iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** X **Month of publication:** October 2023

DOI: https://doi.org/10.22214/ijraset.2023.56077

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 11 Issue X Oct 2023- Available at www.ijraset.com

Machine Learning based Movie Recommendation System

Sahil Pawar[1], Girik Tripathi[2], Harsh Patel[3], Anant Singh[4]
[1, 3, 4]*Department of CSE MIT World Peace University*
[2]*Department of E&TC University of Mumbai*

Abstract: ***Machine Learning (ML) could be a fashionable engineering technique to form machines suppose or use their intelligence like humans by mimicking traits and by learning to require acceptable choices and to perform appointed tasks properly. a number of the businesses that have done outstanding add the sphere of ML (AI) ar Facebook, Google, Microsoft, IBM, etc. that ar investment millions and billions during this terribly field of ML development and analysis. presently there's a large market and want for building Intelligent Systems for Recommendation. To counter this, one amongst the simplest and most desirable System is Recommendation System (RS). Recommendation Systems had well-tried to play a crucial role within the field of E-Commerce websites, on-line searching, chemical analysis Apps, Social-Networking, Digital selling, on- line Advertisements, etc. by providing customized recommends and feedback to users in step with their preferences and selections. the subject of this report is AI primarily based picture show Recommendation System. because the topic of this paper suggests we have a tendency to ar planning to discuss concerning varied ways in which and approaches of ML (AI) to make a picture show Recommendation System (RS) application. There ar several approaches to make a recommendation system in step with one's want like cooperative Filtering, Content primarily based Recommendation Systems, K-Means algorithmic rule. which can be in brief mentioned during this report, however we are going to primarily discuss and work on K suggests that algorithmic rule approach.***
Keywords: ML; Movie Recommendation Systems; Collaborative Filtering; Content Based Filtering, K-Means algorithm;

I. INTRODUCTION

The quick broadening of software industry has guided to evolution of customized M R.S.s. A R.S. inspects individual predilection of individual customer & recognizes a lay of stuffs which can be of attention to a definite customer. R.S.s were previously utilized in various websites to endorse numerous stuffs namely news articles, music, movies, books, etc. By implementing M R.S.s, individual may suppose to surge trades by imparting correct info to the correct customer. There've been numerous procedures recommender techniques. Content-based recommender technique study the documented data about chosen stuffs & acclaim new stuffs through discovering stuffs with analogous info . Subsequently C.B. technique remains suitable where rich illustrative info is present. There stand structures which utilizes C.B. procedure & C.F. procedure collectively . KMeans algorithm. analyze likenesses among consumers according to predilection statistics, like movie ratings, & streaming history.

Numerous variations had been anticipated for advancement of presentation. Demographic M.R.S.s stay alike to C.F.B.R.S.s, but they analyze likenesses among consumers according to their demographic statistics, like occupation, sex, age. The detachments among users are calculated by means of that info, and a user's predilection on certain movie is predicted by using the neighbor's preferences. Movie Recommender Systems are not just about suggesting Movies to users, but sometimes it can be about suggesting users for Movies(content). Many times, movie recommender systems are not about what we think of them as a product. For example, when google maps suggests a toll-free route to avoid toll that's also a recommender system. So, with this we can conclude that Movie Recommendation System is about personalization. There are mainly three methods to construct a M.R.S. which are C.B.R.S, C.F.B.R.S.s K-Means.

II. LITERATURE REVIEW

As per reference number [1], a survey was performed in which they briefly discussed about various recommender systems like Content based Recommender Technique, Collaborative Filtering based RS, Hybrid Recommender System and spoke about the method/technique to figure a Hybrid System by merging both Collaborative Filtering and Content based RS. They also discussed about various problems which arrive when creating a new recommender technique like Cold Start, Sparsity, Limited Coverage, Diversity and Over- Specialization. The paper also included and discussed about various AI techniques like Fuzzy Sets, ANN, AIS, Swarm Intelligence and Evolutionary Computing.

©IJRASET: All Rights are Reserved

As per reference number [2], a system was introduced to learn preferences of new users in Recommendation Systems. They proposed a Collaborative Filtering method to counter problems like Cold-Start which generally comes in the instance of a new user profile. For a new user, his/her proclivities are not available or not known to the system. They used various strategies to learn about new user profiles by taking in account various aspects like Popularity of the item/product, Entropy and Entropy0: as there are many available items which are never rated by any user and if we assume a scale out of 10 they will come at 0. So, they characterized this type of product in Entropy0.

As per reference number [3], a work was developed mainly about works related to Recommender Systems on E- Commerce Websites using implicit feedbacks as there is very few Explicit information present on E-Commerce websites. The researchers used pseudo rating data from the implicit feedback information. The paper talks about the kind of data used to build a pseudo rating array, like time of purchase of the items, the launch time of items, feedbacks provided by the users to the system. The authors established a recommendation system utilizing time-based technique.

As per reference number [4], a technique was proposed to build a Collaborative Filtering based Recommender System while using Artificial Neural Network. The paper validated the result obtained by Collaborative Filtering by cross checking it by Artificial Neural Network. The paper briefly discussed about ANN, Support Vector Machines and the types of approaches for building a RS. The paper also included work related to tackle problem like Complete and Incomplete Cold- Start. This work included the use of CF approach, LSVM and the ANN.

As per reference number [7],paper described an efficient implementation of Lloyd's k-means clustering algorithm.

The algorithm is easy to implement and only requires that a kd-tree be built once for the given data points. Efficiency is achieved because the data points do not vary throughout the computation and, hence, this data structure does not need to be recomputed at each stage. Since there are typically many more data points than ªqueryº points (i.e., centers), the relative advantage provided by preprocessing in the above manner is greater.

[5], a system was projected which offers an inclusive review of the tourism recommender systems. Several aspects were analyzed from these recommender systems like their interface, their functionality, their recommender mechanism and the AI methods and techniques employed. They also provided the merits of Semantic way of representing information and moving away from pure textual information. A thorough review of the field, in view of the unlike interface was provided in the paper.

As per reference number [6], a system was performed for decision making process using Multiple-Criteria Decision Analysis. The researchers used hybrid recommender system with collaborative filtering technique for identifying most preferred unknown items for every user. They also used Utilities Additives (UTA) algorithm demonstrated the accuracy and efficiency of the proposed methodological framework and compared it to several CF techniques. They used three phase framework which included Data Acquisition, Multicriteria User Modelling, Clustering and at last Recommending.

III. METHODOLOGY USED

With the growing setting and enhancements over discipline of IT sector, info/data burden has become a giant drawback, creating the entire web setting mussy and to urge obviate this drawback. M.R.S.s deliver nice facilitate to customers. moving-picture show Recommender Systems deliver the users with the utmost precise and pertinent feedback that area unit of users' interests and assist/help users in creating an improved call. M.R.S.s area unit creating by mental acts tools that expect & advise stuffs by scheming similarity from interactions between moving-picture show contents and user. to make such software package programs by means that of C.B.F., C.F. and K-Means etc. varied algorithms area unit needed.

To build a Movie Recommendation System, numerous kinds of methods and approaches are present. Some of them are procured and conferred briefly in this segment.

A. Content–Based (CB) R.S.

The things that area unit analogous to the sooner chosen items by definite user area unit counseled in content-based recommender technique. the fundamental principle concerned in content-based recommender system are: The items' description that area unit most popular by the user area unit analyzed to see the principle common attributes that area unit accustomed differentiate between the things and therefore the preferences area unit then saved to user profiles. In order to advocate, every items' attributes area unit compared with the user profile in order that the things with higher degree/number of similarities area unit counseled.

©IJRASET: All Rights are Reserved

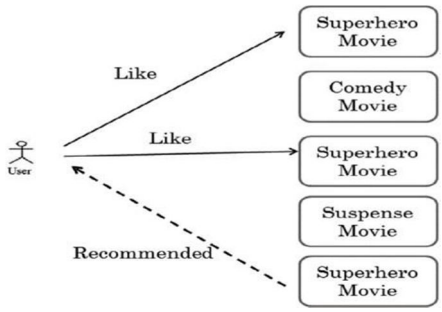
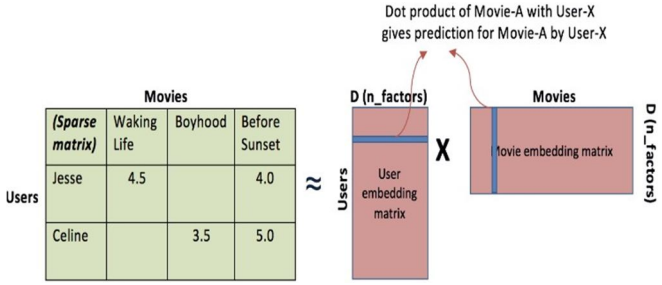


Fig. 1. A content-based Recommender System

B. Collaborative Filtering (CF)–Based R.S.

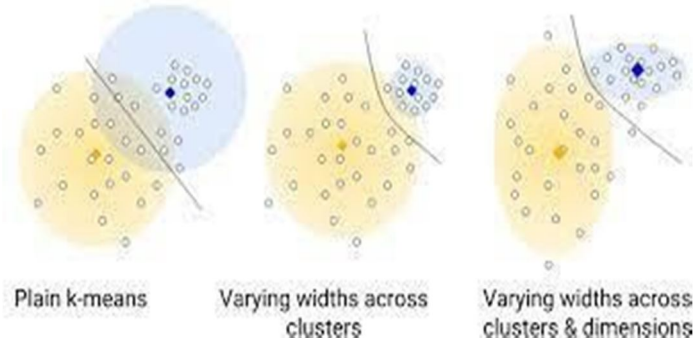
The C.F. based R.S. is split into 2 parts- The User-Based cooperative Filtering- In user-based approach, those things are counseled that are standard, liked and are in nice demand in similar reasonably users. The Item-Based cooperative Filtering- In item-based C.F. method, those types of things are prompt that are similar to the things that are admired and are largely browsed by the users.



C. K-Means Algorithm

K-means cluster algorithmic program computes the centre of masss and iterates till we have a tendency to it finds optimum centroid. It assumes that the quantity of clusters area unit already noted. it's conjointly referred to as flat cluster algorithmic program. the quantity of clusters known from knowledge by algorithmic program is portrayed by 'K' in K-means. In this algorithmic program, the information points area unit allotted to a cluster in such a way that the add of the square distance between the information points and centre of mass would be minimum. it's to be understood that less variation among the clusters can cause additional similar knowledge points among same cluster

1) *Step-1:* choose the K to choose the quantity of clusters.
2) *Step-2:* choose random K points or centroids. (It is often different from the input dataset).
3) *Step-3:* Assign every datum to their nearest center of mass, which is able to type the predefined K clusters.
4) *Step-4:* Calculate the variance and place a brand-new center of mass of every cluster.
5) *Step-5:* Repeat the third steps, which suggests delegate every datapoint to the new nearest center of mass of every cluster.
6) *Step-6:* If any duty assignment happens, then move to step-4 else move to end
7) *Step-7:* The model is prepared.



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue X Oct 2023- Available at www.ijraset.com

IV. RESULT

A. *Challenges and Problems in Recommendation System*

There are numerous problems and difficulties which arose during building recommender systems.

1) *Cold-Start:* While building a new RS, the system doesn't have any prior knowledge of user's preferences. The system also does not have any prior familiarity about new items or the items which are never rated by any user. So, the system faces many difficulties while recommending to the users. This problem is called Cold-Start.
2) *Sparsity:* When users does not rate various items, the user- item array turn out to be inadequate and large. When users rate very few items on the internet, the RS faces problem to recommend properly as very few information about users' preferences is available to the system.
3) *Limited Coverage:* For providing recommends to users, many times RS systems compare the ratings of similar kind of users or dissimilar users on analogous type of items. Therefore, if the users have same type of proclivity for usual items then they can be said adjoining. But when user don't rate items, this approach becomes unsuccessful which is termed as Limited Coverage.
4) *Diversity:* If there is multiplicity in recommended items, then the client may discover suitable as well as relevant items. If a user is not getting diversified recommends by the system, then there is no merit to build a RS.
5) *Over-Specialization:* When users chose preferences, the items are also recommended according to those preferences. So, it narrows down all the stuff present on the internet and only those items are left which are of user's interest. This creates dearth of variation of items in the customer's emprise. This kind of problem is named as Over- Specialization.

B. *Advantages of Recommender System*

There are numerous edges of Recommendation Systems but some of the greatest significant advantages of RSs system are: Any information that is difficult for machine to automate content analysis, it helps to filter them. Recommendations Systems can advise new items to users without past predilections of users. It shares/uses the experience of other users to avoid incomplete or inaccurate filtering. It has ability to filter items based on complex, hard-to- interpret concepts.

C. *Disadvantages of Recommender System*

Users' data is used for recommending movies, if this data is not stored properly, data privacy can be breached very easily. Sites can sell this data to other companies and private data can go in wrong hands.
Recommended Movies can be wrong some times due to inaccurate data, and data scarcity, new user do not get accurate result.
In this we have chosen 2 datasets i.e., movies and ratings for those movies given by users. In movies dataset there are 3883 movies and it also contains list of genre of movies in '|' separated form. So in beginning we have to perform some data cleaning to check for any missing genre for movies and if found drop that movie. Remove year of release from movie if present and add it to new column. Ratings dataset contains 1000209 ratings given by different users to different movies . we donot check for missing values in dataset as we need missing values for operation. Firstly, we performed the movie recommendation system using collaborative approach and content-based approach. And we also tried to implement the recommender engine using K-Means algorithm.
Number of K used for decision making is 8. As it was found optimal during the execution.

	movieId	title	year
2200	2284	Bandit Queen	1994
3243	3329	Year My Voice Broke, The	1987
3669	3759	Fun and Fancy Free	1947
3679	3769	Thunderbolt and Lightfoot	1974
3685	3775	Make Mine Music	1946
4978	5073	Son's Room, The (Stanza del figlio, La)	2001
6563	6672	War Photographer	2001
6667	6776	Lagaan: Once Upon a Time in India	2001
9064	26801	Dragon Inn (Sun lung moon hak chan)	1992
18106	90531	Shame	2011

Fig: result of collaborative filtering

©IJRASET: All Rights are Reserved

	movieId	title	genres	year
664	673	Space Jam	[Adventure, Animation, Children, Comedy, Fanta...	1996
1824	1907	Mulan	[Adventure, Animation, Children, Comedy, Drama...	1998
2902	2987	Who Framed Roger Rabbit?	[Adventure, Animation, Children, Comedy, Crime...	1988
4923	5018	Motorama	[Adventure, Comedy, Crime, Drama, Fantasy, Mys...	1991
6793	6902	Interstate 60	[Adventure, Comedy, Drama, Fantasy, Mystery, S...	2002
8605	26093	Wonderful World of the Brothers Grimm, The	[Adventure, Animation, Children, Comedy, Drama...	1962
8783	26340	Twelve Tasks of Asterix, The (Les douze travau...	[Action, Adventure, Animation, Children, Comed...	1976
9296	27344	Revolutionary Girl Utena: Adolescence of Utena...	[Action, Adventure, Animation, Comedy, Drama, ...	1999
9825	32031	Robots	[Adventure, Animation, Children, Comedy, Fanta...	2005
11716	51632	Atlantis: Milo's Return	[Action, Adventure, Animation, Children, Comed...	2003
11751	51939	TMNT (Teenage Mutant Ninja Turtles)	[Action, Adventure, Animation, Children, Comed...	2007

Fig: result of content-based filtering

```
Select The Movies Id you would like to watch:
957--->Scarlet Letter, The (1926)
416--->Bad Girls (1994)
3790--->Groove (2000)
3609--->Regret to Inform (1998)
61--->Eye for an Eye (1996)
3417--->Crimson Pirate, The (1952)
3255--->League of Their Own, A (1992)
2175--->Déjà Vu (1997)
1789--->Sadness of Sex, The (1995)
1018--->That Darn Cat! (1965)
3879--->Art of War, The (2000)
2650--->Ghost of Frankenstein, The (1942)
2245--->Working Girl (1988)
1758--->Dangerous Beauty (1998)
3555--->U-571 (2000)
----------------------------------------------------------------
957
Recommended Movies -
22                                Assassins (1995)
1176                Pump Up the Volume (1990)
1477                               187 (1997)
1483                          Edge, The (1997)
1467                Desperate Measures (1998)
384                        Coldblooded (1995)
1144                 Harold and Maude (1971)
1199    Indiana Jones and the Last Crusade (1989)
1225      Amityville 1992: It's About Time (1992)
1390                          Traveller (1997)
```

Fig: K-Means clustering (K=8)

The interpretation of suggestions is scaled as Hit Ratio (% of matches amid suggested stuffs). Hit ratio characterizes the number of suggested stuffs which are ideal stuffs. As the no. of feedbacks on suggested stuffs rises, the performance of total suggestions also rises as the coordinator agent gives more weightage to that agent which receives more positive feedback.

For this set of data, on an average, the C.F.B.R.S.s receives an increase in positive responses, and the C.F. agent increases to 0.55. Contrarily, the C.B.R. decreases to 0.09. The reason is that Movie set of data doesn't have sufficient info for demographic agent and C.B.R. agent. The demographic agent discovers neighbors by using gender & age. The C.B.R. agent receives keywords only from the cinema labels; therefore, it always suggests cinemas which consists the words found in the chosen labels. Though, the demographic agent is in peak in initial phase since there is not sufficient data for the collaborative agent.

.

V. CONCLUSION

In this paper the various ways of building a recommendation system are discussed according to the different needs of users. Detailed information about various Recommender techniques currently in use namely Collaborative Filtering, Content based filtering, K-means algorithm technique are also discussed. The paper also debates about the applications of Recommendation Systems, the challenges involved while building a Recommender system and its pros and cons. In this paper, an Experimental work was also performed by applying the K- Means algorithm on Movie Dataset and ratings dataset. From the experiment we came to a result that with increasing number of neighbors, there increases the number of Hit Ratio of recommendation systems with coordinator agent. The paper also discusses about the inability of Collaborative Filtering to recommend users without prior information about preferences of users.

VI. FUTURE WORK

From the above discussion we can say that in Hybrid M.R.S.s uses both C.B. and C.F. method, has ample of possibility to give a better Movie Recommendation Systems as they are free from various challenges faced while creating a recommender system. Also, in the field of M.R.S.s there is a limited use of ML sets like ANN, Fuzzy Sets,etc. which might upswing in future years.

REFERENCES

[1] Gabrani, Goldie, Sangeeta Sabharwal, and Viomesh Kumar Singh. "Artificial intelligence based recommender systems: A survey." International Conference on Advances in Computing and Data Sciences. Springer, Singapore, 2016

[2] Rashid, Al Mamunur, George Karypis, and John Riedl. "Learning preferences of new users in recommender systems: an information theoretic approach." Acm Sigkdd Explorations Newsletter 10.2 (2008): 90-100.

[3] Lee, Tong Queue, Young Park, and Yong-Tae Park. "A time-based approach to effective recommender systems using implicit feedback." Expert systems with applications 34.4 (2008):3055-3062.

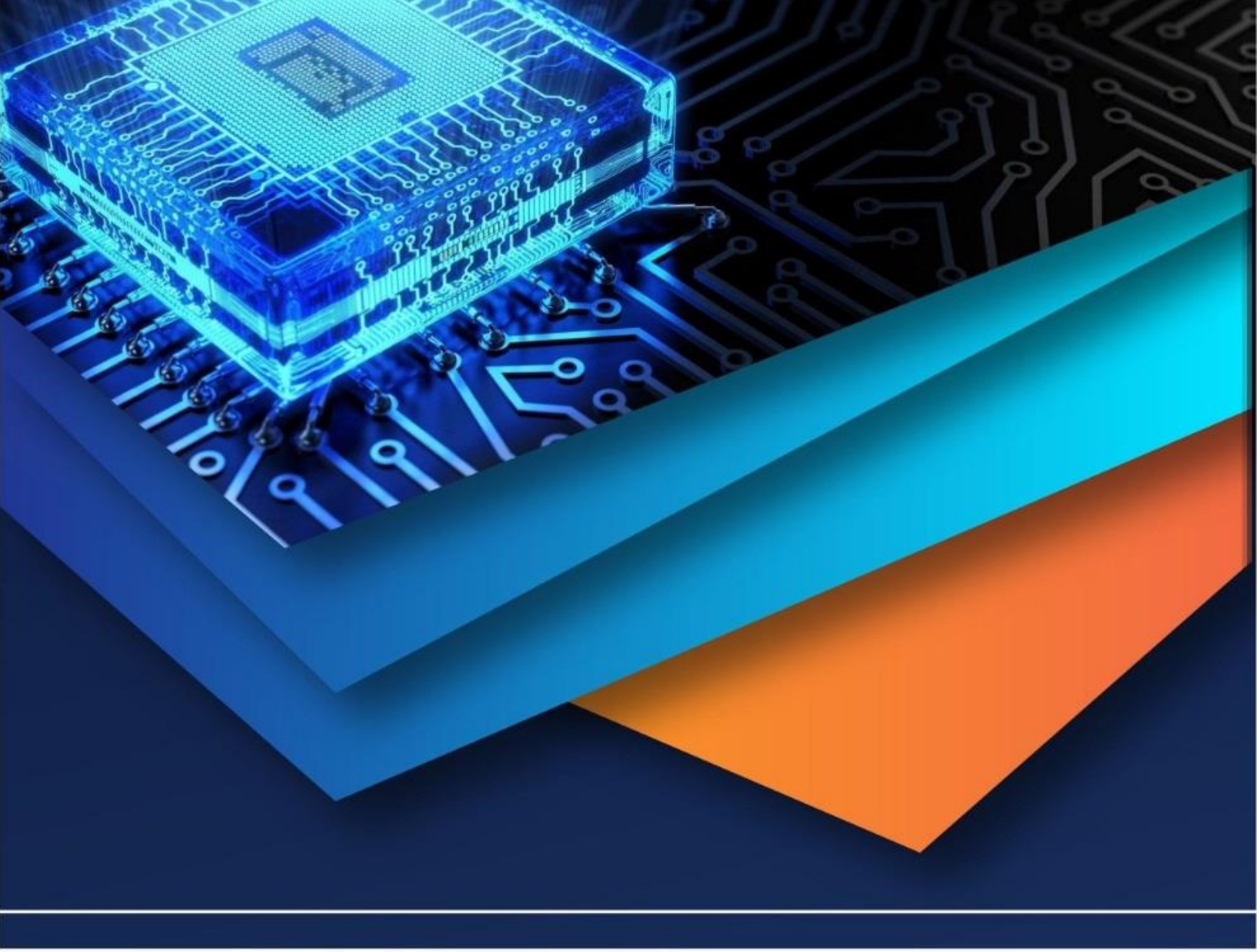
[4] Mylavarapu, Bharadwaja Krishnadev. "Collaborative Filtering and Artificial Neural Network Based Recommendation System for Advanced Applications." Journal of Computer and Communications 6.12 (2018): 1-14.

[5] Borràs, Joan, Antonio Moreno, and Aida Valls. "Intelligent tourism recommender systems: A survey." Expert Systems with Applications 41.16 (2014): 7370-7389.

[6] Lakiotaki, Kleanthi, Nikolaos F. Matsatsinis, and Alexis Tsoukias. "Multicriteria user modeling in recommender systems." IEEE Intelligent Systems 26.2 (2011): 64-76.

[7] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 881-892, July 2002, doi: 10.1109/TPAMI.2002.1017616..

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



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)