Jurnal Informatika dan Tekonologi Komputer

E-ISSN: 2809-9230
P-ISSN: 2809-9249

Klasifikasi Konten *Thumbnail* TikTok untuk Deteksi Kata Kasar Menggunakan *Support Vector Machine* (SVM)

Muhammad Fathurrahman[1]*, Ryan Ari Setiawan [2], and Fatsyarina Fitriastuti [3]

[1] Universitas Janabadra; Kota Yogyakarta, Daerah Istimewa Yogyakarta; e-mail : fathurrahmanmuhammad32@gmail.com
[2] Universitas Janabadra; Kota Yogyakarta, Daerah Istimewa Yogyakarta; e-mail : ryan@janabadra.ac.id
[3] Universitas Janabadra; Kota Yogyakarta, Daerah Istimewa Yogyakarta; e-mail : fitri@janabadra.ac.id
* Corresponding Author : Muhammad Fathurrahman

Abstract: The rapid growth of the social media platform TikTok has introduced new challenges in content moderation, particularly in detecting offensive language that appears not only in text form but also in visual elements such as video thumbnails. This study aims to develop a classification model capable of detecting offensive content in TikTok thumbnails using the Support Vector Machine (SVM) algorithm. Data were collected through web scraping of 4,153 TikTok videos containing offensive elements, which were then processed and manually labeled into 24 classes of offensive words. The dataset was divided into training and testing sets with a ratio of 20:80. Model performance was evaluated using AUC, accuracy, precision, recall, F1-Score, and Matthews Correlation Coefficient (MCC). The results show that the SVM model achieved an AUC of 0.791, indicating a reasonably good ability to distinguish between classes. However, accuracy (0.340), precision (0.293), recall (0.340), F1-Score (0.298), and MCC (0.264) indicate that the classification performance remains low. These findings suggest the need to improve Preprocessing quality, select more representative visual features, and develop more advanced classification methods. This research contributes to expanding the detection approach of harmful content from text-based to visual-based domains and lays the groundwork for more comprehensive automated content moderation systems in the future.

Keywords: TikTok; thumbnail; offensive language; visual classification; Support Vector Machine (SVM)

Abstrak: Pertumbuhan pesat platform media sosial TikTok telah menghadirkan tantangan baru dalam moderasi konten, terutama dalam mendeteksi ujaran kasar yang muncul tidak hanya dalam bentuk teks tetapi juga dalam elemen visual seperti *thumbnail* video. Penelitian ini bertujuan untuk mengembangkan model klasifikasi yang mampu mendeteksi konten kasar pada *thumbnail* TikTok menggunakan algoritma *Support Vector Machine* (SVM). Data dikumpulkan melalui proses web scraping terhadap 4.153 video TikTok yang mengandung unsur kata kasar, kemudian diproses dan dilabeli secara manual ke dalam 24 kelas kata kasar. Data selanjutnya dibagi menjadi data pelatihan dan pengujian dengan rasio 20:80. Evaluasi performa model dilakukan menggunakan metrik AUC, akurasi, presisi, *recall*, *F1-Score*, dan *Matthews Correlation Coefficient* (MCC). Hasil menunjukkan bahwa model SVM memiliki nilai AUC sebesar 0.791, menandakan kemampuan cukup baik dalam membedakan antara kelas. Namun, nilai akurasi (0.340), presisi (0.293), *recall* (0.340), *F1-Score* (0.298), dan MCC (0.264) menunjukkan performa klasifikasi yang masih rendah. Hal ini mengindikasikan perlunya peningkatan kualitas *Preprocessing*, pemilihan fitur visual yang lebih representatif, serta pengembangan metode klasifikasi yang lebih canggih. Penelitian ini memberikan kontribusi dalam memperluas pendekatan deteksi konten negatif dari berbasis teks ke domain visual, serta menjadi dasar bagi pengembangan sistem moderasi konten otomatis yang lebih komprehensif di masa mendatang.

Kata Kunci: TikTok; *thumbnail*; kata kasar; klasifikasi visual; *Support Vector Machine* (SVM)

Received: May 28, 2025
Revised: June 4, 2025
Accepted: July 18, 2025
Published: July 19, 2025
Curr. Ver.: July 19, 2025

Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/)

1. Pendahuluan

Seiring pesatnya pertumbuhan pengguna TikTok di Indonesia—yang kini menempati peringkat kedua dalam jumlah pengguna secara global—platform ini telah menjadi salah satu media sosial yang paling digemari, terutama oleh kalangan remaja dan dewasa muda. Fenomena ini mencerminkan pergeseran perilaku konsumsi konten digital masyarakat yang semakin intens, dengan TikTok bukan hanya digunakan untuk hiburan semata, tetapi juga sebagai medium ekspresi diri, interaksi sosial, bahkan promosi komersial. Akan tetapi, di tengah arus konten yang begitu deras, tidak sedikit pula ditemukan bentuk-bentuk interaksi negatif seperti ujaran kasar, perundungan siber (*cyberbullying*), dan konten ofensif lainnya [1]. Persoalan ini menjadi semakin kompleks ketika ujaran tersebut tidak hanya muncul dalam bentuk teks pada kolom komentar, tetapi juga tersirat dalam elemen visual seperti *thumbnail* video, yang sering kali diabaikan dalam proses moderasi konten.

Upaya deteksi konten negatif secara otomatis pun menjadi krusial, mengingat jumlah konten yang diunggah ke platform seperti TikTok setiap detiknya sangat besar. Salah satu pendekatan yang banyak digunakan dalam berbagai studi untuk mengklasifikasi konten digital adalah algoritma *Support Vector Machine* (SVM). SVM merupakan salah satu metode supervised learning yang telah terbukti efektif dalam menangani berbagai jenis permasalahan klasifikasi, termasuk dalam konteks analisis sentimen dan deteksi ujaran kebencian. Studi dalam [2] menunjukkan bahwa penerapan SVM dalam menganalisis sentimen komentar pada TikTok dapat mencapai akurasi hingga 84 %, menjadikannya sebagai salah satu algoritma yang cukup andal untuk tugas semacam ini. Sementara itu, [3] membandingkan performa SVM dengan algoritma Random Forest, dan hasilnya menunjukkan bahwa SVM mampu memperoleh akurasi hingga 90 % dalam mengklasifikasikan ulasan aplikasi TikTok, memperkuat validitas algoritma ini dalam ranah media sosial.

Penelitian lain juga menunjukkan hasil yang sejalan. Dalam [4], analisis terhadap komentar TikTok yang membahas produk skincare menunjukkan bahwa SVM masih unggul dibandingkan dengan Naïve Bayes, meskipun akurasi yang diperoleh sekitar 60 %. Ini menandakan bahwa SVM tetap relevan, terutama ketika dioptimalkan dengan teknik praproses data atau peningkatan kualitas fitur. Sementara itu, di luar TikTok, SVM juga telah diterapkan dalam mendeteksi ujaran kasar di media sosial lain seperti Twitter. Penelitian oleh [5] berhasil membangun sistem yang mampu mengidentifikasi ujaran kasar dalam bahasa Indonesia dengan akurasi yang sangat tinggi, yakni mencapai 99 %. Dalam [6], SVM dibandingkan kembali dengan *Naïve Bayes* dalam konteks deteksi *cyberbullying* di Twitter, dan hasilnya menunjukkan bahwa SVM lebih unggul dengan akurasi mencapai 88 %, memperlihatkan konsistensi performa SVM dalam mendeteksi konten negatif berbasis teks.

Berbagai penelitian juga mencoba menggabungkan metode SVM dengan pendekatan lain untuk meningkatkan akurasi. Misalnya, dalam [7], integrasi Word2Vec sebagai metode representasi kata dan SMOTE sebagai teknik penyeimbangan data digunakan untuk memperkuat performa SVM, dan terbukti berhasil meningkatkan akurasi hingga 88 %. Pendekatan hibrid seperti ini membuka peluang besar dalam pengembangan sistem klasifikasi yang lebih cerdas dan adaptif terhadap dinamika data media sosial. Di platform Instagram, studi [8] menggunakan SVM untuk mendeteksi komentar spam dan berhasil mencapai akurasi tinggi sebesar 96,8 %, menunjukkan bahwa algoritma ini juga fleksibel dalam mengklasifikasi jenis konten lain di luar ujaran kasar.

Studi lain yang berfokus pada ekosistem TikTok seperti dalam [9] membandingkan kinerja SVM dengan CNN dan *Naïve Bayes* dalam analisis sentimen pada ulasan Tokopedia Seller Center, yang berkaitan dengan penggunaan TikTok untuk promosi. Hasilnya kembali menunjukkan keunggulan SVM dengan akurasi 90 %. Dalam konteks analisis reaksi publik terhadap kebijakan penutupan TikTok Shop, penelitian [10] juga menggunakan SVM dan berhasil memperoleh akurasi sebesar 91,3 %, memperlihatkan bahwa SVM tidak hanya mampu mendeteksi sentimen umum, tetapi juga menginterpretasi opini publik terhadap isu-isu spesifik yang sedang tren.

Studi lain yang berfokus pada ekosistem TikTok seperti dalam [9] membandingkan kinerja SVM dengan CNN dan *Naïve Bayes* dalam analisis sentimen pada ulasan Tokopedia Seller Center, yang berkaitan dengan penggunaan TikTok untuk promosi. Hasilnya kembali menunjukkan keunggulan SVM dengan akurasi 90 %. Dalam konteks analisis reaksi publik terhadap kebijakan penutupan TikTok Shop, penelitian [10] juga menggunakan SVM dan berhasil memperoleh akurasi sebesar 91,3 %, memperlihatkan bahwa SVM tidak hanya mampu mendeteksi sentimen umum, tetapi juga menginterpretasi opini publik terhadap isu-isu spesifik yang sedang tren.

Berangkat dari berbagai temuan dan keterbatasan tersebut, penelitian ini bertujuan untuk mengembangkan model klasifikasi konten berbasis visual, yaitu *thumbnail* TikTok, guna mendeteksi adanya kata kasar atau isyarat kasar secara implisit. Dengan menggabungkan teknik ekstraksi fitur visual dan algoritma klasifikasi *Support Vector Machine*, penelitian ini diharapkan mampu mengisi celah dalam literatur yang selama ini lebih banyak berfokus pada teks. Selain itu, pendekatan ini juga mendukung pengembangan sistem moderasi otomatis yang lebih komprehensif dan adaptif terhadap bentuk-bentuk konten baru yang berkembang di media sosial. Dengan demikian, hasil penelitian ini tidak hanya diharapkan memberi kontribusi ilmiah, tetapi juga praktis dalam membangun ruang digital yang lebih sehat dan aman bagi penggunanya, khususnya di platform dengan pertumbuhan pesat seperti TikTok.

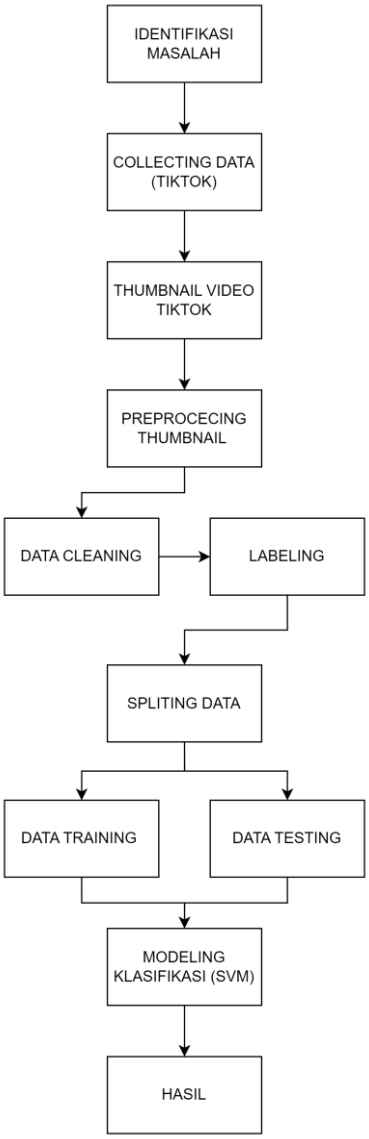
2. Metode yang Diusulkan

Penelitian ini menerapkan algoritma *Support Vector Machine* (SVM) sebagai metode utama dalam melakukan klasifikasi terhadap konten visual, thumbnile yang mewakili isi dari sebuah konten video di TikTok , guna mendeteksi adanya kata-kata kasar.

Ilustrasi 1 menggambarkan tahapan alur penelitian ini, dimulai dari proses identifikasi masalah, diikuti dengan pengumpulan data berupa *thumbnail* dari platform TikTok. Data yang diperoleh kemudian melalui tahapan *Preprocessing*. Berdasarkan konten visual tersebut, dilakukan proses pelabelan data secara manual ke dalam 24 kelas kata kasar.

Selanjutnya, data dibagi menjadi dua bagian, yaitu data training dan data testing, guna memastikan evaluasi model secara objektif agar distribusi data menjadi lebih seimbang sebelum proses pelatihan SVM. Setelah pelatihan, model diuji pada data yang belum pernah dilihat sebelumnya, dan performanya dievaluasi menggunakan metrik seperti akurasi, presisi, *recall*, dan *F1-Score*. Penelitian ini diakhiri dengan analisis terhadap seberapa efektif model SVM dalam mengenali konten kasar berbasis visual, serta potensi penerapannya dalam sistem moderasi konten otomatis pada platform media sosial seperti TikTok.

Penelitian ini menerapkan algoritma *Support Vector Machine* (SVM) untuk mengklasifikasi konten visual *thumbnail* video TikTok sebagai bentuk upaya pendeteksian kata-kata kasar. Metode ini serupa dengan penelitian Romindo et al. [11] juga menerapkan SVM berbasis TF-IDF untuk mendeteksi komentar *cyberbullying* di TikTok, mencapai akurasi 88% dan F1-score 92%.



Gambar 1. Alur Penelitian

2.1. Collecting Data

Pada penelitian ini, pengumpulan data dilakukan melalui teknik web scraping menggunakan pustaka *Python* seperti *BeautifulSoup* dan *Selenium*, serta aplikasi *TubeTrendy*. *Tube-Trendy* adalah sebuah dekstop apps yang dapat memudahkan seseorang untuk mengoptimasi kan konten yang dapat di kelola dari berbagai chanel . Pengguna dapat mengakses berbagai alat dan fitur yang dirancang untuk membantu dalam mengelola dan meningkatkan kinerja konten di platform-platform media sosial, aplikasi ini mungkin menyediakan analisis statistik, alat manajemen posting, atau fitur pencarian kata kunci untuk membantu pengguna dalam mengidentifikasi tren yang sedang populer di media sosial Menurut penelitian Ulfah & Najiah, kombinasi *Selenium* dan *BeautifulSoup* efektif untuk scraping website seperti Jurnal SINTA [12]. Penelitian serupa oleh Pratama dkk. menunjukkan bahwa *BeautifulSoup* memiliki kecepatan eksekusi rata-rata sekitar 1,93 detik, lebih tinggi dibanding *Selenium* (~3,59 detik) saat scraping situs dinamis [13]. *TubeTrendy* digunakan untuk mengambil metadata seperti profil pengguna, hashtag, dan postingan relevan, serta mendukung berbagai platform termasuk TikTok.

Dataset dikumpulkan antara 1 April 2024 hingga 9 Juni 2024, dan terdiri atas 4,153 *thumb-nail* video TikTok yang mengandung bahasa kasar. Semua data masih mentah dan belum men-jalani proses *Preprocessing* selanjutnya.

2.2. Preprocessing Data

Pada bagian ini terdiri dari beberapa tahapan normalisasi data yang perlu dilakukan agar data dapat lebih mudah dilakukan serangkaian tahapan normalisasi data agar data mentah yang diperoleh dapat diproses lebih mudah dan menghasilkan keluaran yang akurat. Tahapan *Pre-processing* sangat penting dalam pemrosesan data karena data yang dikumpulkan dari berbagai sumber umumnya belum siap digunakan secara langsung dalam pemodelan atau analisis. Proses ini mencakup beberapa langkah, antara lain:

2.2.1. *Data Cleaning*

Tahapan ini mencakup penghapusan gambar *thumbnail* yang tidak valid, blur, duplikasi, serta pengisian atau penghilangan missing value. Teknik ini sejalan dengan praktik umum dalam pengolahan teks dan citra, di mana *case folding, stopword removal, dan stemming* menunjukkan peningkatan akurasi klasifikasi sentimen [14]. Pada data visual (*thumbnail*), proses *cleaning* mencakup penghilangan noise dan penyesuaian ukuran gambar untuk menjamin integritas data sebelum ekstraksi fitur [15].

2.2.2. *Labelling*

Proses labelling dataset berupa tumbnile dan audio dari video yang di peroleh dan mendapatkan 24 class name. Berikut adalah Tabel 1. Class Name yang di temukan:

Tabel 1. Class Name

ASU	KERE	NDASMU
ANJING	KEPLE	PANTEK
BACOT	GENDENG	SILIT
BAJINGAN	ITIL	TELANJANG
JANCOK	TAI	TOBRUT
BANGSAT	ICLIK	TOLOL
BEDES	GOBLOK	GATHEL

Pelabelan dilakukan pada setiap video dianotasi ke dalam 24 kelas kata kasar — seperti ASU, KERE, NDASMU, ANJING, dan seterusnya. Pelabelan yang teliti sangat penting sebagai basis data berkualitas tinggi untuk melatih model klasifikasi secara akurat, sejalan dengan praktik umum dalam literatur klasifikasi ujaran kebencian berbasis teks/audio[16][17].

2.3. *Spliting* Data

Pada tahap *splitting data*, dataset yang telah melalui proses pra-pemrosesan dibagi menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*). Pembagian ini bertujuan untuk melatih model pada sebagian data, dan menguji akurasi serta performa model menggunakan data yang tidak dilatih sebelumnya. Dalam penelitian ini, digunakan rasio pembagian sebesar 20% untuk data latih dan 80% untuk data uji. Pemilihan rasio ini didasarkan pada pertimbangan efektivitas evaluasi model terhadap data yang belum dikenali, serta sesuai dengan pendekatan yang juga dilakukan dalam penelitian serupa sebelumnya[18].

2.4. *Support Vector Machine* (SVM)

Support Vector Machine (SVM) adalah metode pembelajaran terawasi yang bertujuan untuk mencari hyperplane yang dapat memisahkan dua atau lebih kategori data yang berlainan[19]. Prinsip utama SVM adalah membangun hyperplane dengan margin yang seimbang sehingga tidak terlalu dekat dengan salah satu kelas[20]. Berikut adalah rumus umum yang digunakan dalam perhitungan *Support Vector Machine* (SVM).

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{sign}\ (\boldsymbol{w}.\ \boldsymbol{x} + \boldsymbol{b}) \tag{1}$$

Keterangan:
$F(x)$ = Fungsi perkiraan
w = Vektor normal hyperplane
x = Vektor fitur masukan(input)

b = bias atau intercept

2.5. Perhitungan Model

Langkah selanjutnya adalah mengukur kinerja algoritma klasifikasi dengan menggunakan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, dan *F1-Score*. Metrik-metrik ini digunakan untuk membandingkan hasil prediksi dengan nilai aktual pada data uji.

True Positive (TP) = Jumlah dokumen dari kelas true yang benar diklasifikasikan sebagai kelas *true*.
True Negative (TN) = Jumlah dokumen yang berasal dari kelas true salah yang diklasifikasikan sebagai kelas *false*.
False Positive (FP) = Jumlah dokumen dari kelas false yang keliru diklasifikasikan sebagai kelas *false*.
False Negative (FN) = Total dokumen dari kategori true yang secara keliru dikelompokkan sebagai kategori *false*.

Rumus matematika untuk metrik *accuracy*, *precision*, *recall*, dan *F1-Score* ditampilkan di bawah ini:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FN}$$

$$Recall = \frac{TP}{TP + FP}$$

$$F1 - Score = 2 \, x \, \frac{recall \; x \; precision}{recall + precision}$$

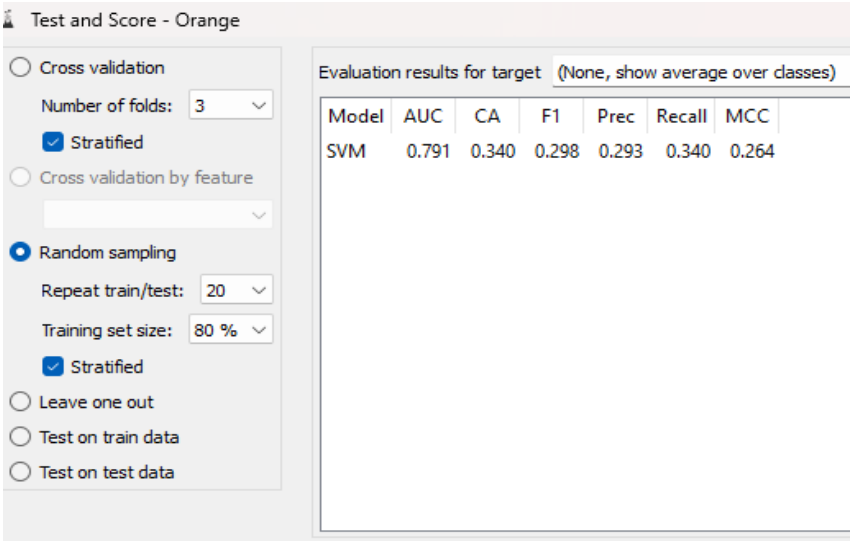
3. Hasil dan Pembahasan

Hasil evaluasi model *Support Vector Machine* (SVM) dalam penelitian ini menunjukkan performa yang bervariasi berdasarkan beberapa metrik pengukuran. Nilai *Area Under Curve* (AUC) sebesar 0.791 mengindikasikan bahwa model memiliki kemampuan yang cukup baik dalam membedakan antara kelas yang berbeda. AUC yang mendekati angka 1 menunjukkan bahwa model memiliki sensitivitas dan spesifisitas yang cukup tinggi dalam memisahkan kelas kasar dan tidak kasar pada konten *thumbnail* TikTok. Namun demikian, akurasi klasifikasi (*Classification Accuracy*/CA) yang hanya mencapai 0.340 menunjukkan bahwa proporsi prediksi benar terhadap keseluruhan data masih tergolong rendah. Hal ini mencerminkan bahwa dari seluruh prediksi yang dihasilkan, hanya sekitar 34% yang sesuai dengan label sebenarnya, yang berarti model belum cukup andal dalam melakukan klasifikasi secara keseluruhan.

Lebih lanjut, nilai F1 Score sebesar 0.298 memperlihatkan bahwa keseimbangan antara presisi dan *recall* belum optimal. Nilai ini menunjukkan bahwa meskipun terdapat beberapa prediksi benar, jumlah prediksi salah masih cukup tinggi sehingga berdampak pada penurunan performa keseluruhan. Hal ini juga diperkuat oleh nilai presisi (*Precision*) sebesar 0.293, yang menandakan bahwa hanya sekitar 29.3% dari seluruh prediksi positif yang benar-benar relevan. Dengan kata lain, sebagian besar prediksi positif yang dihasilkan model merupakan false positive. Selain itu, nilai *recall* sebesar 0.340 menunjukkan bahwa hanya 34% dari data positif sebenarnya yang berhasil dikenali oleh model, yang berarti masih banyak data positif yang tidak terdeteksi atau terklasifikasi secara salah. Kinerja model juga dievaluasi melalui *Matthews Correlation Coefficient* (MCC) dengan nilai sebesar 0.264. Nilai ini menggambarkan korelasi lemah antara hasil prediksi dan label sebenarnya, mengindikasikan bahwa model belum mampu menangkap pola klasifikasi secara konsisten dan akurat.

Secara keseluruhan, meskipun model SVM menunjukkan performa AUC yang cukup baik, nilai-nilai lain seperti akurasi, presisi, *recall*, F1 score, dan MCC masih menunjukkan bahwa kinerja model perlu ditingkatkan. Hal ini mengindikasikan perlunya perbaikan dalam

aspek *Preprocessing* data, pemilihan fitur, atau tuning parameter model agar dapat meningkatkan akurasi dan efektivitas klasifikasi kata kasar pada konten *thumbnail* TikTok secara lebih optimal.



Gambar 2. Hasil

4. Kesimpulan

Penelitian ini mengkaji penerapan algoritma *Support Vector Machine* (SVM) dalam melakukan klasifikasi konten berbasis visual, yaitu *thumbnail* video TikTok, untuk mendeteksi adanya kata-kata kasar. Dataset yang digunakan berasal dari hasil scraping video TikTok yang mengandung ujaran kasar dalam periode April hingga Juni 2024. Setelah melalui tahapan *Preprocessing* yang meliputi data *cleaning* dan pelabelan manual terhadap 24 kelas kata kasar, dilakukan pemisahan data menjadi data pelatihan dan data pengujian dengan rasio 20:80. Evaluasi terhadap model dilakukan menggunakan metrik AUC, akurasi, presisi, *recall*, *F1-Score*, dan MCC.

Hasil evaluasi menunjukkan bahwa model SVM memiliki nilai AUC sebesar 0.791, yang menunjukkan kemampuan model dalam membedakan kelas kasar dan tidak kasar tergolong cukup baik. Namun, metrik lainnya seperti Classification *Accuracy* (0.340), *Precision* (0.293), *Recall* (0.340), *F1-Score* (0.298), dan *Matthews Correlation Coefficient* (0.264) mengindikasikan bahwa performa model secara keseluruhan masih rendah dan perlu ditingkatkan. Akurasi dan *F1-Score* yang rendah menunjukkan bahwa model belum optimal dalam menyeimbangkan antara presisi dan *recall*, sehingga berisiko tinggi menghasilkan false positive maupun *false negative* dalam jumlah signifikan.

Temuan ini mengindikasikan bahwa meskipun SVM memiliki potensi dalam klasifikasi visual berbasis *thumbnail*, efektivitasnya masih sangat bergantung pada kualitas *Preprocessing*, representasi fitur visual yang digunakan, serta proporsi data pelatihan yang memadai. Untuk itu, pengembangan sistem klasifikasi yang lebih akurat di masa mendatang dapat dilakukan dengan memperluas jumlah dan variasi data, menerapkan teknik augmentasi gambar, memperkaya representasi fitur menggunakan model *deep learning* seperti CNN, serta menerapkan teknik balancing data seperti SMOTE.

Secara umum, penelitian ini memberikan kontribusi awal dalam memperluas ranah deteksi ujaran kasar dari analisis berbasis teks ke domain visual. Hal ini membuka peluang besar untuk pengembangan sistem moderasi otomatis yang lebih menyeluruh dan adaptif di platform media sosial, khususnya TikTok, guna menciptakan ekosistem digital yang lebih sehat dan aman bagi seluruh penggunanya.

Daftar Pustaka

[1] M. Attar Jibran, A. Eviyanti, dan Y. Findawati, "Deteksi Ujaran Kebencian Menggunakan Metode *Support Vector Machine* (SVM)," KESATRIA: Jurnal Penerapan Sistem Informasi, vol. 4, no. 4, pp. 890–900, Okt. 2023. doi: 10.30645/kesatria.v4i4.239

[2] W. Ayu, R. Abdulhakim, Y. Umaidah, dan J. H. Jaman, "Optimasi *Support Vector Machine* Berbasis Particle Swarm Optimization Untuk Mendeteksi Hate Speech Pilkada Karawang," J. Appl. Informatics Comput., vol. 5, no. 2, pp. 190–201, 2021. DOI: 10.30871/jaic.v5i2.3473

[3] L. P. A. S. Tjahyanti, "Pendeteksian Bahasa Kasar (Abusive Language) Dan Ujaran Kebencian (Hate Speech) Dari Komentar Di Jejaring Sosial," J. Chem. Inf. Model., vol. 7, no. 9, pp. 1689–1699, 2020.

[4] W. A. Luqyana, I. Cholissodin, dan R. S. Perdana, "Analisis sentimen *cyberbullying* pada komentar instagram dengan metode klasifikasi *Support Vector Machine*," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 2, no. 11, pp. 4704–4713, 2018.

[5] A. Abdurohim et al., "Penerapan SVM untuk Klasifikasi Komentar Spam di Instagram," Jurnal JURTIK, vol. 2, no. 1, pp. 13–20, 2024.

[6] H. Saputra et al., "Klasifikasi Analisis Sentimen Menggunakan Word2Vec, SMOTE dan *Support Vector Machine*," Jurnal INSTEK, vol. 8, no. 2, pp. 157–166, 2022. DOI: 10.24252/instek.v10i1.54889

[7] E. Indriyani et al., "Analisis Sentimen Komentar Pengguna TikTok Menggunakan Algoritma SVM," Jurnal Teknologi dan Informatika, vol. 6, no. 1, pp. 19–27, 2023.

[8] S. Tangke, A. Syaiful, dan A. Karim, "Analisis Sentimen Komentar Pengguna Aplikasi TikTok Menggunakan SVM dan Random Forest," Jurnal TIMES, vol. 12, no. 1, pp. 45–51, 2024.

[9] R. C. Liem et al., "Klasifikasi Komentar Produk Skincare di TikTok Menggunakan SVM dan Naïve Bayes," Jurnal AICoMS, vol. 3, no. 1, pp. 88–95, 2024.

[10] A. Kurniawan dan M. Al Qorni, "Analisis Sentimen terhadap Penutupan TikTok Shop Menggunakan SVM," Jurnal MATRIK, vol. 25, no. 2, pp. 91–98, 2023.

[11] Romindo, J. J. P., & Barus, O. P. (2023). Implementasi Algoritma TF-IDF dan *Support Vector Machine* terhadap Analisis Pendeteksi Komentar *Cyberbullying* di Media Sosial TikTok. Device, 13(1).

[12] Ulfah & Najiah, "Implementasi Web Scraping pada Situs Jurnal SINTA menggunakan *Python*, *Selenium*, dan *BeautifulSoup*," J. Informatika Indones., vol. 7, no. 1, pp. 29–36, Feb. 2023.

[13] K. Pratama et al., "Implementasi Teknik Web Scraping dan Fitur Data Eksternal pada …," Transient, Universitas Diponegoro, 2021.

[14] A. Tri Jaka H., "*Preprocessing* Text untuk Meminimalisir Kata yang Tidak Berarti dalam Analisis Sentimen," J. Bus. Audit Inf. Syst., vol. 4, no. 2, pp. 16–22, 2021.

[15] D. S. Y. Kartika dan H. Maulana, "Preprosesing serta Normalisasi pada Dataset Kupu-Kupu untuk Ekstraksi Fitur Warna, Bentuk dan Tekstur," *Complete: Journal of Computer, Electronic, and Telecommunication*, vol. 1, no. 2, pp. 1–8, 2019, doi: 10.52435/complete.v1i2.76.

[16] A. Rizky, M. H. Saputra, dan A. I. Nugroho, "Penerapan Deep Learning untuk Klasifikasi Teks Ujaran Kebencian pada Media Sosial Twitter," Jurnal Informatika dan Rekayasa Perangkat Lunak, vol. 3, no. 1, pp. 45–52, 2021.

[17] L. P. Handayani dan R. K. Wardhani, "Klasifikasi Ucapan Kata Dengan *Support Vector Machine* (SVM)," Jurnal Masyarakat Informatika, vol. 3, no. 6, 2020. doi: 10.14710/jmasif.3.6.8456

[18] F. R. A. Harianto, Z. Alawi, dan I. A. Sa'ida, "Pengaruh Komposisi Split Data pada Akurasi Klasifikasi Penderita Diabetes Menggunakan Algoritma Machine Learning," Jurnal Sistem Informasi dan Informatika (Simika), vol. 8, no. 1, pp. 36–44, Jan. 2025. doi: 10.47080/simika.v8i1.3663

[19] P. K. Handayani, "Penerapan Algoritma *Support Vector Machine* (Svm) Untuk Analisis Pola Klasifikasi Pada Parkinson'S Dataset," Indones. J. Technol. Informatics Sci., vol. 3, no. 1, pp. 31–35, 2021, doi:10.24176/ijtis.v3i1.7530

[20] W. Zhong and L. Du, "Predicting Traffic Casualties Using *Support Vector Machine*s with Heuristic Algorithms: A Study Based on Collision Data of Urban Roads," Sustain., vol. 15, no. 4, 2023, doi: 10.3390/su15042944