Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi
https://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/index
Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru
Loktabat – Banjarbaru (Tlp. 0511 4782881), e-mail: puslit.stmikbjb@gmail.com
e-ISSN: 2685-0893

Klasifikasi Risiko Keamanan Siber Pada Ulasan Pengguna Tiktok Dengan Mengunakan *Support Vector Machine*

DOI: http://dx.doi.org/10.35889/jutisi.v15i1.3484

Jaya Sandra[1*], RG Guntur Alam[2],Gunawan[3]
Sistem Informasi, Universitas Muhammadiyah Bengkulu, Bengkulu, Indonesia
*e-mail *Corresponding Author:* jayasandra572004@gmail.com

Abstract

The development of social media platforms such as TikTok encourages users to openly share experiences, complaints, and opinions, including issues related to cybersecurity risks. The large volume of user reviews makes manual analysis inefficient, thereby requiring a computational approach for automatic risk classification. This study aims to classify cybersecurity risks in TikTok user reviews using the Support Vector Machine (SVM) algorithm. The research method includes data collection, text preprocessing based on Natural Language Processing, risk labeling, and classification using SVM. Model evaluation was conducted using two data split scenarios, namely 80:20 and 70:30. The experimental results show that SVM achieves consistent performance with an accuracy above 93%, particularly in classifying the No Risk and Medium Risk categories. However, the model shows limitations in detecting the High Risk category due to class imbalance. The classification results can be used as an initial basis for identifying cybersecurity risks in social media user reviews.

Keywords: *Cybersecurity; TikTok; Text Mining; Support Vector Machine; Classification*

Abstrak

Perkembangan platform media sosial seperti TikTok mendorong pengguna untuk menyampaikan pengalaman, keluhan, dan opini secara terbuka, termasuk yang berkaitan dengan risiko keamanan siber. Banyaknya ulasan yang dihasilkan menyebabkan analisis manual menjadi tidak efisien, sehingga diperlukan pendekatan komputasional untuk melakukan klasifikasi risiko secara otomatis. Penelitian ini bertujuan untuk mengklasifikasikan risiko keamanan siber pada ulasan pengguna TikTok menggunakan algoritma *Support Vector Machine* (SVM). Metode penelitian meliputi pengumpulan data ulasan, *preprocessing* teks berbasis *Natural Language Processing*, pelabelan risiko, serta proses klasifikasi menggunakan SVM. Evaluasi model dilakukan dengan dua skenario pembagian data, yaitu 80:20 dan 70:30. Hasil pengujian menunjukkan bahwa SVM mampu memberikan kinerja yang konsisten dengan tingkat akurasi di atas 93%, terutama dalam mengklasifikasikan kategori Tidak Berisiko dan Risiko Menengah. Namun, performa model pada kategori Risiko Tinggi masih terbatas akibat ketidakseimbangan distribusi data. Hasil klasifikasi ini dapat dimanfaatkan sebagai dasar identifikasi awal risiko keamanan siber pada ulasan media sosial.

Kata kunci*: Keamanan Siber; TikTok; Text Mining; Support Vector Machine; Klasifikas*

1. Pendahuluan

Perkembangan teknologi informasi dan komunikasi telah mendorong meningkatnya penggunaan media sosial sebagai sarana interaksi, hiburan, dan pertukaran informasi. Di balik manfaat tersebut, media sosial juga menghadirkan berbagai tantangan, khususnya yang berkaitan dengan keamanan siber, seperti peretasan akun, kebocoran data pribadi, dan penyalahgunaan informasi digital. Isu keamanan siber menjadi semakin penting untuk diteliti karena dapat berdampak langsung terhadap privasi, keamanan, dan kepercayaan pengguna dalam memanfaatkan layanan digital [1].

TikTok merupakan salah satu platform media sosial dengan jumlah pengguna dan tingkat interaksi yang sangat tinggi. Pengguna secara aktif menyampaikan pengalaman, keluhan, serta opini melalui ulasan pada platform digital, yang berpotensi mengandung informasi terkait permasalahan keamanan siber yang dialami pengguna. Namun, jumlah ulasan yang sangat besar dan sifatnya yang tidak terstruktur menyebabkan proses analisis secara manual menjadi tidak efisien serta berpotensi menimbulkan subjektivitas [2], sehingga diperlukan pendekatan yang lebih sistematis untuk mengidentifikasi risiko keamanan siber secara akurat.

Untuk mengatasi permasalahan tersebut, diperlukan pendekatan otomatis berbasis *text mining* dan *machine learning* dalam menganalisis ulasan pengguna. Sejumlah penelitian menunjukkan bahwa algoritma *Support Vector Machine* (SVM) memiliki kinerja yang baik dalam mengklasifikasikan data teks berdimensi tinggi [3]. Selain itu, SVM juga dinilai mampu memberikan performa yang relatif stabil pada kondisi distribusi data yang tidak seimbang dibandingkan metode klasifikasi lainnya [4]. Oleh karena itu, penerapan SVM dinilai tepat sebagai solusi untuk mengklasifikasikan risiko keamanan siber berdasarkan ulasan pengguna TikTok.

Penelitian ini difokuskan pada pengelompokan risiko keamanan siber dalam ulasan pengguna TikTok ke dalam empat tingkat, yaitu Risiko Tinggi, Risiko Menengah, Risiko Rendah, dan Tidak Berisiko, dengan memanfaatkan algoritma Support Vector Machine. Melalui pendekatan tersebut, penelitian ini diharapkan mampu memberikan gambaran yang lebih sistematis mengenai potensi ancaman keamanan yang dirasakan pengguna, sekaligus menjadi bahan pertimbangan bagi pengembang dan pihak terkait dalam meningkatkan kualitas serta keamanan layanan digital.

2. Tinjauan Pustaka

Berbagai studi terdahulu menunjukkan bahwa algoritma *Support Vector Machine* (SVM) memiliki performa yang kompetitif dalam tugas klasifikasi teks, termasuk dalam analisis komentar atau ulasan pengguna pada aplikasi berbasis digital. Dalam studi yang dilakukan oleh Maulana dan Bagas Akbar Fahmi [5] SVM dibandingkan dengan *Naïve Bayes* dalam mengklasifikasikan ulasan aplikasi investasi Pluang di Google Play Store. Berdasarkan hasil evaluasi, model SVM mencatat tingkat ketepatan sebesar 99,50%, sedikit melampaui performa *Naïve Bayes* yang berada pada 99,25%. Perbedaan ini menunjukkan bahwa SVM memiliki kemampuan yang baik dalam memproses dan mengklasifikasikan ulasan berbentuk teks bebas, termasuk komentar yang mengandung keluhan sistem dan indikasi risiko keamanan. Amal, Rahmasita, Suryaputra, dan Rakhmawati [6] melakukan klasifikasi teks media sosial yang memuat opini dan keluhan pengguna terkait isu kebocoran data menggunakan *algoritma machine learning*, termasuk *Support Vector Machine* (SVM). Hasil penelitian menunjukkan bahwa SVM mampu mengelompokkan teks yang mengandung indikasi risiko keamanan data secara efektif. Temuan ini relevan dengan kajian klasifikasi ulasan pengguna terkait keamanan sistem karena menunjukkan bahwa pola bahasa dalam opini pengguna dapat dimanfaatkan untuk mendeteksi potensi ancaman terhadap sistem. Rahbini dan Fahlapi [7] melakukan klasifikasi ulasan pengguna aplikasi OVO di Google Play Store menggunakan algoritma *Naïve Bayes*. Hasil penelitian menunjukkan bahwa metode klasifikasi teks mampu mengelompokkan keluhan pengguna yang berkaitan dengan gangguan layanan dan isu keamanan sistem secara terstruktur. Temuan ini memperkuat bahwa pola bahasa dalam ulasan pengguna dapat dimanfaatkan untuk mengidentifikasi potensi risiko terhadap keamanan sistem.

Siti Aminah dan Riza Fahlapi [8] mengkaji pengelompokan ulasan pengguna aplikasi Getcontact yang berkaitan dengan upaya perlindungan dari penipuan daring menggunakan *algoritma Naïve Bayes* dan *Support Vector Machine* (SVM). Hasil penelitian menunjukkan bahwa teknik klasifikasi teks dapat dimanfaatkan untuk mengenali tanggapan pengguna terhadap aspek keamanan dan fitur proteksi sistem. Temuan tersebut mendukung penelitian mengenai klasifikasi ulasan pengguna yang berfokus pada identifikasi indikasi risiko keamanan sistem. Penelitian lain yang berfokus pada ulasan pengguna aplikasi digital juga memperlihatkan efektivitas SVM. Maharani dkk. [9] menerapkan SVM pada analisis sentimen ulasan aplikasi Shazam di Google Play Store dan memperoleh tingkat akurasi sebesar 84% dalam membedakan ulasan positif dan negatif. Hasil tersebut menunjukkan bahwa SVM mampu mengolah data teks ulasan pengguna secara efektif untuk mengidentifikasi keluhan yang berkaitan dengan gangguan fungsi dan stabilitas sistem aplikasi, yang merupakan bagian

dari aspek keamanan sistem. Selain analisis sentimen umum, algoritma SVM juga digunakan untuk mendeteksi konten bermasalah pada media sosial. Sahli Andrean, Agustin, Susanti, Rahmiati, dan Hamdani [10] menganalisis ulasan pengguna aplikasi perbankan digital menggunakan metode klasifikasi teks, termasuk SVM dan *Naïve Bayes*. Hasil penelitian menunjukkan bahwa algoritma tersebut mampu mengidentifikasi keluhan yang berkaitan dengan gangguan layanan dan keamanan data. Temuan ini relevan dengan klasifikasi ulasan pengguna terkait keamanan sistem karena menunjukkan bahwa pola bahasa dalam ulasan dapat digunakan untuk mendeteksi indikasi risiko keamanan.

Pendekatan *text mining* tidak hanya digunakan untuk analisis sentimen, tetapi juga untuk memahami aspek keamanan siber secara lebih spesifik. Saju dkk. [11] menjelaskan bahwa metode *text mining* dengan algoritma klasifikasi mampu menangani data teks berdimensi tinggi dan membentuk batas pemisah kelas yang optimal, sehingga efektif dalam pengelompokan informasi berbasis teks yang mengandung indikasi risiko keamanan sistem. Selanjutnya, Lahitani, Aesyi, Wulandari, dan Santosa [12] menganalisis tingkat kesadaran pengguna terhadap isu keamanan perangkat lunak dengan memanfaatkan komentar pengguna YouTube. Penelitian ini menggunakan pembobotan TF-IDF dan Cosine Similarity untuk mengukur tingkat kemiripan teks pada komentar pengguna, sehingga mampu mengidentifikasi pola ulasan yang mengandung indikasi risiko keamanan sistem. Selain penelitian yang berfokus pada analisis sentimen, kajian yang secara khusus membahas klasifikasi ancaman keamanan sistem juga telah dilakukan. Irwan Budianto, dkk. [13] mengkaji klasifikasi ancaman keamanan siber berbasis teks menggunakan algoritma *Naïve Bayes* untuk mengidentifikasi pola ancaman pada sistem informasi. Hasil penelitian tersebut menunjukkan bahwa pendekatan klasifikasi berbasis *machine learning* mampu mendukung proses identifikasi risiko keamanan sistem secara lebih terstruktur dan sistematis.

Berdasarkan hasil kajian terhadap penelitian-penelitian sebelumnya, dapat disimpulkan bahwa mayoritas studi yang ada masih berfokus pada analisis sentimen secara umum atau identifikasi isu keamanan secara terbatas pada platform media sosial dan aplikasi digital tertentu. Sebagian besar penelitian tersebut mengklasifikasikan opini pengguna hanya ke dalam kategori sentimen dasar, seperti positif, negatif, dan netral, atau menelaah aspek keamanan tanpa membedakan tingkat keparahan risikonya secara sistematis. Di sisi lain, pemanfaatan ulasan pengguna aplikasi TikTok sebagai sumber data utama dalam kajian risiko keamanan siber masih jarang ditemukan. Oleh sebab itu, kebaruan penelitian ini terletak pada pengembangan pendekatan klasifikasi risiko keamanan siber berbasis ulasan pengguna TikTok menggunakan algoritma *Support Vector Machine* (SVM) dengan pengelompokan risiko ke dalam empat tingkat, yaitu Risiko Tinggi, Risiko Menengah, Risiko Rendah, dan Tidak Berisiko. Pendekatan ini memungkinkan analisis yang tidak hanya menilai sentimen pengguna, tetapi juga memberikan pemetaan tingkat risiko keamanan secara lebih terstruktur dan mendalam, sehingga menghasilkan kontribusi yang lebih komprehensif dibandingkan penelitian terdahulu.

3. Metodologi

Penelitian ini menggunakan pendekatan metodologis yang disusun secara bertahap dan terintegrasi. Seluruh proses penelitian dilaksanakan melalui lima tahapan utama yang dimulai dari proses pengumpulan data dan diakhiri dengan evaluasi terhadap hasil pemodelan yang diperoleh [14]. Tahapan-tahapan penelitian yang digunakan akan dijelaskan secara bertahap pada bagian berikut.

Gambar 1. Tahapan Penelitian

3.1 Pengumpulan Data

Penelitian ini memanfaatkan dataset publik yang diperoleh dari Kaggle, yang berisi 3.199 ulasan pengguna aplikasi TikTok dari Google Play Store. Data tersebut disajikan dalam format CSV agar dapat diproses dan dianalisis secara sistematis pada tahap selanjutnya. [15]

.

3.2 Processing Data

Data yang telah dikumpulkan kemudian melalui proses pra-pengolahan sebagai tahap persiapan sebelum pemodelan dilakukan. Proses ini difokuskan pada penyempurnaan struktur teks dengan membuang tanda baca, simbol, dan karakter yang tidak penting, menyeragamkan huruf ke format *lowercase*, serta menghapus kata-kata umum yang tidak berkontribusi terhadap pembentukan pola klasifikasi [14]. Setelah tahap pra-pemrosesan, data teks direpresentasikan ke dalam bentuk numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). *Term Frequency* (TF) digunakan untuk mengukur frekuensi kemunculan suatu kata dalam dokumen, yang dirumuskan sebagai berikut:

$$\mathrm{TF(t,d)} = 1 + \frac{f(t,d)}{\sum f(d)} \quad (1)$$

Inverse Document Frequency (IDF) merepresentasikan bobot pentingnya suatu istilah terhadap seluruh kumpulan dokumen dan dinyatakan pada Persamaan.

$$\mathrm{IDF\ (t)} = \log\left(\frac{N}{df(t)}\right) \quad (2)$$

Bobot akhir TF-IDF diperoleh melalui kombinasi antara nilai TF dan IDF, sebagaimana ditunjukkan pada Persamaan.

$$\mathrm{TF-IDF(t,d)} = \mathrm{TF(t,d)} \times \mathrm{IDF(t)} \quad (3)$$

Representasi ini digunakan sebagai fitur masukan pada proses klasifikasi.

3.3 Pelabelan Data

Setelah proses pra-pemrosesan selesai, data ulasan yang telah dibersihkan selanjutnya digunakan untuk tahap pelabelan. Pada tahap ini, setiap komentar pengguna dianalisis berdasarkan kemunculan kata dan frasa yang berkaitan dengan isu keamanan siber. Nilai risiko ditentukan dari kecenderungan isi ulasan yang menunjukkan adanya masalah keamanan, seperti indikasi peretasan, kebocoran data, atau gangguan sistem. Berdasarkan hasil analisis tersebut, setiap ulasan kemudian diklasifikasikan ke dalam empat kategori tingkat risiko, yaitu Risiko Tinggi, Risiko Sedang, Risiko Rendah, dan Tanpa Risiko [16]. Penentuan kategori risiko pada penelitian ini dilakukan secara konsisten dengan tujuan penelitian, sehingga hasil pelabelan dapat merepresentasikan kondisi risiko keamanan siber yang dirasakan pengguna secara lebih objektif dan relevan sebagai dasar proses klasifikasi menggunakan *Support Vector Machine.*

3.4 Klasifikasi SVM

Tahap klasifikasi dilakukan dengan memanfaatkan data ulasan yang telah diberi label risiko keamanan siber sebagai dasar pembentukan model prediksi. Algoritma *Support Vector Machine* (SVM) digunakan untuk melakukan proses pembelajaran dengan cara mengidentifikasi pola pembeda antar kelas risiko berdasarkan representasi fitur teks. Model dilatih menggunakan data latih pada proporsi tertentu untuk menghasilkan fungsi pemisah yang optimal antar kategori risiko. Selanjutnya, model yang telah dilatih diuji menggunakan data uji guna mengukur kemampuan prediksi terhadap data baru, serta dievaluasi untuk mengetahui tingkat keandalan model dalam mengklasifikasikan risiko keamanan siber. Untuk mengukur tingkat kinerja model klasifikasi, penelitian ini menerapkan beberapa metrik evaluasi sebagai dasar penilaian. Penilaian difokuskan pada kemampuan algoritma *Support Vector Machine* dalam mengklasifikasikan tingkat risiko keamanan siber berdasarkan ulasan pengguna. Parameter evaluasi yang digunakan meliputi akurasi, *precision*, *recall*, dan *F1-score* yang diperoleh dari hasil *confusion matrix*. Metrik tersebut digunakan untuk menilai tingkat ketepatan dan konsistensi model dalam melakukan pengelompokan data secara keseluruhan [17]. SVM bekerja dengan membentuk sebuah *hyperplane* optimal yang memisahkan data ke dalam kelas-kelas yang berbeda dengan margin maksimum. Fungsi keputusan SVM dapat dirumuskan sebagai berikut:

$$f(x) = w \cdot x + b \quad (4)$$

Dengan w sebagai *vektor bobot*,x sebagai *vektor fitur* hasil *TF-IDF*, dan b sebagai *bias*. Model SVM dilatih menggunakan data latih untuk mempelajari pola risiko keamanan siber, kemudian digunakan untuk mengklasifikasikan data uji.

3.5 Evaluasi

Data yang telah diberi label selanjutnya dibagi menjadi data latih dan data uji dengan dua skenario pengujian, yaitu rasio 80:20 dan 70:30. Pembagian data tersebut digunakan untuk menguji konsistensi kinerja model pada proporsi data yang berbeda. Evaluasi kinerja model dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score* untuk memberikan penilaian yang lebih menyeluruh, terutama pada kondisi distribusi data yang tidak seimbang. Evaluasi dilakukan untuk mengukur kemampuan model klasifikasi dalam menghasilkan prediksi yang akurat sebelum digunakan pada konteks nyata. Dalam penelitian ini, *confusion matrix* digunakan sebagai pendekatan evaluasi guna menilai performa prediktif algoritma *Support Vector Machine* (SVM) [16]. Akurasi digunakan untuk mengukur tingkat ketepatan klasifikasi secara keseluruhan dan dirumuskan sebagai:

$$Accuracy = \frac{TF+TN}{TF+TN+FP+FN} \quad (5)$$

Precision merepresentasikan tingkat ketepatan hasil prediksi model terhadap suatu kelas tertentu, sebagaimana dinyatakan pada Persamaan.

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

Recall digunakan untuk mengukur kemampuan model dalam mendeteksi data yang relevan dan dirumuskan sebagai:

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

F1-score digunakan untuk menilai keseimbangan antara *precision* dan *recall*, yang dirumuskan sebagai:

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision \times Recall} \quad (8)$$

Metrik-metrik tersebut digunakan untuk memberikan penilaian yang lebih komprehensif terhadap kinerja model *Support Vector Machine,* khususnya pada kondisi distribusi data yang tidak seimbang.

4.Hasil dan Pembahasan

4.1 Pengambilan Data

Data penelitian diperoleh dari *dataset* sekunder yang bersumber dari platform Kaggle. *Dataset* tersebut berisi kumpulan ulasan pengguna aplikasi TikTok yang relevan dengan fokus penelitian. Seluruh data dikumpulkan dalam format file CSV dan digunakan sebagai bahan analisis untuk mengklasifikasikan risiko keamanan siber menggunakan metode *Support Vector Machine* (SVM).

	A	B	C	D	E	F
4	876c3be7-	Pengguna	1	Tiktok sekarang udah jelek kebanyakan di update jadi lemot ditambah HP ku	19/02/2025 07.33	
5	9eac358b-	Pengguna	4	Gabisa masuk tik tok najis	19/02/2025 07.32	
6	147b41cc-	Pengguna	3	Karena akun ku ga fyp fyp kan cape bikin video mulu ga fyp lagi	19/02/2025 07.29	
7	237e1d0b-	Pengguna	5	Tik tok kenapa aku mau main Tik tok kenapa nggak bisa main Tik tok semoga	19/02/2025 07.28	
8	6676a09a-	Pengguna	5	Saya sedikit kecewa dengan tiktok karna pas mau masuk lama banget sampe	19/02/2025 07.28	
9	73acd005-	Pengguna	5	Bagus dan sangat menghibur	19/02/2025 07.28	
10	95f50f31-3	Pengguna	1	karna tidak bisa di buka	19/02/2025 07.28	
11	586287e6-	Pengguna	1	gak bisa masuk	19/02/2025 07.27	
12	b32d538c-	Pengguna	3	Karenan kadang dak bisa di pencet	19/02/2025 07.26	
13	d15792e8-	Pengguna	1	Elor sekarang tiktok, harus install mulu	19/02/2025 07.25	
14	9bcf9cef-(	Pengguna	1	stuck di logo	19/02/2025 07.24	
15	fc57e210-(	Pengguna	1	Tiktok gmana sieh	19/02/2025 07.23	
16	7749ac94-	Pengguna	2	Aku sangat suka sama. tiktok tapi kenapa aku ga bisa login	19/02/2025 07.20	

Gambar 2. Data Mentah Hasil Pengumpulan

Data penelitian diperoleh dari kumpulan ulasan pengguna aplikasi TikTok yang tersedia pada repositori Kaggle. *Dataset* ini memuat teks komentar yang berkaitan dengan fokus penelitian dan telah tersaji dalam bentuk data terstruktur. Jumlah data yang dianalisis sebanyak 3.199 ulasan, yang disimpan dalam format berkas CSV. Penyimpanan dalam format tersebut

memudahkan proses lanjutan, mulai dari pembersihan data, penentuan kategori risiko keamanan siber, hingga tahap pembentukan model klasifikasi dengan pendekatan *Support Vector Machine* (SVM).

4.2 Pra-pemerosesan data

Sebelum memasuki tahap klasifikasi, data terlebih dahulu melalui proses pra-pemrosesan guna menghilangkan elemen yang tidak diperlukan dan menyusun kembali data agar lebih teratur. Langkah ini dilakukan agar data siap dianalisis serta memenuhi standar kebersihan dan konsistensi yang dibutuhkan.

4.2.1 Pembersihan

Proses pembersihan data dilakukan dengan menghilangkan berbagai komponen teks yang tidak berhubungan langsung dengan kebutuhan analisis, seperti tautan URL, mention akun pengguna (@username), tanda petik(”), karakter tidak terbaca serta karakter selain huruf yang tidak memiliki kontribusi terhadap analisis.

Tabel 1. Hasil Pembersihan Data

Text_Lengkap	Pembersihan
@mujianto28 Suka nglag” smpe ga mau buka terpaksa udh bbrpa kali hapus habis tu instal ulg,,udh lebih 10x kek gnii,ini aja baru habis hapus aplikasi terus instal lagiðŸ˜□	Suka ngelag sampe ga mau buka terpaksa udh beberpa kali hapus habis tu instal ulang udah lebih 10x kek gni,ini aja baru habis hapus aplikasi terus instal lagi

4.2.2 Penyeragaman huruf

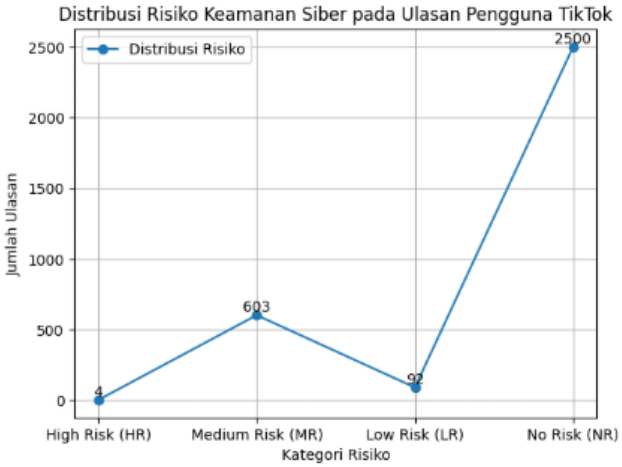
Normalisasi dilakukan dengan mengubah seluruh karakter alfabet menjadi huruf kecil. Langkah ini memastikan tidak ada perbedaan representasi kata akibat penggunaan huruf kapital, sehingga pengolahan teks dapat berlangsung secara seragam pada tahap selanjutnya.

Tabel 2. Hasil Penyeragaman Huruf

Pembersihan	Penyeragaman huruf
Banyak bug setelah update □	banyak bug setelah update
Aplikasi sering error	aplikasi sering error

4.3 Pelabelan Data Keamanan Siber

Pelabelan data dilakukan untuk mengklasifikasikan ulasan pengguna TikTok ke dalam kategori risiko keamanan siber sesuai dengan tingkat ancaman yang terkandung di dalam teks. Proses ini dilaksanakan setelah tahap *preprocessing* selesai, sehingga data yang dianalisis telah melalui pembersihan dan siap digunakan dalam proses klasifikasi.



Gambar 3. Distribusi Risiko Keamanan Siber

Penentuan label dilakukan dengan menganalisis isi ulasan serta menelusuri keberadaan kata atau frasa yang mengindikasikan isu keamanan siber, seperti peretasan akun, kebocoran informasi pribadi, penyalahgunaan data, gangguan sistem aplikasi, maupun bentuk ancaman keamanan digital lainnya.

Gambar 3 memperlihatkan sebaran tingkat risiko keamanan siber pada ulasan pengguna TikTok yang diperoleh melalui tahap *preprocessing* dan proses pelabelan data. Kategori Tidak Berisiko mendominasi dengan jumlah 2.500 ulasan atau sebesar 78,15%, disusul oleh kategori Risiko Menengah sebanyak 603 ulasan atau 18,85%. Selanjutnya, kategori Risiko Rendah tercatat sebanyak 92 ulasan atau 2,88%, sedangkan kategori Risiko Tinggi memiliki jumlah paling sedikit, yaitu 4 ulasan atau 0,13%. Temuan ini menunjukkan bahwa mayoritas ulasan pengguna tidak berkaitan dengan risiko keamanan siber, sementara laporan yang mengindikasikan risiko tinggi relatif jarang ditemukan.

4.4 Hasil *Preprocessing* dan Pelabelan Risiko Keamanan Siber

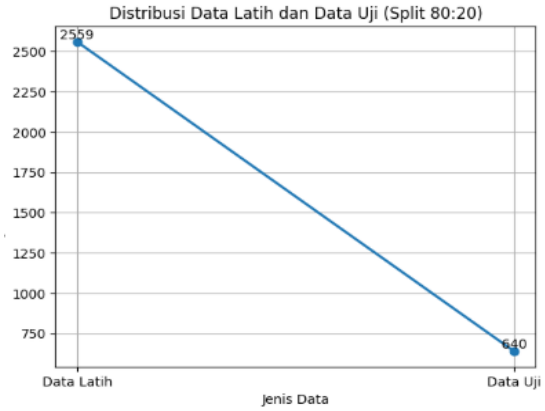
Tabel 3. Tabel *Preprocessing Result* And *Security Risk Labeling*

No	Original Sentence	*Preprocessing* Resul	*Risk Labe*l
1	Gabisa masuk tik tok najis	gabisa masuk tik tok najis	MR
2	Akun saya seperti di hack	akun saya seperti di hack	HR
3	Aplikasi sering error	aplikasi sering error	MR
4	Update malah makin jelek	update malah makin jelek	LR
5	Tidak ada masalah sejauh ini	tidak ada masalah sejauh ini	NR
6	Video tidak bisa diputar	video tidak bisa diputar	MR
7	Banyak bug setelah update	banyak bug setelah update	MR
8	Aman dan nyaman digunakan	aman dan nyaman digunakan	NR
9	Takut data dicuri	takut data dicuri	HR
10	Aplikasi lambat sekali	aplikasi lambat sekali	LR

Tabel 3 menyajikan hasil *preprocessing* dan pelabelan risiko keamanan siber terhadap sepuluh ulasan pengguna TikTok yang paling relevan. Tabel ini menunjukkan perbedaan antara teks ulasan asli dan hasil *preprocessing,* serta kategori risiko keamanan siber yang diberikan. Data pada tabel ini merupakan bagian dari data latih yang digunakan dalam proses *training* model *Support Vector Machine,* sehingga merepresentasikan dasar pembelajaran model dalam mengenali pola risiko keamanan siber.

4.5 Klasifikasi Algoritma

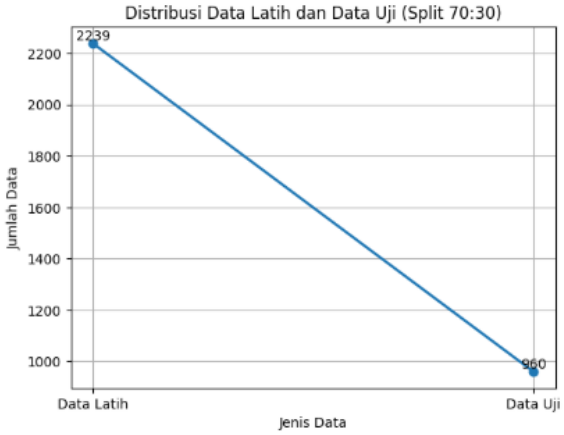
Pada tahap klasifikasi, data ulasan dikelompokkan ke dalam empat tingkat risiko keamanan siber, yaitu Risiko Tinggi, Risiko Menengah, Risiko Rendah, dan Tidak Berisiko. Seluruh data hasil pra-pemrosesan tetap digunakan tanpa penghapusan kelas tertentu agar distribusi data mencerminkan kondisi asli.Pengembangan model *Support Vector Machine* (SVM) dilakukan melalui dua skema pembagian data, yakni 80:20 dan 70:30.



Gambar 4. Data Latih dan Data Uji (Split 80:20)

Pada skema 80:20, dari total 3.199 ulasan, sebanyak 2.559 data dialokasikan sebagai data pelatihan dan 640 data sebagai data pengujian. Sementara itu, pada skema 70:30, sejumlah 2.239 data digunakan untuk pelatihan dan 960 data untuk pengujian.Data pelatihan dimanfaatkan untuk membangun model agar mampu mengenali pola risiko keamanan siber, sedangkan data pengujian digunakan untuk mengevaluasi kemampuan model dalam mengklasifikasikan ulasan yang belum pernah diproses sebelumnya.

Gambar 4 menunjukan skema pembagian data latih dan data uji dengan perbandingan 80:20. Dari keseluruhan 3.199 data ulasan yang telah diproses pada tahap *preprocessing*, sebanyak 2.559 data dimanfaatkan sebagai data latih, sedangkan 640 data lainnya digunakan sebagai data uji. Proporsi ini diterapkan untuk memastikan model *Support Vector Machine* (SVM) memperoleh data pelatihan yang cukup dalam mempelajari pola risiko keamanan siber. Sementara itu, data uji berperan sebagai acuan dalam menilai performa model serta kemampuannya dalam mengklasifikasikan data yang belum pernah dipelajari.



Gambar 5. Data Latih dan Data Uji (Split 70:30)

Gambar 5 menggambarkan komposisi data pelatihan dan pengujian pada skema pemisahan 70:30. Dari keseluruhan 3.199 ulasan yang dianalisis, 2.239 data dimanfaatkan untuk proses pelatihan model, sedangkan 960 data sisanya digunakan sebagai data evaluasi. Skema ini diterapkan untuk menilai konsistensi serta stabilitas performa *Support Vector Machine* (SVM) ketika porsi data pengujian lebih besar, sehingga kemampuan generalisasi model dapat diamati secara lebih menyeluruh.

Tabel 4. Tabel Distribusi Kelas Setelah *Preprocessing*

Label	Kategori Risiko	Jumlah Data Awal	Jumlah Data Setelah *Preprocessing*	Status Sampling
HR	Risiko Tinggi	4	4	Tidak Seimbang
MR	Risiko Menengah	603	603	Tidak Seimbang
LR	Risiko Rendah	92	92	Tidak Seimbang
NR	Tidak Berisiko	2500	2500	Tidak Seimbang

Berdasarkan tabel distribusi kelas, jumlah data pada setiap kategori risiko tidak tersebar secara merata. Kategori Tidak Berisiko mendominasi sebagian besar *dataset*, sementara kategori Risiko Tinggi hanya muncul dalam jumlah yang sangat terbatas. Kondisi ini menunjukkan adanya ketidakseimbangan data yang berpotensi memengaruhi kinerja model klasifikasi.

Evaluasi model *Support Vector Machine* dilakukan menggunakan dua skema pembagian data, yaitu 80:20 dan 70:30. Pada skema pembagian data 80:20, model mencapai tingkat akurasi sekitar 93% dan menunjukkan kinerja yang sangat baik dalam mengklasifikasikan kategori Tidak Berisiko dan Risiko Menengah. Namun, kemampuan model dalam mengenali kategori Risiko Tinggi masih terbatas karena jumlah data pada kategori tersebut sangat sedikit. Sementara itu, pada skema pembagian data 70:30, model memperoleh

tingkat akurasi sebesar 93,23% dengan hasil yang relatif konsisten dibandingkan skema sebelumnya. Meskipun demikian, nilai *recall* untuk kategori Risiko Rendah dan Risiko Tinggi masih tergolong rendah akibat ketidakseimbangan distribusi data.

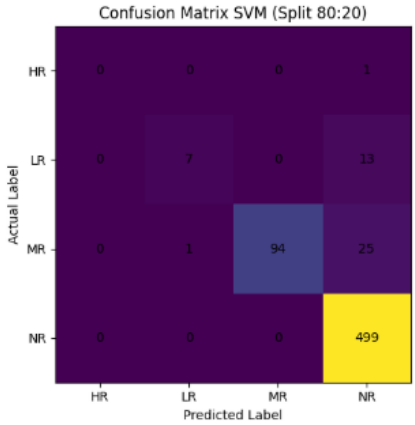
Tabel 5. Distribusi Risiko Keamanan Siber

Skor	Jumlah Ulasan	Risiko Tinggi	Risiko Menengah	Risiko Rendah	Tidak Berisiko
1	1.272	1	359	45	867
2	384	1	100	15	268
3	407	0	102	10	295
4	290	1	49	6	234
5	846	1	77	12	756

Tabel distribusi risiko keamanan siber berdasarkan skor ulasan menunjukkan adanya hubungan antara tingkat kepuasan pengguna dan potensi risiko yang teridentifikasi. Ulasan dengan skor rendah, khususnya skor 1 dan 2, memiliki jumlah yang lebih besar dan didominasi oleh kategori risiko menengah serta tidak berisiko, meskipun masih ditemukan risiko tinggi dalam jumlah yang sangat terbatas. Pada skor 3, jumlah ulasan relatif lebih sedikit dan sebagian besar berada pada kategori tidak berisiko dan risiko menengah, tanpa adanya risiko tinggi. Sementara itu, ulasan dengan skor tinggi, yaitu skor 4 dan 5, hampir seluruhnya termasuk dalam kategori tidak berisiko dengan proporsi risiko menengah dan risiko rendah yang kecil. Temuan ini mengindikasikan bahwa semakin rendah skor ulasan yang diberikan pengguna, semakin besar kemungkinan munculnya indikasi risiko keamanan siber dalam ulasan tersebut.

4.6 Evaluasi Model SVM

Tahap selanjutnya dilakukan evaluasi kinerja model *Support Vector Machine* menggunakan tingkat ketepatan, tingkat sensitivitas, nilai F1, dan jumlah data pada setiap kategori risiko keamanan siber. Pengujian diterapkan dengan dua skema pembagian data, yaitu 80:20 dan 70:30, untuk menilai kestabilan performa model. Hasil evaluasi ini digunakan untuk mengukur tingkat ketepatan serta kemampuan model dalam mengklasifikasikan ulasan pengguna.



Gambar 6. *Confusion Matrix* Model SVM Dengan Split Data 80:20

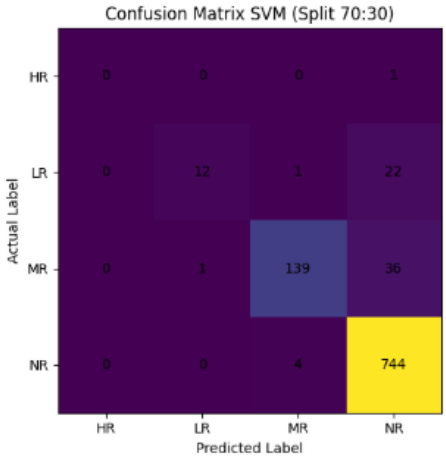
Gambar 6 menyajikan hasil evaluasi model *Support Vector Machine* pada skema pembagian data 80:20 dalam bentuk *confusion matrix*. Dari hasil tersebut terlihat bahwa kelas Tidak Berisiko (NR) memiliki tingkat prediksi paling tinggi dengan 499 ulasan teridentifikasi secara tepat. Kinerja pada kategori Risiko Menengah (MR) juga tergolong baik, ditunjukkan oleh 94 data yang berhasil dikenali dengan benar. Sebaliknya, model belum mampu mengenali kelas Risiko Tinggi (HR), karena seluruh data pada kategori tersebut terklasifikasi ke kelas lain. Kondisi ini berkaitan dengan distribusi data yang tidak merata antar kelas, sehingga pembelajaran model terhadap kelas minoritas menjadi kurang optimal.

Tabel 6. Sampel Hasil Klasifikasi Risiko Keamanan Siber Pada Data Uji (Split 80:20)

No	Teks Ulasan	Hasil *Preprocessing*	Label Aktual	predeksi SVM	Status klasifikasi
1	“harus 8gb baru bisa buka“	harus gb baru bisa buka	MR	NR	Tidak Benar
2	Tolong perbaiki bug tiktok	tolong perbaiki bug tiktok	MR	MR	Benar
3	Seharian nonton tiktok	seharian nonton tiktok	NR	NR	Benar
4	Tidak bisa mengundang teman	Tidak bisa mengundang teman	MR	NR	Tidak Benar
5	Aplikasi sering bug	aplikasi sering bug	MR	MR	Benar

Berdasarkan tabel 6, sampel data uji pada skema pembagian 80:20, ulasan yang berhasil dikenali (diklasifikasikan secara benar) umumnya mengandung kata kunci yang secara jelas merepresentasikan pernyataan penggunaan normal, maupun gangguan sistem aplikasi, seperti bug dan tidak bisa dibuka. Keberadaan kata-kata tersebut membentuk pola fitur yang kuat pada ruang *vektor* TF-IDF, sehingga model *Support Vector Machine* (SVM) mampu memisahkan kelas risiko dengan lebih akurat.

Sebaliknya, ulasan yang tidak berhasil dikenali (diklasifikasikan secara tidak benar) cenderung memiliki konteks keluhan yang bersifat umum dan tidak secara eksplisit menunjukkan indikasi risiko keamanan siber. Ulasan seperti “harus 8gb baru bisa buka” lebih berkaitan dengan keterbatasan perangkat pengguna dibandingkan ancaman keamanan sistem, sehingga model cenderung mengelompokkannya ke kategori Tidak Berisiko. Hal ini menunjukkan bahwa pada skema 80:20, kesalahan klasifikasi lebih dipengaruhi oleh ambigu-itas makna teks daripada kegagalan model dalam mempelajari pola data secara keseluruhan.



Gambar 7. *Confusion Matrix* Model SVM Dengan Split Data 70:30

Gambar 7 memperlihatkan matriks kebingungan (*confusion matrix*) sebagai hasil pengujian model *Support Vector Machine* (SVM) pada skema pembagian data 70:30. Dari hasil evaluasi tersebut, model mampu mengidentifikasi kelas Tidak Berisiko (NR) dengan tingkat ketepatan yang tinggi, ditunjukkan oleh 744 prediksi yang sesuai dengan label sebenarnya. Performa yang cukup baik juga terlihat pada kategori Risiko Menengah (MR), dengan 139 data berhasil dikenali secara akurat. Sebaliknya, pada kelas Risiko Rendah (LR) dan Risiko Tinggi (HR), jumlah prediksi yang keliru masih cukup besar. Hal ini kemungkinan disebabkan oleh kemiripan karakteristik teks antar kategori serta ketidakseimbangan jumlah data pada masing-masing tingkat risiko keamanan siber.

Tabel 7. Sampel Hasil Klasifikasi Risiko Keamanan Siber Pada Data Uji (Split 70:30)

No	Teks Ulasan	Hasil *Preprocessing*	Label Aktual	Predeksi SVM	Status klasifikasi
1	”harus update biar bisa dipakai”	harus update biar bisa dipakai	MR	NR	Tidak Benar
2	Bug ga bisa buka tiktok	bug ga bisa buka tiktok	MR	MR	Benar
3	Nonton vidio lancar	nonton vidio lancar	NR	NR	Benar
4	Aplikasi sering keluar sendiri	aplikasi sering keluar sendiri	MR	NR	Tidak Benar
5	Terima kasih tiktok	terima kasih tiktok	MR	MR	Benar

Berdasarkan tabel 7, Data uji yang berhasil dikenali (diklasifikasikan secara benar) umumnya memiliki kata kunci dan struktur kalimat yang jelas serta konsisten dengan pola yang dipelajari pada data latih. Ulasan tersebut mengandung istilah yang secara eksplisit merepresentasikan kondisi tertentu, seperti gangguan aplikasi, kesalahan sistem, atau pernyataan penggunaan normal, sehingga model *Support Vector Machine* (SVM) mampu menempatkannya pada sisi *hyperplane* yang sesuai dengan kategori risiko yang tepat.

Sebaliknya, data uji yang tidak berhasil dikenali (diklasifikasikan secara tidak benar) cenderung memiliki karakteristik linguistik yang ambigu atau memiliki kemiripan dengan lebih dari satu kategori risiko. Kondisi ini banyak terjadi pada ulasan yang berada di perbatasan antara Risiko Menengah dan Tidak Berisiko. Selain itu, ketidakseimbangan jumlah data antar kelas menyebabkan kelas minoritas tidak memiliki representasi yang cukup dalam proses pelatihan, sehingga model belum mampu membentuk batas pemisah yang optimal untuk mengenali seluruh variasi pola teks secara akurat.

Tabel 8. Tabel SVM *Classification* Results Across Two Data Split Ratios Split 80:20

Risk Category	*Precision*	*Recall*	*F1-Score*	Support
HR	0.00	0.00	0.00	1
LR	0.93	0.78	0.85	120
MR	0.97	0.79	0.87	200
NR	0.93	0.99	0.96	499

Pada rasio pembagian data 80:20, model memperlihatkan kinerja yang sangat baik pada kategori Tidak Berisiko dan Risiko Menengah, ditandai dengan nilai *precision* dan *recall* yang tinggi pada kedua kelas tersebut. Sebaliknya, pada kategori Risiko Tinggi, nilai *precision* dan *recall* tercatat 0,00 karena jumlah data yang sangat terbatas. Kondisi ini menunjukkan bahwa keterbatasan data pada kelas minoritas memengaruhi kemampuan model dalam mengenali pola risiko secara maksimal.

Tabel 9. Tabel SVM *Classification* Results Across Two Data Split Ratios Split 70:30

Risk Category	*Precision*	*Recall*	*F1-Score*	Support
HR	0.00	0.00	0.00	1
LR	0.92	0.34	0.50	35
MR	0.9	0.79	0.87	176
NR	0.93	0.99	0.96	176

Pada skema pembagian data 70:30, performa model relatif konsisten dibandingkan skema sebelumnya. Nilai *recall* pada kategori Risiko Rendah mengalami penurunan, yang menunjukkan bahwa peningkatan proporsi data uji turut memengaruhi kemampuan model dalam mendeteksi kelas minoritas. Temuan ini menegaskan bahwa ketidakseimbangan data berpengaruh signifikan terhadap kinerja klasifikasi.

4.6 Pembahasan

Analisis terhadap 3.199 ulasan pengguna TikTok memperlihatkan bahwa penerapan algoritma *Support Vector Machine* (SVM) menghasilkan kinerja klasifikasi yang konsisten pada dua skema pembagian data, yakni 80:20 dan 70:30, dengan akurasi yang tetap berada di atas 93%. Model bekerja sangat baik dalam mengenali kategori Tidak Berisiko (NR) dan Risiko

Menengah (MR) yang jumlahnya dominan dalam *dataset*. Nilai *recall* yang tinggi pada kelas NR menunjukkan kemampuan model dalam mengidentifikasi ulasan tanpa indikasi ancaman keamanan siber. Namun demikian, performa pada kategori Risiko Rendah (LR) dan Risiko Tinggi (HR) masih terbatas, terutama karena jumlah data pada kedua kelas tersebut jauh lebih sedikit dibandingkan kategori lainnya.

Kondisi ini menunjukkan adanya *accuracy paradox*, yaitu ketika nilai akurasi yang tinggi belum tentu mencerminkan kemampuan model dalam mengenali kelas dengan jumlah data terbatas. Meskipun SVM memiliki performa yang cukup stabil, rendahnya nilai *recall* dan *F1-score* pada kategori Risiko Tinggi menandakan bahwa model masih kurang optimal dalam mendeteksi ancaman keamanan siber yang serius. Oleh sebab itu, penilaian kinerja model perlu mempertimbangkan metrik lain seperti *precision*, *recall*, dan *F1-score agar* evaluasi dilakukan secara lebih menyeluruh.

Temuan dalam penelitian ini konsisten dengan studi yang dilakukan oleh Ahmad Turmudi Zy dan Wahyu Hadikristanto [1] yang mengungkapkan bahwa algoritma *Support Vector Machine* (SVM) dapat menghasilkan tingkat akurasi yang tinggi dalam analisis sentimen kebocoran data di Twitter, namun performanya masih dipengaruhi oleh kondisi ketidakseimbangan kelas. Hal ini menguatkan bahwa penilaian kinerja model klasifikasi risiko keamanan siber perlu melibatkan metrik lain selain akurasi agar kemampuan deteksi risiko dapat dievaluasi secara lebih tepat. Hasil penelitian ini konsisten dengan temuan Alvali Zaqi Taufan dan Wahyu Wibowo [4] yang menunjukkan bahwa algoritma *Support Vector Machine* (SVM) mampu memberikan performa klasifikasi yang baik pada analisis persepsi keamanan data. Kesamaan ini menegaskan bahwa SVM efektif digunakan untuk analisis risiko keamanan siber berbasis teks, meskipun tetap dipengaruhi oleh ketidakseimbangan distribusi data.

Penelitian oleh Andri Wijaya dkk. [9] menunjukkan bahwa algoritma *Support Vector Machine* (SVM) mampu mengklasifikasikan ulasan pengguna aplikasi Shazam dengan tingkat akurasi sebesar 84%, di mana model lebih efektif dalam mengenali sentimen positif dibandingkan sentimen negatif pada data ulasan Google Play Store. Studi yang dilakukan oleh Muhamad Aji Afani Maldini dan Septi Andryana [10] melaporkan bahwa penerapan algoritma *Support Vector Machine* (SVM) pada ulasan aplikasi perbankan digital menghasilkan tingkat akurasi sebesar 88%. Data yang dianalisis mencakup berbagai komentar pengguna yang mengandung keluhan mengenai gangguan layanan serta isu perlindungan data. Hasil tersebut mendukung temuan dalam penelitian ini, bahwa teknik klasifikasi berbasis teks dapat dimanfaatkan untuk mendeteksi potensi risiko keamanan sistem melalui analisis pola bahasa pada ulasan pengguna. Penelitian oleh Muhammad Zidna Mubarok dkk. [15] menunjukkan bahwa algoritma *Support Vector Machine* (SVM) Linear memiliki kinerja klasifikasi sentimen yang sangat tinggi pada ulasan aplikasi Claude di Google Play Store, dengan nilai akurasi, *precision*, *recall*, dan *F1-score* yang unggul dibandingkan *Naïve Bayes*. Temuan ini menegaskan efektivitas SVM dalam menangani data teks berdimensi tinggi dan distribusi kelas yang tidak seimbang.

Temuan penelitian ini sejalan dengan hasil studi Frencis Matheos Sarimole dan Kudrat [18] yang membandingkan SVM dan *Naïve Bayes* pada analisis sentimen aplikasi SATU SEHAT di Twitter. Penelitian tersebut menunjukkan bahwa SVM menghasilkan akurasi lebih tinggi (87,95%) dibandingkan Naïve Bayes, namun tetap menghadapi keterbatasan dalam mendeteksi kelas minoritas akibat distribusi data yang tidak seimbang, sehingga menegaskan pentingnya evaluasi berbasis *precision* dan *recall* selain akurasI. Kontribusi utama penelitian ini terletak pada penerapan algoritma *Support Vector Machine* (SVM) untuk mengklasifikasikan risiko keamanan siber berdasarkan ulasan pengguna TikTok secara efektif. Hasil analisis menunjukkan bahwa SVM mampu memberikan performa klasifikasi yang stabil dengan tingkat akurasi di atas 93% pada dua skema pembagian data, serta sangat baik dalam mengenali kategori Tidak Berisiko dan Risiko Menengah. Hasil penelitian ini menunjukkan bahwa metode klasifikasi teks dapat digunakan sebagai langkah awal dalam mendeteksi indikasi risiko keamanan siber berdasarkan ulasan pengguna. Selain itu, penelitian ini menekankan perlunya penggunaan *metrik evaluasi* yang beragam dalam mengukur performa model pada data yang tidak seimbang, guna mendukung pengembangan sistem deteksi risiko keamanan yang lebih akurat dan sistematis.

5. Simpulan

Penelitian ini menunjukkan bahwa *algoritma Support Vector Machine* (SVM) layak diterapkan untuk mengidentifikasi tingkat risiko keamanan siber pada ulasan pengguna TikTok melalui pendekatan analisis teks. Dari pengolahan 3.199 data ulasan, penerapan skema pembagian 80:20 dan 70:30 menghasilkan performa model yang relatif konsisten dengan akurasi melebihi 93%. Model bekerja sangat baik pada kategori Tidak Berisiko dan Risiko Menengah yang mendominasi dataset.Nilai kebaruan penelitian ini terletak pada pengembangan skema klasifikasi risiko ke dalam empat level, yaitu Risiko Tinggi, Risiko Menengah, Risiko Rendah, dan Tidak Berisiko, sehingga analisis tidak terbatas pada sentimen umum, melainkan mampu memberikan pemetaan risiko yang lebih sistematis. Kinerja SVM yang stabil pada data teks berdimensi tinggi serta kemampuannya beradaptasi pada distribusi data yang tidak seimbang menjadi keunggulan utama dalam penelitian ini.Berbeda dengan studi terdahulu yang lebih menitikberatkan pada analisis sentimen, penelitian ini secara khusus menyoroti aspek keamanan siber pada ulasan pengguna TikTok. Namun demikian, keterbatasan dalam mengenali kelas risiko dengan jumlah data minoritas menunjukkan perlunya penelitian lanjutan, khususnya dalam penanganan ketidakseimbangan data dan eksplorasi metode klasifikasi alternatif guna meningkatkan sensitivitas model terhadap potensi ancaman keamanan siber.

Daftar Refensi

[1] A. T. Zy and W. Hadikristanto, “Implementasi Algoritma Metode Naive Bayes dan Support Vector Machine Tentang Pembobolan dan Kebocoran Data di Twitter,” *Bull. Inf. Technol. (BIT*, vol. 4, no. 1, pp. 49–56, 2023, doi: 10.47065/ bit.v3i1. 493.

[2] P. R. Artika and A. H. Lubis, “Implementasi Algortima Support Vector Machine Dalam Klasifikasi Komentar Pengguna Produk Skintific di E-Commerce,” *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 14, no. 2, p. 1349, 2025, doi: 10.35889/jutisi.v14i2.3137.

[3] K. Al Mas Ud, M. I. Fieldi, M. H. Al-Farisy, M. Alfarizi, F. Fathoni, and A. I. Ibrahim, “Analisis Sentimen Deepseek Berdasarkan Ulasan Google Play Store Menggunakan Metode Naïve Bayes,” *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 14, no. 2, p. 1020, 2025, doi: 10.35889/jutisi.v14i2.2778.

[4] Y. Khoiruddin, "Analisis Sentimen Gojek Indonesia Pada Twitter Menggunakan Algoritma Naive Bayes dan Support Vector Machine. Progresif: Jurnal Ilmiah Komputer, vol. 19, no. 1, pp. 391-400. doi:http://dx.doi.org/10.35889/progresif.v19i1.1173

[5] I. Guspian and M. H. Basri, “Comparison of Naive Bayes and SVM Algorithms in Sentiment Analysis for the Optimization of Hotel Operational Services in Central Bangka Regency,” *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 14, no. 2, p. 1315, 2025, doi: 10.35889/jutisi.v14i2.3107.

[6] R. Artikel, M. I. Amal, E. S. Rahmasita, E. Suryaputra, and N. A. Rakhmawati, “Analisis Klasifikasi Sentimen Terhadap Isu Kebocoran Data Kartu Identitas Ponsel di Twitter,” *J. Tek. Inform. dan Sist. Informas*, vol. 8, no. September, pp. 645–660, 2022, doi: 10.28932/jutisi.v8i3.5483.

[7] M. Ramdan, A. Surya, and U. Hayati, “Analisis Sentimen Ulasan Pengguna Ovo Menggunakan Algoritma Naive Bayes Pada Google Play Store,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 2780–2786, 2024.

[8] R. Fahlapi, A. Y. Kuntoro, and T. Asra, “Perbandingan Algoritma Klasifikasi Analisis Sentimen Pengguna Aplikasi Getcontact Dalam Pencegahan Penipuan Online,” *J-INTECH (Journal Inf. Technol.*, vol. 2, no. 2, pp. 158–167, 2024.

[9] A. Wijaya, Meilinda, and M. Maharani, “Analisis Sentimen Ulasan Aplikasi Shazam Di Google Play Store Menggunakan Support Vector Machine,” *Zo. J. Sist. Inf.*, vol. 6, no. 1, pp. 197–207, 2024, doi: 10.31849/zn.v6i1.17994.

[10] M. Aji, A. Maldini, and S. Andryana, “Analisis Sentimen Ulasan Pengguna Aplikasi Perbankan Menggunakan Algoritma Support Vector Machine Dan Naive Bayes,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 3, pp. 4098–4105, 2025.

[11] A. S. Kirana, R. Roeswidiah, and A. Pudoli, “Analisis Sentimen Pada Media Sosial Terhadap Layanan Samsat Digital Nasional,” *Idealis Indones. J. Inf. Syst. Vol.*, vol. 8, no. 1, pp. 53–63, 2025.

[12] A. Lahitani, U. S. Aesyi, N. Wulandari, and B. D. Santosa, “Cosine Similarity untuk Mengukur Tingkat Kesadaran pada Topik Software Security Berbasis Teks Komentar di

Media Sosial Youtube," *J. Sains dan Inform.*, vol. 8, no. 2, pp. 107–115, 2022, doi: 10.34128/jsi.v8i2.535.
[13] I. Budianto, Nurchim, and H. Permatasari, "Klasifikasi Ancaman Keamanan Siber Menggunakan Algoritma Naive Bayes," *Indones. J. Appl. Informatics*, vol. 9, no. 2, pp. 431–439, 2025.
[14] M. Iranda and N. Huda, "Analisis Kinerja Algoritma SVM dan Naïve Bayes untuk Klasifikasi Sentimen Program Makan Gratis," *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 14, no. 3, pp. 1452–1464, 2025, doi: 10.35889/jutisi.v14i3.3183.
[15] M. Z. Mubarok, S. Rizqi, R. M. Rizal, A. B. N. Syalim, and N. Iksan, "Klasifikasi Sentimen Ulasan Aplikasi Claude Menggunakan Naive Bayes dan Support Vector Machine," *JUSIFOR J. Sist. Inf. dan Inform.*, vol. 4, no. 1, pp. 110–119, 2025, doi: 10.70609/jusifor.v4i1.7054.
[16] L. F. S. Minggow, A. V. Vitianingsih, S. Kacung, A. L. Maukar, and J. F. Rusdi, "Sentiment Analysis on Ajaib App Using the SVM Method," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 14, no. 4, pp. 551–556, 2025, doi: 10.32736/sisfokom.v14i4.2402.
[17] T. Tan, H. Sama, G. Wijaya, and O. E. Aboagye, "Studi Perbandingan Deteksi Intrusi Jaringan Menggunakan Machine Learning: (Metode SVM dan ANN)," *J. Teknol. dan Inf.*, vol. 13, no. 2, pp. 152–164, 2023, doi: 10.34010/jati.v13i2.10484.
[18] F. M. Sarimole and Kudrat, "Analisis Sentimen Terhadap Aplikasi Satu Sehat Pada Twitter Menggunakan Algoritma Naive Bayes Dan Support Vector Machine," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 3, pp. 783–790, 2024, doi: 10.57152/malcom.v4i2.1206.