iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** XI **Month of publication:** November 2021

DOI: https://doi.org/10.22214/ijraset.2021.38916

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: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

A Short Video Recommendation System

Ragu G[1], Sai Bhargav V[2], Sujith Vignesh P[3]
[1]*Assistant Professor, SRM Institute of Science and Technology, Ramapuram, Tamil Nadu.*
[2, 3]*Student, SRM Institute of Science and Technology, Ramapuram, Tamil Nadu.*

Abstract: ***With the development of the Internet and social networking service, the micro-video is becoming more popular, especially for youngers. However, for many users, they spend a lot of time to get their favourite micro-videos from amounts videos on the Internet; for the micro-video producers, they do not know what kinds of viewers like their products. Therefore, we propose a micro-video recommendation system. The recommendation algorithms are the core of this system. Traditional recommendation algorithms include content-based recommendation, collaboration recommendation algorithms, and so on.***
At the Big Data times, the challenges what we meet are data scale, performance of computing, and other aspects. Thus, we improve the traditional recommendation algorithms, using the popular parallel computing framework to process the Big Data. Slope one recommendation algorithm is a parallel computing algorithm based on MapReduce and Hadoop framework which is a high-performance parallel computing platform. The other aspect of this system is data visualization. Only an intuitive, accurate visualization interface, the viewers and producers can find what they need through the micro-video recommendation system.
Keywords: Short, video, recommendation , machine learning

I. INTRODUCTION

Video is a new form of information media. With the development of the Internet, 3G (the 3rd Generation mobile communication technology), and 4G (the 4th Generation mobile communication technology) network, the bandwidth and speed of network become faster and faster. These technologies provide conditions for dissemination of information media. Video is a short time video, which lasts for 30 seconds to 300 seconds. The short time micro-videos are popular with young people, because the teenagers prefer to watch the micro-video on their comfortable time through mobile devices. For micro-video producers, the problem is they do not know how many people like their products, and do not know how many times their video has been watched. Therefore, this proposes a video recommendation system (MRS).

One of the purposes is an overview of videos for the producer. In this way, the producer knows how many users love their video, and how many times their videos are on- demand. Another purpose is for users.

The system can analyse the users' favourites and watching history, automatically push appropriate video to the users. It is becoming more popular with the internet technology development, which means the data sets whose size is beyond the ability of current technology, method and theory to capture, manage, and process the data within a tolerable elapsed time [3].

In order to enhance the MRS accuracy, we need to collect large volume data sets about who and when watched the micro-video, how many times the micro- video on demanded, and how many people love the micro-video. Therefore, the MRS, proposing in this paper, use technology to process the collected data sets.

Data sets are the foundation of the recommendation system. The first step of video recommendation is to collect data as far as possible from the Internet. We download data from video websites, video forum, video online chat websites, and so on. Web crawlers, one of basic data sets collection, can download resources from Internet. The web crawlers originally used for search engine. In this paper, the results of crawler directly affect the accuracy of recommendation system.

II. RELATED WORK

Micro-video is a new form of information media. With the development of the Internet, 3G (the 3rd Generation mobile communication technology), and 4G (the 4th Generation mobile communication technology) network, the bandwidth and speed of network become faster and faster. These technologies provide conditions for dissemination of information media. Micro-video is a short time video [1]. Big Data is becoming more popular with the internet technology development, which means the data sets whose size is beyond the ability of current technology, method and theory to capture, manage, and process the data within a tolerable elapsed time [2]. In order to enhance the MRS accuracy, we need to collect large volume data sets about who and when watched the micro-video, how many times the micro-video on- demanded, and how many people love the micro-video. Therefore, the MRS, proposing in this paper, use the Big Data technology to process the collected data sets.

The system can analyse the user's favourites and watching history, automatically push appropriate video to the users. It is becoming more popular with the internet technology development, which means the data sets whose size is beyond the ability of current technology, method and theory to capture, manage, and process the data within a tolerable elapsed time [3]

We download data from video websites, video forum, video online chat websites, and so on. Web crawlers [4], one of basic data sets collection, can download resources from Internet. The web crawlers originally used for search engine. In this paper, the results of crawler directly affect the accuracy of recommendation system.

A. Problem Statement

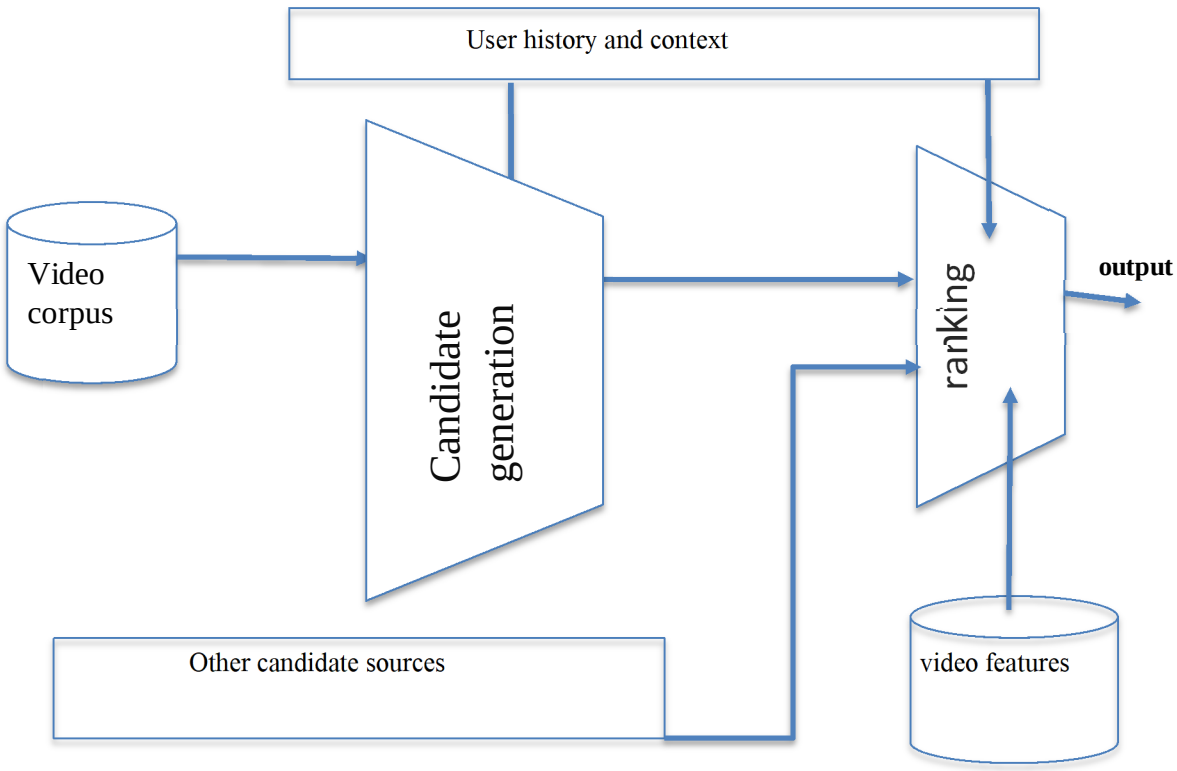
With the development of the Internet and social networking service, the micro-video is becoming more popular, especially for youngers. However, for many users, they spend a lot of time to get their favorite micro-videos from amounts videos on the Internet; for the micro-video producers, they do not know what kinds of viewers like their products. Therefore, this paper proposes a micro-video recommendation system. The recommendation algorithms are the core of this system. Traditional recommendation algorithms include content-based recommendation, collaboration recommendation algorithms, and so on. At the Bid Data times, the challenges what we meet are data scale, performance of computing, and other aspects. Thus, this paper improves the traditional recommendation algorithms, using the popular parallel computing framework to process the Big Data. Slope one recommendation algorithm is a parallel computing algorithm based on MapReduce and Hadoop framework which is a high-performance parallel computing platform. The other aspect of this system is data visualization. Only an intuitive, accurate visualization interface, the viewers and producers can find what they need through the micro-video recommendation system

III. MODULES

A. Data source is the basic of MRS, since all the data what we need are download from the web sites. The video websites are the main data source from which we can download, such as youku, iQiyi, and so on. The contents that we need download include the video ID, brief description, click rate, ranking list and so on. We also download video comments from social networking service web sites, for example, Weibo, We chat, and micro-video forum.

B. With the development of the new generation of information network technology represented by mobile Internet, big data, cloud storage, the traditional advertising micro video creation has a fundamental change. This project studies the design of the IT architecture and the division of the system service level in the era of big data, and designs the structure of the system.

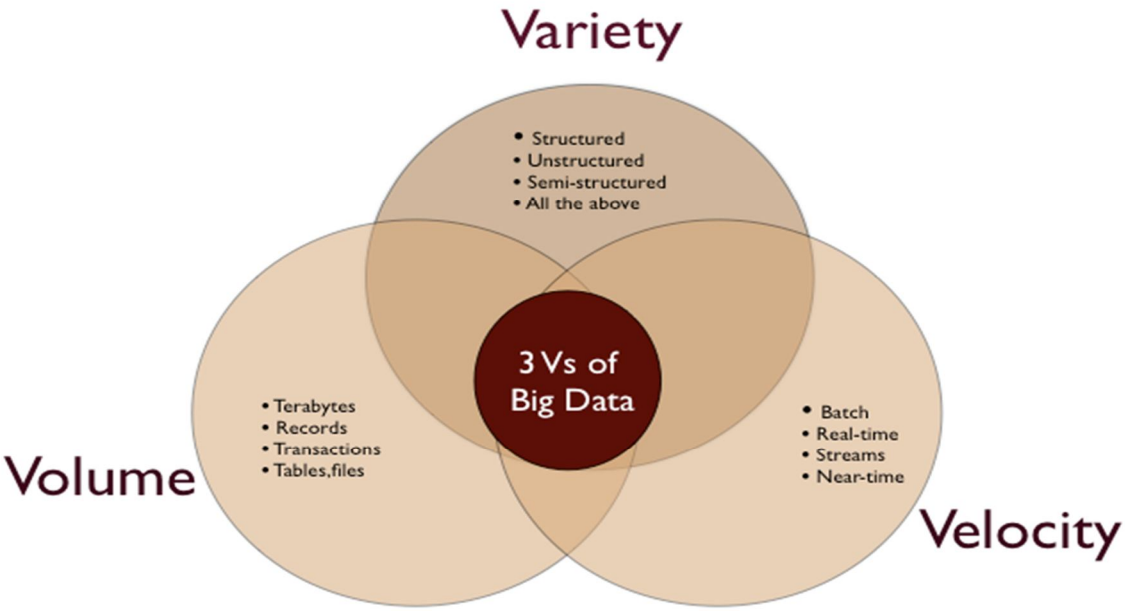
Architecture Diagram



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

IV. METHODOLOGY

1) Focus on "HOW" the data can be analyse using 3V's of big data.
2) Break the large data into smaller pieces.
3) Decide how each component works and how they Work together.
4) The data can be combined by data deduplication To eliminate the similar data.
5) By this process the large data can be compressed And get valuable and creative data information



A. Volume

The most obvious one is where we'll start. Big data is about volume. Volumes of data that can reach unprecedented heights in fact. It's estimated that 2.5 quintillion bytes of data is created each day, and as a result, there will be 40 zettabytes of data created by 2020 – which highlights an increase of 300 times from 2005. As a result, it is now not uncommon for large companies to have Terabytes – and even Petabytes – of data in storage devices and on servers. This data helps to shape the future of a company and its actions, all while tracking progress.

B. Velocity

The growth of data, and the resulting importance of it, has changed the way we see data. There once was a time when we didn't see the importance of data in the corporate world, but with the change of how we gather it, we've come to rely on it day to day. Velocity essentially measures how fast the data is coming in. Some data will come in in real-time, whereas other will come in fits and starts, sent to us in batches. And as not all platforms will experience the incoming data at the same pace, it's important not to generalise, discount, or jump to conclusions without having all the facts and figures.

C. Variety

Data was once collected from one place and delivered in one format. Once taking the shape of database files - such as, excel, csv and access - it is now being presented in non-traditional forms, like video, text, pdf, and graphics on social media, as well as via tech such as wearable devices. Although this data is extremely useful to us, it does create more work and require more analytical skills to decipher this incoming data, make it manageable and allow it to work.

Big Data is much more than simply 'lots of data'. It is a way of providing opportunities to utilise new and existing data, and discovering fresh ways of capturing future data to really make a difference to business operatives and make it more agile.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

V. DATA DEDUPLICATION

A. Data De-Duplication

Data deduplication refers to a technique for eliminating redundant data in a data set. In the process of deduplication, extra copies of the same data are deleted, leaving only one copy to be stored. Data is analysed to identify duplicate byte patterns to ensure the single instance is indeed the single file. Then, duplicates are replaced with a reference that points to the stored chunk.

Given that the same byte pattern may occur dozens, hundreds, or even thousands of times think about the number of times you make only small changes to a PowerPoint file or share another important business asset the amount of duplicate data can be significant. In some companies, 80% of corporate data is duplicated across the organization. Reducing the amount of data to transmit across the network can save significant money in terms of storage costs and backup speed in some cases, savings up to 90%.

VI. ALGORITHM

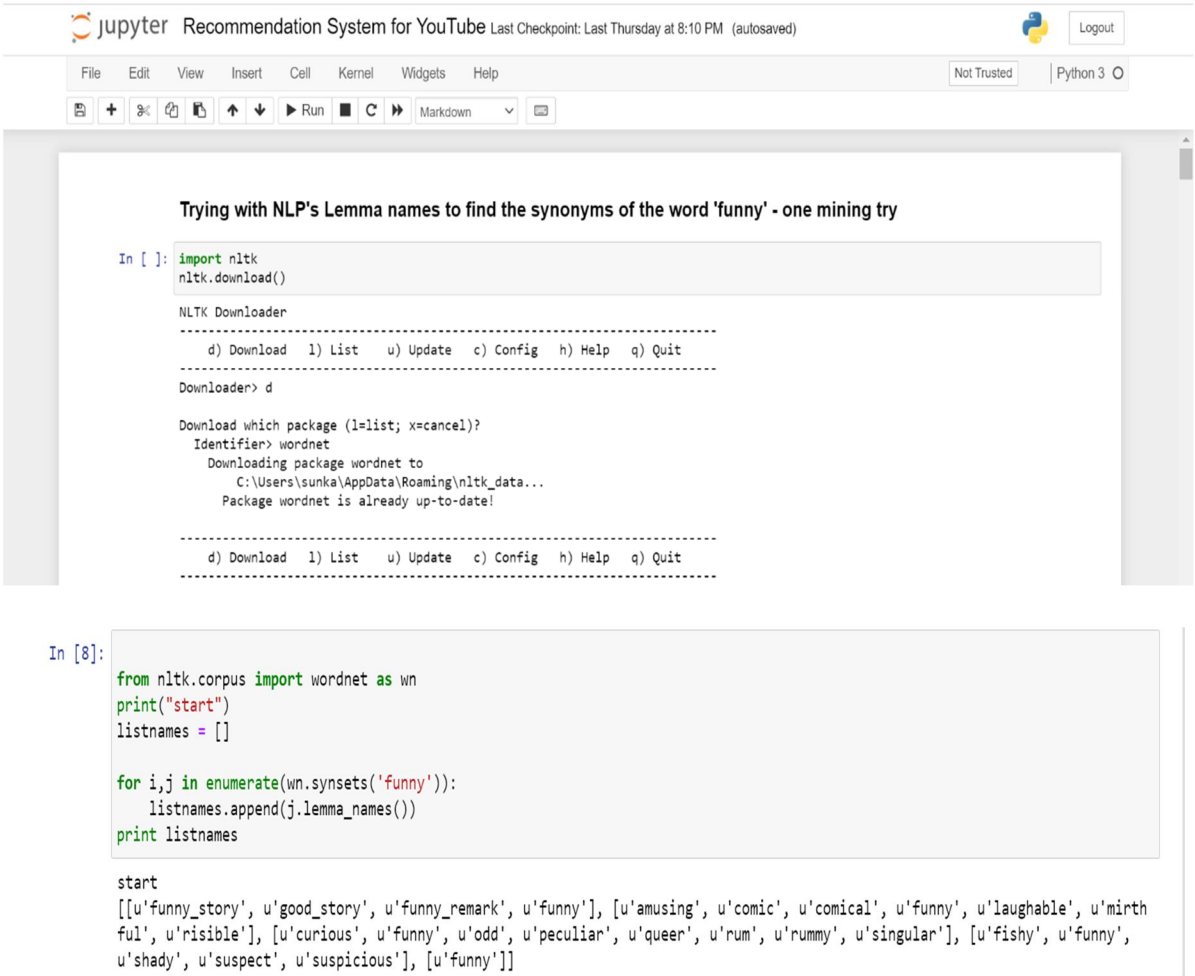
A. Collaborative Filtering

Collaborative filtering is a technique that can filter out items that a user might like on the basis of reactions by similar users.

It works by searching a large group of people and finding a smaller set of users with tastes similar to a particular user. It looks at the items they like and combines them to create a ranked list of suggestions.

There are many ways to decide which users are similar and combine their choices to create a list of recommendations. This article will show you how to do that with Python.

VII. OUTPUT




```
In [8]:
from nltk.corpus import wordnet as wn
print("start")
listnames = []

for i,j in enumerate(wn.synsets('funny')):
    listnames.append(j.lemma_names())
print listnames
```

```
start
[[u'funny_story', u'good_story', u'funny_remark', u'funny'], [u'amusing', u'comic', u'comical', u'funny', u'laughable', u'mirthful', u'risible'], [u'curious', u'funny', u'odd', u'peculiar', u'queer', u'rum', u'rummy', u'singular'], [u'fishy', u'funny', u'shady', u'suspect', u'suspicious'], [u'funny']]
```

Popularity Model

```
In [6]: train = newdf.ix[:250,:]
        test = newdf.ix[250:,:]
        train_data = graphlab.SFrame(train)
        test_data = graphlab.SFrame(test)
        popularity_model = graphlab.popularity_recommender.create(train_data, user_id='users', item_id='v_title', target='Liked')

        Recsys training: model = popularity

        Preparing data set.

            Data has 572 observations with 6 users and 99 items.

            Data prepared in: 0.004009s

        572 observations to process; with 99 unique items.

In [16]: #Get recommendations for first 5 users and print them
         #users = range(1,6) specifies user ID of first 5 users
         #k=5 specifies top 5 recommendations to be given
         user_names = ['Kathir','MSD','MarkZ','Chris','Sundhar','Patrick']
         popularity_recomm = popularity_model.recommend(users=user_names,k=10)
         popularity_recomm.print_rows(num_rows=25)
```

Evaluation metrics for this Recommendation System

1. Recall

What ratio of items that a user likes were actually recommended. If a user likes say 5 items and the recommendation decided to show 3 of them, then the recall is 0.6

2. Precision

Out of all the recommended items, how many the user actually liked? If 5 items were recommended to the user out of which he liked say 4 of them, then precision is 0.8

VIII. CONCLUSION

With the Development, the micro-video are increasingly common, especially young teenagers are likely to watch videos on mobile device According to the viewers browsing or watching history, this system can recommend the videos to the viewers.

Big Data has just started, in which Data deluge is going to keep on increasing throughout the next years, and each data scientist will have to handle much more quantity of data every year. This data is becoming more diverse, larger, and faster.

The core function of the system is the recommendation algorithms. The commonly recommendation algorithms are suite for tradition date sets, such as content-based recommendation, collaboration recommendation, and so on. However, with the development of Big Data, the recommendation algorithms should have the ability to deal with the Big Data.

The Slope one algorithm is a Big Data micro video topic recommendation system, including data layer, interface layer, core layer, service layer, business layer. Then research on the key algorithms involved in the system, and pick up a suitable algorithm. Finally, introduce the implementation of the system.

Video recommender system, i.e., when a new coming video is added to the library, the recommender system has to bootstrap the video relevance score with very few user behavior with respect to the newly added video. To solve this problem, we proposed a content-based video recommendation approach by taking the advantage of deep convolutional neural networks.

As per the watcher's looking at or watching history, genres of videos, mouse movements, recommendation will be generated. With the help of big data, we are going to handle huge amount of video data set.

REFERENCES

[1] Y. Z. Li, T. Gao, and X. Y. Li, "Design of video recommender system based on cloud computing", Journal on Communications, Vol. 34, No. Z2, pp. 138-140, 147, 2013.

[2] Y. Li, "Development mode of micro-video communication", Academic Exchange, Vol. 248, pp.177-181, 2014.

[3] S. M. Meng, W. C. Dou, X. Y. Zhang, et al., "KASR: a keyword-aware service recommendation method on MapReduce for Big Data application", IEEE Transactions on parallel and distributed system, Vol. 25, No. 12, 2014.

[4] D. M. Zhou, Z. J. Li, "Survey of high-performance web crawler", Computer Science, Vol. 36, No. 8, pp.26-29, 53, 2009.

[5] G. Y. Su, J. H. Li, and Y. H. Ma, et al., "New focused crawling algorithm", Journal of Systems Engineering and Electronics, Vol. 16, No. 1, pp.199-203, 2005.

[6] X. W. Meng, X. Hu, and L. C. Wang, et al., "Mobile recommender system and their applications", Journal of Software, Vol. 24, No.1, pp. 91-108, 2013.

[7] G. X. Wang, H. P. Liu, "Survey of personal recommendation system", Computer Engineering and Applications", Vol. 48, No. 7, pp. 66-76, 2012.

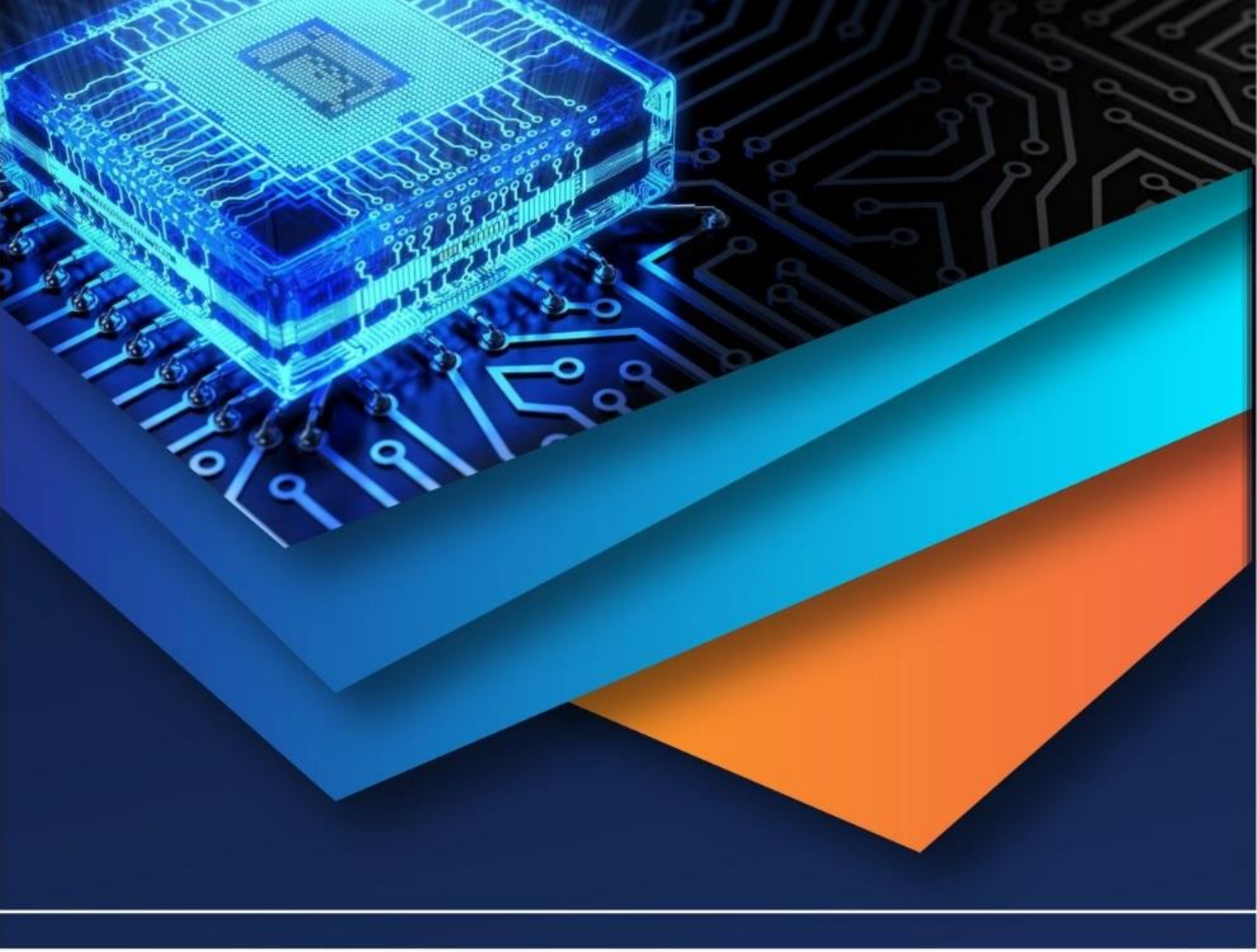
[8] G. F. Sun, L. Wu, and Q. Liu, et al., " Recommendations based on collaborative filtering by exploiting sequential behaviors", Journal of Software, Vol. 24 No. 11, pp.2721-2733, 2013.

[9] L. Guo, J. Ma, and Z. M. Chen, et al., "Incorporating item relations for social recommendation", Chinese Journal of Computers, Vol. 37, No. 1, pp. 218-228, 2014.

[10] Y. R. Wang, M. Chen, and H. H. Wang, et al., "A content-based filtering algorithm for scientific literature recommendation", Computer Technology and Development, Vol .21, No. 2, pp. 66-69, 2011.

[11] W. F. Liu, J. Z. Gu, and X. Lin, et al., " A Big Data management and analysis system based on Hadoop and Mahout", Computer Application and Software, Vol. 32, No. 1, pp. 47-50, 2015.

[12] E. Jain, S. K. Jain, "Twitter users on the basis of their interests using Hadoop/Mahout platform", Industrial and Information Systems (ICIIS), 2014 9th International Conference, pp. 15-17, 2014.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY