iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.44739

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 VI June 2022- Available at www.ijraset.com

Knowledge Integrating System Using Video Mining

Monali Rajput[1], Samiksha Umbarje[2], Pranoti Shinde[3], Pranjali Shivsharan[4], Prof Mr. A.S. Shimpi[5]
[1, 2, 3, 4]*Department of Computer Science and Engineering , Brahmadevdada Mane Institute of Technology, Solapur*
[5]*Prof in Department of Computer Science and Engineering, Brahmadevdada Mane Institute of Technology, Solapur.*

Abstract: A system will be developed in which only admin can upload videos on Cloud Server. Videos are sorted based on category and videos are automatically uploaded on Cloud Server on the schedule provided by the admin. Users can watch videos online. Users can rate videos that will be analyzed by the system. This system automatically removes videos that are less popular.

I. INTRODUCTION

The multimedia data including text, image, audio, and video have been produced massively. The digital videos rapidly became an important source for education, advertisements, promotions, and entertainment. Looking at the need of users at large digital videos, the system proposes to maintain the repository of videos on Cloud Server.

Private cloud is a model of cloud computing where IT services are provisioned over private IT infrastructure for the dedicated use of a single organization. A private cloud is a computing model that offers a proprietary environment dedicated to a single business entity. The private cloud provides extended, virtualized computing resources via physical components stored on-premises or at a vendor's data centre.

Videos are automatically uploaded on Cloud Server on the schedule provided by the admin and videos are sorted based on the categories. Users can search for videos by keywords like title, date, and author so that content-based videos will be retrieved. Users can also decide to download the videos based on the summary provided for each video. The users can also watch videos online. Users can rate the videos so that the system can analyse the popularity of videos and this analysis is used further for deleting the least popular videos. Users can share videos with other registered users without using another platform in the proposed system.

The following algorithms are used for implementing the above features are as follows:

- The clustering algorithm is used for sorting the videos based on categories.
- Clustering is the process of a grouping of objects that belongs to the same class. Similar objects are grouped in one cluster and remaining dissimilar objects are grouped in another cluster.
- Indexing and prediction algorithms are used for searching and retrieving content-based videos from huge video datasets.
- An index or database index is a data structure that is used to quickly locate and access the data from a database .
- Predictive analytics is the practice of extracting useful information from existing data sets to determine patterns and predict future outcomes and trends.

The above data mining algorithms can be used for implementing a proposed system and algorithms can vary as per the need for implementing the proposed system in the future.

II. LITERATURE REVIEW

1) In [1] Zhong Hanyang, Song Xin*, Yan Zhenguo, “Vessel Sailing Patterns Analysis from S-AIS Data Dased on K-means Clustering Algorithm”. In this paper, a typical clustering algorithm called K-means is applied to deal with the Spacebased AIS(S-AIS) data received by “TianTuo-3” satellite developed by National University of Défense Technology. They used Elbow Rule to determine the optimal number of clusters and calculate the normalized standard deviation of COG(Course Over Ground) and SOG(Speed Over Ground) of vessels in south Africa area as their features to conduct clustering. This method is supposed to evaluate vessels’ sailing stability and used in detection of low-likelihood behaviours or anomalies of vessels. The real-time performance of one single AIS satellite is still poor, which shows the importance of establishing AIS satellite constellations in the future.

2) In [2] Shalini L1, Gopali Naga Sravya2, "Analysis of Health-Tweets using K-means clustering". In this paper, they analysed tweets from several records and using several sources like BBC health, CBC health, CNN health, everyday health, fox news health, GDN health care, good health, etc. By combining the tweets from these 16 records, the top 5 clusters of positive and negative words are obtained by using K-means clustering, the analysis is done on the number of positive and negative words present in those records, and the corresponding plot for the frequency of top 40 words is obtained. They find the views of different people related to the news on health in different channels.

3) In [3] Igor V. Anikin, Rinat M. Gazimov, "Privacy Preserving Data Mining in Terms of DBSCAN Clustering Algorithm in Distributed Systems". In this paper, they developed DBCSAN clustering algorithm, which can provide security of information during all stages of distributed data mining process. It could be very useful for data mining techniques in distributed systems with bigdata. 5. In [4] Uma Ojha, Dr. Savita Goel, "A Study On Prediction Of Breast Cancer Recurrence Using Data Mining Techniques". The main objective of this paper is to find how precisely these data mining algorithms can predict the probability of recurrence of the disease among the patients on the basis of important stated parameters. The research highlights the performance of different classification and clustering algorithms on the dataset. Experiments show that classification algorithms are better predictors than clustering algorithms. The result indicates that the decision tree (C5.0) and SVM is the best predictor with 81% accuracy on the holdout sample and fuzzy c-means came with the lowest accuracy of 37% among the algorithms used in this paper.

In previous researches of the video data mining any users can upload the videos, they can search and retrieves videos based on the keywords with less accuracy and are not able to remove the least popular videos automatically from a system. YouTube allows users to share only the video link not the actual video and uses another platform for sharing the link.

III. SYSTEM DESIGN

A. *System Architecture*

System Architecture Diagram

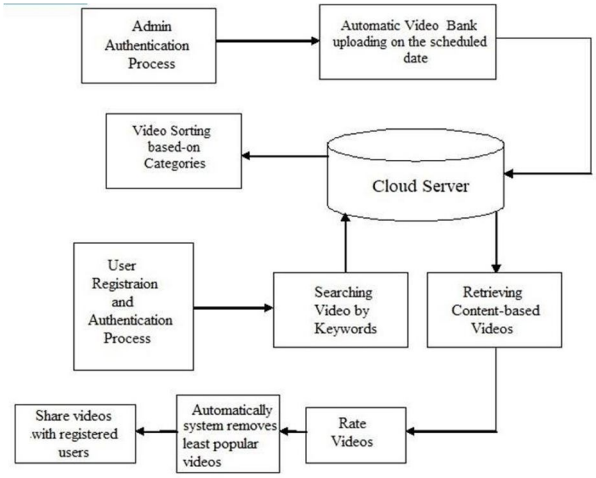


Fig 3.1: Cloud based platform for video-on-demand service

B. *Data flow Diagram*

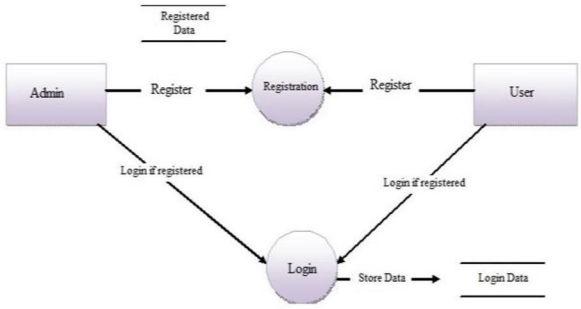


Figure 3.2.1 DFD level 0 of System Architecture

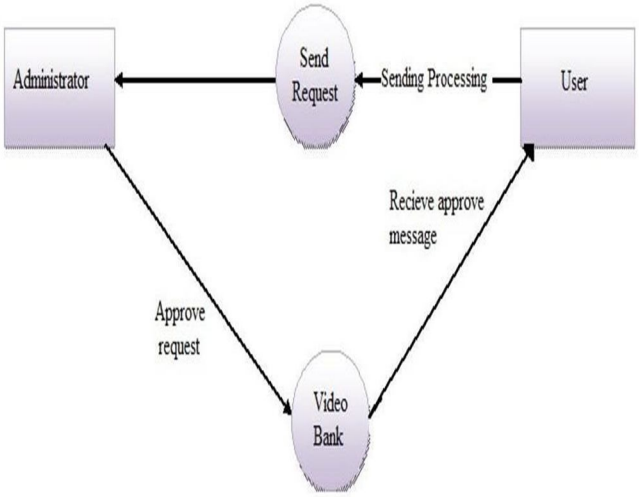


Figure 3.2.2 DFD level 1 of System Architecture

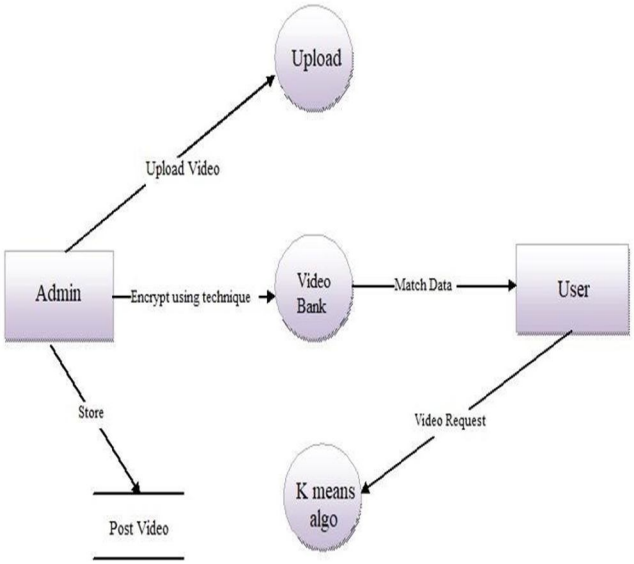


Figure 3.2.3 DFD level 2 of System Architecture

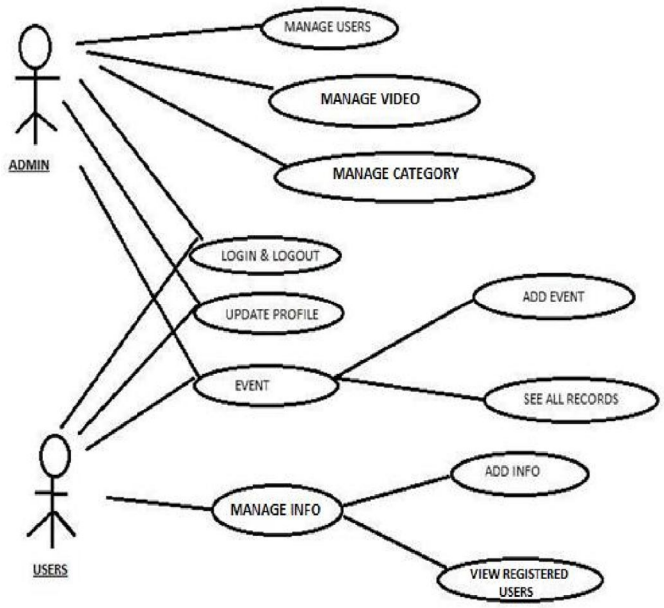


Figure 3.3.1 Use Case Diagram of System Architecture

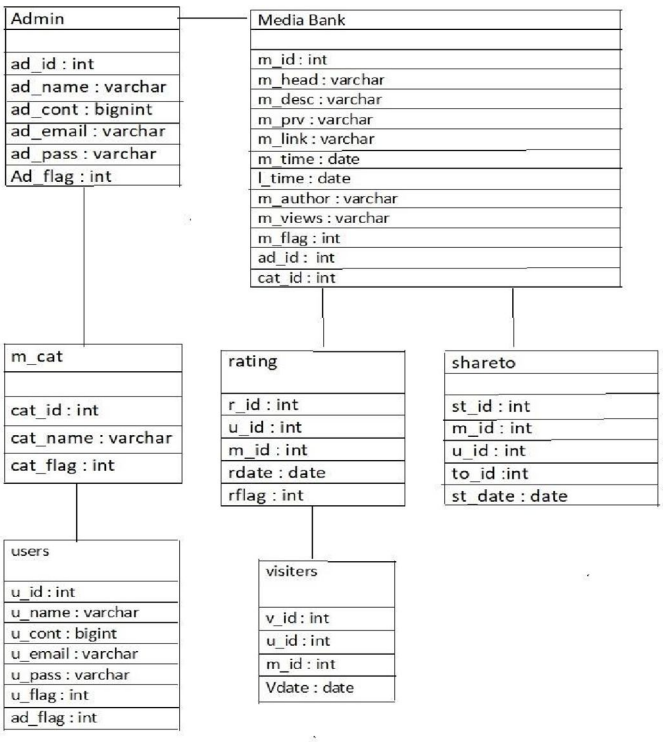


Figure 3.3.2 Class Diagram of System Architecture

C. UML Diagrams

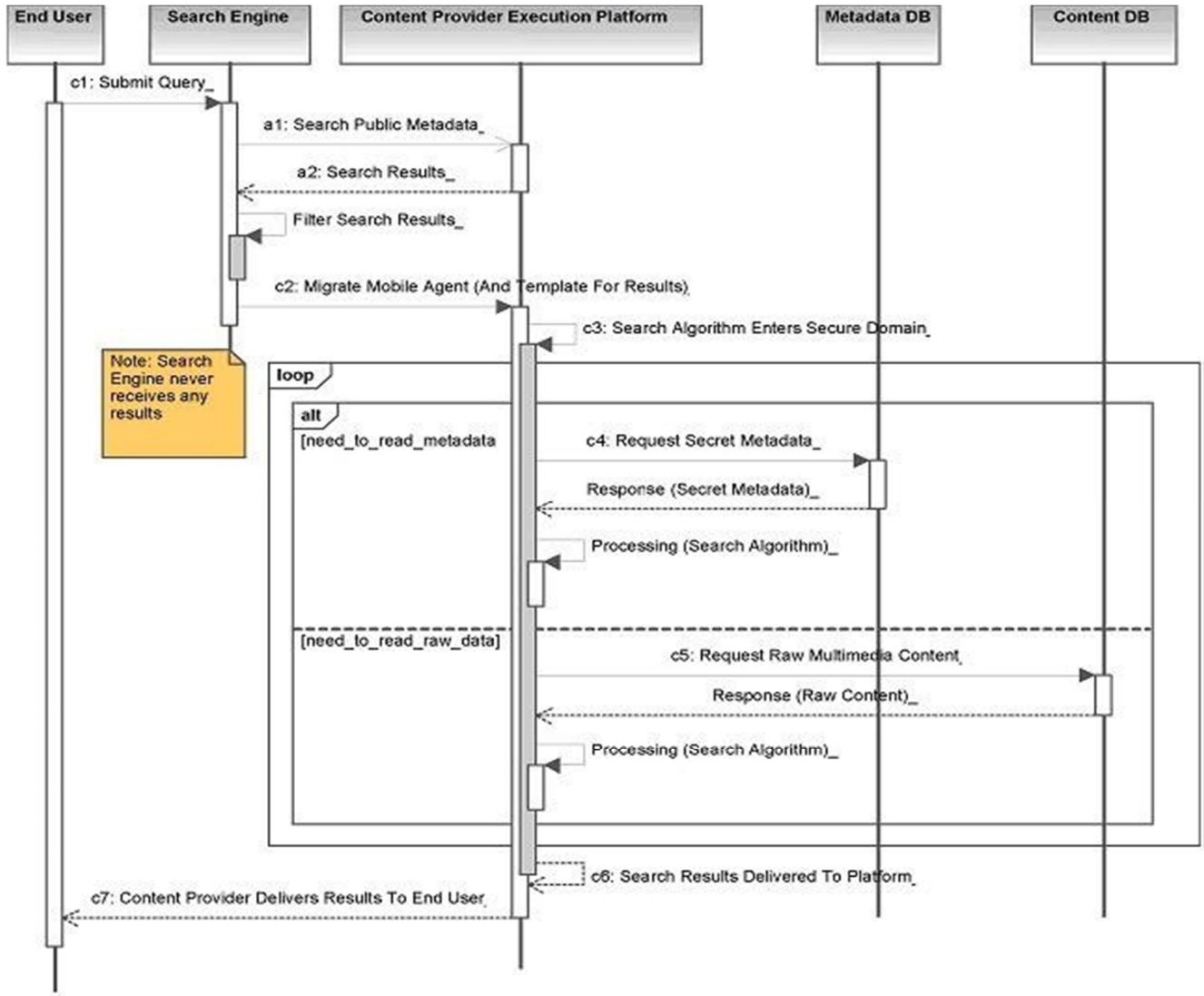


Figure 3.3.3 Sequence Diagram of System Architecture

IV. TECHNOLOGY & PROGRAMMING LANGUAGE

PHP is a server-side scripting language designed for web development but also used as a general-purpose programming language. Originally created by Rasmus Lerdorf in 1994, the PHP reference implementation is now produced by The PHP Group. While PHP originally stood for Personal Home Page, it now stands for the recursive backronym PHP: Hypertext Preprocessor.

PHP code may be embedded into HTML code, or it can be used in combination with various Web template systems and web frameworks. PHP code is usually processed by a PHP interpreter implemented as a module in the web server or as a Common Gateway Interface (CGI) executable. The web server combines the results of the interpreted and executed PHP code, which may be any type of data, including images, with the generated web page. PHP code may also be executed with a command-line interface (CLI) and can be used to implement standalone graphical applications.

Standard PHP interpreter, powered by the Zend Engine, is free software released under the PHP License. PHP has been widely ported and can be deployed on most web servers on almost every operating system and platform, free of charge.

The PHP language evolved without a written formal specification or standard until 2014, leaving the canonical PHP interpreter as a de facto standard. Since 2014 work has been ongoing to create a formal PHP specification.

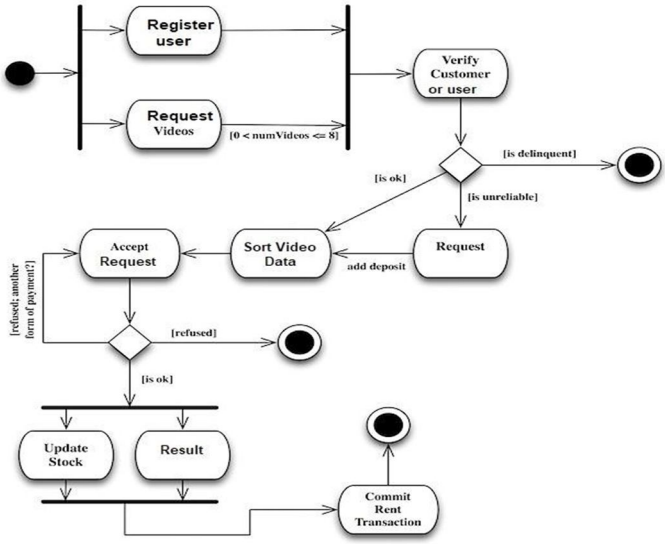


Figure 3.3.3 Activity Diagram

A. Server: XAMPP

XAMP is a free and open source cross-platform web server solution stack package developed by Apache Friends, consisting mainly of the Apache HTTP Server, MariaDB database, and interpreters for scripts written in the PHP and Perl programming languages. XAMPP stands for CrossPlatform (X), Apache (A), MariaDB (M), PHP (P) and Perl (P). It is a simple, lightweight Apache distribution that makes it extremely easy for developers to create a local web server for testing purposes. Everything needed to set up a web server – server application (Apache), database (MariaDB), and scripting language (PHP) – is included in an extractable file. XAMPP is also cross-platform, which means it works equally well on Linux, Mac and Windows. Since most actual web server deployments use the same components as XAMPP, it makes transitioning from a local test server to a live server extremely easy as well.

XAMPP is regularly updated to incorporate the latest releases of Apache, MariaDB, PHP and Perl. It also comes with a number of other modules including OpenSSL, phpMyAdmin, MediaWiki, Joomla, WordPress and more. Self-contained, multiple instances of XAMPP can exist on a single computer, and any given instance can be copied from one computer to another. XAMPP is offered in both a full and a standard version (Smaller version).

Officially, XAMPP's designers intended it for use only as a development tool, to allow website designers and programmers to test their work on their own computers without any access to the Internet. To make this as easy as possible, many important security features are disabled by default. XAMPP has the ability to serve web pages on the World Wide Web. A special tool is provided to password-protect the most important parts of the package. XAMPP also provides support for creating and manipulating databases in MariaDB and SQLite among others. Once XAMPP is installed, it is possible to treat a localhost like a remote host by connecting using an FTP client. Using a program like FileZilla has many advantages when installing a content management system (CMS) like Joomla or WordPress. It is also possible to connect to localhost via FTP with an HTML editor.

The default FTP user is "newuser", the default FTP password is "wampp". The default MariaDB user is "root" while there is no default MariaDB password.

B. Database: MySQL

MySQL is an open-source relational database management system (RDBMS), in July 2013, it was the world's second most widely used RDBMS, and the most widely used open-source client–server model RDBMS. It is named after co-founder Michael Widenius's daughter, My. The SQL acronym stands for Structured Query Language. The MySQL development project has made its source code available under the terms of the GNU General Public License, as well as under a variety of proprietary agreements.MySQL was owned and sponsored by a single for-profit firm, the Swedish company MySQL AB, now owned by Oracle Corporation. For proprietary use, several paid editions are available, and offer additional functionality.

MySQL is a popular choice of database for use in web applications, and is a central component of the widely used LAMP open source web application software stack (and other "AMP" stacks). LAMP is an acronym for "Linux, Apache, MySQL, Perl/PHP/Python. "Free-software-open source projects that require a full-featured database management system often use MySQL. Applications that use the MySQL database include: TYPO3, MODx, Joomla, WordPress, phpBB, MyBB, Drupal and other software. MySQL is also used in many high-profile, large-scale websites, including Google (though not for searches), Facebook, Twitter, Flickr, and YouTube.

On all platforms except Windows, MySQL ships with no GUI tools to administer MySQL databases or manage data contained within the databases. Users may use the included command line tools, or install MySQL Workbench via a separate download. Many third party GUI tools are also available.

MySQL is written in C and C++. Its SQL parser is written in yacc, but it uses a home-brewed lexical analyzer. MySQL works on many system platforms, including AIX, BSDi, FreeBSD, HP-UX, eComStation, i5/OS, IRIX, Linux, OS X, Microsoft Windows, NetBSD, Novell NetWare, OpenBSD, OpenSolaris, OS/2 Warp, QNX, Oracle Solaris, Symbian, SunOS, SCO Open Server, SCO UnixWare, Sanos and Tru64. A port of MySQL to OpenVMS also exists.

C. Web browser: Google Chrome

Google Chrome is a freeware web browser developed by Google. It used the WebKit layout engine until version 27 and with the exception of its iOS releases, from version 28 and beyond uses the WebKit fork Blink. It was first released as a beta version for Microsoft Windows on September 2, 2008, and as a stable public release on December 11, 2008.

As of December 2015, Stat Counter estimates that Google Chrome has a 58% worldwide usage share of web browsers as a desktop browser. It is also the most popular browser for smartphones, and combined across all platforms at about 45%. Its success has led to Google expanding the 'Chrome' brand name on various other products such as the Chromecast. Google releases the majority of Chrome's source code as an open-source project Chromium.

V. SCOPE AND FUTURE WORK

A. Project Scope

The Proposed system maintains video datasets on the cloud server from which only available content-based videos are accessed by the registered users.

1) *Subtitle:* We provide Subtitle to our videos.
2) *Different Languages:* In Future Scope we can make our videos in different language like (Hindi, Marathi, Tamil, Telugu, English) with their subtitles.
3) *Online Test:* We can provide Online Exams/Tests in our portal at future scope.

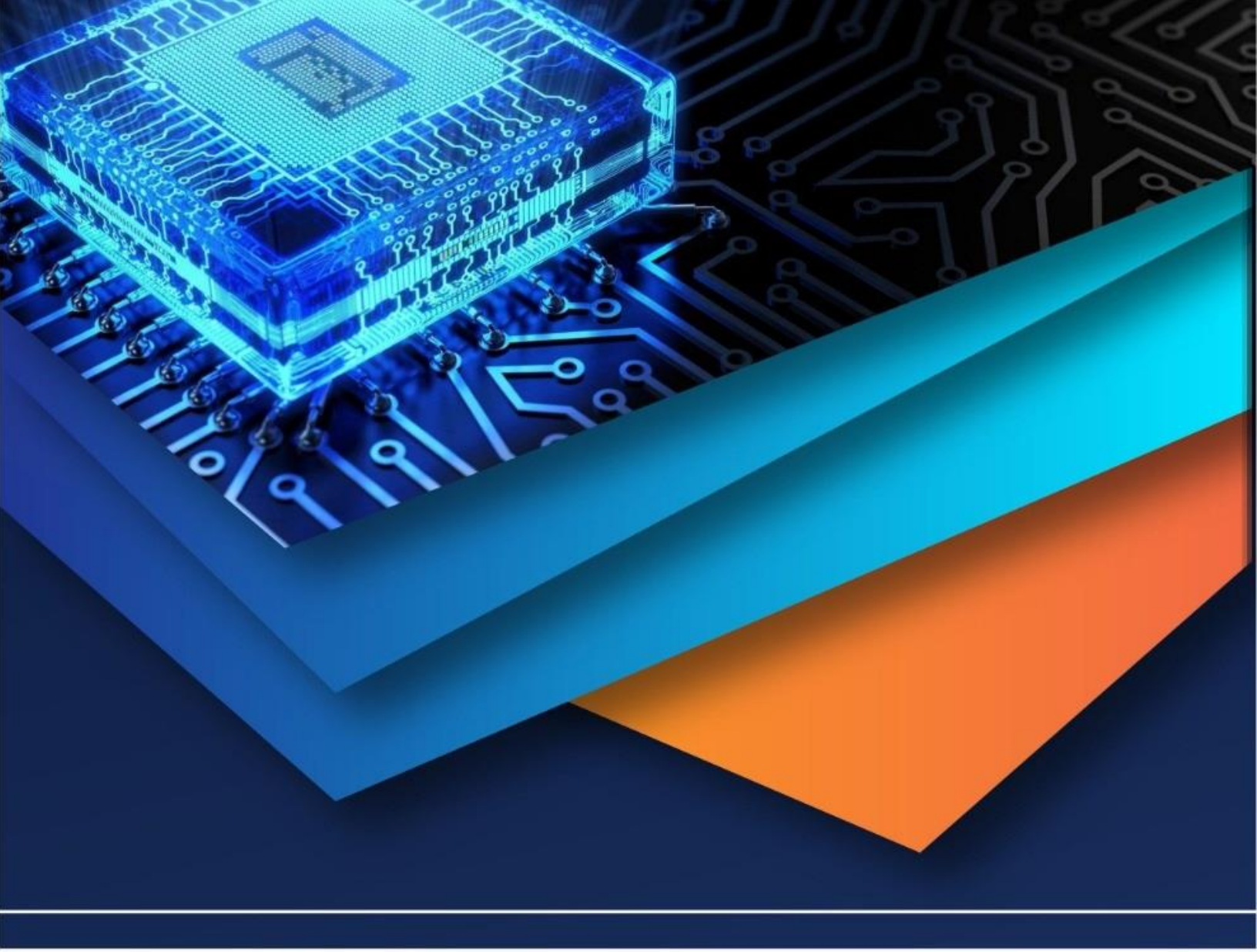
VI. CONCLUSION

1) The proposed system which maintains huge amount of multimedia using cloud server from which users can quickly and easily access content-based videos based on keywords.
2) Users can also watch the videos online to reduce disk space required for storing videos.
3) Videos can be downloaded based on video summary and least popular videos are automatically removed based on rating.

REFERENCES

[1] Zhong Hanyang, Song Xin*, Yan Zhenguo, “Vessel Sailing Patterns Analysis from S-AIS Data Dased on K-means Clustering Algorithm”, 978-1-7281-1282-4/19/$31.00 ©2019IEEE.
[2] Shalini L1, Gopali Naga Sravya2, “Analysis of HealthTweets using K-means clustering”, IEEE 2018 , Volume: 05 Issue: 03 , e-ISSN: 2395-0056, p-ISSN:2395-0072.
[3] Igor V. Anikin, Rinat M. Gazimov, “Privacy Preserving Data Mining in Terms of DBSCAN Clustering Algorithm inDistributed Systems”, 978-1-5386-4307-5/18/$31.00 a. ©2018 IEEE.
[4] Uma Ojha, Dr. Savita Goel, “A Study On Prediction Of Breast Cancer Recurrence Using Data Mining Techniques”, 978-1-5090-3519-9/17/$31.00 ©2017IEEE.
[5] M. Sivasakthi, “Classification and Prediction based Data Mining Algorithms to Predict Students’ Introductory programming Performance”, 978-1-5386-4031-9/17/$31.00©2017 IEEE.
[6] Peter Brauna, Alfredo Cuzzocreab, Lam M.V. Doana, SuyoungKima, Carson K. Leunga,*, Jose Francisco A. Matundana, Rashpal Robby Singha , “Enhanced Prediction of User-Preferred YouTube Videos Based on Cleaned Viewing Pattern History ”, International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France.

[7] Brian McClanahan and Swapna S. Gokhale, “Interplay Between Video Recommendations, Categories, and Popularity on YouTube”, DOI: 10.1109/UIC-ATC.2017.8397661©2017 IEEE.
[8] MdShamsur Rahim, AZM Ehtesham Chowdhury, Md. Asiful Islam, Mir Riyanul Islam, "Mining Trailers Data from YouTube for Predicting Gross Income of Movies", DOI: 10.1109/R10-HTC.2017.8289020©2017 IEEE.
[9] Ms.Tismy Devasia1 ,Ms.Vinushree T P2, Mr.Vinayak Hegde3,”Prediction of Students Performance using Educational Data Mining”, DOI: 10.1109/SAPIENCE.2016.7684167©2016 IEEE.
[10] CHENYU LI, (Student Member, and IEEE), JUN LIU, (Member, IEEE), AND SHUXIN OUYANG, “Characterizing and Predicting the Popularity of Online Videos”, DOI: 10.1109/ACCESS.2016.2552218©2016 IEEE.
[11] Anand Bhave, Himanshu Kulkarni, Vinay Biramane, Pranali Kosamkar, “Role of Different Factors in Predicting Movie Success”, DOI: 10.1109/PERVASIVE.2015.7087152©2015 IEEE.
[12] Ronit Nossenson, Shuval Polacheck, “On-line flows Classification of Video Streaming Applications”, 978-1-50901849-9/15 $31.00 ©2015 IEEE, DOI 10.1109/NCA.2015.51.
[13] Xu Cheng, Student Member, IEEE, Jiangchuan Liu, Senior Member, IEEE, and Cameron Dale, “Understanding the Characteristics of Internet Short Video Sharing: A YouTubeBased Measurement Study”, 1520-9210/$31.00 © 2013 IEEE.
[14] GONCA GURSUN, MARK CROVELLA IBRAHIM MATTA, “Describing and Forecasting Video Access Patterns”,978-1-4244-9921-2/11/$26.00 ©2011 IEEE.
[15] Milan Mirkovic, Dubravko Culibrk, Symeon Papadopoulos, Christos Zigkolis, Yiannis Kompatsiaris, Gavin McArdle, Vladimir Crnojevic “A comparative study of spatial, temporal and content-based patterns emerging in YouTube and Flickr”, 978-1-4577-1133-6/11/$26.00 c 2011 IEEE.
[16] Gloria Chatzopoulou , Cheng Sheng , Michalis Faloutsos , “A First Step Towards Understanding Popularity in YouTube”, DOI: 10.1109/INFCOMW.2010.5466701©2010 IEEE.
[17] Tom Broxton, Yannet Interian, Jon Vaver, Mirjam Wattenhofer, “Catching a Viral Video”, 978-0-7695-42577/10 $26.00 © 2010 IEEE, DOI 10.1109/ICDMW.2010.160.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY