20	3	2	2	2	1	1	2	4	3
31	8	29	2	20	3	4	2	2	1	1	2	2	4
32	12	29	1	20	2	3	2	3	2	2	3	3	4
21	7	29	1	10	3	4	2	2	1	1	2	3	3
28	14	32	2	30	3	3	1	3	1	1	7	2	5
31	17	29	1	10	3	3	1	2	1	5	1	5	3
17	13	32	1	20	3	3	1	3	1	1	2	4	4
29	19	32	1	20	2	2	2	2	1	2	1	3	4
22	16	-1	2	3	5	4	1	3	1	1	2	3	5

Table 2. The LDOS-CoMoDa dataset: A brief description

Variable Name	No. of Category	Missing Value Ratio	Description
Time	4	0.019	1-morning, 2-afternoon, 3-evening, 4-night
Day type	3	0.017	1-working day, 2-weekend, 3-holiday
Season	4	0.019	1-spring, 2-summer, 3-autumn, 4-winter
Region	3	0.018	1-house, 2-public space, 3-friend's residence
Climate	5	0.023	1-sunny/clear, 2-rainy, 3-stormy, 4-snowy, 5-cloudy
Social	7	0.015	1-single, 2-partnered, 3-buddies, 4- coworkers, 5-grandparents, 6-accessible, 7-relatives
End Emo	7	0	1-depressed, 2-content, 3-alarmed, 4-astonished, 5-furious, and 6-offended, 7-Ambivalent
Prevailing Emo	7	0	1-depressed, 2-content, 3-alarmed, 4-astonished, 5-furious, 6-dismayed, 7-Ambivalent
Mental State	3	0	1-unfavorable, 2-unbiased, 3-beneficial
Physically	2	0.024	1-well, 2-unwell
Choice	2	0.023	1-healthy, 2- ill
Collaboration	2	0.022	1- user-chosen, 2- provided by another

An intuitive website created especially for this reason was put into place as soon as individuals had completed viewing the targeted movies using the database in order to collect feedback and contextual information. In the first step, ratings for films are determined by 14 contextual elements. General information about users, material, and documentation are the other 20 elements that determine whether to proceed to the recommended step. Table 2 describes the LDOS-CoMoDa information.

The first set of ratings, which are either explicitly stated by users or inferred from them, starts the suggestions procedure. The approach suggests estimating the categorization functional when these preliminary categories are established.

The grade of R movie (ALI, Titanic)=4 denotes that ALI Intelligence gave the film "Titanic" a 4 (out of 5) rating.

R: User×Item (Movie) Rating →

The relationships between consumers and their evaluations of films were stored in a matrix-like data arrangement. For the combinations (user, item) that individuals haven't yet rated. The attributes of the object and the user's demographics are used to address the issue of item categorization. Assessment is a logical collection of actual numbers within a given range, and R is a two-dimensional shapes association or matrices.

R: User×Item (Movie)×Context Rating →

In this instance, rating is based on the element's and the material's historical context.

Table 2 displays the historical data as follows:

(1) User surroundings (sentimental levels).

(2) Dependent Assessments (connected to the surroundings, including time, place, and climate).

Ratings are defined as follows: (place, the climate, length of sentence, sociability).

User: referred to as User (User Identification, Name, Sex, Age), this refers to the individuals to whose movie recommendations are sent.

Relevant data can take on a multitude of forms depending on the differences.

Algorithm 1 illustrates the phases of the suggested strategy, and Figure 3 displays the instruction phase of the suggested model.

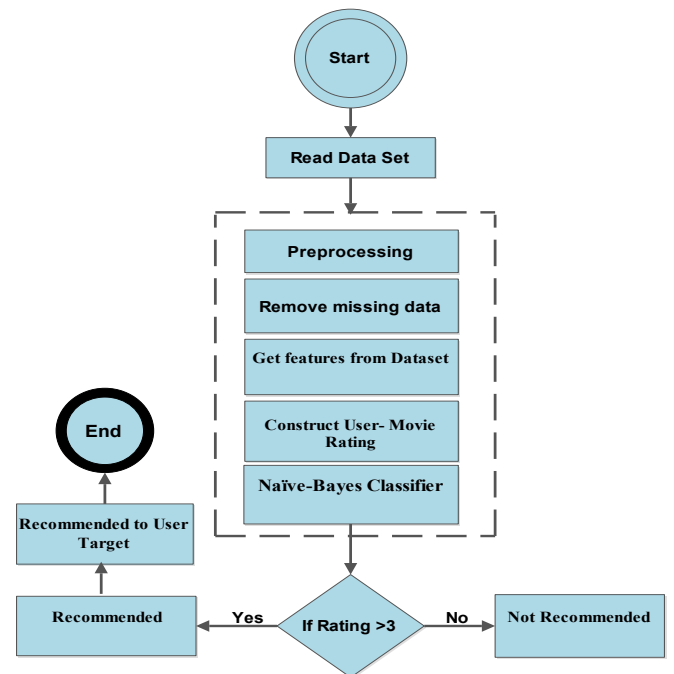


Figure 3. The flow diagram of the proposed training system

Algorithm (1) A Proposed Recommender System

Input: The array-formatted LDOSCOMODA dataset

Output: Assessing for User Objective

Begin

Bring in the required modules and frameworks.

Use python to generate the CSV file.

All instances of -1 in the uploaded data should be replaced with 1.

Prepare the characteristic and variables to be targeted by removing particular information columns.

Use the train assess split() function in learning to split the information into a 70% training set and a 30% testing set.

Set the PCA object's initial four components.

Utilizing the PCA object, fit and modify the training set.

Utilizing the PCA product, modify the experimental data.

Set up the object for the Gaussian Naive Bayes algorithm.

Utilizing the experimental data, fit the framework.

Utilizing the model that was trained, predict the class labels for the testing data.

To evaluate the model's predictive effectiveness, use sklearn's f1 score, reliability score, precise result, and recall rating algorithms.

Utilizing the corr () method from pandas, determine the relationship between the variable 'rating' and each of the columns in the data.

Output the 'rating' and the corresponding correlation factors for every column.

End

4.1 Data preparation

Data preliminary processing, or gathering data, is a crucial stage in creating regression models (RSs) and accounts for 65-75% of the overall time needed. Three phases of data transformation, combining data, and data cleansing are involved in this step.

4.1.1 Data cleaning

Correcting missing information, locating and removing outlier's data, and handling disagreements over data are all part of the data purification process. Furthermore, it is vital to remove any irrelevant or unnecessary fields or characteristics at this point, and if needed, create additional variables from the original dataset. At this point, the amount of data information drop from sixteen thousand to 15,472, and the total amount of clients drops from 3,742 to 3,068 as insufficient data and clients with inaccurate data are removed from the primary dataset. The RFM parameters are computed for every product category.

4.1.2 Data integration

The second step in collecting data is the incorporation of data, which is the process of merging and merging multiple datasets in order to better comprehend the data and produce a better scientific administration for them within each dataset. At this point, the profile data of the customers—including their

gender, age, marital status, level of learning, and employment—was combined with the data from their purchases.

4.1.3 Data conversion

At this point, the data needs to be transformed into a format that can be used for data mining and RS design. In the collection of data, character elements like gender are transformed into numerical parameters and results.

4.2 Features extraction

The LRFM model (length, recency, frequency, monetary, and volume) is a simple feature but powerful tool for market segmentation. Table 3 shows the evaluation of these features for a sample of customers. LRFM analysis will segment the customer base and maximize the purchase response rates of marketing efforts, according to this paper. LRFM research enhances market segmentation by looking The LRFM model. According to the study, customers who had recently invested a lot of money and purchased a lot of items were much more likely to react to potential promotions. As a result, the scope of LRFMV research has been broad. Through length, recency, frequency, monetary and volume, the customer relationship matrix assists management in identifying the characteristics of four different types of customer traits. The volume highlights the customers who provide more profit to the organization as their buying habit is larger than any other customer segment. where, the length (L) indicated to the Time since first purchase or account creation. was calculated for each agent from the equation as follows,

$$L = pl - pf \quad (3)$$

The Recency (R) indicates the days which exist after any valuable customer's last purchase to find the irregularity following that visit. The Recency (R) value was calculated. Mathematically, If the most recent date of the dataset is denoted by Dr and the last purchase date of a particular customer is Cr then Recency, R can be calculated as,

$$R = Dr - Cr \quad (4)$$

Frequency (F): The number of purchases made by a customer in a customer life cycle is referred to as frequency. Counting the number of times, a customer purchased any service from the superstore yielded the Frequency (F) value. Mathematically, If the purchase for a customer is denoted by pf then Frequency, F for that particular customer will be,

$$F = cont(pf) \quad (5)$$

For the Monetary (M), its number of transactions and total expenditure were essential to calculate the monetary. In a customer's life cycle, the total amount of his complete transactions has been divided by the number of those transactions to find the monetary value. Mathematically, if the total spending on purchasing of a customer is ps, x is the total number of transactions denoted here. Monetary, M can be calculated as,

$$M = \frac{\sum_{n=1}^x ps}{x} \quad (6)$$

Table 3. Features for the LRFM model-based client categorization

Cust.ID	L	R	F	M
AB-10225	1362.0	8	19	732.751
AB-10235	1344.0	6	23	256.965
CC-10225	1232.0	125	20	1020.855
CC-10225	1407.2	28	36	459.452
YS-10225	1300.0	2	7	325.755
YS-10225	1300.0	9	26	806.952
YS-10225	12.0	199	1	7.152
ZZ-10225	1192.0	4	37	789.605

4.3 Segmentation based on the LRFM model

This work uses a two-stage grouping method for division, integrating K-means (which creates clusters) with SOM (which determines the appropriate number of groups). Client segmentation is done at this step using the L factor (length of the client connection) and (recency, frequency, monetary) factors of categories of products (the amounts for RFM characteristics at each group of products are determined individually) as features extraction of these factors for a specific consumer, depending on which segmentation procedure is done.

LRFM method was employed in the present investigation for three reasons, including the reasons that follow, for client segmentation at the item group level:

1. Improving the suggested method's effectiveness and productivity by lowering the user-item matrix's client count, which will lessen the matrix's limited space and problems with scalability.

2. This model can accurately represent the tastes and interests of the clients in addition to determining their value. For example, if two consumers have similar buy frequencies for a particular product category, it can be assumed that both of them are equally passionate about acquiring these goods. M and Rare's factors also suggested the same thing.

3. The group of products in the framework, or factor R, takes into account the customer's most recent choices and tastes. Since recent expenditures more accurately represent the clients' present desires and tastes, it will make suggestions that are more precise and better aligned with the client's present passions.

4.4 Extracting association rules based on product category-user matrix

This work uses the a priori method to extract the principles of association. Using financial information from product group-user matrices (the frequency with which a customer makes purchases from a certain product group), the rules for association are retrieved at the level of every cluster of the preceding stage in this step. For every target consumer, a list of the Top-N suggested categories of products is generated using the extracted criteria. Each suggested product category's priority is determined by the rule's approval and trust metrics. Product categories with higher trust as well as backing scores will be given preference when it comes to recommendations. The anticipated types of goods for every one of the intended clienteles, which are the result of the gathered rules in this step, will serve as the phase's input as indicated in Table 4. Reducing the amount of goods in the matrix of user items is the goal of extraction rules at the item category level, which

successfully mitigates the matrix's limited space and scalability issues as shown in Figure 4.

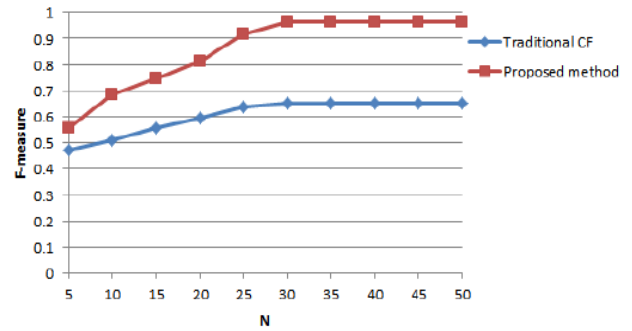


Figure 4. Contrasting the effectiveness of the suggested approach with the usual approach for various values of N

Table 4. Association rules based on product category-user matrix

Item ID	1	2	3
1	(LT ₁ ,PT ₁)		(LT ₃ ,PT ₂)
2		(LT ₂ ,PT ₃)	(LT ₃ ,PT ₄)
3		(LT ₂ ,PT ₅)	

4.5 Changing the user-item matrix

User-item matrix structures, comprising binary information on products purchased or not, or user-assigned ratings to items, form the foundation for consumer preference grouping in CF algorithms and help identify the target consumer's neighborhoods. However, because online shopping items vary widely, this matrix frequently encounters the issue of scarce data. Research indicates that the number of items of the user-item matrices influences the quality of suggestions. The outcomes of the classification and association criteria stages are utilized in this matrix to make it thicker. Specifically, the elements in the item-user matrices used to find the target clients relatives will only be those anticipated for the intended consumer, and the target consumer will only be contrasted with users in their own cluster. This method reduces the number of objects and users, which lessens the scalability issue and improves RS speed as well as precision. Thus, according to Table 5, the matrix of user-items and two parameters the time a user bought a product (PT) and the period when it became accessible for purchase (LT) are taken into account for each item.

In order to calculate the evaluations for the PT and LT factors, these two characteristics are initially split into five groups according to their latency in this modified matrix. Purchases made in past generations (score 1), purchases made in the past (score 2), purchases made with ordinary frequency (score 3), purchases made recently (score 4), and purchases made most recently (score 5). These are the results for the PT factor. The scores are comparable for the LT factor. According to Table 6, an amalgamated score is utilized in the matrix of user items in place of the user-provided scores. The result for an arrangement of (2, 5) is 7 (five plus two). Both of these characteristics have the same weight when determining the hybrid score.

As a result, 25(5×5) double possibilities are made, with every client represented by an assortment of 55, 54, 53, 11; the

set of numbers that best fits the client's needs is 55 (which represents the group of products for the latest launch time and purchase time) plus a score of 10. The client's poorest category, which is represented by 11 (the earliest launch and purchase dates) and gets a score of 2, is as follows. Clients can be divided into groups and their degree of resemblance can be computed using this scoring approach. This new customer-item matrix was designed with the intention of improving suggestion reliability over time by taking customer preferences into account.

Table 5. Modifying the matrix of user items in light of temporal data

ID	L (days)	R (days)						
		C _{1R}	C _{2R}	C _{3R}	C _{4R}	C _{5R}	C _{6R}	C _{7R}
1890	95	-	20	-	-	97	53	-
ID	L (days)	F						
1890	95	C _{1F}	C _{2F}	C _{3F}	C _{4F}	C _{5F}	C _{6F}	C _{7F}
		-	3	-	-	1	3	-
ID	L (days)	M(\$)						
		C _{1M}	C _{2M}	C _{3M}	C _{4M}	C _{5M}	C _{6M}	C _{7M}
1750	95	-	59	-	-	15	39	-

Table 6. Scoring the LT and PT parameters

PT LT	1	2	3	4	5
1	2	3	4	5	6
2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10

4.6 Calculating similarity based on CF

There are numerous ways to determine how comparable a target client is to other consumers in CF networks. When contrasted to other conventional statistical approaches, the Pearson's correlation value yields the best forecast and suggestion results, making it one of the most widely used methods [53]. Table 7 indicates the Contrasting the suggested system's assessment metrics with those of the conventional CF recommendations system. Eq. (7) is used in the Pearson's correlational approach to calculate the similarities between two given consumers, u and u' :

$$Sim(u, u') = \frac{\sum_{i=1}^n (r_{ui} - \bar{r}_u)(r_{u'i} - \bar{r}_{u'})}{\sqrt{\sum_{i=1}^n (r_{ui} - \bar{r}_u)^2 \sum_{i=1}^n (r_{u'i} - \bar{r}_{u'})^2}} \quad (7)$$

Table 7. Contrasting the suggested system's assessment metrics with those of the conventional CF recommendations system

	Suggested Method			Standard CF		
	Precision	Coverage	F-Measure	Precision	Coverage	F-Measure
T-10	63.7	60	56.41	52.73	38.11	44.32
T-15	76.54	64.66	69.43	60.31	48.73	53.77
T-20	79.83	71.43	75.41	62.03	52.08	57.41
T-25	86.73	78.41	82.3	69.41	58.14	63.16
T-30	93.08	92.2	92.42	73.64	61.73	67.31
T-35	97.09	98.2	97.46	75.04	64.61	69.61
T-40	97.02	97.93	97.51	74.21	64.34	69.12
T-45	97.15	97.83	97.47	75.23	64.15	69.35
T-50	97.92	98.04	97.43	75.31	64.21	69.32
T-55	97.08	97.78	97.35	75.32	65.13	69.12

The variables in the above formula are: n is the total number of things that two users have rated; i is a collection of items that both users have rated; r_{ui} is the value of the item i score that user u has provided; and r_u is the standard deviation of the customer score. The resemblance between what consumers want is determined in the present investigation using Pearson's correlation coefficient correlation approach according to the new user-item matrices. The outcomes of this phase (resemblance dependent on CF) will then be utilized to anticipate the tastes of the consumers.

4.7 Segmentation based on demographic attributes

Utilizing the two-stage grouping technique, the consumers are divided into groups according to marital status, age, gender, work, and academic achievement. Their within-cluster resemblance is then determined. Consequently, two clients with comparable demographic characteristics will view a few of the products similarly, which isn't captured in CF approaches. In this research, the goal of segmenting consumers based on these criteria is to more accurately choose a new user's fellow citizens, hence enhancing the process of suggestion by resolving the cold start issue.

4.8 Determining neighbors based on the new similarity function

In order to lessen the issue of cold begin, the group resemblance that was acquired from the initial phase (based on demographics variables) and the CF resemblance are now integrated in a new resemblance function. In situations where a person has just recently entered the system, the new resemblance algorithm does not exclusively rely on the person-item matrices. By means of this purpose, we can acquire a resemblance. Eq. (8) following calculates the resemblance functional ($Hsim$) by summing the within-cluster resemblance ($clusSim$) and CF-based similarities (Sim) weightedly.

$$Hsim(u, u') = (1-\alpha) \times Sim(u, u') + \alpha \times clusSim(u, u') \quad (8)$$

The K consumers that have the greatest resemblance to the intended client are chosen as this client's neighbors once the resemblance between the consumers is determined using the formula above. Figure 5 analyzes how different K values affect the suggested technique's efficiency.

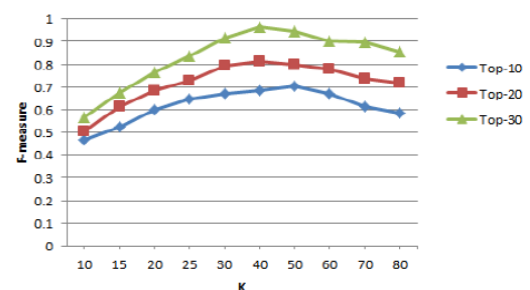


Figure 5. Analyze how different K values affect the suggested technique's efficiency

4.9 Providing recommendations

The intended client's tastes are forecasted by computing a

weighted mean of the ratings provided by those nearby who were identified in the preceding stage. The following formula predicts user u 's rating for item i (ru,i) depending on the weighted mean of neighboring ratings (ru',i) Eq. (9):

$$r_{u,i} = \frac{1}{n} \sum_{u' \in U} Hsim(u, u') r_{u',i} \quad (9)$$

The surrounding individuals of user u who evaluated item i are indicated by U . The more comparable individuals u and u' are, the higher the calculated rate at which (ru',i) is utilized to forecast (ru,i). Lastly, the intended consumer will be offered the N products with the greatest anticipated ratings.

5. PERFORMANCE EVALUATION

The F-measure, accuracy, and completeness metrics of the recommendation system will be evaluated for various values of N (the number of suggested items) based on the results of the tests. Six clusters are the ideal number for classification based on LRFM factors, according to the SOM method, and five groups are created based on demographic data. Table 6 shows the precision of the suggested and conventional approaches based on metrics of assessment for various values of N . The number of relatives, or characteristic K , has a value of 40 for each of them. Based on every criterion, Table 6's results show that the suggested method performs better than the conventional system. Actually, the five suggested items in the previous system had an accuracy of 43.27% (depending on the F-measure), but the five suggested items in the new strategy had an efficiency of 55.59%. The precision of the conventional approach has been 62.43% in 20 suggested products and 81.4% in its suggested system. The precision of the suggested system was found to be 96.59% in 30 things, compared to 68.54% in the conventional system. This means that the precision of the suggested method is almost 28% greater across 30 items than the conventional method.

According to the findings, both approaches perform better when the number of suggested items is increased to $N=30$. However, there is no discernible improvement in the precision of the findings with an increase in data volume. As a result, raising the number of suggested items to a threshold=30 can enhance the caliber of the suggestions that are made. Figure 6 illustrates the shifting trend of the F-measure criterion to help with a more accurate comparison of the two approaches' performances. Figure 7 shows the coefficients of the F-measure for three distinct values of N in order to assess the impact of modifications in variable K on the effectiveness of the investigation's suggested methodology. The suggested strategy will work better if the total amount of neighbor is increased to $K=40$, as shown in Figure 7 for each of the three settings of the N variable. Nevertheless, raising this value more will result in the precision of the approach being diminished, which will lower the standard of the suggestions.

The degree of accuracy, accuracy, recollection rate, and F1-score are the four primary indices used to evaluate the work: -

Accuracy: Efficiency is defined as the proportion of the specimens that, in accordance with the subsequent the formula, will yield an outcome that is favorable [54].

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} * 100\%$$

Expertise is an indicator of the accuracy rates, or how well an equation forecasts test findings.

$$Precision = \frac{TP}{TP + FP} * 100\%$$

Recall ratio: The recall rates can be calculated by splitting the total amount of positive samples that fit into the correct category by the total number of affirmative examples that need to be categorized.

$$Recall = \frac{TP}{TP + FP} * 100\%$$

F1-score: This evaluation statistic considers recall and precision simultaneously.

$$F1 - score = \frac{2 * Recall * precision}{Recall + Precision} * 100\%$$

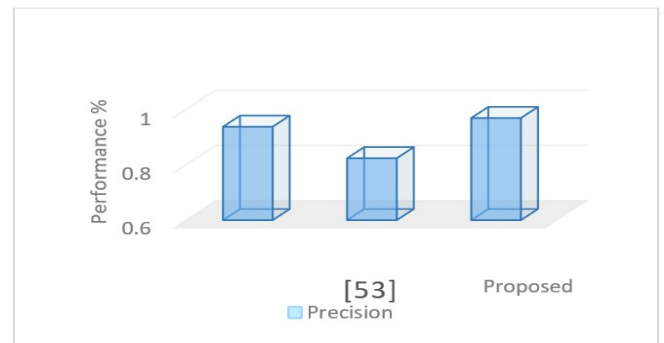


Figure 6. Precision evaluation analysis

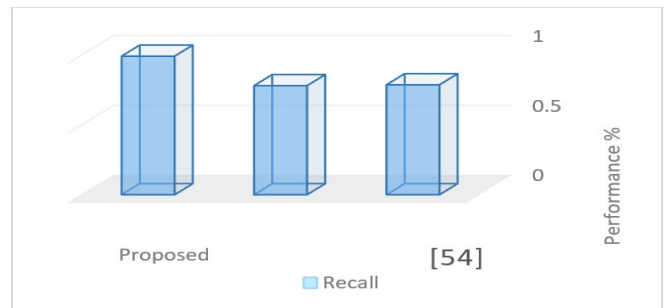


Figure 7. Recall evaluation analysis

6. RECOMMENDATION PERFORMANCE EVALUATION

The foundation of film algorithmic recommendations is an ordered list of recommended movies for a specific audience, together with accurate forecasts. The efficacy of the predictions and recommendations was measured using assessment scales. With eight contextual parameters, the LDOS-CoMoDa database is the biggest geographically aware database in terms of factors related to context, which is why it was examined. The outcomes from the two earlier investigations by studies [55, 56], as well as the evaluation and research of many matrices for the recommended technique, are highlighted in Table 8. The inquiry used precision, recall, accuracy, and the F1-score as evaluation metrics.

Table 8. Evaluation analysis of different matrices

	Accuracy	Precision	Recall	F1-Score
[53]	NA	0.9384	0.7854	0.8690
[54]	0.754	0.8244	0.7792	NA
Proposed	0.98	0.98	0.99	0.96

The precision, recall, F1-score, and accuracy statistics are not available (NA) (NA) as of the inquiry conducted in reference [55]. On the other hand, their recall rates are 0.7854, their F1-score is 0.8690, and their precision score is 0.9384. Recall, reliability, and specificity were all 0.7792, 0.754, and 0.8244 in the research conducted by the study [56]. It was not possible to determine the F1-score for this investigation.

The proposed method achieves excellent levels of specificity (0.98), recall (0.99), reliability (0.97), and F1-score (in comparison to state-of-the-art methods) (0.96). These results show that the recommended approach performs exceptionally well for all evaluation criteria. High reliability indicates a sizable number of accurate predictions (both true-positive and true-negative), while high retrieval frequencies and specificity demonstrate that the movies that are suggested are related to the users' likes and that a sizable number of pertinent materials are being recalled. The system's overall efficacy is measured by the F1-score, which is an inverse mean of precision and recall that shows a good balance between the two measures.

Figures 6 and 7 show the results of the recall and precision assessment experiments, accordingly. These results highlight the great recall rates and precision reached by the proposed method, indicating its value in offering accurate film suggestions to users.

Overall, the results show that the method proposed outperforms the previous studies in terms of F1-score, accuracy, precision, and recall. The suggested method is a significant addition to the suggestion system field because of its excellent performance and well-balanced precision-recall trade-off, which suggest that it can offer users accurate and timely movie suggestions. The fact that the compromise between recall and precision is balanced suggests this.

7. CONCLUSIONS

The goal of this work is to address or minimize the basic and typical drawbacks of RS in online shopping, namely in relation to cold begin, flexibility, dense user-item matrices, and customer interest changes. Consequently, an extensive online shopping suggestion engine has been proposed to overcome these restrictions through a suite of data mining techniques. Based on the study's findings, the suggested system performs better than conventional CF systems. Naive Bayes was used to develop a sophisticated system of suggestions. This program used the LDOSCOMODA dataset to classify movies according to user expectations. The precision of the algorithm was enhanced and scores for unrated films were predicted with the use of contextual information. The suggested system constructed a state-of-the-art recommendation for movies framework using our research's advanced Naive Bayes algorithm. Using user-driven forecasts from carefully selected inputs from the LDOSCOMODA dataset, a complex classification framework was developed that accurately discovers ideal viewing recommendations that are tailored to each user's preferences. Our suggestion method and discriminating acuity were enhanced by using historical

context to produce more reliable rating assessments and prognostications for movies that were not yet assessed. Extensive investigation afterwards produced encouraging data that validated remarkable performance.

Here are some ideas for additional study:

1. Using the CBF and CF approaches to combine in creating the suggested model.
2. Examining how the efficacy of the suggested system is affected by segmenting clients according to personality characteristics, additional socioeconomic variables like location and revenue, and additional behavioral parameters like the overall quantity of things bought.
3. Using datasets from various industries to test the suggested approach in order to increase generalization.
4. Big data, which can be sourced from sizable online retailers, are utilized to further assess the suggested model.

Like the majority of earlier studies, this one had certain operational and assessment issues. These issues include:

1. The dataset that is being used has a one-year time constraint.
2. The study's data is restricted to an internet clothing retailer.
3. The small sample size of the shop being studies' patrons.

REFERENCES

- [1] Raghuwanshi, S.K., Pateriya, R.K. (2019). Collaborative filtering techniques in recommendation systems. *Data, Engineering and Applications*, 1: 11-21. https://doi.org/10.1007/978-981-13-6347-4_2
- [2] Chen, W.H., Hsu, C.C., Lai, Y.A., Liu, V., Yeh, M.Y., Lin, S.D. (2020). Attribute-aware recommender system based on collaborative filtering: Survey and classification. *Frontiers in Big Data*, 2: 49. <https://doi.org/10.3389/fdata.2019.00049>
- [3] Juan, W., Yue, X.L., Chun, Y.W. (2019). Survey of recommendation based on collaborative filtering. *Journal of Physics: Conference Series*, 1314(1): 012078. <https://doi.org/10.1088/1742-6596/1314/1/012078>
- [4] Zhu, L., Li, H., Feng, Y. (2019). Research on big data mining based on improved parallel collaborative filtering algorithm. *Cluster Computing*, 22(Suppl 2): 3595-3604. <https://doi.org/10.1007/s10586-018-2209-9>
- [5] Ajaegbu, C. (2021). An optimized item-based collaborative filtering algorithm. *Journal of Ambient Intelligence and Humanized Computing*, 12(12): 10629-10636. <https://doi.org/10.1007/s12652-020-02876-1>
- [6] Afoudi, Y., Lazaar, M., Al Achhab, M. (2018). Collaborative filtering recommender system. In *International Conference on Advanced Intelligent Systems for Sustainable Development*, pp. 332-345. <https://doi.org/10.1007/978-3-030-11928-7-30>
- [7] Sun, Z., Zhang, J., Sun, H., Zhu, X. (2020). Collaborative filtering based recommendation of sampling methods for software defect prediction. *Applied Soft Computing*, 90: 106163. <https://doi.org/10.1016/j.asoc.2020.106163>
- [8] Najafabadi, M.K., Mohamed, A.H., Mahrin, M.N.R. (2019). A survey on data mining techniques in recommender systems. *Soft Computing*, 23(2): 627-654. <https://doi.org/10.1007/s00500-017-2918-7>
- [9] Anitha, J., Kalaivasu, M. (2021). Retracted article: Optimized machine learning based collaborative

- filtering (OMLCF) recommendation system in e-commerce. *Journal of Ambient Intelligence and Humanized Computing*, 12(6): 6387-6398. <https://doi.org/10.1007/s12652-020-02234-1>
- [10] Nallamala, S.H., Bajjuri, U.R., Anandarao, S., Prasad, D.D., Mishra, P. (2020). A brief analysis of collaborative and content based filtering algorithms used in recommender systems. *IOP Conference Series: Materials Science and Engineering*, 981(2): 022008. <https://doi.org/10.1088/1757-899X/981/2/022008>
- [11] Valdiviezo-Diaz, P., Ortega, F., Cobos, E., Lara-Cabrera, R. (2019). A collaborative filtering approach based on Naïve Bayes classifier. *IEEE Access*, 7: 108581-108592. <https://doi.org/10.1109/ACCESS.2019.2933048>
- [12] Singh, P.K., Pramanik, P.K.D., Choudhury, P. (2020). Collaborative filtering in recommender systems: Technicalities, challenges, applications, and research trends. In *New Age Analytics*, pp. 183-215.
- [13] Panda, S.K., Bhoi, S.K., Singh, M. (2020). A collaborative filtering recommendation algorithm based on normalization approach. *Journal of Ambient Intelligence and Humanized Computing*, 11(11): 4643-4665. <https://doi.org/10.1007/s12652-020-01711-x>
- [14] Widayanti, R., Chakim, M.H.R., Lukita, C., Rahardja, U., Lutfiani, N. (2023). Improving recommender systems using hybrid techniques of collaborative filtering and content-based filtering. *Journal of Applied Data Sciences*, 4(3): 289-302. <https://doi.org/10.47738/jads.v4i3.115>
- [15] Bobadilla, J., Ortega, F., Gutiérrez, A., Alonso, S. (2020). Classification-based deep neural network architecture for collaborative filtering recommender systems.
- [16] Ahmed, S.S., Shakir, H.R. (2024). Enhancement of secure hospital healthcare monitoring system based-software defined network (SDN) with machine learning. *JOIV: International Journal on Informatics Visualization*, 8(4): 2305-2315. <https://dx.doi.org/10.62527/joiv.8.4.2425>
- [17] Fkih, F. (2022). Similarity measures for collaborative filtering-based recommender systems: Review and experimental comparison. *Journal of King Saud University-Computer and Information Sciences*, 34(9): 7645-7669. <https://doi.org/10.1016/j.jksuci.2021.09.014>
- [18] Shakir, H.R., Mehdi, S.A., Hattab, A.A. (2022). RGB image encryption approach built on a new chaotic system with DNA coding. In *2022 Fifth College of Science International Conference of Recent Trends in Information Technology (CSCTIT)*, Baghdad, Iraq, pp. 236-241. <https://doi.org/10.1109/CSCTIT56299.2022.10145642>
- [19] Li, J., Zhang, K., Yang, X., Wei, P., Wang, J., Mitra, K., Ranjan, R. (2019). Category preferred canopy-K-means based collaborative filtering algorithm. *Future Generation Computer Systems*, 93: 1046-1054. <https://doi.org/10.1016/j.future.2018.04.025>
- [20] Wang, T., Manogaran, G., Wang, M. (2020). Retracted article: Framework for social tag recommendation using lion optimization algorithm and collaborative filtering techniques. *Cluster Computing*, 23(3): 2009-2019. <https://doi.org/10.1007/s10586-019-02980-8>
- [21] Ramakrishnan, G., Saicharan, V., Chandrasekaran, K., Rathnamma, M.V., Ramana, V.V. (2019). Collaborative filtering for book recommendation system. *Soft Computing for Problem Solving: SocProS 2018*, 2: 325-338. https://doi.org/10.1007/978-981-15-0184-5_29
- [22] Jiang, M., Zhang, Z., Jiang, J., Wang, Q., Pei, Z. (2019). A collaborative filtering recommendation algorithm based on information theory and bi-clustering. *Neural Computing and Applications*, 31(12): 8279-8287. <https://doi.org/10.1007/s00521-018-3959-2>
- [23] Rezaimehr, F., Dadkhah, C. (2021). A survey of attack detection approaches in collaborative filtering recommender systems. *Artificial Intelligence Review*, 54(3): 2011-2066. <https://doi.org/10.1007/s10462-020-09898-3>
- [24] Salloum, S., Rajamanthri, D. (2021). Implementation and evaluation of movie recommender systems using collaborative filtering. *Journal of Advances in Information Technology*, 12(3): 189-196.
- [25] Kumar, P., Kumar, V., Thakur, R.S. (2019). A new approach for rating prediction system using collaborative filtering. *Iran Journal of Computer Science*, 2(2): 81-87. <https://doi.org/10.1007/s42044-018-00028-5>
- [26] AL-Bakri, N.F., Hashim, S.H. (2019). Collaborative filtering recommendation model based on k-means clustering. *Al-Nahrain Journal of Science*, 22(1): 74-79.
- [27] Liu, X. (2019). A collaborative filtering recommendation algorithm based on the influence sets of e-learning group's behavior. *Cluster Computing*, 22(Suppl 2): 2823-2833. <https://doi.org/10.1007/s10586-017-1560-6>
- [28] Gupta, M., Thakkar, A., Gupta, V., Rathore, D.P.S. (2020). Movie recommender system using collaborative filtering. In *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, pp. 415-420. <https://doi.org/10.1109/ICESC48915.2020.9155879>
- [29] Koren, Y., Rendle, S., Bell, R. (2021). Advances in collaborative filtering. In *Recommender Systems handbook*. Springer, New York, pp. 91-142. https://doi.org/10.1007/978-1-0716-2197-4_3
- [30] Ambulgekar, H.P., Pathak, M.K., Kokare, M.B. (2018). A survey on collaborative filtering: Tasks, approaches and applications. In *Proceedings of International Ethical Hacking Conference 2018: eHaCON 2018*, Kolkata, India, pp. 289-300. https://doi.org/10.1007/978-981-13-1544-2_24
- [31] Srifi, M., Oussous, A., Ait Lahcen, A., Mouline, S. (2020). Recommender systems based on collaborative filtering using review texts-a survey. *Information*, 11(6): 317. <https://doi.org/10.3390/info11060317>
- [32] Sharma, S., Rana, V., Malhotra, M. (2022). Automatic recommendation system based on hybrid filtering algorithm. *Education and Information Technologies*, 27(2): 1523-1538. <https://doi.org/10.1007/s10639-021-10643-8>
- [33] Davagdorj, K., Park, K.H., Ryu, K.H. (2019). A collaborative filtering recommendation system for rating prediction. In *Advances in Intelligent Information Hiding and Multimedia Signal Processing: Proceedings of the 15th International Conference on IHH-MSP in Conjunction with The 12th International Conference on FITAT*, Jilin, China, pp. 265-271. https://doi.org/10.1007/978-981-13-9714-1_29
- [34] Kommineni, M., Alekhya, P., Vyshnavi, T.M., Aparna, V., Swetha, K., Mounika, V. (2020). Machine learning based efficient recommendation system for book

- selection using user based collaborative filtering algorithm. In 2020 Fourth International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, pp. 66-71. <https://doi.org/10.1109/ICISC47916.2020.9171222>
- [35] Li, L., Zhang, Z., Zhang, S. (2021). Hybrid algorithm based on content and collaborative filtering in recommendation system optimization and simulation. *Scientific Programming*, 2021(1): 7427409. <https://doi.org/10.1155/2021/7427409>
- [36] Deng, J., Guo, J., Wang, Y. (2019). A novel K-medoids clustering recommendation algorithm based on probability distribution for collaborative filtering. *Knowledge-Based Systems*, 175: 96-106. <https://doi.org/10.1016/j.knosys.2019.03.009>
- [37] Iftikhar, A., Ghazanfar, M.A., Ayub, M., Mehmood, Z., Maqsood, M. (2020). An improved product recommendation method for collaborative filtering. *IEEE Access*, 8: 123841-123857. <https://doi.org/10.1109/ACCESS.2020.3005953>
- [38] Han, X., Wang, Z., Xu, H.J. (2021). Time-weighted collaborative filtering algorithm based on improved mini batch K-means clustering. *Advances in Science and Technology*, 105: 309-317. <https://doi.org/10.4028/www.scientific.net/AST.105.309>
- [39] Alhijawi, B., Al-Naymat, G., Obeid, N., Awajan, A. (2021). Novel predictive model to improve the accuracy of collaborative filtering recommender systems. *Information Systems*, 96: 101670. <https://doi.org/10.1016/j.is.2020.101670>
- [40] Margaris, D., Kobusińska, A., Spiliotopoulos, D., Vassilakis, C. (2020). An adaptive social network-aware collaborative filtering algorithm for improved rating prediction accuracy. *IEEE Access*, 8: 68301-68310. <https://doi.org/10.1109/ACCESS.2020.2981567>
- [41] Sallam, R.M., Hussein, M., Mousa, H.M. (2022). Improving collaborative filtering using lexicon-based sentiment analysis. *International Journal of Electrical and Computer Engineering*, 12(2): 1744-1753. <https://doi.org/10.11591/ijece.v12i2.pp1744-1753>
- [42] Logesh, R., Subramaniaswamy, V., Malathi, D., Sivaramakrishnan, N., Vijayakumar, V. (2020). Enhancing recommendation stability of collaborative filtering recommender system through bio-inspired clustering ensemble method. *Neural Computing and Applications*, 32(7): 2141-2164. <https://doi.org/10.1007/s00521-018-3891-5>
- [43] Aljunid, M.F., Dh, M. (2020). An efficient deep learning approach for collaborative filtering recommender system. *Procedia Computer Science*, 171: 829-836. <https://doi.org/10.1016/j.procs.2020.04.090>
- [44] Li, J., Ye, Z. (2020). Course recommendations in online education based on collaborative filtering recommendation algorithm. *Complexity*, 2020(1): 6619249. <https://doi.org/10.1155/2020/6619249>
- [45] Iwanaga, J., Nishimura, N., Sukegawa, N., Takano, Y. (2019). Improving collaborative filtering recommendations by estimating user preferences from clickstream data. *Electronic Commerce Research and Applications*, 37: 100877. <https://doi.org/10.1016/j.elerap.2019.100877>
- [46] Neysiani, B.S., Soltani, N., Mofidi, R., Nadimi-Shahraki, M.H. (2019). Improve performance of association rule-based collaborative filtering recommendation systems using genetic algorithm. *International Journal of Information Technology and Computer Science*, 11(2): 48-55. <https://doi.org/10.5815/ijitcs.2019.02.06>
- [47] Benkessirat, S., Boustia, N., Nachida, R. (2021). A new collaborative filtering approach based on game theory for recommendation systems. *Journal of Web Engineering*, River Publishers, 20(2): 303-326. <https://doi.org/10.13052/jwe1540-9589.2024>
- [48] Sallam, R.M., Hussein, M., Mousa, H.M. (2020). An enhanced collaborative filtering-based approach for recommender systems. *International Journal of Computer Applications*, 176(41): 9-15.
- [49] Anwar, T., Uma, V. (2021). Comparative study of recommender system approaches and movie recommendation using collaborative filtering. *International Journal of System Assurance Engineering and Management*, 12(3): 426-436. <https://doi.org/10.1007/s13198-021-01087-x>
- [50] Chen, J., Wang, B., Ouyang, Z., Wang, Z. (2021). Dynamic clustering collaborative filtering recommendation algorithm based on double-layer network. *International Journal of Machine Learning and Cybernetics*, 12(4): 1097-1113. <https://doi.org/10.1007/s13042-020-01223-2>
- [51] Khojamli, H., Razmara, J. (2021). Survey of similarity functions on neighborhood-based collaborative filtering. *Expert Systems with Applications*, 185: 115482. <https://doi.org/10.1016/j.eswa.2021.115482>
- [52] Wang, F., Wen, Y., Guo, T., Liu, J., Cao, B. (2020). Collaborative filtering and association rule mining-based market basket recommendation on spark. *Concurrency and Computation: Practice and Experience*, 32(7): e5565. <https://doi.org/10.1002/cpe.5565>
- [53] Qasem, M.H., Obeid, N., Hudaib, A., Almaiah, M.A., Al-Zahrani, A., Al-Khasawneh, A. (2021). Multi-agent system combined with distributed data mining for mutual collaboration classification. *IEEE Access*, 9: 70531-70547. <https://doi.org/10.1109/ACCESS.2021.3074125>
- [54] Shen, J., Zhou, T., Chen, L. (2020). Collaborative filtering-based recommendation system for big data. *International Journal of Computational Science and Engineering*, 21(2): 219-225. <https://doi.org/10.1504/IJCSE.2020.105727>
- [55] Li, M., Wen, L., Chen, F. (2021). A novel collaborative filtering recommendation approach based on soft co-clustering. *Physica A: Statistical Mechanics and Its Applications*, 561: 125140. <https://doi.org/10.1016/j.physa.2020.125140>
- [56] Duan, R., Jiang, C., Jain, H.K. (2022). Combining review-based collaborative filtering and matrix factorization: A solution to rating's sparsity problem. *Decision Support Systems*, 156: 113748. <https://doi.org/10.1016/j.dss.2022.113748>