Machine learning-based movie recommendation engine

[1]Sonu Airen, [2]Puja Gupta

Assistant professor
Department of Information Technology
Shri G.S. institute of technology & science, Indore, India

Abstract:
We propose a movie recommendation engine that can offer films to both new and existing customers as part of this study. The system searches movie databases to gather relevant information, including popularity and beauty, needed to make a recommendation. We use content-based filtering and collaborative filtering, and we evaluate the advantages and disadvantages of each strategy. We use hybrid filtering, combining the outcomes of two algorithms to create a system that offers more precise movie recommendations. Recommendation engines are used for commercial purposes and aid in developing strategies for businesses. Recommendation systems are crucial due to the growing demands of customers and user suggestions. Recommender systems help us find more relevant searches, improving our time management in a busy workplace. It is typical to use these systems alongside movie websites or other business apps due to their high utility. Particular results may be achieved with the use of this kind of recommendation system. Movie suggestions will become more tailored to customers' needs as a consequence of this.

Keywords: Recommendation System, Machine Learning, content-based filtering.

Introduction:
Recommendation systems are becoming more crucial in today's exceptionally hectic society. People are constantly pressed for time due to the many responsibilities they must do within the limited 24-hour day. Recommendation structures are crucial as they assist individuals in making informed decisions without depleting their cognitive resources. A recommendation system's primary purpose is to identify information that would be exciting to a certain user. It also offers several features to provide personalized lists of useful and engaging material tailored to each user. Recommendation structures are algorithms based on Artificial Intelligence that analyze all potential options and provide a personalized list of items that might be interesting and pertinent to an individual.
The primary objective of this research paper is to develop an enhanced recommender system that offers accurate suggestions to the consumer. Recommendation engines[1] often combine content-based filtering as well as collaborative filtering techniques. Collaborative Filtering relies on the user's past queries and experiences. We will predict what other users with comparable experiences would see and provide those suggestions to the user. Collaborative filtering involves generating recommendations and suggestions for consumers or users based on their prior experiences and behaviors. The premise is that it operates when you pick an item in the basket or make a transaction on the website. It will provide similar suggestions based on the products you have selected in the basket. Content-based [2] filtering approaches provide suggestions by analyzing movie attributes including genre, director, actor, and narrative. To enhance the advice, we need to focus more on a certain quality. We will also include a popularity and rating component into this. A content-based filtering strategy is implemented using Cosine similarity and Term Frequency-Inverse Document Frequency (TF-IDF) Vectorizer. Content-based filtering provides suggestions according to the user's interests. If the user has searched the item's previous history. This filter works by recommending information similar to what the user has already seen. This strategy relies entirely on the preferences of the consumers. Currently, we use a combination of filtering methods known as hybrid filtration. The combination of both forms of filtering will create a new strategy known as the hybrid model. This model will provide us with a more accurate method for filtering processes.

Literature Review:

This portion of this chapter provides a comprehensive examination and evaluation of the existing literature pertaining to the chosen issue, elucidating how the findings and practical implications of these studies will contribute to the current investigation.

singh,P.K et.al[3] provide an in-depth overview of movie system recommendations, emphasizing the methods used, obstacles encountered, and potential research paths in the area. Movie recommendation systems are essential for assisting consumers in finding new films that match their interests, thereby improving their watching experience.

The authors in reference [5] proposed a practical approach for using data mining to generate a suggestion list. Their approach was based on pairs of things that were quicker than the standard ARM. The average score for the suggestions was 88.94%.

Privacy-preserving collaborative filtering is gaining more attention due to the growing need for safeguarding sensitive data while providing recommendations. To enhance the experience of statistics owners and provide forecasts, many systems were suggested to estimate indicators while safeguarding anonymity. These solutions use advanced privacy-preserving algorithms to address and minimize privacy, financial, and legal issues of data owners [4].

collaborative filtering,[6] which is based on the assumption that customers who have purchased a certain product would have comparable demands with other users who have also purchased the same product. The author analyzes and evaluates three types of a collaborative filtering-based recommender system using customer previous purchases, consumer browsing patterns, and user groups.

It has become more common to make use of the suggestion machine as a result of the development of the Internet and online shopping. A similar approach[7] is used in this work by the electronic commerce recommendation system, which also makes a specialty of the collaborative algorithm for filtering in the context of the application of the customized film recommendation system.

This article[8] aims to minimize human work by recommending movies according to the user's preferences. We devised a paradigm that integrates both content-based and collaborative approaches to address these issues. It will provide more detailed results in comparison to other systems that rely on a content-based approach. Content-based recommendation systems are limited to individuals and do not provide suggestions beyond what is already known, so restricting the opportunity to discover new options. Therefore, we have concentrated on a system that addresses these problems.

This article[9] presents MovieGEN, an expert system designed for recommending movies. We use machine learning and cluster analysis in our system, using a hybrid recommendation technique. The technology collects users' personal data and forecasts their movie choices using highly trained support vector machine (SVM) models. The SVM prediction is used to choose movies form the dataset, cluster them, and create questions for the users.

The author in [10] analyzed E-commerce large data, focusing on the K-means clustering technique. The research utilizes geographic location and the customer's unique identification number as clustering limitations. Mining such data is a challenging task. Clustering similar items or data is a crucial mining process that provides considerable benefits for classification and modeling. The K-means clustering method is a leading partition-based clustering technique known for generating high-quality results.

A movie recommendation[11] system was conceptualized and executed in reference [9]. There are a variety of international film genres, cultures, and languages from which to choose. Users may receive movie recommendations predicated on the films' notoriety or their personal interests. An average of 600 films are published annually in Hollywood, according to a survey. Important components of recommendation algorithms are streaming movie services such as Netflix. By facilitating the process of consumers finding fresh films to view. A considerable quantity of effort has been expended thus far in this domain. Nevertheless, one can always find room for enhancement.

The authors implemented[12] a movie recommendation utilizing collaborative filtering in [10]. This system is constructed utilizing Apache Mahout and provides movie recommendations by analyzing the evaluations. The system presented the unprocessed results obtained from the collaborative filtration method. The system provides users with a recommendation of ten films and returns the nearest neighbors whose taste preferences are most similar to the user's.

Proposed Methods:

There are classification algorithms that are used in the field of machine learning. These algorithms use several methods of arranging and categorizing information.

- Collaborative Filtering

Collaborative filtering is a widely used recommendation method that examines users' previous interactions with objects, such as movie reviews, to anticipate current preferences. The system discovers commonalities among users via their ratings and suggests things favored by people with similar preferences. Collaborative filtering utilizes the combined knowledge of a vast user community to provide tailored suggestions, even for users with limited or missing information. This method is adaptable and may be used to a variety of recommendation domains, such as movies, music, and e-commerce, without needing specific information on the objects.

Figure 1: shows collaborative filtering

The Singular Value Decomposition (SVD) is a linear algebra technique often used in machine learning for dimensionality reduction. SVD is a matrix factorization method that reduces the number of features in a dataset by reducing the space dimension from N to K, where K<N. The SVD is a collaborative filtering method used in recommender systems. The data is structured in a matrix format, where each row corresponds to a user and each column corresponds to an item. Users' ratings for things are the components of this matrix.

- Content Based Filtering

Content-based filtering suggests movies by analyzing their characteristics and consumers' tastes. The system examines movie attributes including genre, director, and actors to provide suggestions. Content-based filtering recommends movies based on users' historical choices and the content of movies to match their likes. It does not depend on user activities or the preferences of comparable users, unlike collaborative filtering. It concentrates on the inherent characteristics of movies and strives to align them with consumers' preferences. Content-based filtering is beneficial for suggesting specialized or less mainstream films that may not have received many ratings.

Content Based Filtering using Cosine Similarity:

Content-Based The process of filtering using Cosine Similarity measures the likeness of movies by considering their characteristics, such genre, actors, and storyline keywords. The algorithm calculates the cosine similarity between two video vectors in a multi-dimensional space, with each dimension corresponding to a feature. Greater cosine similarity values indicate stronger resemblance between the movies. This method suggests films that closely match the user's tastes, prioritizing content above user engagement.

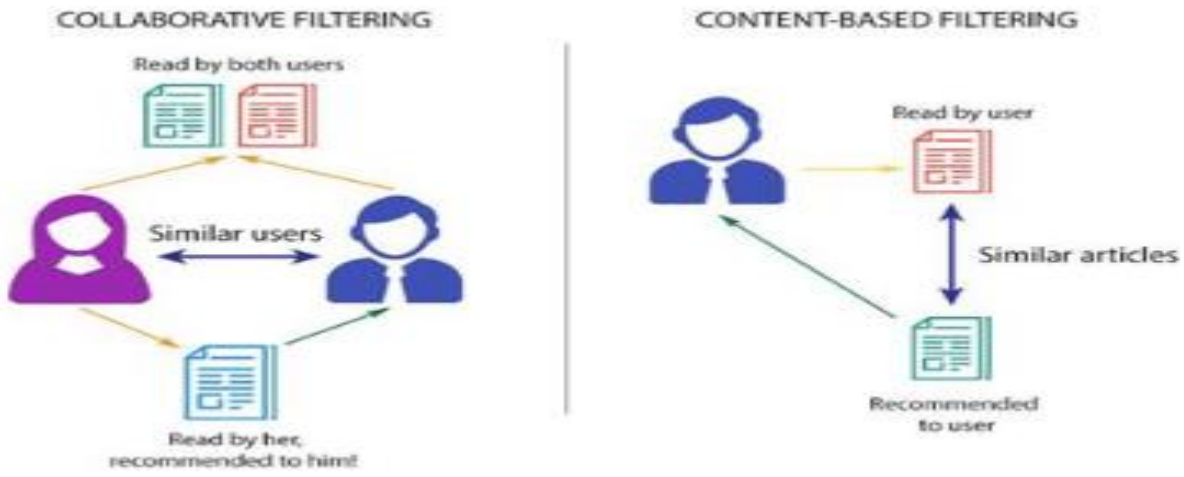

Figure 2 for Content Based Filtering using Cosine Similarity:

Content-Based TF-idf filtering adds weights to terms in movie descriptions based on their frequency in the description and rarity in the full corpus. TF-idf considers the significance of words in depicting the content of a movie. It computes a score for each word, indicating its significance to the movie. Movies that have identical TF-idf scores for common terms are deemed similar, whereas those with different word distributions are regarded dissimilar. This approach suggests movies by analyzing their textual content to identify semantic similarities with user preferences.

Results and Discussion

Following our learning and examination of the aforementioned strategies, we have made an effort to put them into practice. All of the implementations have been carried out on the datasets that were provided. We have obtained the dataset from the movies lens website, which is where we obtained it. There is a massive database of movies that it contains. The TMBD ratings have been taken into consideration for every one of our implementations. In our evaluation, we have taken into account the many genres, cast members, crew members, reviews, and ratings. We have used a variety of filtering strategies, including content-based filtration and collaborative filtration, among others.Graph of the Frequency of Movie Ratings, Figure is a bar graph that illustrates the number of reviews.A graph is shown in Figure 6 that illustrates the number of ratings that were assigned to the entire number of movies that were included in the database. The graph shown in Figure is used for the purpose of analyzing the ratings and the quantity of reviews that a certain movie is receiving. In essence, this is a bar graph that illustrates the number of reviews that a certain rating might have.

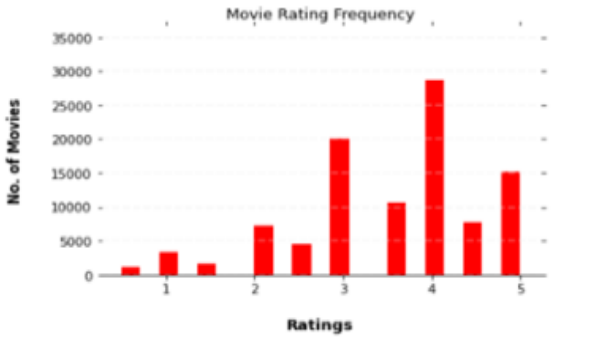

Figure 3 : Movie Rating frequency

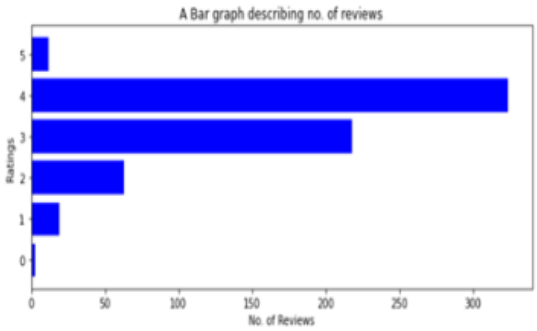

Figure 4:Movie recommendation ratings

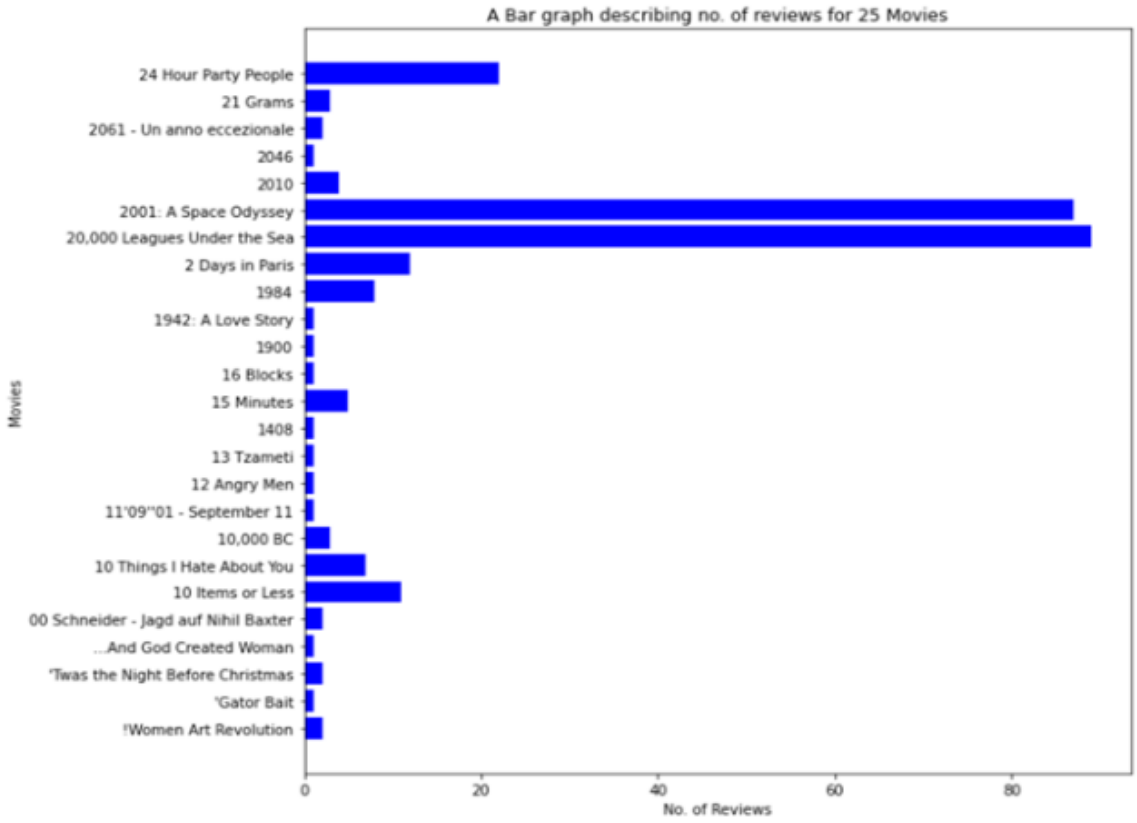

Figure 5: No of reviews for 25 movies

Conclusion

The purpose of this article was to demonstrate the use of a movie recommendation system that was built on machine learning methods. Therefore, consumers obtain better recommendations as a result of collaborative filtering, which is based on their previous experiences and activities. This is because of the fact that each user's experiences and activities are taken into consideration. In the process of Collaborative Filtering, we used the SVD algorithm in order to provide movie recommendations to the user. The underlying difficulty with collaborative filtering is that if a new user has no past experience, the recommender cannot deliver meaningful suggestions. It is also likely that collaborative filtering will not be able to provide recommendations that are useful if the amount of data that is being processed is significantly increased. In content-based filtering, recommendations are generated by comparing the characteristics of the object that is being defined with those of other things. In order to do content-based filtering, a TF-IDF vectorizer and a cosine similarity were used. The results that the TF-IDF vectorizer produces are superior to those that are produced by the cosine similarity algorithm since it is able to count each and every word in movie genres, actors, and directors.

REFERENCES:

[1] Sudhanshu Kumar, Kanjar De and Partha Pratim Roy, "Movie Recommendation System Using Sentiment Analysis From Microblogging Data", IEEE Transactions on computational social systems., vol. 7, no. 4, pp. 915-923, Aug. 2020.

[2] Mooney, R.J. and Roy, L., 2000, June. Content-based book recommending using learning for text categorization. In Proceedings of the fifth ACM conference on Digital libraries (pp. 195-204).

[3] Singh, P.K., Pramanik, P.K.D., Dey, A.K. and Choudhury, P., 2021. Recommender systems: an overview, research trends, and future directions. International Journal of Business and Systems Research, 15(1), pp.14-52.

[4] Bilge, A., Kaleli, C., Yakut, I., Gunes, I., Polat, H.: A survey of privacy-preserving collaborative filtering schemes. Int. J. Softw. Eng. Knowl. Eng. 23(08), 1085–1108 (2013)CrossRefGoogle Scholar.

[5] Hosein Jafarkarim i; A.T.H. Sim and R. Saadat Boost A Naïve Recommendation Model for Large Databases, International Journal of Information and Education Technology, June 2012
[6] Malik, S., 2020. Movie Recommender System Using Machine Learning. EAI Endorsed Transactions on Creative Technologies, 9(3).
[7] Peng, Xiao, Shao Liangshan, and Li Xiuran. "Improved Collaborative Filtering Algorithm in the Research and Application of Personalized Movie
[8] Furtado, F. and Singh, A., 2020. Movie recommendation system using machine learning. International journal of research in industrial engineering, 9(1), pp.84-98.
[9] Eyjolfsdottir, E.A., Tilak, G. and Li, N., 2010. Moviegen: A movie recommendation system. UC Santa Barbara: Technical Report.
[10] Shaik, I.; Nittela, S.S.; Hiwarkar, T.; Nalla, S.; "K-means Clustering Algorithm Based on E-Commerce Big Data", International Journal of Innovative Technology and Exploring Engineering, vol: 8, 2019,pp: 1910-191
[11] Pavithra, M.; Sowmiya, S.; Tamilmalar, A.; Raguvaran,S.; "Searching an Optimal Algorithm for Movie Recommendation System", International Research Journal of Engineering and Technology, vol: 6, 2019,pp:216-221
[12] Wu, C.-S. M., Garg, D., &B handary, U. (2018). Movie Recommendation System Using Collaborative Filtering, 23-25 Nov, 2018, IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2018.
[13] P Gupta, U Singh "A Review of Friend Recommendation on Social MInternational Journal of Science, Engineering and Technology 157 2020,
[14] Gupta, P. and Kulkarni, N., 2013. An introduction of soft computing approach over hard computing. International Journal of Latest Trends in Engineering and Technology (IJLTET), 3(1), pp.254-258.
[15] Gupta, P., Sharma, V. and Varma, S., 2020. People detection and counting using YOLOv3 and SSD models. Materials Today: Proceedings. https://doi. org/10.1016/j. matpr, 562.