

Video Data Mining

Ms. Karishma Apparao Kavle¹, Ms. Shrutika Govind Jadhav²

Lecturer, Computer Technology, Solapur Education Society's Polytechnic, Solapur¹

Lecturer, Computer Technology, Solapur Education Society's Polytechnic, Solapur²

Abstract: *In recent years, the exponential growth of multimedia content—particularly video—has created a demand for efficient storage, management, and retrieval systems. This paper proposes a real-time video streaming web portal that leverages private cloud infrastructure to store and manage a large video repository. The system enables automatic video uploads based on a predefined schedule, categorizes videos using clustering algorithms, and supports content-based retrieval through indexing and prediction techniques. Users can watch videos online, download them based on summarized metadata, and share them directly with other registered users within the platform. The system also features a user rating mechanism that assesses video popularity, enabling the automatic removal of low-engagement content to optimize storage usage and improve overall user experience. This integrated approach improves accessibility, content organization, and system scalability for video-based applications across various domains.*

Keywords: Video streaming, cloud computing, clustering, indexing, prediction, content-based retrieval, private cloud, video sharing.

I. INTRODUCTION

The rise of digital videos has transformed the way users consume content across domains such as education, marketing, and entertainment. To manage this growing demand, the proposed system offers a cloud-based solution that stores, categorizes, and streams video content efficiently.

The private cloud infrastructure ensures data security and organization-specific access. The system automatically uploads videos to the cloud on a predefined schedule and categorizes those using clustering algorithms. Users can search and retrieve videos based on content features such as title, author, and date. The system allows video streaming, downloading, rating, and intra-platform sharing, while less popular videos are periodically removed.

II. LITERATURE REVIEW

Several studies have contributed to the foundation of this project:

Zhong et al. [1] Utilized the K-means clustering algorithm on AIS vessel data to identify navigational patterns and detect anomalies.

Shalini and Gopali [2] used K-means to classify health tweets based on sentiment.

Igo [3] proposed the use of the DBSCAN clustering algorithm within a distributed environment to ensure data privacy throughout the clustering process.

Ojha and Goel [4] explored predictive models for breast cancer recurrence.

Sivasakthi [5] implemented supervised learning models to predict student performance in programming.

Braun et al. [6] cleaned YouTube viewing history for enhanced video recommendation.

McClanahan and Gokhale [7] studied how categories and popularity metrics affect video recommendations on YouTube.

These works highlight the effectiveness of clustering and predictive algorithms for content analysis, anomaly detection, and user-centric recommendations.

III. PROBLEM STATEMENT

All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

“A Real-Time Video Streaming Web Portal”



To develop a system that maintains a large collection of videos on a cloud server, enabling efficient content-based search, streaming, downloading, user-based ratings for popularity analysis, and direct sharing of videos among users.

IV. OBJECTIVES AND SCOPE

OBJECTIVES

- Automatically upload videos to a private cloud based on admin-defined schedule.
- Use clustering algorithms to categorize videos.
- Employ indexing and prediction techniques for content-based retrieval.
- Enable online viewing and downloading of videos based on summaries.
- Enable users to provide ratings for videos, with the system analyzing these ratings to identify and automatically remove the least popular content.
- Facilitate sharing of videos between registered users within the system.

Scope

- The system restricts access to registered users and maintains a secure repository on a private cloud server. It provides only content-based video access and omits third-party sharing platforms.

V. HARDWARE AND SOFTWARE REQUIREMENTS

COMPONENT	SPECIFICATION
Operating System	Windows XP/8/10
Programming Language	PHP (5.5.6)
Server	XAMPP (1.8.3)
Database	MySQL (5.6.14)
Front-End Technologies	HTML5, CSS3, JavaScript, Ajax, JQuery
Browser	Google Chrome

VII. METHODOLOGY

System Flow:

- Admin authentication and scheduled upload of videos to the cloud.
- Video categorization using clustering algorithms.
- User registration and authentication.
- Keyword-based video search using indexing and prediction algorithms.
- Video playback, download, and summary display.
- Rating-based analysis to determine and remove less popular videos.
- Secure sharing between registered users within the portal.

VIII. ALGORITHMS USED

VIII. I. Clustering Algorithm

Used to group similar videos under the same category. It enables efficient sorting and browsing.

VIII II. Indexing and Prediction

Predictive models help determine the relevance of content and forecast user interest trends..



IX. CONCLUSION

This paper presents a secure, efficient, and intelligent video streaming web portal leveraging cloud storage and data mining techniques. The system enhances user experience through smart search, categorization, sharing, and popularity analysis. Future scope includes integration with AI for better personalization and support for real-time video analytics.

REFERENCES

- [1]. Shalini L1, Gopali Naga Sravya2, “Analysis of Health-Tweets using K-means clustering”, IEEE 2018 , Volume: 05 Issue: 03 , e-ISSN: 2395-0056, p-ISSN: 2395-0072.
- [2]. Igor V. Anikin, Rinat M. Gazimov, “Privacy Preserving Data Mining in Terms of DBSCAN Clustering Algorithm in Distributed Systems”, 978-1-5386-4307-5/18/\$31.00 ©2018 IEEE
- [3]. 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 ©2017 IEEE.
- [4]. 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.
- [5]. Peter Brauna, Alfredo Cuzzocrea, 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

