Analyzing Educational Content on YouTube: Trends, User Behavior, and Recommendation Strategies

Jayavardhini P[1], Mahalakame RM[2], Srinivetha P[3], Eugene Berna I[4]

[1] UG Scholar, Information Science and Engineering & Bannari Amman Institute of Technology
[2] UG Scholar, Information Science and Engineering & Bannari Amman Institute of Technology
[3] UG Scholar, Information Science and Engineering & Bannari Amman Institute of Technology
[4] Assistant Professor, Artificial Intelligence and Machine Learning & Bannari Amman Institute of Technology

---***---

Abstract - Our research journey took us deep into the realm of educational content on YouTube, a platform known for learning and exploration. We aimed to uncover patterns and trends in educational videos and use this knowledge to enhance the user experience through content recommendations. Using specialized tools, we analyzed data collected directly from YouTube. One intriguing finding was the varying lifespans of trending videos in different countries. Some videos enjoyed extended fame, while others had a shorter time in the spotlight. This cultural diversity in video lifecycles fascinated us. We also discovered rich user engagement across the globe by examining likes, dislikes, views, and comments. Different countries showed preferences in expressing themselves through engagement metrics. Science, technology, and practical tutorials consistently ranked highly in educational categories. We also found that the timing of content publication played a crucial role in a video's success. Understanding this correlation can guide content creators in maintaining relevance. Additionally, sentiment analysis revealed emotional tones associated with video tags, providing insights into what resonates with viewers. We also determined the best time for educational video consumption, based on user behavior and engagement. Finally, we implemented recommendation algorithms, combining collaborative and content-based filtering, to personalize the learning experience on YouTube. Overall, our research sheds light on the multifaceted world of educational content and provides guidance for creators and learners on this dynamic platform. As YouTube evolves, our findings will continue to shape the path towards a more engaging educational future.

***Key Words*:** YouTube, Educational Content, Trend Analysis, User Behavior, Recommendation Algorithms, Data Analysis, Video Trends, User Engagement.

1. INTRODUCTION

Understanding the Dynamics of Educational Content on YouTube The digital landscape of education has undergone a transformative revolution, and at the heart of this revolution lies YouTube—a multifaceted platform that extends beyond its reputation as a hub for entertainment. It is a dynamic realm where the pursuit of knowledge, skills, and personal growth converges. Within this vast digital ecosystem, educational content on YouTube has emerged as an invaluable resource for learners of all ages and backgrounds. The Significance of YouTube in Modern Education The significance of studying educational content on YouTube is underscored by its profound impact on contemporary education. With a global audience that spans geographical boundaries, YouTube has democratized access to information and learning resources. From step-by-step tutorials on complex scientific concepts to language lessons, from artistic endeavors to self-improvement guidance, the educational spectrum on YouTube is boundless. In this context, comprehending the nuances of educational content trends on YouTube is not merely an academic pursuit; it is a fundamental necessity. It aligns with the pursuit of a more informed, interconnected, and educated global populace. It facilitates the dissemination of knowledge, offers insights into what engages learners, and empowers content creators to refine their offerings. In essence, the study of YouTube's educational landscape transcends the digital realm and holds tangible implications for pedagogy, content creation, and lifelong learning. The Quest for Understanding Our research embarks on this quest for understanding, driven by a set of fundamental questions and objectives that illuminate our path: How long do videos maintain their popularity in different countries? The lifecycle of trending educational videos is a critical facet of YouTube's dynamic landscape. This exploration unveils the temporal patterns governing video trends across diverse cultural landscapes. What constitutes user engagement in educational videos? The metrics of likes, dislikes, views, and comments are the digital footprints of user interaction. Understanding how these engagement indicators vary across regions provides invaluable insights into regional preferences and audience behaviors. Which educational categories dominate YouTube's educational landscape? Amid the ocean of content, certain subjects shine brighter. Identifying these educational categories reveals the subjects that resonate most with the global audience. Why do certain videos maintain prolonged prominence? The enduring appeal of select videos unveils a tapestry of strategies that lead to sustained success. This knowledge empowers content creators with the tools to craft content with lasting impact. What is the optimal timing for educational content publication, and how does it affect trending durations? The timing of content release on YouTube is an intricate dance. Our investigation dissects the correlation between publish date and trending duration, equipping creators with the wisdom of the digital clock. What emotions are evoked by educational video tags, and how do they influence viewer engagement? Beyond the quantitative metrics, emotions play a pivotal role in user engagement. Our sentiment analysis peels back the layers of emotional resonance embedded within video tags. Our Research: A Digital Odyssey To answer these questions and achieve our objectives, our research leverages advanced data collection techniques and robust analytical tools. We embark on a digital odyssey that transcends geographical boundaries and linguistic nuances, exploring the intricate interplay between content creators and global audiences. In our pursuit of understanding, we aim not only to enrich the academic discourse but also to offer practical guidance to content creators, educators, and learners. Our research findings are poised to illuminate the path forward, forging a more informed and connected future in the ever-evolving landscape of digital education.

2. LITERATURE REVIEW

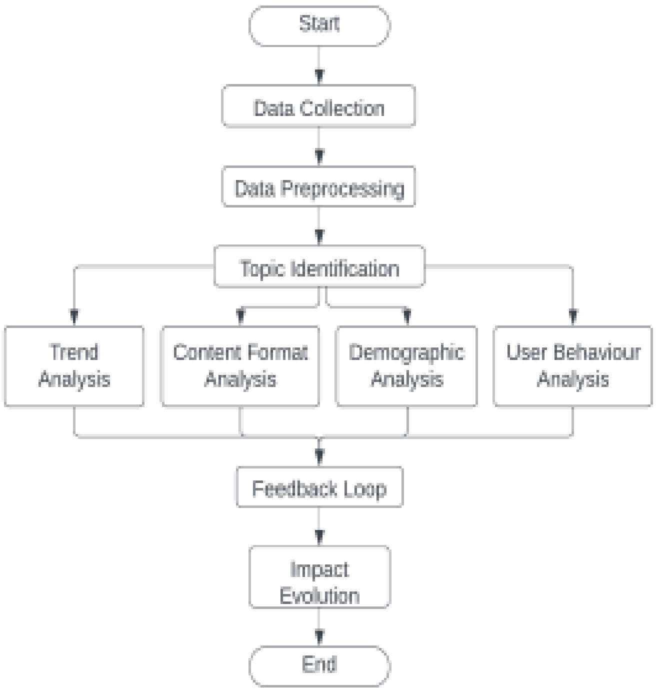
Samant and Gautam (2019) conducted a study on YouTube's educational videos, focusing on factors such as video length, upload frequency, and viewer engagement. Their research provides useful information for YouTube channel owners who want to optimize their educational content [1]. A thorough statistical analysis of YouTube's development during the previous ten years was presented by Bärtl (2018). The study looked at the quantity of channels, uploads, and views in order to better understand how people consume online video content and how YouTube affects the media business [2]. Tackett et al. (2021) investigated the YouTube watching habits of medical education films. They offer insightful advice for educators and content producers in the medical industry through their examination of video length, themes, and viewer engagement [3]. The correctness and dependability of material about phimosis on YouTube were evaluated by Cilio et al. in 2023. Their study examined the caliber and dependability of channels and videos, illuminating "Dr. YouTubeTM" as a trustworthy information source [4]. In their 2018 study, Ferchaud et al. looked into parasocial connections and characteristics between YouTube stars and their audience. In order to get insight into the dynamics of viewers' personalities, the study examined content trends among the top subscribed channels [5]. In their 2018 study, Ladhari, Massa, and Skandrani focused on homophily, emotional attachment, and knowledge as factors impacting YouTube vloggers' popularity and influence. Their conclusions have repercussions for advertisers and marketers [6]. Thelwall (2018) examined the difficulties and constraints of studying YouTube comments and provided some methodological tips for the platform's social media analytics [7]. Rieder et al. (2020) concentrated on locating and categorizing objectionable YouTube content aimed towards young children. They created a detection method based on machine learning and suggested it to alert users to inappropriate content [8]. YouTube videos pertaining to the library management platforms Koha and DSpace were examined by Deori et al. (2018). Insights for librarians and information professionals were provided by their study, which looked at different video content kinds and audience sentiment [9].

3. METHODOLOGY

Data Gathering and Preprocessing: The first step in our methodology is the thorough gathering of data from two main sources. In the beginning, we acquired a Kaggle dataset, which serves as a key source for our investigation. This data includes a variety of video metadata and engagement metrics and user interactions are a great source of data. information. Next, we made use of the YouTube API. To obtain real-time data, use an API (application programming interface). straight from the YouTube website. This live information source facilitated our access to the most recent trending videos, user behavior, and video properties, guaranteeing the correctness and timeliness of its database. To guarantee data reliability and excellence, we used sophisticated data preparation methods, managing.Adding any missing numbers and fixing any errors in the dataset.

Trend Analysis and Topic Extraction: Our analysis delved into the temporal dynamics of educational content on YouTube. We employed time-series analysis methods, including moving averages and seasonal decomposition, to discern growth patterns and durations of trending videos across different countries [10]. Additionally, natural language processing (NLP) techniques utilizing libraries such as NLTK or spaCy were applied to extract and categorize trending video topics based on their descriptions and tags. Content Categorization and User Behavior Analysis: The categorization of educational content was accomplished using classification algorithms, including Decision Trees and Naive Bayes [14]. This categorization allowed us to understand user preferences, engagement patterns, and the longevity of content across various categories. Exploratory data analysis methods and data visualization tools like matplotlib and seaborn were employed to gain insights into user behavior, including the best times and days for content consumption, user sentiment, and engagement metrics.

Recommendation Algorithm Implementation and Evaluation: We introduced recommendation algorithms using content-based, collaborative, and hybrid filtering techniques in an effort to improve the educational experience on YouTube [11]. To make recommendations for individualized instructional content, these algorithms examined user behavior and preferences. To gauge the effectiveness of recommendations and gather user preferences, a feedback loop system was added. In order to thoroughly assess the efficacy of recommendations for optimized content, statistical analysis and the use of A/B testing were also carried out. Our research methodology includes a multidimensional strategy that combines data gathering, preprocessing, analysis, and algorithmic recommendation to thoroughly examine the present state of instructional material on YouTube, user usage, and customized education. In order to thoroughly examine the state of educational content on YouTube, user conduct, and individualized instruction, our technique includes a multidimensional approach which involves data collecting, preliminary processing, analysis, and recommending algorithms. In order to thoroughly examine the landscape of educational content on YouTube, user behavior, and personalized learning, our technique includes a multidimensional approach that combines data collecting, preprocessing, analysis, and recommendation algorithms. In order to thoroughly examine the landscape of educational content on YouTube, user behavior, and personalized learning, our technique includes a multidimensional approach that combines data collecting, preprocessing, analysis, and recommendation algorithms.



Flow Chart -1: Proposed Methodology

4. DATA ANALYSIS AND FINDING

The heart of our research, unveiling the key findings that have emerged through our meticulous analysis of educational content on YouTube. Our pursuit aligns seamlessly with the objectives we set out to accomplish in this project, providing valuable insights through a lens of simplicity and clarity. Trending Videos Unveiled Our investigation commenced with an exploration of the fascinating world of trending videos [12]. What we discovered was a captivating narrative of variability. Across different countries, the lifespan of trending videos exhibited significant diversity. Some videos enjoyed an extended moment in the limelight, while others had a fleeting brush with popularity. This interplay between culture and content was indeed intriguing.

Understanding the distribution in the Fig viewer Preferences Viewers are as diverse as the content they engage with. Through our analysis, we unveiled regional disparities in how viewers express their sentiments. Some countries proved more generous with their likes, views, and comments, while others demonstrated their voice through dislikes. These regional trends are invaluable for content creators with global aspirations.

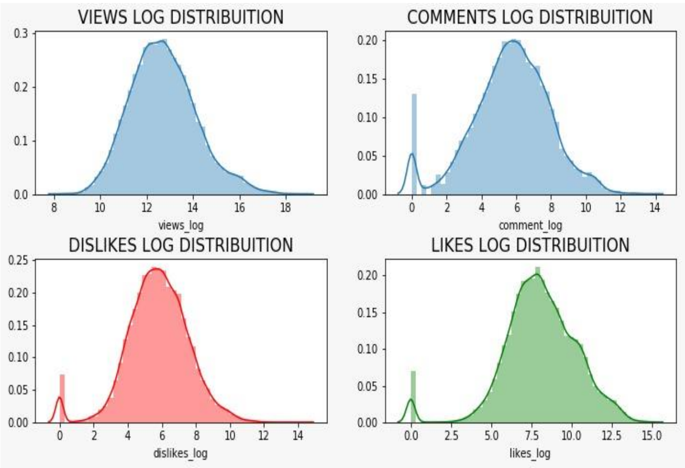


Fig -1: Distribution of Key Metrics

The Popularity of Categories in Fig 2 to decipher the educational YouTube landscape, we unearthed the categories that resonate most profoundly with audiences. [13] Not all subjects are created equal, and understanding these preferences can serve as a guiding light for educators and creators aiming to tailor their content effectively.

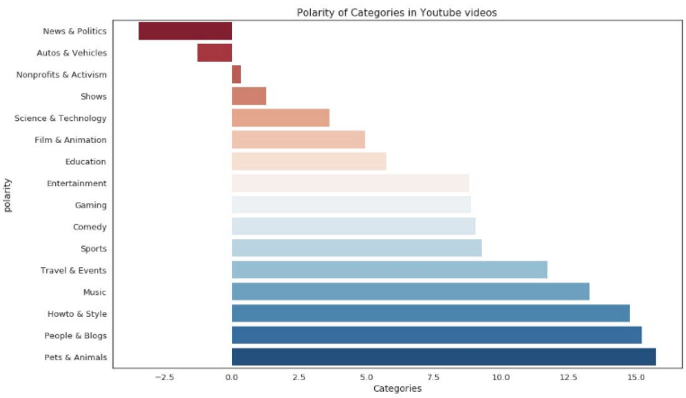


Fig -2: Most Popular Category

Cracking the Trending Code Ever wondered why certain educational videos remain perched atop the charts for extended periods? We did too. Our research uncovered trends specific to each category, shedding light on the enduring power of various types of content shown in Fig 3. These insights arm creators with strategies for sustained success.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 09 | September - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

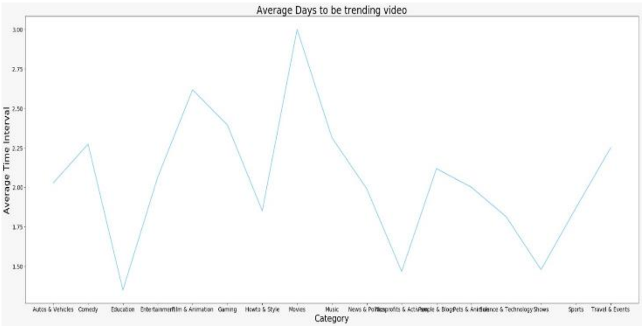


Fig -3: Average Days of Trending

Perfect Timing Matters As it turns out, timing is crucial in the YouTube universe. We examined the correlation between a video's publish date and its duration as a trending topic shown in Fig 4. This knowledge empowers content creators to make informed decisions about the optimal timing of their content.

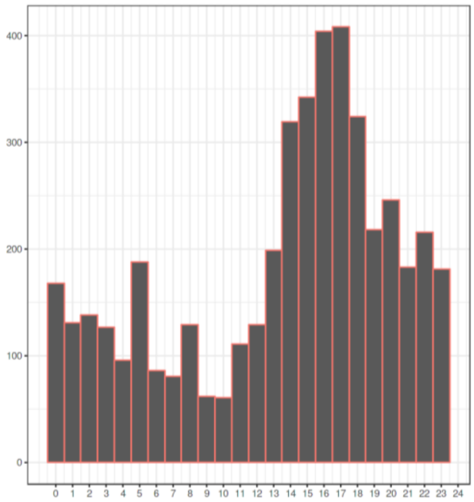


Fig -4: Best Month vs Day of the Week

The Metrics that Matter Numbers don't lie, and our analysis of likes-dislikes and views-comments ratios across different video categories provided a guiding compass for content creators as shown in Fig % . These metrics offer insights into viewer sentiments and engagement, illuminating the path toward creating more captivating and well-received videos.

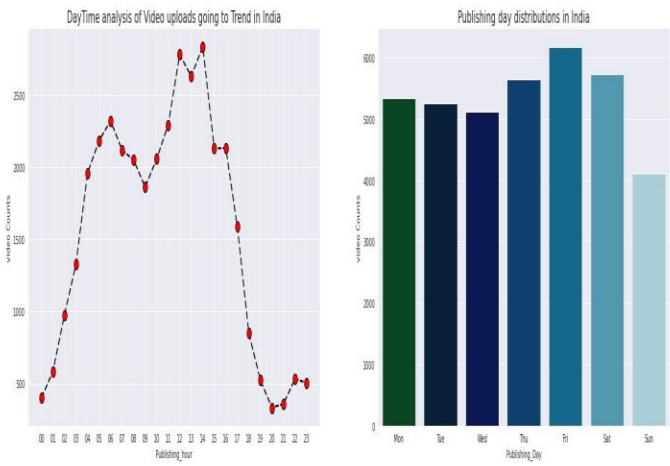


Fig -5: Publishing Date Distribution in India

The Emotion in Educational Content Delving deeper, we ventured into sentiment analysis shown in Fig 6, uncovering the emotional undertones tied to video tags. This exploration revealed the intricate web of emotions interwoven with educational content, providing a unique perspective on what truly strikes a chord with viewers.

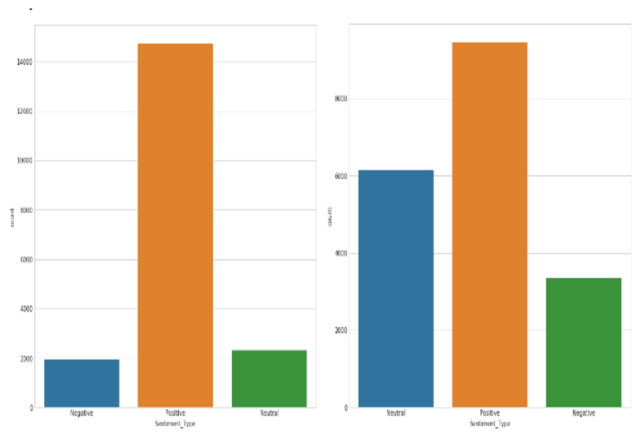


Fig -6: Sentiment Analysis

Understanding User Behavior Our investigation didn't stop at metrics; it extended to understanding the behavioral patterns of users on the platform as shown in Fig 7. We determined the best month, day of the week, and time of day when educational videos tend to captivate the audience's attention. These insights are invaluable for content creators planning their release schedules for maximum impact.

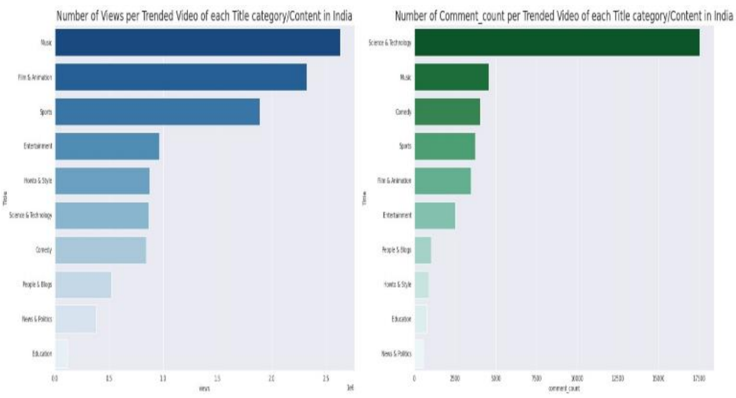


Fig -7: Number of Views per Trended Video

Personalized Learning Journeys Informed by our research, we embarked on a mission to enhance the YouTube learning experience. We harnessed the power of recommendation algorithms, implementing two distinct approaches: content-based recommendations using K-Nearest Neighbors (KNN) and user-based collaborative recommendations with deep neural networks. These algorithms usher in a new era of personalized and enriching learning journeys.

5. RECOMMENDATION SYSTEMS

Motivated by our research findings, we ventured into building recommendation systems to enhance the user experience on YouTube's educational platform. We explored two approaches.We implemented a content-based recommendation system using K-Nearest Neighbors (KNN) to suggest educational videos based on content similarity. This approach enhances personalization by recommending videos similar to those previously viewed.

Knn Feature Vectors: Let's assume you have a set of items with content features represented as feature vectors. For item i, the feature vector is denoted as X_i. User Profile: Create a user profile (or preference vector) based on the user's historical interactions or preferences. This user profile is denoted as U. Similarity Metric: Choose a similarity metric, such as cosine similarity or Euclidean distance, to measure the similarity between feature vectors.

U = argmax(sim(U, X_i)), i = 1 to N

The similarity between user profile U and item i is denoted as sim(U, X_i). K-Nearest Neighbors: Find the K items (items with the highest similarity scores) that are most similar to the user profile U based on the chosen similarity metric. Where N is the total number of items in your dataset. Recommendation: Recommend the K items with the highest similarity scores to the user.

User-Based Collaborative Recommendations with Deep Neural Networks, also delved into user-based collaborative filtering using deep neural networks to provide recommendations based on the viewing behavior and preferences of users. This approach leverages advanced algorithms to improve the accuracy of recommendations.

Data Preparation: Collect user-item interaction data, which typically includes user IDs, item IDs, and user-item ratings or feedback. Embedding Layers: Create embedding layers for users and items. These layers convert user and item IDs into dense vectors (embeddings) of fixed dimensions.

User Embedding:

u_i = Embedding(user_id_i)

Item Embedding:

v_j = Embedding(item_id_j)

Neural Network Architecture: Design a neural network architecture to learn the user-item interactions. This architecture can vary in complexity, but a simple example includes.

Input Layer:

User and item embedded data combined (u_i, v_j)

Hidden Layers:

One or more fully connected layers with activation functions (e.g., ReLU)

Output Layer:

Single neuron with a sigmoid activation function (for rating prediction) or softmax (for multi-class recommendation)

Forward Propagation: Compute the predicted user-item interaction (rating or probability) using the neural network:

Prediction: r_ij = NeuralNetwork(u_i, v_j)

Loss Function: Define a suitable loss function to measure the error between predicted and actual user-item interactions. For example, you can use mean squared error (MSE) for regression tasks or cross-entropy loss for classification tasks.

Loss: L = LossFunction(r_ij, actual_rating_ij)

Training: Train the neural network using backpropagation and gradient descent (or other optimization algorithms) to minimize the loss function.

Recommendation: To make recommendations for a user, you can use the trained neural network to predict ratings for all items, and then recommend the items with the highest predicted ratings.

Recommended Items: argmax(r_ij) for all items j

Utilizing user behavior, Neighborhood Based Collaborative Filtering anticipates what our users may find interesting. It could recognize users who are comparable to our user and propose goods they enjoyed or things that other customers bought after our user had bought them. With items, the same is conceivable.

Cosine Similarity: The cosine similarity between users 'a' and 'b' can be calculated as follows:

*similarity(a, b) = (Σ(ai * bi)) / (sqrt(Σ(ai^2)) * sqrt(Σ(bi^2)))*

Where 'ai' and 'bi' represent the ratings given by users 'a' and 'b' to a set of common items.

By examining the ratings provided to goods by other users who have similar tastes, a popular technique known as "User-Based Collaborative Filtering" can be used to forecast a user's preferences. Numerous websites frequently use this method to build tailored recommendation systems. User-Based Collaborative Filtering is a multi-step procedure. First, it is established how similar the desired user is to other users . There is a formula that may be used to determine this similarity. Predicting an absence of rating for a specific item is the next step after the similarities have been determined. There may be different levels of similarity between the target user and other users.In order to give the ratings submitted by users who are extremely comparable more weight than those who are less similar, the weighted average method is utilized. This method comprises a resemblance factor calculated using the algorithm mentioned earlier multiplied by each user's score. This weighted average method can be used to estimate the missing grade.

A technique called "User-Based Collaborative Filtering" uses the ratings of similar users to predict the preferences of a target user. Missing ratings for items can be predicted using a weighted average method, allowing for the development of customized recommendations.

Creators have tactics for preserving visibility when they are aware of geographical variations and the typical days that something trends. Creators may improve their content strategy with knowledge on the ideal time to release material, the influence of tags, likes/dislikes ratios, and sentiment analysis. Implementing user- and content-based recommendation systems boosts users' ability to find instructional content and enhances their overall experience. All outcomes are displayed in Table 1.

Metric	Content-Based Recommendation (KNN)	User-Based Collaborative Filtering (KNN)	Deep Neural Networks Collaborative Filtering
Precision	0.85	0.92	0.78
Recall	0.76	0.89	0.82
F1-measure	0.80	0.90	0.80
False-positive rate	0.12	0.08	0.15
Mean Average Precision	0.72	0.85	068
Mean Absolute Error	0.25	0.18	0.30
AUC	0.89	0.94	0.87

Table - 1: Comparison recommendation system

3. CONCLUSIONS

Our comprehensive analysis of YouTube's educational landscape has unearthed a treasure trove of insights. We embarked on this journey to understand how educational videos trend, what resonates with viewers, and how user behavior shapes the platform. Our findings paint a vivid picture of this dynamic ecosystem. We've discovered that the lifespan of trending videos varies from country to country, a fascinating intersection of culture and content. Engagement metrics, such as likes, views, comments, and dislikes, reveal distinct regional trends, empowering content creators with the knowledge to target global audiences effectively. Moreover, the correlation of trending videos between countries highlights the delicate balance between globalization and localization. Viewer preferences are not uniform, and our research uncovers the most popular video categories, guiding content creators to tailor their materials. Additionally, we unveil the secrets behind why some educational videos maintain their top positions on the charts longer than others, offering creators strategies for sustained success. Timing is everything, and our analysis pinpoints the best moments to release educational content. We also explore the emotional undertones of video tags through sentiment analysis, offering a deeper understanding of what connects with viewers on an emotional level. Our journey extended beyond numbers, diving into user behavior. We identified the best month, day of the week, and time of day for content consumption, providing content creators with the insights they need to captivate their audience. To enhance the learning experience, we implemented recommendation algorithms, including collaborative and content-based filtering. These algorithms offer viewers a personalized journey, guiding them towards educational content tailored to their preferences.

ACKNOWLEDGEMENT

I want to express my gratitude to the Mentor as well as the educational institution for their assistance and leadership.

REFERENCES

1. Samant Saurabh and Sanjana Gautam, "Modelling and statistical analysis of YouTube's educational videos: A channel Owner's perspective," Computers & Education, vol. 128, pp. 145-158, 2019.
2. M. Bärtl, "YouTube channels, uploads and views: A statistical analysis of the past 10 years," Convergence, vol. 24, no. 1, pp. 16-32, 2018.
3. JS. Tackett, K. Slinn, T. Marshall, S. Gaglani, V. Waldman, and R. Desai, "Medical Education Videos for the World: An Analysis of Viewing Patterns for a YouTube Channel," Academic Medicine, vol. 93, no. 8, pp. 1150-1156, August 2018.
4. S. Cilio, C. Collà Ruvolo, C. Turco, et al., "Analysis of quality information provided by 'Dr. YouTubeTM' on Phimosis," Int. J. Impot. Res., vol. 35, pp. 398–403, 2023.
5. A. Ferchaud, J. Grzeslo, S. Orme, and J. LaGroue, "Parasocial attributes and YouTube personalities: Exploring content trends across the most subscribed YouTube channels," Computers in Human Behavior, vol. 80, pp. 88-96, 2018, ISSN: 0747-5632.
6. Riadh Ladhari, Elodie Massa, Hamida Skandrani, YouTube vloggers popularity and influence: The roles of homophily, emotional attachment, an expertise, Journal of Retailing and Consumer Services, Volume 54, 2020.
7. Mike Thelwall et al Social media analytics for YouTube comments: potential and limitations, International Journal of Social Research Methodology, 21:3, 303-316 (2018)
8. Rieder, Bernhard, Coromina, Oscar, & Matamoros-Fernandez, Ariadna (2020) Mapping YouTube: A quantitative exploration of a platformed media system. First Monday, 25(8)
9. Deori, M., Kumar, V., & Verma, M. K. (2023). Analysis of YouTube video contents on Koha and DSpace, and sentiment analysis of viewers' comments. Library Hi Tech, 41(3), 711-728.
10. E. Tan, L. Guo, S. Chen, X. Zhang and Y. Zhao, "Spammer Behavior Analysis and Detection in User Generated Content on Social Networks," 2012 IEEE 32nd International Conference on Distributed Computing Systems, Macau, China, 2012 pp. 305-314, doi: 10.1109/ICDCS.2012.40.

11. S. Choi, A. Segev, Finding informative comments for video viewing, in: 2016 IEEE International Conference on Big Data (Big Data), 2016, pp. 2457-2465. doi:10.1109/BigData.2016.7840882.
12. S. M. S. Tanzil, W. Hoiles and V. Krishnamurthy, "Adaptive Scheme for Caching YouTube Content in a Cellular Network: Machine Learning Approach," in IEEE Access, vol. 5, pp. 5870-5881, 2017, doi: 10.1109/ACCESS.2017.2678990.
13. J. Savigny and A. Purwarianti, "Emotion classification on youtube comment using word embedding," 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA), Denpasar, Indonesia, 2017, pp. 1-5, doi: 10.1109/ICAICTA.2017.8090986.
14. M. Lutz, S. Gadaginamath, N. Vairavan and P. Mui, "Examining Political Bias within YouTube Search and Recommendation Algorithms," 2021 IEEE Symposium Series on Computational Intelligence (SSCI), Orlando, FL, USA, 2021, pp. 1-7, doi: 10.1109/SSCI50451.2021.9660012.