**DEMOGRAPHY-BASED RECOMMENDER SYSTEM**

**Abstract**

The exponential increase of digital data is increasing the problem of too much online information. This complicates the user's decision-making. Most online merchants and service providers utilize recommender systems to solve this problem and meet customer needs. In this research, a demography-based recommender system is proposed based on demographic filtering. Firstly, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are calculated using cosine similarity based on rating filtering at 20, 40, and 60% no. of ratings then it is calculated by using normal Pearson correlatioyn. After this MAE and RMSE are calculated by using both cosine similarity and Pearson correlation based on demographic filtering. At last, the results obtained from the above calculation are compared and the result shows that the MAE and RMSE values are higher in the case of only considering rating filtering on both cosine similarity and normal Pearson correlation on the other hand by utilizing the demographic filtering the value of MAE and RMSE are lower, hence it is concluded that the accuracy of the proposed system is higher than existing methods.

**Keywords**: Recommender system, Collaborative Filtering (CF), Hybrid recommendation, MAE, RMSE, Normal Pearson correlation.

1. **Introduction**

The problem of having too much information available on the internet is becoming a much more widespread issue, as a result of the exponential growth in the amount of digital data. This difficulty presents additional difficulties for the user in terms of the decision-making process. As a result, the majority of online retailers and service providers make use of the idea of recommender systems to tackle this issue and satisfy user demands most practically and efficiently [1]. The English word "recommend is derived from the Latin word recommendare, which means to recommend, approve, endorse, praise, or acclaim anything. The process of making decisions is facilitated more effectively by these kinds of systems, which provide customers with the most relevant options [2].

The Xerox Palo Alto Research Centre introduced the world's first recommender system, known as Tapestry, in the year 1992. Later on, as Tapestry gained popularity, recommender systems were developed for a variety of different applications. These applications include Amazon, CiteSeer, Docear, Ebay, Hulu, Jester, LinkedIn, Netflix, Pandora, Trip advisor, Yahoo, and YouTube. The recommendations for these apps can be seen in a variety of different ways, such as "You may enjoy," "People you may know," "Suggested movies," "Trending now," "Similar products," and "Suggestions for you," "People who purchased this also bought this," and so on [2].

Content-based filtering, collaborative filtering (CF), knowledge-based systems, and hybrid systems are just a few of the many successful recommendation methods that have been created over the years [3-4]. Machine learning algorithms are the backbone of recommender systems, allowing them to automatically learn about users' preferences and behaviors, identify people with similar tastes, and discover previously unseen connections between users, products, and services [5].

Recommended Systems are used to reduce the difficulty of data excess in numerous industries based on a person's intuitive and explicit preferences [6]. It is common for these algorithms to provide recommendations based on user ratings, hotel attributes, and user reviews to particular individuals or groups [7]. A certain amount of relevant data is typically necessary to make smart decisions in any situation. The internet, in particular, makes it simpler to acquire more information, because of technological advances. Recommendation Systems play a key part in screening the destinations that fit the user's demands and interests [8]. Figure 1 shows the sparse dataset with a comparison rating prediction that meets three similarities using recommendation systems which are similarity calculation using the traditional method, similarity calculation using demographic profile (distance), and a similarity calculation using profile rating.

Diagram

Description automatically generated

Figure 1. Rating prediction of using the sparse dataset

Recommendation systems are used by a wide range of travel enterprises to improve customer happiness and enjoyment. Research suggests that a platform that takes into consideration customer evaluations and the surrounding facilities of hotels should be created as the main target of this approach. For better recommendations, attributes from reviews must be extracted [9]. A recommendation system is the core of a customized recommendation system that uses recommendation technology. There is a vast selection of recommendation technologies on the market today, including content-based recommendation, CF recommendation, association rule recommendation, hybrid recommendation technology, knowledge-based recommendation, and effectiveness-based recommendation. One of the most important and useful technologies now in use is CF. It analyzes the data of users with similar interests or personality traits. Consequently, the group's favorite hotel might be recommended to target consumers. There are three primary problems with the CF method [10]:

1. **Data sparsity**: A sparse user-item matrix is the effect of a lack of user feedback. This is a common problem when the number of items is big, but the number of rating values is low, or when the recommendation system is still in its infancy [10].
2. **Cold start:** New users and new items/hotels both have difficulties. The lack of transaction data and the inability to determine individual purchase patterns make it difficult to give customized recommendations to new customers at the beginning of their usage. If a product has never been rated, it can't be recommended to users [10].
3. **Scalability:** In point of fact, consumers are unable to provide ideas if the databases of people and things are sufficiently extensive [10].
4. **Mobile Data**

The collecting of mobility data is a difficult task because of three reasons [11].

* **Fears of people about their privacy**

Users do not want to reveal their location information, according to several studies.

* **Business considerations**

Just a few organizations that deliver Location-Based Services (LBS) applications gather and own high-quality mobility information. Companies that are of concern, such as the catering chain, have a restricted approach to it, which is a direct result of the significance of mobility data in the process of selecting a site for the firm.

* **Cost of implementations**

For even the government, developing urban devices and collecting data on a broad scale takes time and money. It is also worth noting that urban applications like urban planning as well as company site collection are more useful as well as helpful through an urban region's early growth phase because changing the plans as well as location site once the area has been developed are more expensive. As a result, estimating a city's human mobility is beneficial [11].

1. **Recommendation System**

Users can utilize the recommender system to discover specific information. The specific information is sought without having to type a long query into the information rescue system. The output feature of any recommended system is solely determined by the amount of information utilized. It is possible to generate the best suggestion outcomes when a large amount of data is accessible. Work parallelism or control parallelism is used to create valuable recommendations. Existing recommender algorithms must be modified to meet the needs of the user. Furthermore, by fully exploiting hybrid computing, high-quality recommender systems are created by combining serial and parallel processing. Delivering outcomes promptly is one of the advantages. The parallel execution of algorithms can make it easier to generate output without sacrificing performance. These benefits are:

* Large volumes of data can be handled quickly, resulting in increased efficiency.
* A wide range of item types can be suggested.
* It is simple to convert an existing method to parallel processing [12]. Figure 2 shows three stages of the recommended system as given below.

Diagram

Description automatically generated

Figure 2. Recommendation system phases

The first phase is the information collection phase in which data is collected from different sources such as environmental data, agency data, similar process data, etc. The second phase is the learning phase in which inputs takes from the information collection phase and then learn features from a different method. The third phase is the recommendation phase which takes inputs from the learning phase and is then used for prediction. After that feedback comes that goes directly to the first phase this process is continued whenever a recommended output is satisfied. This whole process is known as the recommendation system phase [13].

1. **Collaborative Filtering (CF)**

A technique known as CF can exclude content that a particular user might otherwise find interesting based on the reviews of other users. It searches through a large number of people to find other people who, share a user's tastes. Material (content) based recommendation and CF recommendation techniques are the two basic kinds of information recommendation methods. CF is the most often utilized method of making recommendations. CF might be distributed into two broad types: item-based and user-based. User-based CF is to locate a cluster of users who share the same interests as the direct users. The similarity among the items is calculated mainly by analyzing the users’ behavior known as item-based CF which is centered on the similarity of the items. However, there are major data sparseness and cold start issues. The issue of information cold starts as well as sparseness in the traditional method. It can be mitigated as more data on the internet is obtained, including words, photographs, and tag data, which include tailored demand information and rich project information. The purpose of a CF process is to recommend new things or estimate the usefulness of a certain item/hotel for a specific user based on the user's prior preferences and the views of other users. There are lists of m users Uand n items in a typical CF scenario. Each user has a collection of items for which he or she has expressed a view, denoted by . Users can express their opinions in the form of an open rating, which is commonly a numerical type of scale, or it might be implicitly formed from information such as purchase data, examination of timing logs, and linking patterns. Mention that and it is viable for to be an empty set. There exists a prominent user also known as the active agent for whom the project of a CF process is to find an item likelihood which has two forms [14].

* **Recommendation:** It's a ranking of the top N hotels that the current user enjoys the most . must include items/hotels that have not formerly been purchased by the current customer in the recommended list. It's also known as a Top-N recommendation algorithm interface [14].
* **Prediction:** is a numeric value that represents the projected possibility of the item for the active user , based on the user's activity. According to the data provided by , this projected value falls within the same range of 1 to 5 [14].

Figure 3 shows the CF process. It has three steps, first is input data then set according to neighbors’ format in the second step. In the third step, it recommends for generation of new items.

Diagram

Description automatically generated

Generation Recommended

Neighbors format

Rate input data

Figure 3. The CF process [15]

**1.4 Demographic Recommender Systems (DRS)**

This stimulates the desire for investigating in some detail the profiling methods utilized by DRS, along with the benefits of utilizing each profiling method and the relevant similarity computation algorithms for each profiled method. Intuitively, the age, gender, and occupation of a population are the most essential demographic data [16]. There are occasions group's Zip Code is regarded by demographic data.  This conducting research in this area is extremely challenging because there are so few possibilities available. The different ways of encoding each element as well as the accompanying similarity methods for comparing the attributes [17].

A condensed user model that makes use of the individual demographic data in addition to features that are rating-driven. The user profited from the inclusion of demographic data in this model, and the solution to the cold-start problem was made possible by the system. In particular, the additional demographic information makes it possible for customers who have a lower total number of ratings to still benefit from the system [18]. Presented two hybrid consensus-based rating systems that in some manner combined DRS ratings. The first concept combined CRS and DRS with some of the contents of the items, whereas the second plan combined demographic, collaborative, and content-based techniques at various components of the user model [19].

Given that it assigns users to categories based on their demographic characteristics, DRS is a conventional system. In the future, DRS can base its suggestions on the users' assessments of the system's components [20]. It's important to remember that DRS and Conversational Recommender Systems (CRS), although collect different data, both rely on correlations between users. For these reasons, DRS offers many of the same benefits as CRS, including the ability to discover previously untapped niches across genres, encourage users to venture beyond their comfort zones, and evolve [21].

1. **Literature of Review**

This section displays multiple similar works by several writers based on the demography-based recommender system.

**Wang et al., (2022) [22]** investigated that recommender systems are indispensable in the age of the mobile internet for helping people locate relevant content. As a result, creating a recommendation system that has high prediction efficiency and robust privacy protection is important and advantageous. For this purpose, a DNN-based recommendation model called PrivRec is proposed, which operates in the distributed Federated Learning (FL) environment to keep users' data safe on their own devices. Finally, in a simulated FL setting, extensive investigations on two datasets are conducted, the results of which prove the effectiveness of the suggested PrivRec.

**Ali et al., (2022) [23]** emphasized that the primary objective of every profit firm is to increase its revenue by offering helpful recommendations to the people. In order to accomplish this goal, Netflix and Amazon are conducting extensive research on improving their recommender system and expanding the range of options available to their customers. The study's goal is to present a comprehensive analysis of many types of recommendation engines. The implementation of a recommendation engine is based on PySpark and utilizing the ALS (Alternate Least Square) approach would be the primary focus. In the conclusion, it was determined that with the help of ALS, a scalable recommendation system can be developed. This system has the potential to be utilized to improve a company's revenues and deliver correct predictions for user questions.

**Ali et al., (2022) [24]** revealed that online educational resources help educators like teachers, students, and administrators gain knowledge and skills. There are some difficulties that users must overcome, such as finding and deciding on the best courses and resources for their individual needs. The study suggests the E-learning Recommendation Architecture (ELRA) solves the challenges of online education. The statistical and experimental results suggest that the virtual agent-based recommendation system increased user learning capabilities and made course selection easier based on users' interests and preferences when compared to existing methodologies.

**Lian and Tang (2022) [25]** It is hypothesized that as more open Application Programming Interfaces (APIs) have become available on the web, developers' interest in reusing or mixing existing APIs to create new apps has increased significantly. Although several different CF approaches have been developed for API selection, the accuracy of their recommendations is restricted and still needs to be improved. In this research, a method for recommending APIs that makes use of the high-order connectivity that exists between APIs and API users is proposed. This method is focused on the neural graph CF technique. The findings indicate that the proposed technique performs better than the methods that are considered to be the current standard for API recommendation.

**Chi et al., (2022) [26]** highlighted that with the arrival of the age of the Internet of Things (IoT), both the types and quantity of web services have been rising at a rapid speed and the choices that users make for web services are frequently made to be more difficult. The traditional approach of CF has several drawbacks, one of which is that it only considers data from centralized user services while ignoring data of scattered quality that comes from multiple platforms. In this research, a novel amplified Locality-Sensitive Hashing (LSH)-based service recommendation approach is proposed. The method is referred to as Service recommendation amplified Locality-Sensitive Hashing (SR Amplified-LSH). In the end, it concluded that SR Amplified-LSH is capable of striking a healthy balance between the precision and efficacy of recommendation and the user's right to personal information.

**Seth et al., (2022) [27]** described that multiple methods of study, such as content-based filtering and CF, have been used in the study of recommender systems. The primary difficulty is in the cold start problem of recommending suggestions to a new member of the network. A system that can provide recommendations based on the user's demographic category by considering the collaborative activities of users in the same category is needed. By combining collaborative work, demographic analysis, and content-based strategies, the suggested hybrid model can address the cold start issue. On the evaluation data, the Singular Value Decomposition (SVD) hybrid model's foundational method yielded an RMSE of 0.92.

**Krishnan et al., (2022) [28]** explained that as a result of the current Covid-19 epidemic, even small-town supermarkets and retail businesses have begun distributing their goods online. Customers are sorted into groups according to their Recency, Frequency, and Monetary (RFM) values, and then product suggestions are made to help keep the most valuable ones around. Customer segments can be created using the RFM ratings to categorize buyers. The cognitive similarity notion underpins the method used to provide product recommendations. Researchers first used CF as the starting point to create this hybrid recommendation algorithm. The suggested approach based on cognitive similarity is 5 percentage points more effective when compared to current approaches.

**Shao et al., (2021) [29]** stated that there are now many more online services available via multiple mobile edge platforms, it is important to rank them according to the quality of the services companies provide to save time and money for customers. However, several subjective and objective factors make it difficult to blindly trust the service quality data released by service providers. This leads to several serious trust-aware service evaluation problems, such as a lack of service quality data and a lack of feedback incentive. As a result, researchers provide a concise summary of the key problems plaguing the study of reliable mobile edge services today. Then explore one of the most common use cases for a  trusted service assessment, namely, recommender systems, and the many classifications that exist within this space. Researchers think that study might aid a mobile edge platform in developing a reliable reputation system for the many intelligent apps as the RS method.

**Hietala et al., (2021) [30]** described that the RS studies concentrate on either content-based, collaborative, or hybrid recommendation methods. The content-based method draws information from user profiles and product descriptions to make suggestions. Rather than relying on the user's private information and a description of the product, collaborative methods leverage data about the user's prior behavior and preferences. Successful suggestions are found in hybrid recommendation systems that use both content-based and collaborative-based tactics. Hybrid recommendation systems even though numerous publications in the literature suggest books, articles, news, etc. use either content-based or collaborative-based strategies. Hybrid recommendation systems are in-depth for their widespread usage in industries as diverse as e-commerce, e-learning, media, etc. The metric used to determine a model's efficacy is a suggestion generator. is one that uses several criteria to evaluate items.

**Suresh et al., (2020) [31]** described that the RS studies concentrate on either content-based, collaborative, or hybrid recommendation methods. in the face of massive data quantities and the development of new advertising tactics. The need for online system managers to look for automated solutions that might also benefit their systems is expanding. Many referencing sources have discussed automatic systems that aim to customize a user's experience while visiting a website. The experimental evidence shows that the Adaptive Genetic Algorithm (AGA) uses a multi-criteria compared to conventional collaborative evaluation methods, which rely on a single criterion, RS shows substantial improvement. A comparison of the filtering RS with the basic multi-criteria RS based on the Genetic Algorithm (GA).

**Sharma et al., (2019) [32]** described that the recommender systems have grown in popularity over the years as a way to help people find information that's useful to them, whether looking for something to read, watch, listen to, eat, or do during their free time. Several methods for making recommendations have been proposed, with content-based filtering, CF, hybrid systems, and knowledge-based systems being the most often used. Hybrid systems use several different hybrid models that pool together different recommendation methods to improve the effectiveness of a single recommendation strategy. The current landscape of Recommender systems with a special emphasis on hybrid recommender systems. Additionally, many hybridization models are explored, and the current literature is organized according to the hybrid model that it employs and the specific Machine learning algorithm that it employs.

**Kaur et al., (2019) [33]** described that during an age of rising information overload, Recommender Systems (RS) have emerged as a useful customizing tool. CF is the most widely used recommendation method. Most of instead of using several factors for making recommendations, CF apps rely on simple star ratings. In this effort, an AGA is based on multicriteria RS. In the modernized AGA layout, Crossover and mutation rates are used dynamically to simulate consumer tastes for evaluating objects using many criteria at once.  The AGA concludes specific to each user, an appropriate weight vector. This method is based on many criteria comparing RS to the standard collaborative. Both the filtering RS and the basic multi-criteria RS based on the GA algorithm are shown. The use of RS has been shown to significantly boost the global expansion of web-based application businesses.  The basic concept behind artificial intelligence is to establish a link between products and consumers to recommend the best product for a certain customer.

**Selvi et al., (2018) [34]** described that the order to provide relevant suggestions to engaged internet users, CF has emerged as a popular method. The CF that uses memory to locate users most similar to the currently logged-in one by traditional similarity metrics is one of the more effective types. The majority of these scales focus on the ratings of a shared product by two individuals.  Rating data from throughout the world is considered by the Matusita coefficient in CF (MCF). User-generated ratings on a variety of scales effectiveness of the planned activities are investigated and validated.  By comparing its results with those from more established metrics utilizing industry standard datasets.  The outcomes of the recommendations show that the suggested measure is superior to the traditional ones in MAE, root means square error, accuracy, precision, recall, and other performance characteristics and protection.

**Yang et al., (2018) [35]** explained that the hybrid recommendation algorithm is presented as a remedy to the existing problems with e-commerce recommendation systems, such as their lack of personalization, rigidity, and precision. In place of a unified method, researchers use a hybrid recommendation algorithm that draws from a variety of sources, such as a content-based algorithm, an item-based CF recommendation algorithm, and a demographics-based recommendation algorithm. Investigators utilize classification and clustering algorithms to sift through objects and users' past data to increase the scope of their recommendations. Researchers conduct an effectiveness analysis, which demonstrates that the enhanced scheme yields superior recommendation outcomes in electronic commerce.

**Sang et al., (2017) [36]** described that the users depend greatly on recommendation systems because algorithms help them sift through massive amounts of data to get the knowledge their need to make informed decisions. The similarity between things is the primary focus of any recommendation system of items. CF is a kind of user-generated rating data used to provide product recommendations. The algorithm suggests based on the user's ratings and the degree of resemblance between themselves and the other user. The data set must be split into an unrated and rated subset to use the system's features. Using Item K-Nearest Neighbors (K-NN) construct a model using the top-rated dataset. Researchers use the unrated movies dataset when putting the model to the test. Finally, a list of unrated films is ranked using the Correlation and Cosine distance formulas.

* 1. **Comparison between reviewed literature**

There is a wide range of authors who used the technique and presented their discoveries, as can be found in Table 1.

Table 1. Comparative analysis of literature review

|  |  |  |
| --- | --- | --- |
| **Authors** | **Techniques** | **Outcomes** |
| **Wang et al., (2022) [22]** | DNN- based Priv Rec | Theoretical and experimental result shows the effectiveness of Priv Rec to achieve accuracy and privacy protection. |
| **Ali et al., (2022) [23]** | ALS | ALS creates a scalable recommendation system that can increase business income and provide accurate user query predictions. |
| **Ali et al., (2022) [24]** | ELRA | The virtualized agent-based recommendation system boosted user learning skills and made course selection easier based on user interests and preferences. |
| **Lian and Tang (2022) [25]** | Neural graph CF | The outcomes demonstrate that the proposed strategy outperforms the state-of-the-art method for API recommendation. |
| **Chi et al., (2022) [26]** | LSH | A good compromise between recommendation accuracy and efficiency and user privacy information can be achieved by SR Amplified-LSH. |
| **Seth et al., (2022) [27]** | SVD | On the evaluation data, the SVD hybrid model's foundational method yielded an RMSE of 0.92. |
| **Krishnan et al., (2022) [28]** | RFM | When compared to current approaches, the suggested approach based on cognitive similarity is 5 percentage points more effective. |
| **Shao et al., (2021) [29]** | E-learning media | Researchers think that study might aid a mobile edge platform in developing a reliable reputation system for the many intelligent apps that run there. |
| **Hietala et al., (2021) [30]** | Hybrid recommendation | Hybrid recommendation systems in depth for their widespread usage in industries as diverse as e-commerce, e-learning, media, etc |
| **Suresh et al., (2020) [31]** | AGA | The experimental evidence shows that the AGA uses  a multi-criteria compared to conventional Collaborative Evaluation methods, which rely on a single criterion, RS shows substantial improvement |
| **Sharma et al., (2019) [32]** | Filtering, | To improve the effectiveness of a single recommendation strategy, hybrid systems use several different hybrid models that pool together different recommendation methods. |
| **Kaur et al., (2019) [33]** | AGA | The basic concept behind artificial intelligence is to establish a link between products and consumers to recommend the best product for a certain customer. |
| **Selvi et al., (2018) [34]** | MCF | The outcomes of the recommendations show that the suggested measure is superior to the traditional ones MAE, root means square error, accuracy, precision, recall, and other show characteristics and protection. |
| **Yang et al., (2018) [35]** | Hybrid recommendation algorithm | Investigators utilize classification and clustering algorithms to sift through objects and users' past data to increase the scope of their recommendations. |
| **Sang et al., (2017) [36]** | K-NN | The algorithm suggests based on the user's ratings and the degree of resemblance between themselves and the other user. |

1. **Background Study**

Using demographic information as a starting point thorough analysis of developments in management research that are prospective in the context of service sectors in India is carried out. As is common knowledge, service-oriented technologies and management have been gaining popularity over the years. This trend holds the promise of a method to establish a foundation for agility, allowing businesses to develop new and more adaptable business processes that capitalize on the value that can be derived from taking a services-based approach to serving customers. The yearly growth rate is between 20-22% in Information Technology (IT) services and over 55% in IT-Enabled Services (ITES), which include Business Process Outsourcing (BPO), Knowledge Process Outsourcing (KPO), and other similar services. Additionally, India's software and service exports are expanding at a fast pace. However, even the most prestigious management educational institutions are provided differently depending on the demographical setting. If all goes according to plan, India's per capita Gross Domestic Product (GDP) growth over the next two decades might be boosted by almost two percentage points per year to the demographic dividend. As a result, the following demonstrates some insights into these research traditions that are followed in the context of the service business and suffer architectural change [37].

1. **Research Gap**

* CF is a kind of user-generated rating data used to provide product recommendations.
* The current landscape of recommender systems with a special emphasis on CF recommender systems.
* Retaining clients requires providing them with better deals and CF products that will interest them.

1. **Research Objectives**

The following are the targets of the investigation:

* To find several serious trust-aware service evaluation problems, such as a lack of service quality data.
* Finding out the similarity between things is the primary focus of any recommendation system of items.
* To check the influence of the custom fit on today's relying on the user's private information and a description of the product industrial environment.

1. **Problem formation**

Online system administrators are under growing obligation to search for automated solutions that could also their systems. Many high-priced items are accessed online for listening. Automatic systems that intend to personalize a user's experience while visiting a website, such as RS discussed in a variety of reference sources. The power of the Internet to influence people's actions. Modern information systems include the likes of social media, online commerce, online enterprises, online transportation, and many more. The user's profile in a content-based RS, for instance, would include a collection of attributes generated from the features of the things the person has shown. A user’s profile consists of characteristics that define the user's demographic category or subgroup, a DRS is in effect.  Developed a hybrid fuzzy CRS for mobile goods and service recommendations by fusing user-based and item-based approaches. Recommender systems attempt to alleviate the burden of information overload experienced by internet users to foster a stronger bond between the system and its users. Consumers' reasonable reticence toward DRSs and the challenges of getting correct demographic information from the users related to their fears about security and privacy are the major reasons DRSs are not more generally utilized. For example, while suggesting movies, age brackets are essential but proposing tourist spots, and income levels are more important.

1. **Research Methodology**

Data is sparse whenever a feature's values are mostly zero. This is a significant distinction when compared to features with blanks. Vectors of one-hot-encoded words or counts of categorical data are two examples of sparse features. Machine learning models often make use of sparse information, often in the form of one encoding. Overfitting, incorrect feature importance, and excessive variance are only some of the problems that might arise from using these characteristics in a machine learning model. Pre-processing sparse features using techniques like feature hashing or omitting the feature is advised to lessen the negative effects on the final product. While sparse features tend to have zero values, dense features often contain many more values that are not zero. Models' space and temporal complexity can rise if include numerous sparse features.

The depth of tree-based models can be bigger to include all characteristics, while linear regression models would be able to fit more coefficients. If the characteristics only contain a little amount of data, it might throw off the model's algorithms and diagnostic tools. The models tend to conform to the imperfections in the training data. Instead of adjusting the data's dimensionality, it is possible to employ a machine learning model that is more forgiving of sparse information. The entropy-weighted k-means approach, for instance, performs better than the standard k-means algorithm used for this issue.

* 1. **Techniques Used**

This section discusses the methods that were used in the suggested approach. Methods such as Principal Component Analysis (PCA), Content-Based Filtering CF, and Hybrid Filtering are discussed below.

* + 1. **Content-Based Filtering**

Content-based filtering is a form of a recommender system that tries to predict whatever a user would appreciate based on that user's previous behavior. It generates suggestions by matching keywords and characteristics linked to objects in a database (for example, products in an online marketplace) to a user profile. The user profile is built using information collected from a user's activities, such as purchases and the evaluation of products. It makes suggestions based on similarities in goods, services, or content characteristics, as well as information gathered about the consumer. Figure 4 shows the comparison of content-based filtering and CF.

Diagram

Description automatically generated

Figure 4. Content-based filtering vs CF [38]

(1)

The circumstance in which both the user embedding and the app embedding are binary vectors is . A characteristic appears in both and adds one point to the total. It refers to the number of characteristics that are present in both vectors at the same time. A large dot product, therefore, suggests a greater number of shared characteristics, and hence, a greater degree of similarity.

* + 1. **Collaborative Filtering (CF),**

The most popular method for developing intelligent recommender systems that can learn to deliver better suggestions over time as more data is gathered about users is called CF. CF is a method that considers the opinions of users who are similar to the one making the filtering decision. Method for narrowing down a big population to a subset sharing a user's preferences. CF is a method of making suggestions to a user based on that person's similarity to other users.

An individual's rating (R) of a particular item (I) is likely to be within one standard deviation of the mean rating (AVERAGE) of the five or ten users (similar to U) who have rated I. The average rating from n users would be calculated as follows: Finding a group of people who are similar to a user U allows us to predict the rating R that U would give to a certain item I. Just like with similarity, there are several approaches to do this.

(2)

* + 1. **Hybrid Filtering**

One subset of recommendation systems, called a hybrid recommendation system, combines the strengths of both the content-based and CF approaches. Some research suggests that combining collaborative and content-based filtering is more successful than each method alone in addressing some of the problems using them alone. There are several ways to implement a hybrid recommender system strategy, such as employing both content and collaborative-based techniques to create predictions independently and then merging them or simply augmenting the capabilities of a content-based approach with those of a collaborative-based. Hybrid techniques have been shown to outperform traditional ones in several experiments, leading researchers to conclude that humans make better decisions by using them. Some recommender systems are known as hybrids recommendation systems. Figure 5 shows the hybrid filtering recommender systems.

Diagram

Description automatically generated

Figure 5.Hybrid Filtering [39]

* 1. **Proposed Methodology**

In this section, various steps of the proposed methodology are discussed below. Figure 6 depicts the flow chart of the intended methodology of the information gathered from the nonlinear rating prediction. The data utilized in the dataset is acquired based on the calculation using the demographic profile ranking below.

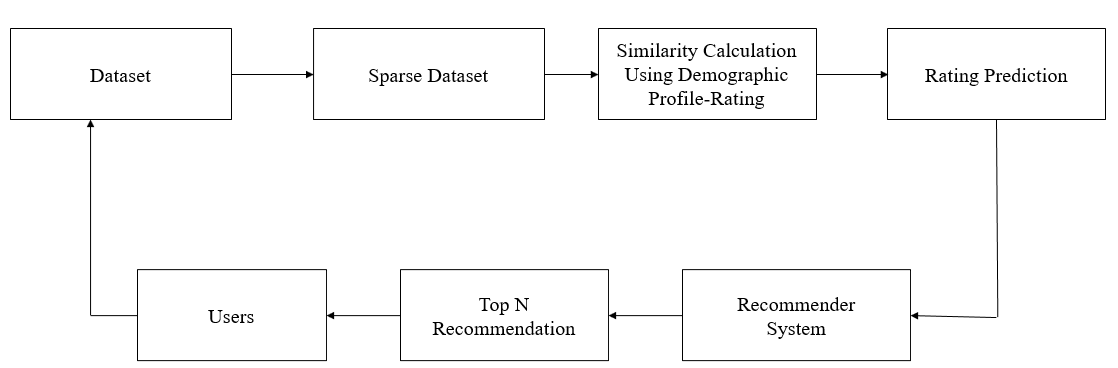


Figure 6 Flowchart of the proposed methodology

**Step 1.** In this step, the dataset is acquired based on the calculation using demographic profile ranking. This dataset is used for generating the list of data handed out for further processing of the methodology.

**Step 2:** The sparse dataset was divided into parts to generate a data similarity calculation using demographic profile rating corresponding based on their profile ranking to their users. In data analysis, sparse data is a variable in which some of the cells do not contain any meaningful information. The values in sparse data are either missing or zero. In contrast to missing data, which doesn't reveal what part or all of the values are, sparse data consists of empty or zero cells.

**Step 3:** Selecting similarity calculation data using the demographic profile rating method that considers the intercorrelation between features and the correlation with the dependent top n recommender variable system help establish the importance of a relevant feature set in a dataset. a similarity algorithm has been devised that factors in both user ratings and user behavior. In evaluating category information, the user score possibility is used to calculate the user behavior value. The algorithm has a drawback in that it only uses category data to derive value from user behavior.

**Step 4:** The number of outcomes getting from this applying extraction using rating prediction data detection transmission possibilities. In order to better anticipate the rating a user's review would get, it is helpful to combine the previously known data about each user's resemblance to other users with the sentiment analysis of the review text itself.

**Step 5:** Users can utilize the recommender system to discover specific information. The specific information is sought without having to type a long query into the information rescue system. The output feature of any recommended system is solely determined by the amount of information utilized to the top n recommender system. It's a ranking of the top N hotels that the current user enjoys the most . must include items/hotels that have not formerly been purchased by the current customer in the recommended list. It's also known as a Top-N recommendation algorithm interf

**Step 6:** Top N recommender systemto generate the best suggestion outcomes when a large amount of data is accessible. Work parallelism or control parallelism is used to create valuable recommendations. Delivering outcomes promptly is one of the advantages. The parallel execution of algorithms can make it easier to generate output without sacrificing performance.

**Step 7:** The last step of this methodology is to test the compute throughput basis on the above output the training of the model is done using the complete to meet the needs of the user, existing recommender algorithms must be modified.

**Results and Analysis**

In this Analysis, MAE and RMSE are used as measures of prediction accuracy. The MAE assesses how far the suggested method's predictions deviate from the actual ratings supplied by the user. Since RMSE squares mistakes before adding them up, it is especially harsh on large errors. This research examines the differences between the results obtained using cosine similarity and Pearson correlation after applying and disabling demographic filtering to determine which of these two measures provides the most reliable assessment of similarity. Both the cosine and normal Pearson formulas are provided below:

(3)

Normal Pearson correlation:

(4)

Where, is the rating of user u on item I, is the rating of user v on item I, , is the average rating of user u, , is an average rating of user v, is the rank of the rating of user u on item I, is the rank of the rating of user v on item I, is the average rank of ratings of the user u, is the average rank of ratings of the user v.

**Result 1: MAE calculation using cosine similarity without demographic filtering**

The resemblance among two vectors in an inner product space is measured by cosine similarity. In this analysis, MAE is calculated with the help of cosine similarity for the general user based on no. of ratings i.e., 20, 40, and 60. First, the MAE is calculated without using demographic filtering then in the second case demographic filtering is considered in the assessment. Figure 7 shows the result of the above calculation which clearly shows that the MAE which is calculated using cosine similarity is lower when demographic filtering is considered in each no. of ratings.

7(a) 7(b)

7(c)

Figure 7(a),7(b), and 7(c) show MAE calculation using cosine similarity at 20,40, and 60% ratings with or without demographic filtering.

**Result 2: MAE calculation using Pearson correlation**

The linear relationship between two variables is assessed using the Pearson correlation. In this analysis, MAE is calculated with the help of normal Pearson correlation for the general user based on no. of ratings i.e., 20, 40, and 60. First, the MAE is calculated without using demographic filtering then in the second case demographic filtering is considered in the assessment. Figure 8 shows the result of the above calculation which clearly shows that the MAE which is calculated using Pearson correlation is lower when demographic filtering is considered in each no. of ratings.

8 (a) 8 (b)

8 (c)

Figure 8 (a),8 (b), and 8 (c) show MAE calculation using Pearson correlation at 20,40 and 60% ratings with or without demographic filtering

**Result 3: RMSE calculation using cosine similarity**

In this analysis, RMSE is calculated with the help of cosine similarity for the general user based on no. of ratings i.e., 20, 40, and 60. First, the RMSE is calculated without using demographic filtering then in the second case demographic filtering is considered in the assessment. Figure 9 shows the result of the above calculation which clearly shows that the RMSE which is calculated using cosine similarity is lower when demographic filtering is considered in each no. of ratings.

9 (a) 9 (b)

9 (c)

Figure 9 (a), 9 (b), and 9 (c) show RMSE calculation using cosine similarity at 20,40, and 60% ratings with or without demographic filtering.

**Result 4: RMSE calculation using Pearson correlation**

In this analysis, RMSE is calculated with the help of normal Pearson correlation for the general user based on no. of ratings i.e., 20, 40, and 60. First, the RMSE is calculated without using demographic filtering then in the second case demographic filtering is considered in the assessment. Figure 10 shows the result of the above calculation which clearly shows that the RMSE which is calculated using Pearson correlation is lower when demographic filtering is considered in each no. of ratings.

10 (a) 10 (b)

10 (c)

Figure 10 (a), 10 (b) and 10 (c) shows RMSE calculation using Pearson correlation at 20,40 and 60% ratings with or without demographic filtering.

* **Comparative analysis**

Based on the above results, it is clear that the proposed method performs better than existing methods because in the proposed method demographic filtering is considered which reduces the MAE and RMSE values in both the cases by using cosine similarity or normal Pearson correlation. Hence it is concluded that the accuracy of the proposed method using demographic filtering is higher than the existing methods that are currently in use. Figure 11 shows a comparison graph based on the accuracy of the proposed method and the existing method.

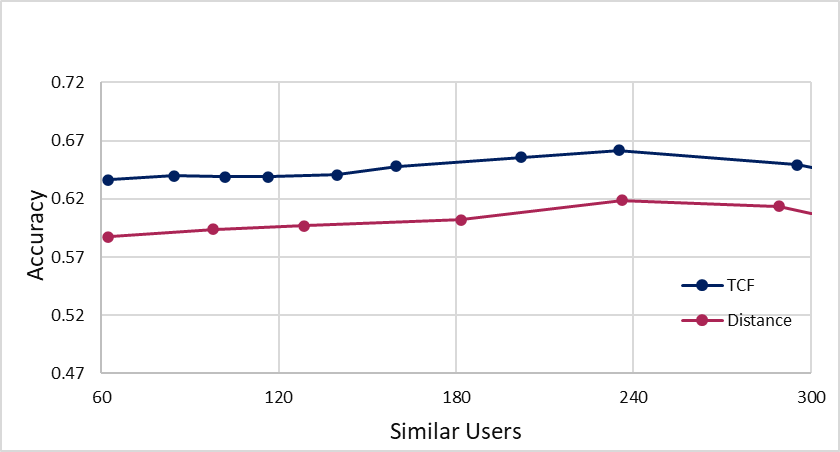


Figure 11 Comparison graph

**Conclusion and future scope**

As digital data grows exponentially, the problem of too much online information becomes increasingly pervasive. This makes decision-making harder for the user. Hence, online retailers and service providers employ recommender systems to solve this problem and meet consumer requests. This research proposes a demographically filtered recommender system. First, MAE and RMSE are determined using cosine similarity then Pearson correlation based on rating filtering at 20, 40, and 60% no. of ratings without considering demographic filtering. Then MAE and RMSE are calculated by considering demographic filtering. The result shows that by considering demographic filtering the MAE and RMSE values are lower which shows the accuracy of the proposed system.

Future studies will involve evaluating the examined methods across a variety of datasets and attempting to investigate further methods in this area. Additionally, it is decided to experiment with hybridization between the demographic recommender system and the collaborative recommender system.

**References**

1. Çano, Erion, and Maurizio Morisio. "Hybrid recommender systems: A systematic literature review." Intelligent Data Analysis 21, no. 6 (2017): 1487-1524.
2. Bobadilla, Jesús, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. "Recommender systems survey." Knowledge-based systems 46 (2013): 109-132.
3. Jannach, Dietmar, and Gerhard Friedrich. "Tutorial: recommender systems." In International Joint Conference on Artificial Intelligence Beijing, pp. 1-128. 2013.
4. Thorat, Poonam B., Rajeshwari M. Goudar, and Sunita Barve. "Survey on collaborative filtering, content-based filtering, and hybrid recommendation system." International Journal of Computer Applications 110, no. 4 (2015): 31-36.
5. Mehta, Pradnya, Onkar Dongare, Rushikesh Tekale, Hitesh Umare, and Rutwik Wanve. "A Survey on Hybrid Recommendation Systems." (2021).
6. Shambour, Qusai, and Salam Fraihat. "An item-based multi-criteria collaborative filtering algorithm for personalized recommender systems." International Journal of Advanced Computer Science and Applications 7, no. 8 (2016).

Kaya, B. (2020). A hotel recommendation system based on customer location: a link prediction approach. Multimedia Tools and Applications, 79(3), 1745-1758.

Sharma, Yashvardhan, Jigar Bhatt, and Rachit Magon. "A multi-criteria review-based hotel recommendation system." In 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, pp. 687-691. IEEE, 2015

Forhad, Md Shafiul Alam, Mohammad Shamsul Arefin, A. S. M. Kayes, Khandakar Ahmed, Mohammad Jabed Morshed Chowdhury, and Indika Kumara. "An Effective Hotel Recommendation System through Processing Heterogeneous Data." Electronics 10, no. 16 (2021): 1920.

Humming, Gao, and Li Weili. "A hotel recommendation system based on collaborative filtering and rank boost algorithm." In 2010 Second international conference on multimedia and information technology, vol. 1, pp. 317-320. IEEE, 2010.

He, Tianfu, et al. "What is the human mobility in a new city: Transfer mobility knowledge across cities." Proceedings of The Web Conference 2020. 2020.

Sivaramakrishnan, N., and V. Subramaniyaswamy. "GPU-based Collaborative Filtering Recommendation System using Task parallelism approach." 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud)(I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics, and Cloud)(I-SMAC), 2018 2nd International Conference on. IEEE, 2018

Sahoo, Abhaya Kumar, et al. "DeepReco: deep learning-based health recommender system using collaborative filtering." Computation 7.2 (2019): 25.

He, Xiangnan, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. "Neural collaborative filtering." In Proceedings of the 26th international conference on world wide web, pp. 173-182. 2017.

Wang, Wei, et al. "Trust-enhanced collaborative filtering for personalized point of interests recommendation." IEEE Transactions on Industrial Informatics 16.9 (2019): 6124-6132.

Jain, Ankush, Surendra Nagar, Pramod Kumar Singh, and Joydip Dhar. "EMUCF: Enhanced multistage user-based collaborative filtering through non-linear similarity for recommendation systems." Expert Systems with Applications 161 (2020): 113724.

Ładyżyński, Paweł, and Przemyslaw Grzegorzewski. "Vague preferences in recommender systems." Expert Systems with Applications 42, no. 24 (2015): 9402-9411.

Singh, Rahul, and Akshama Rani. "A survey on the generation of recommender systems." International Journal of Information Engineering and Electronic Business 9, no. 3 (2017): 26.

Choudhary, Priyankar, Vibhor Kant, and Pragya Dwivedi. "A particle swarm optimization approach to multi-criteria recommender system utilizing effective similarity measures." In Proceedings of the 9th International Conference on Machine Learning and Computing, pp. 81-85. 2017.

Amannejad, Yasaman, Diwakar Krishnamurthy, and Behrouz Far. "Managing performance interference in cloud-based web services." IEEE Transactions on Network and Service Management 12, no. 3 (2015): 320-333.

Tao, Longquan, Jinli Cao, and Fei Liu. "Quantifying textual terms of items for similarity measurement." Information Sciences 415 (2017): 269-282.

Wang, Qinyong, Hongzhi Yin, Tong Chen, Junliang Yu, Alexander Zhou, and Xiangliang Zhang. "Fast-adapting and privacy-preserving federated recommender system." The VLDB Journal 31, no. 5 (2022): 877-896.

Akbar, Ali, Parul Agarwal, and Ahmed J. Obaid. "Recommendation engines-neural embedding to graph-based: Techniques and evaluations." International Journal of Nonlinear Analysis and Applications 13, no. 1 (2022): 2411-2423.

Ali, Sadia, Yaser Hafeez, Mamoona Humayun, Nor Shahida Mohd Jamail, Muhammad Aqib, and Asif Nawaz. "Enabling recommendation system architecture in a virtualized environment for e-learning." Egyptian Informatics Journal 23, no. 1 (2022): 33-45.

Lian, Sixian, and Mingdong Tang. "API recommendation for Mashup creation based on neural graph collaborative filtering." Connection Science 34, no. 1 (2022): 124-138.

Chi, Xiaoxiao, Chao Yan, Hao Wang, Wajid Rafique, and Lianyong Qi. "Amplified locality‐sensitive hashing‐based recommender systems with privacy protection." Concurrency and Computation: Practice and Experience 34, no. 14 (2022): e5681

Seth, Rakhi, and Aakanksha Sharaff. "A Comparative Overview of Hybrid Recommender Systems: Review, Challenges, and Prospects." Data Mining and Machine Learning Applications (2022): 57-98.

Krishnan, Rahul, and Prashant R. Nair. "RFM-Based Customer Analysis and Product Recommendation System." In Advanced Computing and Intelligent Technologies, pp. 159-164. Springer, Singapore, 2022.

Shao, Tingting, Xuan Yang, Fan Wang, Chao Yan, and Ashish Kr Luhach. "Trusted Service Evaluation for Mobile Edge Users: Challenges and Reviews." Complexity 2021 (2021).

Hietala, Joona. "Matrix factorization algorithms for personalized product recommendation: a case study." (2021).

Suresh, A., and Carmel Mary Belinda MJ. "A comprehensive study of hybrid recommendation systems for e-commerce applications." Int. J. Adv. Sci. and Technol 29, no. 3 (2020): 4089-4101.

Sharma, Richa, Shalli Rani, and Sarvesh Tanwar. "Machine Learning Algorithms for building Recommender Systems." In 2019 International Conference on Intelligent Computing and Control Systems (ICCS), pp. 785-790. IEEE, 2019.

Kaur, Gursimarpreet, and Saroj Ratnoo. "Adaptive Genetic Algorithm for Feature Weighting in Multi-Criteria Recommender Systems." Pertanika Journal of Science & Technology 27, no. 1 (2019).

Selvi, C., and E. Sivasankar. "A novel similarity measure towards effective recommendation using Matusita coefficient for Collaborative Filtering in a sparse dataset." Sādhanā 43, no. 12 (2018): 1-13

Yang, Fan. "A hybrid recommendation algorithm–based intelligent business recommendation system." Journal of Discrete Mathematical Sciences and Cryptography 21, no. 6 (2018): 1317-1322.

Sang, Anshu, and Santosh K. Vishwakarma. "Design and implementation of a collaborative filtering approach for movie recommendation system." International Journal of Computer Applications 167, no. 12 (2017): 18-24.

Kaudan, Aditya, Siddhant Bansal, and Yashi Sahu. "Demography Based Trends in Management Research in Service Industries in India."(2014).

Liao, Mengqi, and S. Shyam Sundar. "When e-commerce personalization systems show and tell: Investigating the relative persuasive appeal of content-based versus collaborative filtering." Journal of Advertising 51, no. 2 (2022): 256-267.

Sharma, Sunny, Vijay Rana, and Manisha Malhotra. "Automatic recommendation system based on hybrid filtering algorithm." Education and Information Technologies 27, no. 2 (2022): 1523-1538.