iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI: https://doi.org/10.22214/ijraset.2022.45578

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Machine Learning to Be Used in Recommendation Systems

Shashank H V[1], Ms. Sindhu D[2], Dr. Ravikumar G K[3]
[1]*Dept. of CSE,* [2]*Dept. of ISE,* [3]*Professor & Head(R&D) Dept. of CSE, BGS Institute of Technology Adichunchanagiri University, BG Nagar, Karnataka,India-571448.*

Abstract: ***Basic information on machine learning and recommendation systems, as well as case studies, are included in the paper. The techniques used in such systems, known as machine learning, were investigated in greater detail. The essay concentrated on content filtering algorithms as well as proximity- based filtering algorithms. There are explanations of the commonalities, restrictions, and benefits of different algorithms, along with measures for evaluating the algorithm and calculating thesampling result of the assessment prediction. The categorization of the used information from the Movie critic portal kicks off the project's planning phase. The technology, design, and deployment of the aforementioned algorithms are then displayed.The next part gives an overview of the findings and suggestions development of computer simulations of how the algorithms work. At the end of the assessment, you'll find a summary, an evaluation of the effectiveness of suggestion methods, and lessons learned from the study, as well as a request for more research on the subject.***
Keywords: Recommendation, Machine Learning (ML),System Filtering.

I. INTRODUCTION

In today's environment, the Internet has grown pervasive. You can use it for shopping, going to the cinema, listening to songs, and connecting with friends, to name a few things. Because of the actions of a vast number of people - Internet users - information from hundreds, lakhs, or even millions and millions of people can be collected. For those that examine this data, it's a once-in-a-lifetime opportunity. This allows them to characterize the under researched society's intellectual capacity, aswell as its actions, preferences, and viewpoint. For example, in the marketing sector, this information is useful. They can help determine whether a company's sales methods for a given social group are productive, as well as the correct approach to reach a specific demographic. The Social-media, on the other hand, allows you to keep track of network users' diverse activities without harming their intents.

You may get data from a wide range of social groupings all across the world. The most challenging challenge that data analysis system developers and inventors encounter is examining the database. Dueto its size, people are unable to do intricate and time-consuming calculations on it. The issue with machine learning is that it need assistance. Statistical methods, such as regression and correlation analysis, are one of the most basic ways of machine learning. Training neural networks or fuzzy logic are examples of more advanced approaches. The designer constructs a recommendation methodology, and the computer uses it to establish the conclusions connected to the qualitiesof a particular set of data. Such systems open up the possibilities. The use of recommendation algorithms is growing in popularity. The rise of the Internet and its outcome, the flow of information, has compelled them to conduct research on their subject. Thousands of films, millions of scientific publications, and an incalculable amount of music are available. These figures indicate that no single person would be able to somehow get through everything in their lifetime. For such people, the tips are invaluable.

II. RELATED WORK

Consumer expectations during Black Friday purchases In our daily lives, we focus on suggestions from colleagues, relatives, the networks, online networks, and other outlets. Recommender systems [2] are now arequired component of any e-commerce site. The question that arises here is why does any organization need to offer suggestions in the first place. The most basic explanation is that recommendation systems subject users to a vast number of items in which they feel they will be interested. To further understand, let's look at a simple example of YouTube suggestions. Theprimary goal of delivering suggestions is to improve the amount of time an individual's spend on the site as well as the number of videos he or she can watch. YouTube's recommender system uses several factorsto provide individualized recommendations.

YouTube's recommender system combines the users' individual behavior on the platform with the rules for associating related videos to provide personalized recommendations.

©IJRASET: All Rights are Reserved

Any video that the user has expressly liked, added to a playlist, or rated serves as a basic seed suggestion component. A recommender system and content-based filtering are two essential concepts associated with creating suggestions for any suggestion system. A suggestion system [6] is a mechanism that matches auser's preferences, such as ratings or purchases, with those of other site users who share those preferences. This can be accomplished by comparing user or itemsimilarities. The two approaches are known as subscriber summarization and item-basedinformation retrieval. Item-to-item collaborative filtering is used by YouTube and Amazon to recommend things. When it comes to larger datasets,item-based techniques outperform their equivalents. Because Items are rarely changed, this could be estimated in the digital platform as well [8].

Various ways can be used to construct recommender systems, such as developing from scratch, Using a pre-existing recommender engine or choosing a framework based on the needs. It's also possible to employ a mix of these methods. One of these systems, Apache Mahout, is the subject of our next piece.

III. RESEARCH METHODOLOGY

Artificial intelligence (AI), machine learning (ML), and deep learning (DL) are three fields that are currently in vogue. Although these names are frequently used interchangeably, this is not always the case. It is important to note that, of all these words, AI is the most generic, with ML being a part of Ai and DL being a subset of ML [7]. Any machine learning algorithm's fundamental goal is to generalize far beyond training samples and successfully grasp and analyze information that it had not seen before. There are several methods for developing a recommender system, each of which is based on various features of the acquired data and the context in which it exists. The use of collaborative filtering algorithms is more common than the use of other algorithms, and it often results in superior predicted performance. Furthermore, each approach has its own set of benefits and drawbacks that must be considered before application.

A. Content-Based Filtering

This filtering is focused on the premise that each item has some characteristics. Recommender systems employ a content-based recommendation technique to examine paperwork and/or summaries of items previously evaluated by the user and construct a model or profiling of user interests based on the characteristics of the articles rated by that user [10]. Users have a set of interests relating to the contents of items. A user's profile might be made expressly either by the user or generated automatically based on his or her previous actions. The biography is a systematic arrangement of a user's interests that is used to suggest new items that are interesting to them. Aligning the characteristics of the particular user also with the attributes of an item can be characterized as the recommendation process.

The result is a relevance score, which indicates how interested the user is in that particular content. It is a big benefit for the efficiency of a knowledge discovery operation if a user profile accurately models user interests. Content-based recommendation necessitates the employment of appropriate techniques for representing items and establishing user profiles, as well as a mechanism for comparing a user profile to an item's representation. The content-based recommendation has a number of advantages, including autonomy from users, transparency, and so on. The content- based strategy, like any other technique, has certain important drawbacks that must be considered: restricted critical review or over-specialization.

There are many apps that use content-based recommender systems. LIBRA [11] recommends books based on product descriptions taken from the Walmart online shop and a naive Bayes document classification algorithm. Another interesting example is Romantic [12], which uses document classification learning to improve from movie synopses gathered from IMDB to generate movie suggestions. To receive recommendations from the algorithm, the user must categorize a minimum number of films as dreadful, bad, below average, above average, good, and exceptional. In news proposals, where Daily Learner must be stated, recommenders are also present. It maintains two user models: one just for short-term interests and the other for long-term interests.

B. Collaborative Filtering

Collaborative filtering involves evaluating the behavior of a group of users in order to offer recommendations to other users. The recommendation is influenced by the choices of other users. The primary premise behind the interactive postprocessing methodology is that when a person shares the same view as just another person on either a topic, he is so much more interested in sharing that other person's perspective on a different topic than a randomly selected individual. One of the clearest examples is if a user receives a movie suggestion from a buddy who gave it a high rating and they also have a comparable rating history.

A matrix can be used to show a series of interactions, with each entry I j) here on the matrix representing communication between the user I and the item "j." Clustering algorithm can also be considered a broadening of analysis and categorization.

When opposed to content-based systems, the collaborative filter- based suggestion has various advantages, including no domain expertise, serendipity, an affinity for nuances, and the benefits of huge user base. Cold start (the system requires sufficient information to perform effectively) and complexity are some of the disadvantages of collaborative filtering-based systems.
Collaborative filtering is one of the most popular methods right now, and it routinely surpasses content-based suggestions. YouTube's recommended algorithm is a great example. Their method is made up of two machine learning that operates together to generate suggestions, one for selecting applicants and the other for ranking.

C. Hybrid System

The simplest and easiest way to create a combination recommender system is to consider taking the independent results of subject matter and a collaborative-based recommendation engine, then combine their predictions methods using a voting scheme, where the combination is done by selecting items that match the user's profile while also receiving favorable reviews from the user's neighbors. The method employed in [16] compares users based on their evidenced profiles and employs a collaborative filtering approach that employs the derived similarity metrics. The content-based predictions are utilized to enhance the rating matrix in [17], and then the clustering method is used. The item-based clustering algorithm is used in [18], but before that, the item's topic descriptions and associated ratings are used.

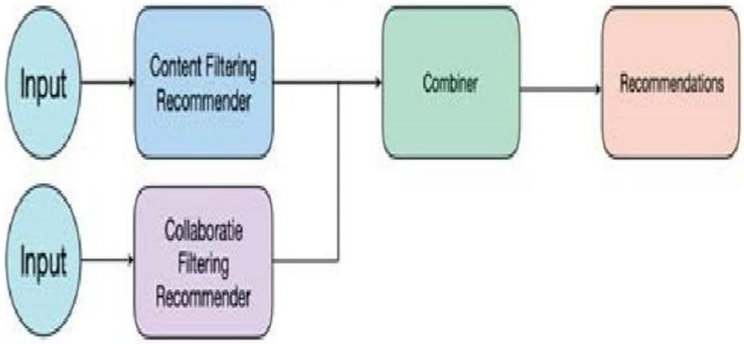


Fig1. The basic architecture of a hybrid recommender system

IV. RESULTS AND DISCUSSIONS

The Python programming language was used tocreate the scripts that calculated the forecasts. For computations on matrices and tables, the libraries pandas and numpy were utilized. Technologies are based on the closeness of persons and objects used to generate predictions. The most difficult task is to create a user-object vector in order to calculate the predicted value(Tab. 1.).

TABLE I. PART OF THE USER -OBJECT MATRIX

title	'Til There Was You (1997)	1-900 (1994)	101 Dalmatians (1996)	12 Angry Men (1957)	187 (1997)	2 Days in the Valley (1996)	20,000 Leagues Under the Sea (1954)	2001: A Space Odyssey (1968)	3 Ninjas: High Noon At Mega Mountain (1998)
user_id									
1	NaN	NaN	2.0	5.0	NaN	NaN	3.0	4.0	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0
3	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	NaN	NaN	2.0	NaN	NaN	NaN	NaN	4.0	NaN

The outcome of the similarity function must then be calculated. If you use the cosine distance to determine the matching factor, it would take too long to calculate all of the numbers. All similarity measures are calculated using Pearson's coefficient of correlation (1) [3,4,5,6] due to limitations imposed by the processing capacity of the hardware on which the scripts were run.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

$$s(u,a)=\frac{\sum_{i\in I_u\cap I_a}(r_{u,i}-\bar{r}_u)(r_{a,i}-\bar{r}_a)}{\sqrt{\sum_{i\in I_u\cap I_a}(r_{u,i}-\bar{r}_u)^2}\sqrt{\sum_{i\in I_u\cap I_a}(r_{a,i}-\bar{r}_a)^2}} \quad (1)$$

where:
I_u, I_a - collections of object rated by users a and u
$r_{u,i}, r_{a,i}$ – rating issued to the object by the user a and u
$\bar{r}_u, \bar{r}_a$ - average of all ratings issued by the user a and u

TABLE II. PART OF THE SIMILARITY FUNCTION MATRIX FOR A CF ALGORITHM BASED ON USER SIMILARITY

user_id	1	2	3	4	5	6	7
user_id							
1	1.000000	0.160841	NaN	NaN	0.420809	0.295410	0.258137
2	0.160841	1.000000	NaN	NaN	NaN	0.446966	0.643675
3	NaN	NaN	1.0000	-0.2626	NaN	-0.109109	0.064803
4	NaN	NaN	-0.2626	1.0000	NaN	NaN	-0.266632
5	0.420809	NaN	NaN	NaN	1.000000	0.241817	0.175630

The coefficient was determined for objects or users, depending upon the nature of the method. Tab. 2 shows the similarity matrix for something like a Participatory Filtering technique based on user similarities.

The analysis and forecast values for the optimization method on object similarity were done assuming different preferences of the minimal number of shared evaluations between the items under examination and a varying collection of instances in the neighborhood. Because of the simulation's computational complexity and length, the following numbers for the minimum number of assessments were supposed: 150, 100, 50, 25,15, 10, 5.

However, the following assumptions were made about the number of things in the neighborhood: 1, 2, 3, 4, 5, 10, 15, and 20 are the numbers. The figure of the minimum number of evaluations was determined for perhaps the MovieLens 1M Dataset database resulting from the operation on matrices with significantly larger dimensions: 50, 20, 5.

The RMSE level, MAE quality, number of suggestions, and script execution time were all examined as part of the examination of the obtained estimated values for the provided data

set. They were all shown in proportion to the amount of goods in the area. The list of suggestions was calculated as a percentage of the total number of grades in the database.

After obtaining matrices reflecting the user-object relationship as well as the parameters of the clustering algorithm, the prediction values had to be calculated. (2) and (3) [4,5] were employed in the instance of a CF method based on user similarity.

$$p_{u,i}=\bar{r}_u+\frac{\sum_{a=1}^{n}s(u,a)(r_{a,i}-\bar{r}_a)}{\sum_{a=1}^{n}|s(u,a)|} \quad (2)$$

where:
$p_{u,i}$ - prediction value for the u user's assessment of the object i
$\bar{r}_u$ - the average of all ratings issued by the user u
$\bar{r}_a$ - average of all ratings issued by user a, user's u neighbour
$r_{a,i}$ - rating of the object i by the user a
$s(u,a)$ – value of the similarity function between the user u and a
n - number of the nearest neighbours of the user u

$$p_{u,i}=\bar{r}_u+\sigma_i\frac{\sum_{a=1}^{n}s(u,a)(r_{a,i}-\bar{r}_a)/\sigma_a}{\sum_{a=1}^{n}|s(u,a)|} \quad (3)$$

where:
σ_i - standard deviation of ratings issued by the user i
σ_a - standard deviation of ratings issued by the user a

$$p_{u,i}=\frac{\sum_{j\in S_i}r_{u,j}s(i,j)}{\sum_{j\in S_i}|s(i,j)|} \quad (4)$$

where:
$r_{u,j}$ - the rating issued by the user u for the object j

It was, however, used (4) in a methodology based on entity similarities, while taking into account objects for which the commonality function was set to (i)0. After it was calculated, the forecast's value was compared to the true estimate provided by the individual under evaluation for a specific item. Then, using (5) and (6), MAE and RMSE faults were determined (6). The length of the script's execution and the quantity of predictions obtained were both noted.

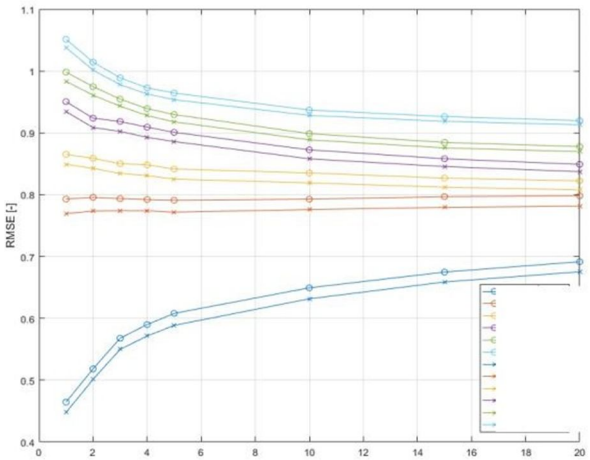


Fig2. The relationship between RMSE andthe number of facilities in the area

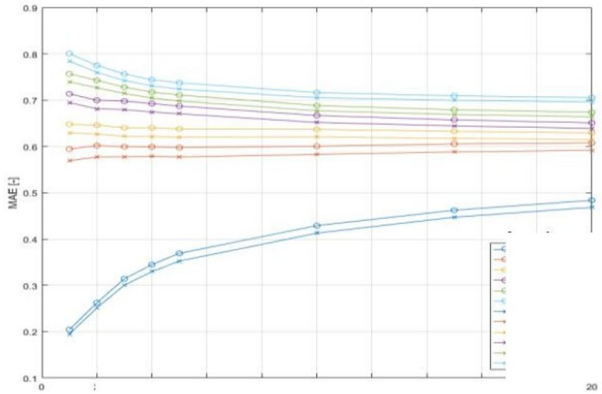


Fig3. MAE's reliance on the quantity ofparticipants in the immediate vicinity

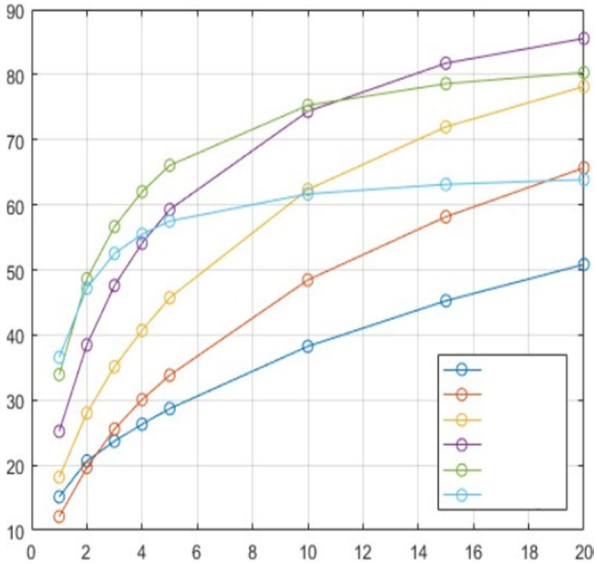


Fig4. The quantity of recommendations isproportional to the number of users in the area.

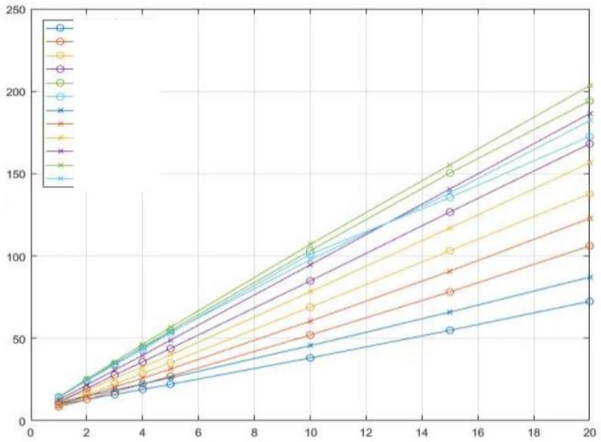


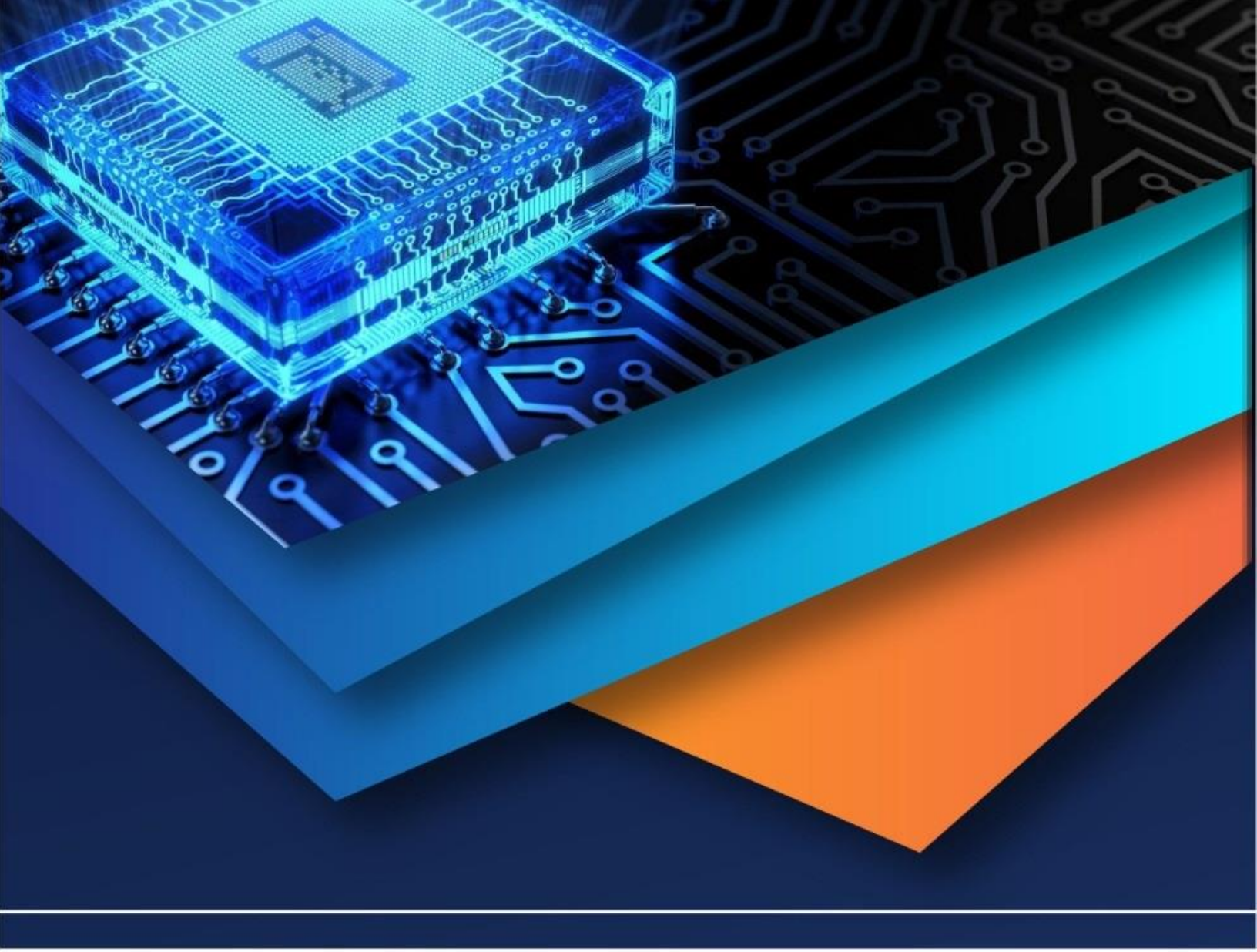
Fig5. The movie's length is determined by theamount of facilities in the area.

V. CONCLUSION

The article highlights the difficulties that come with using machine learning in recommendation systems. After that, mechanisms for generating review prediction performance were incorporated for movie portal visitors. These days, recommendation engines are a common marketing tactic. Only a few of the approaches utilised to create suggestions are described in this study. Because these are the most generally utilised techniques of creating instructions, The emphasis is on content and algorithms that are based on the neighborhood. Using account information databases, an attempt was made to create algorithms that calculate the prediction rating. This objective was accomplished using an algorithm based on user and object similarities. The technique based on user similarity produced better results for the studied data sets. For this type of procedure, the errors assessed - RMSE and IEA - were lower than for a method based on similar items. Furthermore, this system enabled more recommendations to be made in less time. However, it is not reasonable to conclude that the strategy related to existing consumers is preferred in all systems derived from the findings of these investigations. The variables were promised a certain quantity of necessary user reviews in the situation under discussion. In actuality, this is a rare occurrence, and you may need to generate a suggestion for individuals who have just given a few ratings. More work on the recommendation system problem has included an attempt to develop hybrid ways to reduce predicted value calculation errors.

REFERENCES

[1] T. Segaran, ,, Nowe usługi 2.0. Przewodnik po analiziezbiorów danych", Gliwice, Wydawnictwo Helion, 2014 r.
[2] M. Mohri, A. Rostamizadeh, A. Talwalkar „Foundations of Machine Learning", Cambridge, London, The MIT Press, 2012 r.
[3] G. Adomavicius, A. Tuzhilin „Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions".http://trouvus.com/wpcontent/uploads/2016/03/reco mmender-systems-survey-2005.pdf
[4] M. Montaner, B. Lopez, J. L De la Rosa. „A Taxonomy of Recommender Agents on the Internet". https://pdfs.semanticscholar.org/f381/f58e6921a372ecf5740fd93 94ec6bf a145c8.pdf
[5] How Reddit algorithms work. Dostępny: https://medium.com/hackingand-gonzo/how-reddit-ranking- algorithms-work-ef111e33d0d9.
[6] B. Sarvar, G. Karypis, J. A. Konstan, J. Riedl „Item-Based Collaborative Filtering Recommendation Algorithms.
[7] Carlos E. Seminario, David C. Wilson. ³Case study evaluation of mahout as a recommender platform´. ACM 2012, pp 45-50
[8] J.L.Herlocker, J.A. Konstan, L.G. Terveen and J. Riedl. ³ Evaluating Collaborative Filtering Recommender Systems´,
[9] ACM Transactions on Information Systems, 2004, pp 5-53
[10] Katrien Verbert, Hendrik Drachsler, Nikos Manouselis, Martin Wolpers, Riina Vuorikari, Erik Duval, ³Dataset driven research for Improving Recommender Systems for Learning´, Proceedings of the 1st International Conference on Learning Analytics and Knowledge, 2011, pp 44-53
[11] Michael D. Ekstrand, John T. Riedl and Joseph A. Konstan,
[12] Collaborative Filtering Recommender Systems ³, Foundations and Trends in Human Computer Interaction, vol. 4, 2011, pp 81- 173



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)