

# Klasifikasi Popularitas Video Roblox Berbasis *K-Means* dan *Support Vector Machine* pada Metadata YouTube

Wayan Praka<sup>\*1</sup>, Arief Hermawan<sup>2</sup>, Donny Avianto<sup>3</sup>

<sup>1,2</sup>Magister Teknologi Informasi, Fakultas Pascasarjana, Universitas Teknologi Yogyakarta

<sup>3</sup>Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Teknologi Yogyakarta

e-mail: <sup>\*1</sup> wayanpraka123@gmail.com, <sup>2</sup>ariefdb@uty.ac.id, <sup>3</sup>donny@uty.ac.id

## Abstrak

Pertumbuhan metadata YouTube yang masif menyebabkan kreator konten Roblox kesulitan menentukan strategi pengembangan konten karena penilaian popularitas masih bergantung pada jumlah penayangan dan belum didukung oleh label popularitas yang objektif. Kondisi ini menyulitkan proses analisis dan pengambilan keputusan, padahal popularitas video dipengaruhi oleh berbagai metrik interaksi. Penelitian ini mengusulkan pendekatan hybrid menggunakan algoritma *K-Means* sebagai sistem pelabelan otomatis dan serta memanfaatkan algoritma *Support Vector Machine* (SVM) guna memprediksi tingkat popularitas video. Dataset yang digunakan sebanyak 1.500 metadata video Roblox yang dikumpulkan melalui YouTube Data API v3 dengan atribut views, likes, comments, dan durasi video. Berdasarkan metode Elbow, diperoleh tiga kluster optimal yang merepresentasikan kategori popularitas Tinggi, Sedang, dan Rendah. Melalui pengujian, integrasi kernel Radial Basis Function pada SVM mampu menghasilkan tingkat akurasi sebesar 99%, yang berarti model mampu mengklasifikasi tingkat popularitas video ke dalam tiga kategori tersebut dengan tingkat ketepatan yang tinggi. Dibuktikan melalui nilai Macro Average F1-Score mencapai 0,98 yang mengindikasikan ketangguhan model meskipun distribusi kelas tidak seimbang. Selain itu, video dengan durasi sekitar 150 detik cenderung berada pada kategori popularitas tinggi. Temuan ini membuktikan efisiensi penggabungan *K-Means* dan SVM dalam melakukan pengelompokan popularitas video secara otomatis serta dapat menjadi dasar rekomendasi bagi kreator Roblox dalam mengoptimalkan jangkauan audiens.

**Kata kunci**— YouTube, Roblox, *K-Means*, *Support Vector Machine*, Video.

## 1. PENDAHULUAN

Platform video YouTube telah menjadi sumber data besar yang mencerminkan tren dan perilaku masyarakat digital, khususnya dalam komunitas *game* seperti Roblox. Masifnya unggahan video setiap harinya menghasilkan metadata yang kompleks, namun tidak disertai dengan label popularitas yang terstruktur. Penentuan tingkat keberhasilan sebuah konten seringkali masih didasarkan pada jumlah penonton, tanpa mempertimbangkan keterkaitan antar variabel metrik interaksi. Kondisi ini menunjukkan bahwa pengolahan data mentah dari media sosial memerlukan pendekatan data mining untuk mengekstraksi informasi yang lebih bermakna [1]. Di samping itu, tata kelola konten serta peningkatan performa sistem rekomendasi memerlukan adanya kategorisasi video secara terstruktur seiring dengan melonjaknya volume unggahan di platform YouTube [2]. Bagi kreator konten Roblox, kondisi tersebut menimbulkan tantangan dalam menentukan strategi pengembangan *channel* secara tepat. Banyak kreator pemula masih mengukur keberhasilan video hanya berdasarkan jumlah penayangan, padahal

popularitas suatu konten juga dipengaruhi oleh interaksi pengguna seperti jumlah *like*, *comment*, dan durasi video.

Beberapa penelitian terdahulu telah mengeksplorasi penggunaan algoritma *clustering* dan klasifikasi dalam analisis data digital. Algoritma *K-Means* telah banyak diimplementasikan untuk berbagai kebutuhan pengelompokan, seperti pemetaan topik *cyberbullying* [3][4], pengelompokan musik [5] [6], hingga segmentasi klien dan aktivitas pengguna [7][8]. Selain pada platform video, *K-Means* juga terbukti handal dalam identifikasi tingkat kebermanfaatan aplikasi berbasis opini publik [9], serta menentukan waktu posting yang optimal berdasarkan pola *engagement* [10]. Sementara itu, untuk keperluan pengelompokan pada data dengan tingkat kompleksitas tinggi dan variabel yang bervariasi, algoritma *Support Vector Machine* (SVM) terbukti menawarkan performa yang unggul. Penggunaan SVM telah mendominasi penelitian sentimen pada kolom analisis komentar *YouTube*, mulai dari topik politik [11][12][13], ulasan produk kecantikan [14] hingga isu sosial seperti kenaikan harga kebutuhan pokok dan sentimen sosial [15][16]. Bahkan pada data kebencian di media sosial lainnya [17].

Pada domain *Roblox*, penelitian mengenai analisis popularitas masih relatif terbatas. Penelitian yang dilakukan [18] mengembangkan sistem pengategorian ketenaran game *Roblox* lewat algoritma *K-Nearest Neighbor* (KNN) yang berbasis variabel numerik, termasuk metrik kunjungan, jumlah pemain aktif, serta rasio rating. Temuan mereka mengonfirmasi efektivitas KNN dalam memetakan popularitas permainan tersebut, yang mengindikasikan bahwa data pada ekosistem *Roblox* memiliki pola yang dapat dipelajari menggunakan pendekatan *machine learning*. Namun, penelitian tersebut berfokus pada klasifikasi popularitas *game* yang telah memiliki kategori kelas tertentu dan belum mengkaji popularitas video *Roblox* pada platform *YouTube* yang memiliki karakteristik interaksi pengguna berbeda, seperti jumlah penayangan, jumlah suka, jumlah komentar, dan durasi video [18]. Di sisi lain, penelitian [2] membuktikan keberhasilan penggabungan *K-Means* dan SVM dalam mengategorikan video *YouTube*. Pendekatan ini mengandalkan *K-Means* untuk proses pelabelan otomatis serta SVM guna mendongkrak ketepatan klasifikasinya. Di samping itu, implementasi arsitektur hibrida sejenis telah terbukti sukses untuk memprediksi risiko kecanduan media sosial pada siswa [19].

Meskipun berbagai penelitian telah menerapkan algoritma *clustering* dan klasifikasi untuk analisis data digital, masih terdapat beberapa kesenjangan penelitian yang perlu diperhatikan. Pertama, penelitian pada domain *Roblox* masih berfokus pada popularitas game *Roblox* dan belum banyak mengkaji popularitas video *Roblox* pada platform *YouTube* [18]. Kedua, sebagian besar penelitian terdahulu menggunakan pendekatan *supervised learning* yang memerlukan label kelas yang telah tersedia sebelumnya [18], sehingga proses pelabelan masih bergantung pada pengetahuan atau penilaian awal peneliti. Ketiga, penelitian terkait klasifikasi popularitas konten digital umumnya menerapkan metode klasifikasi tunggal, sementara integrasi antara *clustering* untuk pembentukan label otomatis dan klasifikasi untuk meningkatkan performa model masih relatif terbatas [2] [19]. Kondisi ini memunculkan celah pembeda antara literatur sebelumnya dengan usulan penelitian saat ini, khususnya pada aspek objek penelitian yang beralih dari *game Roblox* menjadi video *Roblox* di *YouTube*, aspek pelabelan yang dibentuk secara otomatis melalui *clustering*, aspek metode yang mengintegrasikan *K-Means* dan SVM, serta aspek tujuan yang berfokus pada klasifikasi popularitas video dibandingkan popularitas *game*.

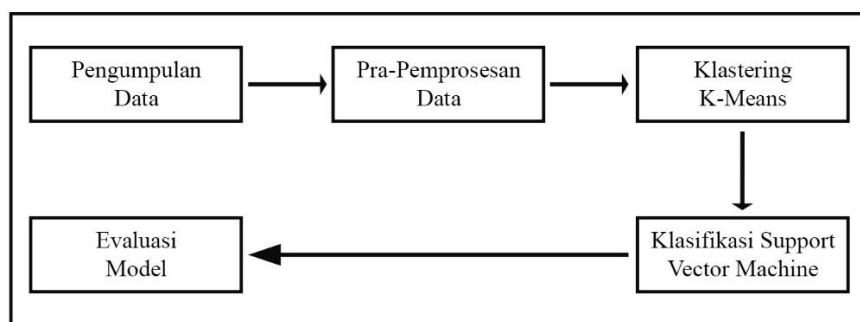
Berdasarkan kesenjangan tersebut, penelitian ini mengusulkan penggunaan *K-Means* sebagai mekanisme pelabelan otomatis dan serta menerapkan SVM selaku algoritma penguji dalam memetakan tingkat popularitas video *Roblox* berdasarkan metrik interaksi pengguna. Pemilihan *K-Means* didasarkan pada kemampuannya mengelompokkan data tanpa memerlukan label awal sehingga mampu menghasilkan segmentasi popularitas yang lebih objektif. Sementara itu, penggunaan SVM didasarkan pada keunggulannya dalam mengonstruksi *hyperplane* terbaik. Konstruksi ini sangat andal ketika dipadukan dengan karakteristik data yang memiliki

keberagaman dimensi dan batas kelas yang kompleks. Selain itu, penelitian sebelumnya juga merekomendasikan pengembangan metode klasifikasi lain seperti SVM guna memperoleh model yang lebih komprehensif [18].

Nilai keterbaruan yang ditawarkan dalam riset ini mencakup tiga aspek penting, di mana poin pertama berfokus pada penggunaan video *Roblox* pada *platform YouTube* sebagai objek penelitian, yang hingga saat ini masih jarang dibahas dalam penelitian klasifikasi popularitas. Kedua, penerapan pendekatan *hybrid K-Means* dan SVM yang menggabungkan *unsupervised learning* dan *supervised learning* dalam satu alur pemrosesan data. Ketiga, pembentukan label popularitas secara otomatis melalui proses *clustering* sebelum dilakukan klasifikasi, sehingga menghasilkan kategori popularitas yang lebih objektif dibandingkan penggunaan label yang telah ditentukan sebelumnya. Dengan memanfaatkan fitur *view\_count*, *like\_count*, *comment\_count*, dan *duration\_sec*, penelitian ini bertujuan untuk mengklasifikasikan tingkat popularitas video *Roblox* ke dalam kategori rendah, sedang, dan tinggi, serta mengungkap pola keterlibatan pengguna yang memengaruhi popularitas konten. Penelitian ini diharapkan mampu memberikan sumbangan ilmiah bagi pengembangan teknik klasifikasi yang memanfaatkan pendekatan data mining sekaligus menjadi dasar rekomendasi bagi kreator *Roblox* dalam menyusun strategi pengembangan konten yang lebih efektif untuk mengoptimalkan jangkauan *audiens*.

## 2. METODE PENELITIAN

Metode eksperimen berbasis data digunakan untuk mengimplementasikan pendekatan kuantitatif dalam riset ini. Kerangka kerja Data Mining yang diusulkan terbagi menjadi empat tahapan inti, yaitu pengumpulan data, pra-pemrosesan data, Klastering algoritma K-Means, dilanjutkan dengan klasifikasi *Support Vector Machine*, serta evaluasi model yang skemanya dipaparkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

### 2.1. Pengumpulan Data

Dataset diperoleh melalui *scraping* menggunakan *YouTube Data API v3* dengan kata kunci sebagai berikut: *Roblox Indonesia*, *Roblox Gameplay*, *Roblox Horror*, *Roblox Blox Fruit*, *Roblox Simulator*, *Roblox Fun moments*, *Roblox Review*. Sebanyak 1.500 baris data dikumpulkan, mencakup fitur numerik berupa *view\_count*, *like\_count*, *comment\_count*, dan *duration\_sec*. Pemilihan fitur ini didasarkan pada relevansinya terhadap indikator popularitas konten digital, di mana jumlah penayangan, suka, komentar, dan durasi video dapat merepresentasikan tingkat keterlibatan *audiens* terhadap suatu video. Sebagian data yang berhasil dikumpulkan disajikan pada Tabel 1.

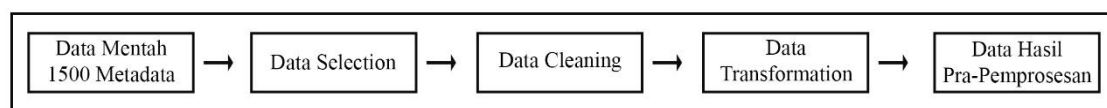
Tabel 1. Sampel Dataset Video Roblox di YouTube

| No | Judul Video                                                                      | View Count | Like Count | Comment Count | Duration (Detik) |
|----|----------------------------------------------------------------------------------|------------|------------|---------------|------------------|
| 1  | MIMIN KAGET NGELIAT AKUN BOCIL!                                                  | 2.997.533  | 24.117     | 923           | 100              |
| 2  | MISSING BACON: The Tower - Roblox Indonesia                                      | 7.1905     | 930        | 114           | 12.690           |
| 3  | Gali Tanah Dan Kabur Dari Sekolah?!! [Roblox Indonesia]                          | 23.951     | 726        | 189           | 7520             |
| 4  | Yuta, Mio Dan Baby Celine Kabur Dari Penjara Barry Perosotan    ROBLOX           | 5.040.718  | 31.226     | 550           | 12.500           |
| 5  | MAKE OVER GEMBEL JADI PRESIDEN?! - Roblox Indonesia Build a Millionaire #2       | 14.425     | 945        | 175           | 6.510            |
| 6  | Gara Gara Liatin Si Razen [Roblox Indonesia] #brorrr #roblox #funnyroblox        | 26.052     | 98         | 2             | 200              |
| 7  | Gw kira beneran, ternyata #roblox #pov #robloxshorts #robloxindonesia            | 2.974.356  | 43.152     | 988           | 180              |
| 8  | Game Roblox Ini Harganya 14 JUTA?                                                | 10.627.181 | 212.678    | 897           | 320              |
| 9  | Kamu Liat Tidak Guys? Coba Komen [Roblox Indonesia] #brorrr #roblox #funnyroblox | 1.478      | 13         | 0             | 130              |
| 10 | LOMBA DI TOWER KUCING OYEN #robloxindonesia #roblox #shorts #leika               | 11.831.664 | 125.660    | 935           | 310              |

Berdasarkan sampel pada Tabel 1, terlihat bahwa dataset memiliki variasi yang cukup besar pada setiap atribut. Jumlah penayangan video berkisar dari ribu hingga puluhan juta, sementara jumlah suka dan komentar juga menunjukkan perbedaan yang signifikan antar video. Variasi tersebut menunjukkan bahwa dataset memiliki karakteristik yang sesuai untuk digunakan dalam proses pengelompokan dan klasifikasi popularitas video.

## 2.2. Pra-pemrosesan Data

Tahap prapemrosesan data dilakukan melalui serangkaian proses untuk memastikan data berada dalam kondisi bersih, konsisten, dan layak digunakan pada tahap analisis serta pemodelan. Pertama, *Data Selection* dilakukan untuk memfilter dan mengambil fitur numerik spesifik, yaitu *views*, *likes*, *comments*, dan *duration\_sec*. Kedua, *Data Cleaning* diterapkan untuk memastikan tidak terdapat *missing values* atau duplikasi data pada 1.500 sampel. Ketiga, *Data Transformation* dilakukan melalui teknik *Min-Max Scaling* untuk menormalisasikan rentang nilai seluruh fitur ke dalam skala 0 hingga 1. Langkah ini penting untuk mencegah fitur yang memiliki rentang nilai besar, seperti *views*, memberikan pengaruh yang lebih dominan dibandingkan fitur dengan nilai relatif kecil, seperti *duration*, pada proses perhitungan jarak dalam algoritma *K-Means* [19]. Tahapan *pre-processing* dipaparkan pada Gambar 2.



Gambar 2. Alur Prosedur Pra-pemrosesan Data

## 2.3. Klastering K-Means

Algoritma *K-Means Clustering* digunakan sebagai metode *unsupervised learning* untuk membentuk kelompok data yang memiliki karakteristik serupa berdasarkan kesamaan nilai fitur yang dianalisis. Pada penelitian yang dilakukan, Algoritma *K-Means* dipakai untuk mengelompokkan video berdasarkan kemiripan karakteristik pada atribut *view\_count*, *like\_count*,

*comment\_count*, dan *duration\_sec*. Untuk mendapatkan hasil pengelompokan yang optimal, jumlah kluster ditetapkan berdasarkan analisis metode *Elbow* yang mempertimbangkan karakteristik data yang ada. Setelah proses klasterisasi selesai, karakteristik setiap kluster dianalisis untuk menentukan kategori tingkat popularitas video, sehingga diperoleh label kelas yang merepresentasikan pola data. Pada tahap berikutnya, label yang diperoleh dari proses klasterisasi dijadikan sebagai variabel target untuk membangun model klasifikasi menggunakan SVM [5][19].

#### 2.4. Klasifikasi Support Vector Machine

Setelah tahap klasterisasi selesai dan setiap video memperoleh label tingkat popularitas, langkah berikutnya membangun model klasifikasi dengan algoritma *Support Vector Machine* (SVM). SVM berfungsi membagi data menjadi kelas berbeda dengan cara menentukan hyperplane terbaik. Pendekatan ini dipilih karena mampu mengatasi data berdimensi tinggi dengan efektif serta memiliki performa klasifikasi yang konsisten dan akurat. Dataset yang telah dilengkapi label hasil klasterisasi dengan membagi menjadi dua bagian, yaitu 80% menjadi data latih dan 20% menjadi data uji. Data latih digunakan untuk membentuk model klasifikasi, sedangkan data uji dimanfaatkan untuk evaluasi kemampuan model saat mengklasifikasikan data yang belum digunakan selama proses pelatihan. Dengan tahapan ini, diharapkan model bisa mempelajari pola hubungan antara jumlah *views*, jumlah *likes*, jumlah *comments*, serta durasi video terhadap kategori popularitas video *Roblox* di platform *YouTube* [14][19].

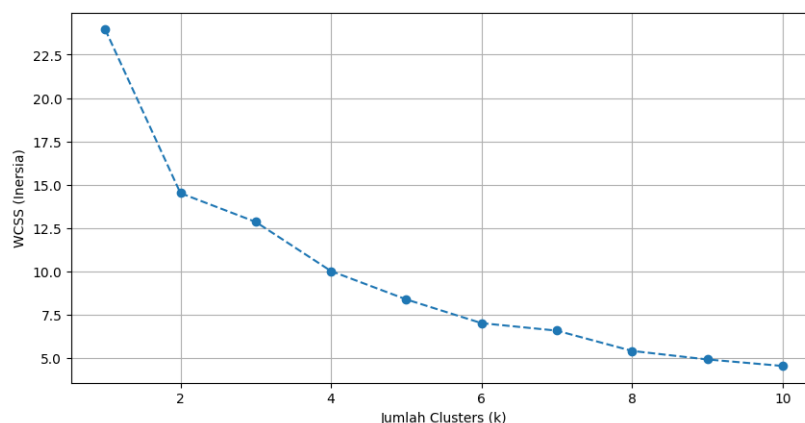
#### 2.5. Evaluasi Model

Pada evaluasi model ini diterapkan untuk menilai kinerja model yang dihasilkan pada proses klasterisasi menggunakan *K-Means* maupun klasifikasi pada *Support Vector Machine* (SVM). Metode *K-Means*, pengujian dilakukan dengan menerapkan *Elbow Method* untuk mengidentifikasi jumlah kluster yang paling optimal berdasarkan karakteristik data yang dianalisis [19]. Selanjutnya, performa dari model SVM dievaluasi memakai *confusion matrix* yang menghasilkan beberapa metrik pengukuran, yaitu *accuracy*, *recall*, *precision*, dan *F1-score*. Metrik *accuracy* berfungsi untuk menunjukkan proporsi prediksi yang benar terhadap keseluruhan data pengujian. *Precision* untuk mengukur performa model dalam memberikan prediksi pada setiap kategori, sedangkan *recall* berguna untuk menilai kinerja model dalam mengidentifikasi semua data termasuk kelas tertentu. Adapun *F1-score* merupakan acuan yang menggabungkan *precision* dan *recall* sehingga menggambarkan yang seimbang terhadap performa model. Hasil evaluasi ini dipakai untuk menetapkan tingkat kinerja model *hybrid K-Means* dan SVM dalam mengklasifikasikan popularitas video *Roblox* secara akurat [14], [19].

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Hasil Klustering K-Means

Tahapan yang memiliki signifikansi tinggi dalam pelaksanaan algoritma *K-Means* sebagai metode *unsupervised learning* adalah proses penentuan jumlah kluster (*k*) yang optimal dan mampu merepresentasikan struktur data secara akurat. Pemilihan nilai *k* dilakukan secara objektif dengan memanfaatkan metode *Elbow*, Parameter *Within-Cluster Sum of Squares* (WCSS) digunakan sebagai metrik evaluasi demi menilai hasil klasterisasi, sebagaimana dipaparkan dalam Gambar 3.

Gambar 3. Grafik Metode *Elbow* Jumlah Kluster Optimal

Merujuk pada Gambar 3, pemilihan jumlah kluster  $k=3$  dilakukan karena pada titik tersebut menunjukkan penurunan nilai *Within-Cluster Sum of Squares* (WCSS), yang paling drastis sebelum kurva mulai melandai secara konsisten yang menandakan tercapainya titik siku. Pemilihan  $k=3$  didasarkan pada prinsip optimasi di mana pembentukan lebih dari tiga kelompok tidak lagi memberikan reduksi variansi internal yang signifikan secara proporsional terhadap kompleksitas model, sehingga tiga kluster dianggap sebagai jumlah paling efisien untuk memisahkan data ke dalam kategori popularitas Rendah, Sedang, dan Tinggi tanpa risiko *overfitting*. Penggunaan titik  $k=3$  ini memberikan struktur label yang stabil bagi algoritma *Support Vector Machine* (SVM). Setelah menentukan jumlah kluster optimal, dilakukan analisis terhadap nilai rata-rata dari masing-masing kluster untuk mendefinisikan label popularitas. Distribusi anggota dan karakteristik tiap kluster yang dipaparkan dalam Tabel 2.

Tabel 2. Karakteristik Tiap Kluster

| Kluster | Jumlah Data | View_count | Like_count | Comment_count | Duration_count | Label  |
|---------|-------------|------------|------------|---------------|----------------|--------|
| 0       | 1425        | 4.211.224  | 54.593     | 1.905         | 542            | Sedang |
| 1       | 55          | 70.124.260 | 927.537    | 13.280        | 150            | Tinggi |
| 2       | 20          | 2.242.304  | 20.498     | 1.570         | 8.659          | Rendah |

Berdasarkan Tabel 2, di mana Kluster 1 popularitas tinggi menunjukkan performa interaksi paling dominan dengan rata-rata views mencapai 70,1 juta dan 927 ribu *likes*, serta memiliki karakteristik durasi video paling singkat yaitu sekitar 150 detik atau sekitar 2,5 menit yang mengindikasikan bahwa konten viral cenderung disajikan secara singkat, padat, dan efisien. sementara itu Kluster 0 popularitas sedang merupakan kelompok mayoritas dengan 1.425 data yang merepresentasikan performa umum video *Roblox* dengan rata-rata durasi sekitar 542 detik dan tingkat interaksi berada pada level menengah dengan rata-rata 4,2 juta views. di sisi lain, Kluster 2 popularitas rendah memiliki tingkat interaksi paling rendah meskipun durasi videonya paling panjang yaitu 8.659 detik, yang menunjukkan bahwa video berdurasi sangat panjang seperti *live streaming* tanpa pengeditan atau konten mentah cenderung memiliki retensi dan minat penonton yang lebih rendah dibandingkan video yang telah melalui proses editing pada durasi video.

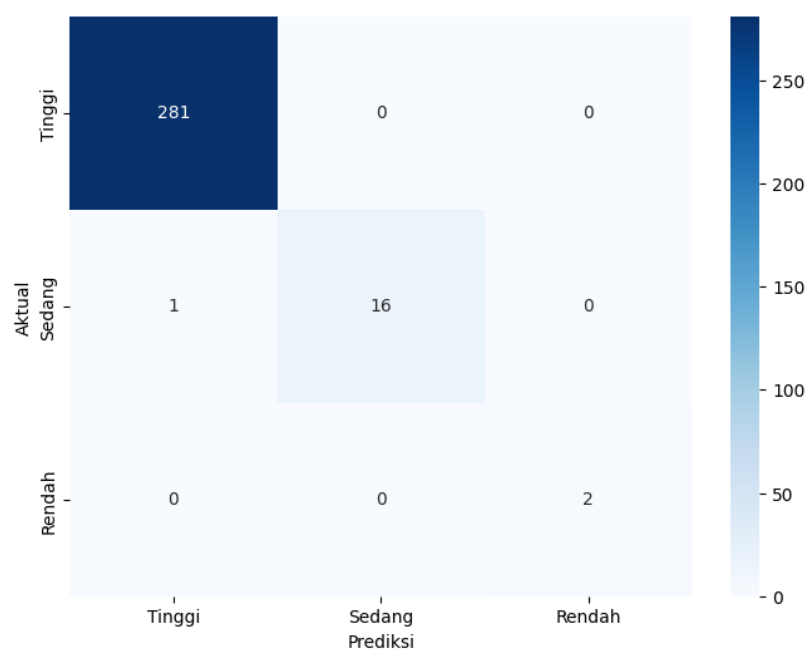
### 3.2. Hasil Klasifikasi *Support Vector Machine* (SVM)

Setelah melalui tahap pelabelan otomatis menggunakan *K-Means*, data diuji menggunakan algoritma *Support Vector Machine* (SVM) dengan mempelajari pola label popularitas yang telah dibentuk oleh metode *K-Means*. Skema klasifikasi ini dijalankan dengan

memisahkan keseluruhan data menggunakan rasio 80% dialokasikan menjadi data latih sedangkan 20% sisanya diposisikan menjadi data uji. Melalui pengelompokan tersebut, dari total keseluruhan 1.500 data, diperoleh 300 sampel data uji yang difungsikan untuk mengukur sejauh mana kapabilitas generalisasi model saat menerapkan kernel *Radial Basis Function* (RBF). Penjelasan mendalam mengenai performa dan capaian evaluasi dari model ini dipaparkan secara detail dalam Tabel 3 dan Gambar 4.

Tabel 3. Hasil Performa Kinerja Model *Support Vector Machine* (SVM)

| Kategori     | Precision | Recall | F1-Score | Support |
|--------------|-----------|--------|----------|---------|
| Tinggi       | 0.99      | 1.00   | 0.99     | 281     |
| Sedang       | 1.00      | 0.94   | 0.96     | 17      |
| Rendah       | 1.00      | 1.00   | 1.00     | 2       |
| Akurasi      |           |        | 0.99     | 300     |
| Macro avg    | 0.99      | 0.98   | 0.98     | 300     |
| Weighted avg | 0.99      | 0.99   | 0.99     | 300     |



Gambar 4 Visualisasi *Confusion Matrix*

Berdasarkan hasil yang disajikan pada Tabel 3, model SVM dapat menghasilkan kinerja klasifikasi sangat baik disertai dengan akurasi sebesar 99%. Meskipun penyebaran data tidak seimbang, sehingga kelas Tinggi lebih dominan sebanyak 281 dari total 300 data uji, nilai *Macro Average F1-Score* sebesar 0,98 menunjukkan performa model tetap konsisten disemua kategori. Keandalan model teruji melalui nilai *Precision* sempurna yaitu 1.00 pada kelas minoritas Sedang dan Rendah, yang menegaskan bahwa model memiliki sensitivitas tinggi tanpa mengalami bias terhadap kelas mayoritas. Meskipun kategori Sedang menjadi kelas yang paling menantang untuk diprediksi dengan nilai *Recall* 0,94 akibat adanya ambiguitas data pada area transisi antara tingkat popularitas sedang dan tinggi, pembentukan *hyperplane* menggunakan kernel RBF terbukti mampu memetakan batas keputusan secara stabil. Hal ini mengonfirmasi bahwa integrasi K-

*Means* sebagai sistem pelabelan otomatis telah menghasilkan densitas klaster yang *solid*, sehingga model SVM yang dihasilkan sangat reliabel untuk digunakan sebagai instrumen prediksi otomatis performa konten *Roblox* di *YouTube* secara saintifik. Dari hasil *confusion matrix* pada Gambar 4, model berhasil mengidentifikasi seluruh 281 data pada kategori Tinggi serta 2 data pada kategori Rendah secara tepat tanpa kesalahan klasifikasi. Kesalahan klasifikasi hanya terjadi pada satu data kategori Sedang yang diprediksi sebagai kategori Tinggi. Jumlah kesalahan yang sangat rendah tersebut menunjukkan bahwa SVM mampu membentuk batas keputusan yang efektif dalam memisahkan karakteristik setiap tingkat popularitas berdasarkan fitur *metadata YouTube* yang digunakan.

Berdasarkan keseluruhan proses analisis yang telah dilakukan, pada penelitian ini menghasilkan temuan penting yang memberikan kontribusi praktis. Temuan ini mengungkapkan bahwa nilai popularitas konten *Roblox* tidak hanya ditentukan oleh akumulasi *views*, melainkan lebih dominan dipengaruhi oleh durasi video yang efisien dan interaksi aktif *audiens*. Melalui analisis karakteristik klaster, teridentifikasi bahwa klaster dengan popularitas tertinggi justru memiliki rata-rata durasi paling singkat, yakni 150 detik, sementara video dengan durasi ekstrem mencapai 8.659 detik secara konsisten menempati klaster popularitas terendah meskipun memiliki potensi jangkauan yang luas. Insight utama bagi kreator konten adalah bahwa strategi optimasi *engagement* lebih efektif dicapai melalui produksi konten yang padat informasi dengan fokus pada pemicu *comment* dan *like*, karena tingginya interaksi dua arah pada video berdurasi pendek terbukti menjadi diskriminator utama yang memisahkan konten viral dari konten dengan retensi rendah. Hal ini memberikan manfaat nyata bagi praktisi pemasaran digital dan kreator untuk memprioritaskan efisiensi durasi guna memaksimalkan algoritma rekomendasi *platform YouTube*.

#### 4. KESIMPULAN

Berdasarkan penelitian yang dilaksanakan berhasil mengimplementasikan pendekatan *hybrid* menggunakan algoritma *K-Means Clustering* dan *Support Vector Machine (SVM)* untuk mengklasifikasikan popularitas konten video *Roblox* pada *platform YouTube* secara objektif. Melalui metode *Elbow*, integrasi *unsupervised learning* terbukti efektif dalam memberikan label otomatis yang stabil pada *metadata* yang tidak terstruktur, berhasil membentuk tiga kategori popularitas yaitu Tinggi, Sedang, dan Rendah dengan karakteristik klaster yang terdefinisi secara jelas. Hasil pengujian membuktikan model SVM melalui kernel *Radial Basis Function (RBF)* memiliki performa sangat handal memperoleh akurasi mencapai 99%. Disisi lain nilai *Macro Average F1-Score* sebesar 0,98 mengindikasikan model memiliki kemampuan yang konsisten dalam mengklasifikasikan seluruh kategori popularitas serta tetap mempertahankan tingkat presisi yang tinggi meskipun terdapat ketidakseimbangan distribusi data antar kelas. Temuan penelitian ini juga mengungkapkan bahwa determinan utama popularitas bukan hanya terletak pada akumulasi *views*, melainkan pada efisiensi durasi video. Dimana video dengan durasi rata-rata 150 detik secara konsisten mendominasi kategori popularitas tertinggi dibandingkan video berdurasi panjang yang memiliki retensi penonton lebih rendah. Secara praktis, hasil ini memberikan kontribusi nyata bagi kreator dalam merancang strategi produksi video yang padat informasi dan interaktif guna memaksimalkan algoritma rekomendasi *YouTube* secara sistematis.

#### 5. SARAN

Penelitian mendatang dapat memperluas fitur dengan menambahkan variabel umur video, jumlah subscriber, dan Engagement Rate untuk menghasilkan prediksi yang lebih komprehensif. Selain itu, penerapan algoritma ensemble learning seperti Random Forest dan XGBoost dapat dikaji sebagai perbandingan terhadap model yang digunakan, mengingat efektivitasnya pada data tabular berdimensi tinggi. Untuk analisis tren berbasis waktu, arsitektur Deep Learning seperti



Artificial Neural Network (ANN) dapat menjadi alternatif guna merepresentasikan hubungan non-linear antar variabel secara lebih akurat.

#### DAFTAR PUSTAKA

- [1] F. Khalish, N. M. Piranti, and O. Martadireja, "Implementasi Data Mining Menggunakan Teknik Clustering dengan Metode K-Means," *Jiip (Jurnal Ilmiah Ilmu Pendidikan)*, vol. 8, no. 5, pp. 5392–5397, May 2025, [Online]. Available: <http://Jiip.stkipyapisdampu.ac.id>
- [2] S. Krisdianto Sitanggang, F. Rakhmat Umbara, and H. Ashaury, "Klasifikasi Video Pada Media Sosial Youtube Dengan Menggunakan Metode K-Means Dan Support Vector Machine," *LOCUS: Penelitian dan Pengabdian*, vol. 2, no. 10, pp. 1027–1032, Oct. 2023, doi: 10.58344/locus.v2i10.1732.
- [3] A. R. Lahitani, A. N. Zhafarina, N. S. W. Oktavia, and N. Jariyah, "Pemetaan Topik Pembicaraan Pada Komentar Live Youtube Menggunakan K-Means Clustering sebagai Identifikasi awal Kejahatan Verbal Cyberbullying," *JTE UNIBA*, vol. 8, no. 2, pp. 399–403, Apr. 2024, [Online]. Available: <https://www.youtube.com/shorts/RABxiBvgxpU>.
- [4] V. P. Ramadhan and G. M. Namung, "Klasterisasi Komentar Cyberbullying Masyarakat di Instagram berdasarkan K-Means Clustering," *J-INTECH (Journal of Information and Technology)*, pp. 32–39, Jun. 2023, [Online]. Available: <https://github.com/rizalespe/Dataset-Sentimen-Analisis-Bahasa-Indonesia>
- [5] B. Hakim, F. J. Kaunang, C. Susanto, J. Salim, and R. Indradjaja, "Implementasi Machine Learning Dalam Pengelompokan Musik Menggunakan Algoritma K-Means Clustering," *Idealis: Indonesia Journal Information System*, vol. 8, no. 1, pp. 74–83, Jan. 2025, [Online]. Available: <http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/indexBhustomyHakim><http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index>
- [6] N. Kusmawanti, N. Suarna, and W. Prihartono, "Pengelompokkan Lagu Tradisional Di Media Sosial Tiktok Menggunakan Algoritma K-Means," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 3, pp. 2962–2967, Jun. 2024, [Online]. Available: <https://forms.gle/JjnNMj4cWhsuwaQE9>.
- [7] A. B. Krisna and D. M. Putri, "Analisis Segmentasi Klien Menggunakan K-Means Clustering Pada Agensi Kreatif Anoorapro," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, pp. 1931–1938, Apr. 2025.
- [8] M. N. Afrilia, N. Rahaningsih, R. D. Dana, and N. D. Nuris, "Optimasi Analisis Clustering Untuk Aktivitas Dan Respon Pengguna Media Sosial Dengan K-Means," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 148–155, Feb. 2024, [Online]. Available: <https://www.kaggle.com/>.
- [9] Kharisma and U. S. Aesyi, "Analisis Tingkat Kebermanfaatan Mypertamina Menggunakan K-Means Clustering," *JOISM (Journal of Information System Management)*, vol. 4, no. 2, pp. 91–96, 2023.
- [10] Fathoni, N. A. Basulina, S. Gustiani, S. Amalia, A. Sabila, and A. Ibrahim, "Identifikasi Waktu Optimal Posting Terhadap Pola Engagement Sosial Media Menggunakan Metode

- K-Means Clustering,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 4, pp. 6237–6244, Aug. 2025.
- [11] E. Tohidi, R. P. Herdiansyah, E. Wahyudin, and Kaslani, “Analisa Sentimen Komentar Video Youtube Di Channel Tvonenews Tentang Calon Presiden Prabowo Subianto Menggunakan Support Vector Machine,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 660–667, Feb. 2024, [Online]. Available: [https://www.youtube.com/watch?v=wu\\_wqHg1YLk](https://www.youtube.com/watch?v=wu_wqHg1YLk)
  - [12] S. A. Rahmadhani, L. D. Rusanti, and H. Al Rosyid, “Klasifikasi Sentimen Komentar Youtube Demonstrasi DPR RI Menggunakan Support Vector Machine,” *Arcitech: Journal of Computer Science and Artificial Intelligence*, vol. 5, no. 2, pp. 357–375, Dec. 2025, doi: 10.29240/arcitech.v5i2.15316.
  - [13] D. Prastyo, I. H. Mursyidin, and D. Irawan, “Klasifikasi Sentimen Komentar YouTube dengan NLP pada Debat Pilkada Banten 2024,” *Bit-Tech (Binary Digital - Technology)*, vol. 7, no. 2, pp. 413–421, Dec. 2024, doi: 10.32877/bt.v7i2.1833.
  - [14] R. T. Adek, Z. Fitri, and S. C. Siregar, “Analisis Sentimen Komentar Pada Saluran Youtube Beauty Vlogger Berbahasa Indonesia Menggunakan Metode Support Vector Machine,” *Jurnal Algoritme*, vol. 5, no. 2, pp. 164–175, Apr. 2025, doi: 10.35957/algoritme.v5i2.9692.
  - [15] R. Asrianto and M. Herwinanda, “Analisis sentimen kenaikan harga kebutuhan pokok dimedia sosial youtube menggunakan algoritma support vector machine,” *CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 431–440, Dec. 2022, doi: 10.37859/coscitech.v3i3.4368.
  - [16] Hidayat, F. Santoso, and L. F. Lidimillah, “Analisis Sentimen Pengguna YouTube Tentang Rohingya Menggunakan Algoritma SVM (Support Vector Machine),” *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 3, pp. 1729–1738, Jun. 2024, doi: 10.33379/gtech.v8i3.4497.
  - [17] O. H. Rahman, G. Abdillah, and A. Komarudin, “Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine,” *RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 17–23, Feb. 2021, doi: 10.29207/resti.v5i1.2700.
  - [18] G. A. Buana, H. Firmansyah, and W. Asriyani, “Analisis Klasifikasi Popularitas Game Roblox Menggunakan Algoritma K-Nearest Neighbor (KNN),” *Jurnal Ilmu Multidisiplin*, vol. 2, no. 1, pp. 1206–1214, Dec. 2026, doi: 10.63822/0dwtj19.
  - [19] S. Saputra, M. F. Hidayat, W. Gunawan, and I. R. Rahadjeng, “Klasifikasi Dampak Kecanduan Media Sosial Mahasiswa Dengan Svm Dan K-Means,” *Journal of Information System Management (JOISM)*, vol. 7, no. 2, pp. 283–291, 2026.