



---

## KLASIFIKASI VIDEO PADA MEDIA SOSIAL YOUTUBE DENGAN MENGGUNAKAN METODE K-MEANS DAN SUPPORT VECTOR MACHINE

Sari Krisdianto Sitanggang<sup>1</sup>, Fajri Rakhmat Umbara<sup>2</sup>, Herdi Ashaury<sup>3</sup>

Universitas Jenderal Achmad Yani Yogyakarta, Indonesia

akunkris1@gmail.com

---

### Abstrak:

Dalam era digital dan sosial media, platform seperti YouTube telah menjadi salah satu sumber utama video konten. Meningkatnya jumlah video di YouTube memunculkan kebutuhan untuk mengklasifikasikan video-video ini, baik untuk tujuan manajemen konten, rekomendasi, atau penegakan hukum. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan metode klasifikasi video yang efektif berdasarkan kontennya. Tujuan dari penelitian ini Mengembangkan sistem klasifikasi video yang dapat mengidentifikasi dan mengkategorikan video di platform media sosial YouTube. Mengimplementasikan metode k-Means dan Support Vector Machine (SVM) sebagai alat utama dalam proses klasifikasi. Meningkatkan akurasi dalam mengklasifikasikan video berdasarkan fitur-fitur kontennya. Meningkatkan pemahaman tentang penggunaan teknik Machine Learning dalam mengelola konten video di platform media sosial. Penelitian ini menggunakan metode k-Means untuk mengelompokkan video berdasarkan kesamaan fitur-fitur kontennya. Selanjutnya, Support Vector Machine (SVM) digunakan untuk melakukan klasifikasi berdasarkan hasil kelompok dari k-Means. Data yang digunakan dalam penelitian ini terdiri dari dataset video dari YouTube, dan fitur-fitur ekstrak yang relevan seperti tag, deskripsi, dan pemrosesan gambar. Proses eksperimen dan pengujian dilakukan untuk mengukur akurasi klasifikasi. Hasil penelitian ini menunjukkan bahwa kombinasi metode k-Means dan SVM dapat digunakan untuk mengklasifikasikan video pada platform media sosial YouTube dengan tingkat akurasi yang baik. Dengan menggunakan fitur-fitur konten yang relevan, sistem ini mampu mengenali kategori video dengan baik, yang bermanfaat untuk manajemen konten, analisis, dan rekomendasi. Kesimpulan dari penelitian ini Dalam era digital yang dipenuhi dengan konten video, penelitian ini menawarkan pendekatan yang efektif dalam mengklasifikasikan video di platform media sosial YouTube. Metode k-Means dan SVM berhasil digunakan untuk mengelompokkan dan mengklasifikasikan video dengan akurasi yang baik.

**Kata kunci:** Youtube; K-Means; SVM

### Abstract:

*In the digital and social media era, platforms like YouTube have become one of the primary sources of video content. The increasing number of videos on YouTube has created a need to classify these videos, for purposes such as content management, recommendations, or law enforcement. Therefore, this research aims to develop an effective video classification method based on their content. The objectives of this research are: Developing a video classification system that can identify and categorize videos on the social media platform YouTube. Implementing the k-Means and Support Vector Machine (SVM) methods as the primary tools in the classification process. Enhancing the accuracy of video classification based on their content features. Increasing understanding of the use of Machine Learning techniques in managing video content on social media platforms. This research utilizes the k-Means method to cluster videos based on the similarity of their content features. Subsequently, Support Vector Machine (SVM) is used to classify them based on the clusters created by k-Means. The data used in this research includes video datasets from YouTube, and relevant extracted features such as tags, descriptions, and image processing. Experimental processes and testing are conducted to measure classification accuracy. The results of this research demonstrate that the combination of the k-Means and SVM methods can be utilized to classify videos on the social media platform YouTube*

*with a high level of accuracy. By employing relevant content features, this system can effectively recognize video categories, which is valuable for content management, analysis, and recommendations. In conclusion, in an era filled with video content, this research offers an effective approach to classify videos on the social media platform YouTube. The k-Means and SVM methods have successfully been used to cluster and classify videos with good accuracy.*

**Keywords:** Youtube; K-Means; SVM

Corresponding: Sari Krisdianto Sitanggang

E-mail: akunkris1@gmail.com



## PENDAHULUAN

Internet pada abad ke 21 dimanfaatkan dalam segala bidang. Mulai dari media sosial, game online, chatting online, shopping online, transportasi online dan masih banyak lagi (Mico, 2021). Orang-orang menggunakan internet tidak hanya karena kebutuhan tapi karena ingin mengisi waktu luang. Salah satunya dengan mengunjungi aplikasi youtube. Berdasarkan survey yang dilakukan oleh Asosiasi Penyelenggara Jasa Internet Indonesia atau APJII tahun 2022 kuartal 1, 63,02% media sosial yang sering digunakan adalah aplikasi youtube (Rofit Fatahillah, 2023).

Setiap pengguna youtube dapat membuat channel sendiri yang memungkinkan pengguna untuk mengunggah video. Pengguna harus membuat konten video yang menarik agar pengguna lain mempunyai rasa ingin tahu dan menonton video yang di-upload (Iskandara & Nadhifab, 2021). Ketika menonton konten video di youtube, pengguna merespon video tersebut dengan memberikan bermacam-macam komentar. Beragamnya komentar tekstual yang disampaikan pengguna membuat kreator kesulitan menentukan jenis video yang diminati pengguna. Sehingga perlu adanya analisis sentimen untuk menganalisis komentar yang beragam untuk mencari informasi penting dibalik komentar setiap pelanggan. Analisis sentimen merupakan komputasional dari opini orang lain, pemberian nilai, serta emosi yang terdapat dalam entitas, event, dan atribut yang dimiliki (Fauzi, n.d.).

Studi tentang analisis sentimen dapat memberi informasi berharga. Analisis sentimen di jejaring sosial, seperti twitter, youtube atau facebook, telah menjadi sarana yang kuat untuk belajar tentang pendapat pengguna. Analisis sentimen adalah metode menganalisis data untuk mengetahui bagaimana perasaan orang. Analisis sentimen dapat dibagi menjadi 3 tugas, yaitu pengenalan teks informasi, ekstraksi informasi, dan klasifikasi sentimental (deteksi emosi, polaritas) (Fernanda, 2022).

Penelitian terdahulu cukup banyak menggunakan Metode Support Vector Machine untuk melakukan klasifikasi, seperti penelitian untuk klasifikasi Ujaran Kebencian pada Media Sosial Twitter mendapat akurasi perhitungan 93% (Rahman, Abdillah, & Komarudin, 2021). Klasifikasi untuk artikel hoax mendapat akurasi 95.8333% (Maulina & Sagara, 2018). Klasifikasi kualitas pengelasan Shield Metal Arc Welding menunjukkan hasil akurasi sebesar 98% (Ritonga & Purwaningsih, 2018). Klasifikasi jenis pantun mendapat nilai akurasi 81,91% (Irmada & Astriratma, 2020). Klasifikasi pada penyakit tuberkulosis memperoleh nilai akurasi 98% (Darsyah, 2014).

Klasifikasi penyakit daun kentang berdasarkan fitur tekstur dan fitur warna mendapat nilai akurasi 80% (Fikry, 2018). Klasifikasi data akreditasi sekolah dasar di kabupaten magelang mendapat nilai akurasi 93,902% (Octaviani, Wilandari, & Ispriyanti, 2014). Klasifikasi Kepribadian Pengguna Twitter, untuk mengklasifikasi kepribadian ekstrover dan introver dengan akurasi sebesar 88,89% (Fikry, 2018). Klasifikasi Data Malaria mendapatkan skor akurasi sebesar 92,3% (Ramadhan & Khoirunnisa, 2021). Klasifikasi SMS spam mendapat nilai akurasi 98,33% (Setiyono & Pardede, 2019).

Penelitian sebelumnya tidak menggunakan algoritma K-Means dalam proses klustering video. Kebaruan penelitian ini adalah penggunaan K-Means, yang merupakan varian yang lebih efisien dan akurat, untuk meningkatkan kualitas klustering video. Dengan K-Means, proses klustering dapat

menghasilkan kluster yang lebih representatif, mengakomodasi variasi yang lebih besar dalam jenis konten video di YouTube (Suryaningsih, Sihwi, & Sulisty, n.d.).

Berdasarkan 10 penelitian sebelumnya yang menggunakan SVM untuk klasifikasi, ditemukan bahwa tingkat akurasi yang dihasilkan cukup tinggi, dengan rata-rata di atas 80%. Metode Support Vector Machine (SVM) telah terbukti berhasil dalam berbagai penelitian klasifikasi, menunjukkan tingkat akurasi yang tinggi dalam berbagai bidang seperti analisis sentimen, klasifikasi dokumen, dan pengenalan pola. Kombinasi antara klusterisasi dengan K-Means dan klasifikasi dengan SVM menjadi pendekatan yang menjanjikan untuk mengklasifikasikan video pada platform YouTube. Tantangan-tantangan spesifik muncul dalam mengklasifikasikan video di YouTube. Keragaman jenis video, variasi metadata, dan pemahaman berbeda terhadap konten oleh pengguna menjadi faktor-faktor yang perlu dipertimbangkan. Oleh karena itu, penelitian ini akan merinci bagaimana teknik pengumpulan dan pengolahan data video diambil dari YouTube, serta bagaimana metode K-Means dan SVM dapat diterapkan untuk mengklasifikasikan konten video secara efektif (Kulsum, Jajuli, & Sulistiyowati, 2022).

## **METODE PENELITIAN**

Penelitian ini dimulai dengan pengumpulan dataset dari website Kaggle, yang terdiri dari 8053 rekaman data dari API YouTube. Tahap selanjutnya adalah preprocessing, di mana data dibersihkan, dipilih, dan diubah menjadi format yang siap diuji. Data dibagi menjadi data latih (70%) dan data uji (30%). Model SVM digunakan untuk melakukan klasifikasi non-linear dengan kernel Linear. Model ini dilatih dengan data latih, yang memungkinkan SVM memahami pola dan keterkaitan antara fitur-fitur dan label dalam data latih. Proses selanjutnya adalah pengumpulan dataset dari website Kaggle, yang terdiri dari 8053 rekaman data dari API YouTube.

## **HASIL DAN PEMBAHASAN**

### **1. Hasil K-Means Clustering**

Setelah semua proses dilakukan maka didapatkan hasil dari klusterisasi dapat dilihat pada tabel, berikut:

**Tabel 1 Hasil dari Klusterisasi**

<b>Kluster</b>	<b>Data</b>
<b>0</b>	<b>1712</b>
<b>1</b>	<b>31</b>
<b>2</b>	<b>1</b>

Dapat dilihat bahwa didapatkan hasil clustering yang sangat tidak seimbang dengan jumlah data yang signifikan dalam kluster 0 dan jumlah yang sangat kecil dalam cluster 1 dan cluster 2.

### **2. Hasil Klasifikasi Dengan SVM**

Berikut merupakan hasil yang didapatkan dalam proses klasifikasi dengan support vector machine:

**Tabel 2 Proses Klasifikasi**

<b>Kelas</b>	<b>precision</b>	<b>recall</b>	<b>f1-score</b>	<b>support</b>
0	0.83	0.62	0.71	8
1	0.99	1.00	1.00	516
accuracy			0.99	524

Kelas	precision	recall	f1-score	support
macro avg	0.91	0.81	0.86	524
weighted avg	0.99	0.99	0.99	524

Hasil evaluasi menunjukkan bahwa model SVM yang telah dibangun memiliki performa yang sangat baik, dan ini dapat disimpulkan dari beberapa metrik evaluasi yang digunakan:

**Akurasi (Accuracy) 0.99**

Akurasi mengukur sejauh mana model mampu mengklasifikasikan data dengan benar. Dalam kasus ini, akurasi sebesar 0.99 berarti model mampu mengklasifikasikan sebagian besar data uji dengan benar, dengan tingkat kesalahan yang sangat rendah (hanya 1% kesalahan).

**Presisi (Precision), Recall, dan F1-score 0.99**

Presisi mengukur kemampuan model untuk mengidentifikasi kelas yang benar dari prediksi positifnya, sedangkan Recall mengukur kemampuan model untuk mengidentifikasi seberapa banyak dari kelas yang sebenarnya berhasil diidentifikasi. F1-score adalah penggabungan dari presisi dan recall, memberikan gambaran keseluruhan tentang keseimbangan antara kedua metrik ini.

Dalam kasus ini, presisi, recall, dan F1-score memiliki nilai yang sangat tinggi (0.99), menunjukkan bahwa model mampu mengidentifikasi kelas yang benar dengan sangat baik dan hampir tidak membuat kesalahan positif. Selain itu, model juga mampu mengidentifikasi sebagian besar kelas yang benar, dengan sangat sedikit kesalahan dalam mengklasifikasikan data negatif.

Hasil ini menunjukkan bahwa model SVM yang dibangun sangat handal dan memiliki performa yang sangat baik dalam mengklasifikasikan data uji. Dengan akurasi yang tinggi dan presisi, recall, serta F1-score yang mendekati 1, model ini sangat cocok untuk tugas klasifikasi yang telah diterapkan dalam penelitian ini. Model ini memiliki keseimbangan yang baik antara kemampuan untuk mengidentifikasi kelas yang benar dan menghindari kesalahan klasifikasi, sehingga dapat diandalkan untuk aplikasi praktis dalam masalah yang berkaitan dengan klasifikasi data.

**KESIMPULAN**

Penelitian ini bertujuan untuk mengklasifikasikan video di media sosial YouTube menggunakan metode K-Means dan Support Vector Machine (SVM). Metode K-Means digunakan untuk mengelompokkan video berdasarkan kesamaan fitur, sedangkan SVM digunakan untuk membangun model klasifikasi (Kusumah, Chandranegara, & Nuryasin, 2023). Dalam penelitian ini, penulis berhasil mengimplementasikan kedua metode tersebut untuk mengklasifikasikan video. Hasil eksperimen menunjukkan bahwa kombinasi K-Means dan SVM dapat memberikan hasil klasifikasi yang akurat dan efektif. Proses klasifikasi dimulai dengan ekstraksi fitur dari setiap video menggunakan metode yang relevan. Fitur-fitur ini kemudian digunakan sebagai input dalam algoritma K-Means untuk mengelompokkan video ke dalam beberapa kelompok berdasarkan kesamaan fitur. Setelah itu, SVM digunakan untuk membangun model klasifikasi dengan menggunakan data latih yang telah diberi label. Hasil evaluasi menunjukkan bahwa metode yang diusulkan mampu mengklasifikasikan video dengan tingkat akurasi yang baik. Pada proses klusterisasi dengan K-mean dimana K=3, didapatkan nilai akurasi sebesar 99%. Hal ini menunjukkan bahwa penggunaan kombinasi K-Means dan SVM dapat menjadi pendekatan yang efektif untuk klasifikasi video pada media sosial YouTube.

## **DAFTAR PUSTAKA**

- Darsyah, Moh Yamin. (2014). Klasifikasi Tuberkulosis Dengan Pendekatan Metode Supports Vector Machine (SVM). *Jurnal Statistika Universitas Muhammadiyah Semarang*, 2(2).
- Fauzi, Raffie Rizky. (n.d.). *Analisis sentimen dampak ekonomi masyarakat Indonesia akibat pandemi covid-19 pada media sosial twitter menggunakan metode naïve bayes classifier, support vector machine dan lexicon*. Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta.
- Fernanda, Paternus Adita Resky. (2022). *Analisis Sentimen Publik Terhadap Ibadah Online Di Masa Pandemi Pada Media Twitter Menggunakan Metode Sentistrength*. Universitas Atma Jaya Yogyakarta.
- Fikry, Muhammad. (2018). Ekstrover atau Introver: Klasifikasi Kepribadian Pengguna Twitter dengan Menggunakan Metode Support Vector Machine. *SITEKIN: Jurnal Sains, Teknologi Dan Industri*, 16(1), 72–76.
- Irmanda, Helena Nurramdhani, & Astriratma, Ria. (2020). Klasifikasi Jenis Pantun Dengan Metode Support Vector Machines (SVM). *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(5), 915–922.
- Iskandara, Trias Pyrenia, & Nadhifab, Rubby Ariza Fadhlani. (2021). Fenomena penggunaan youtube channel pada anak usia dini di masa pandemic covid-19. *Jurnal Ilmiah LISKI (Lingkar Studi Komunikasi) Vol*, 7(2).
- Kulsum, Ummi, Jajuli, Mohamad, & Sulistiyowati, Nina. (2022). Analisis Sentimen Aplikasi WETV di Google Play Store Menggunakan Algoritma Support Vector Machine. *Journal of Applied Informatics and Computing*, 6(2), 205–212.
- Kusumah, Elsandy Wirahadi, Chandranegara, Didih Rizki, & Nuryasin, Ilyas. (2023). Pengembangan Sistem Informasi Geografis Pemetaan daerah Prioritas Tujuan Vaksin Berbasis Website. *Jurnal Impresi Indonesia*, 2(3), 236–245.
- Maulina, Dina, & Sagara, Rofie. (2018). Klasifikasi artikel hoax menggunakan support vector machine linear dengan pembobotan term frequency–Inverse document frequency. *Jurnal Mantik Penusa*, 2(1).
- Mico, Sastra. (2021). Pengaruh Promosi Online Terhadap Keputusan Pembelian Jasa Transportasi Online Grab di Kota Pagar Alam. *Jurnal Aktiva: Riset Akuntansi Dan Keuangan*, 3(1), 29–43.
- Octaviani, Pusphita Anna, Wilandari, Yuciana, & Ispriyanti, Dwi. (2014). Penerapan Metode Klasifikasi Support Vector Machine (SVM) pada Data Akreditasi Sekolah Dasar (SD) di Kabupaten Magelang. *Jurnal Gaussian*, 3(4), 811–820.
- Rahman, Oryza Habibie, Abdillah, Gunawan, & Komarudin, Agus. (2021). Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 17–23.

Ramadhan, Nur Ghaniaviyanto, & Khoirunnisa, Azka. (2021). Klasifikasi Data Malaria Menggunakan Metode Support Vector Machine. *Jurnal Media Informatika Budidarma*, 5(4), 1580–1584.

Ritonga, Alven Safik, & Purwaningsih, Endah Supeni. (2018). Penerapan Metode Support Vector Machine (SVM) Dalam Klasifikasi Kualitas Pengelasan Smaw (Shield Metal Arc Welding). *Jurnal Ilmiah Edutic: Pendidikan Dan Informatika*, 5(1), 17–25.

Rofit Fatahillah, Mohamad. (2023). *Youtube Sebagai Stimulus Media Sosial Dalam Pengambilan Keputusan Pembelian (Analisis Youtube Channel HSR\_Wheel)*. Universitas Islam Sultan Agung Semarang.

Setiyono, Agus, & Pardede, Hilman F. (2019). Klasifikasi Sms Spam Menggunakan Support Vector Machine. *Jurnal Pilar Nusa Mandiri*, 15(2), 275–280.

Suryaningsih, Vera, Sihwi, Sari Widya, & Sulisty, Meiyanto Eko. (n.d.). Analisis Clustering Dokumen Menggunakan Algoritma Self-Organizing Map (SOM)(Studi Kasus: Dokumen Skripsi Di Fakultas Pertanian UNS). *Kata Pengantar*.