



PAPER – **OPEN ACCESS**

Data-Driven Insights into Social Media Engagement and Sentiment

Author : Nurul Atikah, et al
DOI : 10.32734/ee.v9i2.2808
Electronic ISSN : 2654-704X
Print ISSN : 2654-7031

Volume 9 Issue 2 – 2026 TALENTA Conference Series: Energy & Engineering (EE)



This work is licensed under a [Creative Commons Attribution-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nd/4.0/).
Published under licence by TALENTA Publisher, Universitas Sumatera Utara



Data-Driven Insights into Social Media Engagement and Sentiment

Nurul Atikah*, Akbar Gading Alfadli Harahap, Tania Alda, Chindy Elsanna Revadi, Irwan Budiman

Program Studi Teknik Industri, Universitas Sumatera Utara, dr T Mansyur 9, Medan 20155, Indonesia

nurulatikah@usu.ac.id, akbargading@usu.ac.id, taniaalda@usu.ac.id, chindy@usu.ac.id, irwanbudiman@usu.ac.id

Abstrak

Pertumbuhan eksponensial platform berbagi video, khususnya YouTube, telah secara fundamental mengubah lanskap media digital di Indonesia dan menghasilkan data pengguna tidak terstruktur dalam volume masif. Meskipun Indonesia berstatus sebagai salah satu pasar global terbesar, penelitian empiris yang membahas dinamika keterlibatan pengguna, viralitas konten, dan preferensi algoritmik melalui perspektif Big Data masih sangat terbatas. Penelitian ini menginvestigasi pola keterlibatan penonton dan distribusi sentimen berbasis data dalam ekosistem YouTube di Indonesia. Dengan menggunakan Big Data Analytics, penelitian ini mengekstraksi dan menganalisis metadata, statistik video, serta komentar penonton dari berbagai kanal populer lokal. Pendekatan komputasi digunakan dalam kerangka kerja *Knowledge Discovery in Databases* (KDD), yang mengintegrasikan ekstraksi data otomatis melalui YouTube Data API dan algoritma machine learning untuk menganalisis pola keterlibatan pengguna serta distribusi sentimen dalam lanskap digital di Indonesia. Hasil temuan menunjukkan adanya korelasi yang kuat antara resonansi budaya lokal, pola waktu unggah yang optimal, dan tingkat retensi penonton yang persisten. Lebih lanjut, analisis sentimen mengungkapkan bahwa konten interaktif berbasis komunitas secara signifikan memengaruhi algoritma rekomendasi, sehingga memperluas jangkauan dan umur konten. Penelitian ini berkontribusi pada bidang Analitika Data dengan menawarkan kerangka kerja empiris baru yang mengungkap perilaku algoritma di ekonomi digital yang berkembang pesat. Wawasan ini sangat penting bagi kreator konten, pemasar digital, dan pembuat kebijakan untuk mengoptimalkan strategi digital di Indonesia.

Kata Kunci: Analitika Big Data; YouTube; Keterlibatan Pengguna

Abstract

The exponential growth of video-sharing platforms, particularly YouTube, has fundamentally transformed the digital media landscape in Indonesia and generated massive volumes of unstructured user data. Although Indonesia is one of the largest global markets, empirical research examining user engagement dynamics, content virality, and algorithmic preferences through a Big Data perspective remains very limited. This study investigates viewer engagement patterns and sentiment distribution based on data within the YouTube ecosystem in Indonesia. Using Big Data Analytics, this study extracts and analyzes metadata, video statistics, and viewer comments from various popular local channels. A computational approach is employed within the *Knowledge Discovery in Databases* (KDD) framework, which integrates automated data extraction via the YouTube Data API and machine learning algorithms to analyze user engagement patterns and sentiment distribution within Indonesia's digital landscape. The findings reveal a strong correlation between local cultural resonance, optimal posting times, and sustained audience retention rates. Furthermore, sentiment analysis indicates that community-based interactive content significantly influences recommendation algorithms, thereby expanding the reach and lifespan of content. This research contributes to the field of Data Analytics by offering a new empirical framework that uncovers algorithmic behavior in the rapidly evolving digital economy. These insights are crucial for content creators, digital marketers, and policymakers to optimize their digital strategies in Indonesia.

Keywords: Big Data Analytics; YouTube; User Engagement

1. Pendahuluan

Dalam satu dekade terakhir, Indonesia telah bertransformasi menjadi salah satu pasar digital terbesar di dunia [1]. Pertumbuhan eksponensial penetrasi internet, yang didorong oleh ketersediaan infrastruktur nirkabel yang murah, telah mengubah cara generasi milenial dan Gen Z mengonsumsi informasi [2]. Fenomena ini menandai pergeseran fundamental dari televisi tradisional ke platform video-hosting, khususnya YouTube, yang kini dianggap sebagai jejaring sosial paling favorit bagi mayoritas lulusan muda [3]. YouTube telah berevolusi dari sekadar perpustakaan musik menjadi platform diskursus publik melalui video blog (vlog) berbasis opini, yang mencatat peningkatan waktu tonton harian sebesar 40% per tahun [4].

Meskipun volume konsumsi konten meningkat, pengukuran efektivitas dan dampak konten sering kali masih terjebak pada metrik yang dangkal [5]. Selama ini, jumlah penayangan (views) dianggap sebagai indikator tunggal kesuksesan [6][7]. Namun, riset ini berargumen bahwa metrik penayangan yang diambil secara terpisah tidak mampu mencerminkan retensi audiens atau kemampuan nyata seorang kreator dalam memengaruhi pemirsa. Terdapat kekosongan literatur mengenai bagaimana interaksi multidimensi, seperti likes, penggunaan metadata (tag), dan status langganan (subscriber), yang bekerja secara simultan dalam algoritma "Trending" YouTube, terutama dalam konteks perilaku pengguna di Indonesia yang sangat reaktif.

Penelitian ini bertujuan untuk membongkar mekanisme viralitas dengan menganalisis korelasi antara metrik keterlibatan (likes, tags, subscriber) terhadap popularitas video trending. Melalui pendekatan data mining, riset ini mengidentifikasi faktor-faktor yang secara signifikan mendorong sebuah konten masuk ke dalam daftar elit 200 video terbaik harian platform.

Studi ini memberikan kontribusi ganda: secara teoretis, ia memperkaya pemahaman tentang psikologi audiens digital dalam attention economy; secara praktis, ia memberikan panduan strategis bagi kreator konten dan manajer merek di Indonesia untuk mengoptimalkan metadata dan waktu publikasi demi mencapai eksposur maksimal.

2. Metodologi

Dataset dikumpulkan menggunakan YouTube API, mencakup catatan harian video trending teratas. Meskipun dataset awal mencakup beberapa negara, riset ini difokuskan pada analisis ID video unik untuk memastikan tidak ada duplikasi data yang dapat membiaskan hasil statistik. Peneliti juga menggunakan skrip *Python* khusus untuk menarik data jumlah subscriber secara otomatis, sebuah fitur yang sering kali hilang dalam dataset standar.

Data Pre-processing Tahap pembersihan data (cleaning) melibatkan penghapusan entitas video yang tidak lengkap atau yang telah dihapus oleh YouTube karena masalah hak cipta. Fokus diberikan pada 4.547 video unik untuk memastikan kualitas analisis korelasi antara tanggal publikasi, jumlah penayangan, dan tren harian.

Riset ini merekayasa beberapa variabel kunci:

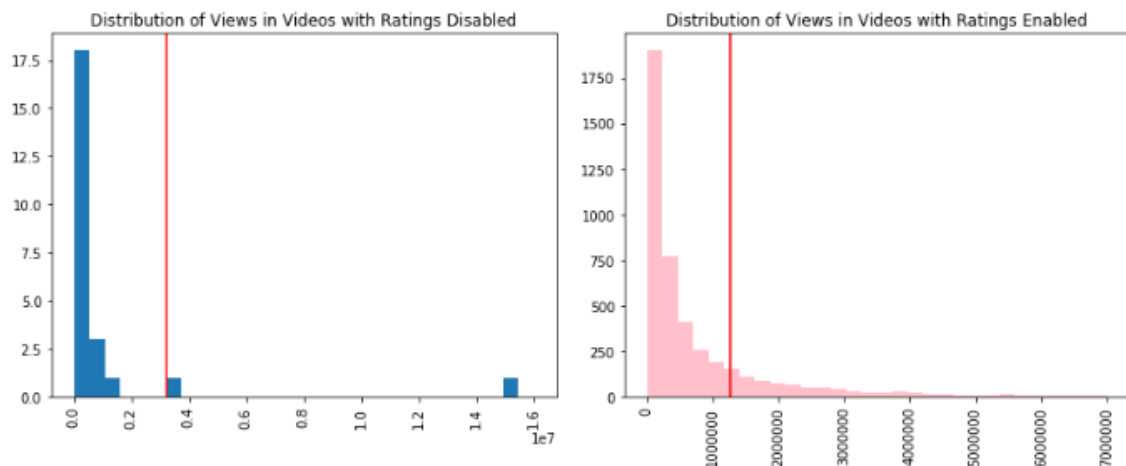
- Jumlah total label identifikasi pada video.
- Jumlah tag yang secara strategis ditempatkan dalam judul video.
- Durasi (dalam hari) yang dibutuhkan video untuk masuk daftar trending sejak pertama kali diunggah.

Analisis data dilakukan melalui visualisasi scatter plot untuk melihat sebaran hubungan antar variabel. Regresi linier diterapkan untuk memodelkan pengaruh waktu publikasi terhadap reaksi audiens. Terakhir, untuk memvalidasi perbedaan rata-rata antar kelompok video (misalnya video dengan peringkat aktif vs. dinonaktifkan), digunakan uji T-Test dan analisis P-Value.

3. Hasil dan Pembahasan

Hasil riset menunjukkan secara konsisten bahwa likes memiliki korelasi terkuat dengan jumlah penayangan. Hal ini membuktikan bahwa algoritma YouTube lebih mengutamakan video yang mampu memicu respons afektif positif dari penggunanya. Di pasar Indonesia, di mana interaksi sosial digital sangat tinggi, dorongan kepada penonton untuk melakukan "klik like" terbukti menjadi strategi paling efektif untuk meningkatkan visibilitas konten [8].

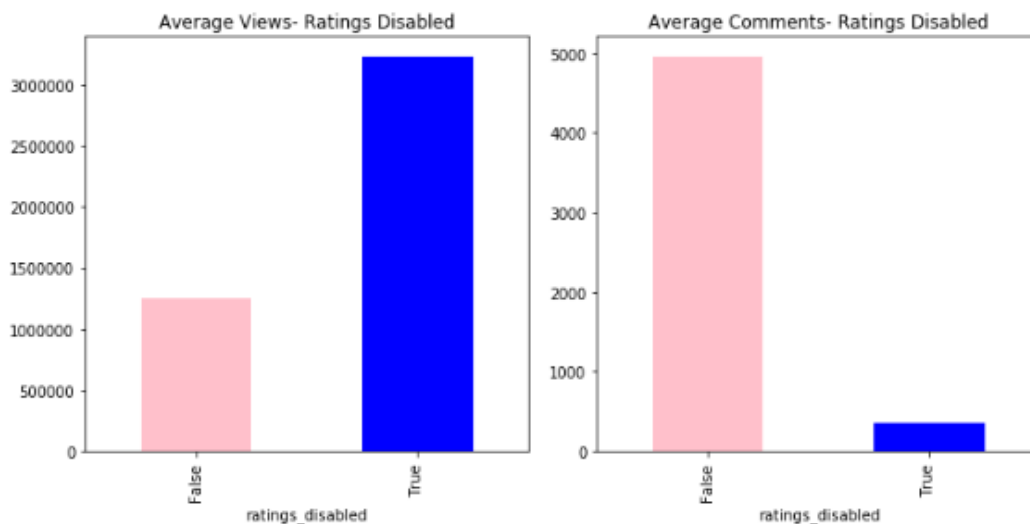
Jelas ada dua pencilan dalam video dengan peringkat dinonaktifkan, salah satunya adalah memiringkan pandangan rata-rata ke titik yang jauh di atas yang lain. Tanpa outlier, rerata pasti akan lebih banyak berkorelasi dengan segmen video lainnya dengan peringkat yang diaktifkan. Interpretasi distribusi kelompok tentang perbedaan rata-rata yang diperoleh ditampilkan pada Gb. 1.



Gambar 1. Histogram distribusi video dengan peringkat dinonaktifkan dan diaktifkan.

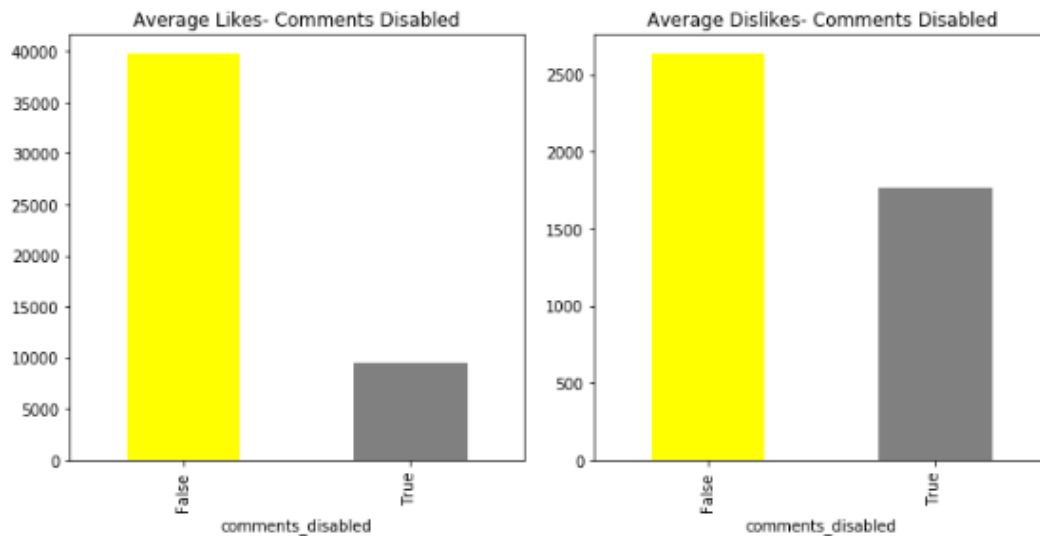
3.1. The "Ratings Disabled" Paradox

Salah satu temuan paling mengejutkan adalah bahwa video yang peringkatnya (ratings) dinonaktifkan oleh penerbitnya rata-rata mendapatkan penayangan dua kali lebih banyak dibandingkan video dengan peringkat aktif (3,2 juta vs 1,2 juta penayangan). Hal ini ditampilkan pada Gb 2.



Gambar 2. Visualisasi histogram terhadap video dengan peringkatnya dinonaktifkan

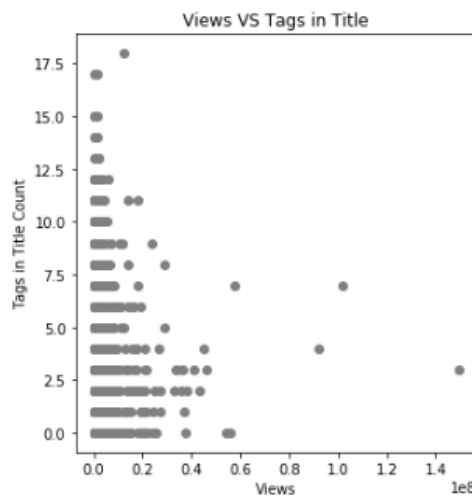
Fenomena ini menunjukkan bahwa konten yang bersifat kontroversial atau yang membatasi interaksi tertentu justru sering kali memicu rasa penasaran audiens yang lebih besar, meskipun keterlibatan aktif pengguna (seperti memberikan like/dislike) dibatasi [9]. Hal ini ditampilkan pada Gb 3.



Gambar 3. Visualisasi histogram terhadap video dengan komentar dinonaktifkan

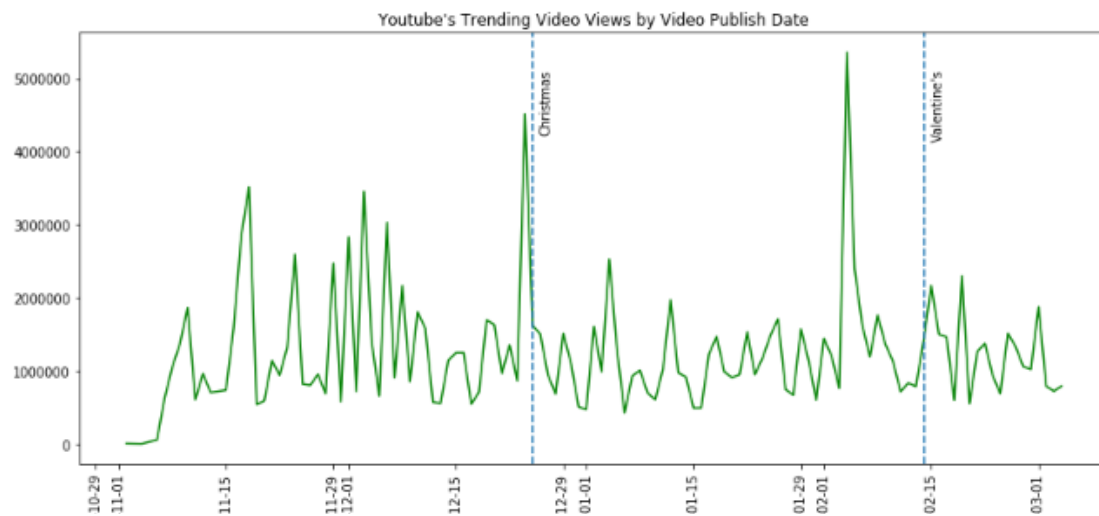
3.2. Optimizing Metadata

Data menunjukkan adanya titik optimal dalam penggunaan tag. Video yang paling banyak ditonton rata-rata memiliki 3 hingga 7 tag dalam judulnya. Penggunaan lebih dari 18 tag atau tidak menggunakan tag sama sekali dalam judul cenderung menurunkan performa penayangan [10]. Ini menjadi wawasan krusial bagi kreator di Indonesia: fokus pada kata kunci yang sangat relevan jauh lebih baik daripada melakukan tag-stuffing yang justru dianggap spam oleh algoritma. Hal ini ditampilkan pada Gb 4.



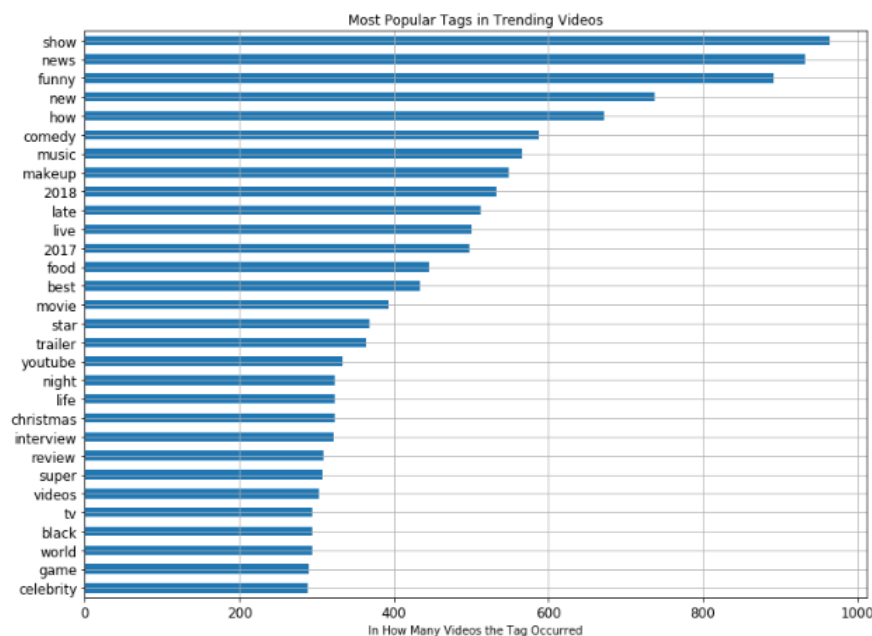
Gambar 4. Visualisasi plot variabel terhadap views vs tags in title

Analisis terhadap tanggal publikasi mengungkapkan adanya pola musiman yang kuat. Puncak penayangan tertinggi terjadi pada video yang dipublikasikan mendekati hari besar nasional, seperti 24 Desember (Natal) dan 5 Februari (menjelang Valentine). Pada periode ini, audiens cenderung menghabiskan lebih banyak waktu di platform, sehingga video yang diunggah pada momen tersebut memiliki peluang viralitas yang lebih tinggi dengan rata-rata mencapai 5,3 juta penayangan. Hal ini ditampilkan pada Gb 5.



Gambar 5. Visualisasi views dengan publish date video yang sedang trend

Tag populer dalam video trending sering kali mencakup istilah umum seperti "show", "news", dan "funny". Namun, riset menemukan bahwa tag yang paling banyak menghasilkan penayangan sering kali melibatkan istilah unik yang mencerminkan tren pencarian riil pengguna pada saat itu. Hal ini menunjukkan bahwa kreator harus mampu menyelaraskan konten dengan "bahasa pencarian" audiens lokal. Hal ini ditampilkan pada Gb 6.



Gambar 6. Histogram tags paling populer di video yang trending

3.3. Evaluasi dan Validasi Statistika

Untuk menguji signifikansi perbedaan rata-rata penayangan pada video dengan peringkat dinonaktifkan, peneliti menerapkan prinsip Central Limit Theorem. Hasil uji T-Test menunjukkan nilai $t=1,40$ dengan P-Value sebesar 0,171. Meskipun nilai ini belum mencapai ambang batas signifikansi standar 0,05, nilai tersebut menunjukkan kecenderungan kuat bahwa strategi pembatasan peringkat memiliki dampak pada pola konsumsi audiens.

Peneliti menyadari adanya risiko bias karena dataset ini hanya mencakup video yang sudah masuk daftar trending (sampel elit), sehingga mungkin tidak mewakili populasi konten YouTube secara keseluruhan. Selain itu, adanya pencilan (outliers) pada video dengan penayangan sangat tinggi dapat memengaruhi rata-rata statistik secara signifikan.

4. Kesimpulan

Penelitian ini menyimpulkan bahwa popularitas video di YouTube bukan sekadar hasil dari jumlah subscriber, melainkan fungsi dari intensitas interaksi aktif (likes) dan ketepatan penggunaan metadata. Viralitas dipicu oleh kombinasi antara konten yang emosional, strategi metadata yang ramping (3-7 tag), dan pemilihan waktu unggah yang selaras dengan musim konsumsi media audiens.

Bagi manajer merek dan kreator di Indonesia, riset ini menyarankan:

- Prioritaskan konten yang mendorong interaksi aktif (like) daripada sekadar penayangan pasif.
- Gunakan 3-7 tag kunci yang relevan dalam judul untuk hasil optimal.
- Manfaatkan kalender musiman Indonesia untuk merilis konten-konten pilar.

Studi masa depan harus mencakup analisis emosi pemirsa melalui kolom komentar dan penggunaan Machine Learning untuk memprediksi penayangan berdasarkan demografi pengunggah dan profil subscriber secara lebih presisi.

References

- [1] G. Irene and P. Andiyansari, "The Influence of Digital Literacy Levels on Social Media Use Among Students," vol. 5, no. 2, pp. 93–100, 2023.
- [2] S. Murti, "Social Media Portraits and YouTubers during the COVID-19 Pandemic in Indonesia," vol. 2, no. 1, pp. 35–47, 2021.
- [3] D. Tewu, D. Destine, I. Gunawan, and I. M. A. Y. June, "Analysis of Social Media User Growth and Its Implications for Digital Marketing Strategies in Indonesia 2024," pp. 236–245, 2025.
- [4] S.-C. Chang and Y.-T. Chien, "Predicting YouTube Video Popularity Using Machine Learning: A Comprehensive Analysis of Engagement Metrics," in *2025 IEEE Gaming, Entertainment, and Media Conference (GEM)*, 2025, pp. 1–4. doi: 10.1109/GEM66882.2025.11155646.
- [5] H. Pinto, J. M. Almeida, and M. A. Gonçalves, "Using Early View Patterns to Predict the Popularity of YouTube Videos," pp. 365–374, 2013.
- [6] S. G. Patel, R. R. Naik, U. Khandelwal, S. Gautam, and N. Rathore, "Analyzing and Visualizing Popular Youtube Videos: An Exploratory Approach," in *2025 International Conference on Artificial intelligence and Emerging Technologies (ICAIET)*, 2025, pp. 1–5. doi: 10.1109/ICAIET65052.2025.11211380.
- [7] A. C. Munaro, R. H. Barcelos, E. Cristine, F. Maffezzolli, S. Rodrigues, and E. C. Paraiso, "To engage or not engage ? The features of video content on YouTube affecting digital consumer engagement," no. September 2020, pp. 1336–1352, 2021, doi: 10.1002/cb.1939.
- [8] H. Nitya, T. Prathima, and K. Sugamy, "Decoding YouTube's Trends: Unveiling Viral Content Secrets," in *2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0*, 2024, pp. 1–13. doi: 10.1109/OTCON60325.2024.10687676.
- [9] W. Hoiles, A. Aprem, and V. Krishnamurthy, "Engagement and Popularity Dynamics of YouTube Videos and Sensitivity to Meta-Data," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 7, pp. 1426–1437, 2017, doi: 10.1109/TKDE.2017.2682858.
- [10] R. P. Singh, A. Kumar, N. Niharika, G. Jain, R. Rastogi, and M. Jain, "Development of a Comprehensive YouTube Analytics Dashboard for Enhanced Performance Insights," in *2024 International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit)*, 2024, pp. 1089–1094. doi: 10.1109/GlobalAISummit62156.2024.10947784.