CyberSecurity dan Forensik Digital e-ISSN: 2615-8442
Vol. 9, No. 1, Mei 2026, hlm. 23-34

Deteksi Dini Indikasi Risiko Keamanan Siber pada Game Online Berdasarkan Ulasan Pengguna Menggunakan Naive Bayes

Jely Estianti[1], RG Guntur Alam[2], Agung Kharisma Hidayah[3]

[1,2,3] Sistem Informasi, Universitas Muhammadiyah Bengkulu, Bengkulu, Indonesia
Email: [1]jelyestianti@gmail.com, [2]rggunturalam@umb.ac.id, [3]kharisma@umb.ac.id

Abstrak

Game online merupakan platform digital yang mengelola data sensitif pengguna, termasuk informasi akun, data pribadi, dan transaksi digital, sehingga rentan terhadap berbagai ancaman keamanan siber. Sebagian besar penelitian sebelumnya memanfaatkan ulasan pengguna untuk analisis sentimen dan kualitas layanan, sementara pemanfaatannya sebagai indikator dini risiko keamanan siber masih terbatas. Penelitian ini bertujuan mengidentifikasi indikasi risiko keamanan siber pada game online berdasarkan ulasan pengguna sebagai bentuk user-reported cybersecurity signals. Sebanyak 3.069 ulasan pengguna Mobile Legends diproses melalui tahapan text mining (case folding, tokenizing, stopword removal, dan stemming), direpresentasikan menggunakan pembobotan TF-IDF, dan diklasifikasikan dengan algoritma Naïve Bayes. Kategori risiko meliputi Account Security Risk, Data Privacy Risk, Phishing & Fraud Risk, Malware Risk, serta Non-Security Issue. Evaluasi menggunakan skenario pembagian data 80:20 menunjukkan akurasi keseluruhan sebesar 76,5% berdasarkan confusion matrix, dengan variasi performa antar kategori. F1-score tertinggi diperoleh pada kategori Non-Security Issue (0,92), sedangkan Malware Risk terendah (0,67) akibat ambiguitas linguistik dalam narasi pengguna. Temuan ini menunjukkan bahwa ulasan pengguna berpotensi dimanfaatkan sebagai mekanisme deteksi dini berbasis komunitas. Secara teoretis, penelitian ini memperkenalkan pendekatan community-based cyber risk identification sebagai bentuk komplementer terhadap mekanisme deteksi teknis dalam manajemen risiko keamanan siber pada platform digital.
***Kata kunci**: keamanan siber, game online, text mining, naïve bayes, deteksi dini risiko*

Early Detection of Cybersecurity Risk Indications in Online Games Based on User Reviews Using Naive Bayes

Abstrak

Online games are digital platforms that manage sensitive user data, including account information, personal data, and digital transactions, making them vulnerable to various cybersecurity threats. Most previous studies have utilized user reviews for sentiment analysis and service quality evaluation, while their use as early indicators of cybersecurity risk remains limited. This study aims to identify indications of cybersecurity risks in online games based on user reviews as user-reported cybersecurity signals. A total of 3,069 user reviews of Mobile Legends were processed using text mining techniques, including case folding, tokenizing, stopword removal, and stemming. The textual data were represented using TF-IDF weighting and classified using the Naïve Bayes algorithm. The risk categories included Account Security Risk, Data Privacy Risk, Phishing & Fraud Risk, Malware Risk, and Non-Security Issue. Evaluation using an 80:20 data split scenario resulted in an overall accuracy of 76.5% based on the confusion matrix, with performance variations across categories. The highest F1-score was achieved in the Non-Security Issue category (0.92), while the Malware Risk category showed the lowest performance (0.67) due to linguistic ambiguity in user narratives. These findings indicate that user reviews have the potential to serve as a community-based early detection mechanism for cybersecurity risks. Theoretically, this study introduces a community-based cyber risk identification approach as a complementary mechanism to technical detection systems in cybersecurity risk management for digital platforms.
Keywords: *cybersecurity; online games, text mining, naïve bayes, early risk detection*

1. PENDAHULUAN

Game online telah berkembang menjadi salah satu layanan digital dengan jumlah pengguna yang besar dan ekosistem yang kompleks. Selain berfungsi sebagai media hiburan, game online juga mengelola berbagai data sensitif pengguna, seperti informasi akun, data pribadi, serta transaksi digital melalui fitur in-app purchase. Kondisi ini menjadikan platform game online rentan terhadap berbagai ancaman keamanan siber, termasuk pencurian akun (account takeover), kebocoran data pribadi, phishing, penipuan transaksi digital, serta penyebaran malware yang menyamar sebagai aplikasi atau pembaruan game(Pongoh et al. 2024). Dalam praktiknya, tidak semua insiden keamanan siber pada game online dapat terdeteksi secara langsung melalui mekanisme teknis seperti log

system atau intrusion detection system. Banyak kasus keamanan siber justru pertama kali terungkap melalui laporan dan keluhan pengguna, baik melalui forum daring, media sosial, maupun ulasan aplikasi (Saputra et al. 2025). Pengguna sering melaporkan pengalaman seperti akun yang diambil alih, transaksi tidak sah, atau adanya tautan mencurigakan yang mengarah pada phishing dan penipuan digital. Oleh karena itu, laporan pengguna dapat dipandang sebagai indikasi awal (early warning) terhadap potensi risiko keamanan siber yang terjadi pada platform game online.

Penelitian sebelumnya banyak memanfaatkan teknik text mining dan analisis sentimen untuk mengevaluasi kepuasan pengguna, kualitas layanan, serta persepsi pengguna terhadap aplikasi digital (Triana et al. 2023). Namun, sebagian besar penelitian tersebut masih berfokus pada aspek non-keamanan, seperti performa aplikasi, fitur, dan pengalaman pengguna (Zy and Hadikristanto 2023). Kajian yang secara khusus memanfaatkan data tekstual ulasan pengguna untuk mengidentifikasi indikasi risiko keamanan siber masih relati f terbatas, terutama pada konteks game online. Berdasarkan kondisi tersebut, terdapat celah penelitian dalam pemanfaatan data ulasan pengguna sebagai sumber informasi untuk mendukung analisis keamanan siber. Ulasan pengguna belum banyak dimanfaatkan sebagai sumber data pendukung deteksi dini risiko keamanan siber, padahal data tersebut bersifat real-time dan mencerminkan pengalaman langsung pengguna terhadap potensi ancaman keamanan yang mereka hadapi (Arie, Suprio, and Najib 2023). Pendekatan ini tidak dimaksudkan untuk menggantikan mekanisme teknis keamanan siber, melainkan sebagai pelengkap dalam meningkatkan kewaspadaan dan respons awal terhadap risiko keamanan siber.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk melakukan deteksi dini indikasi risiko keamanan siber pada game online berdasarkan ulasan pengguna menggunakan pendekatan text mining dan algoritma Naïve Bayes. Fokus penelitian diarahkan pada klasifikasi ulasan yang mengandung indikasi ancaman keamanan siber, seperti pencurian akun, kebocoran data pribadi, phishing dan penipuan transaksi digital, serta malware, dengan memisahkan secara tegas antara isu keamanan dan keluhan non-keamanan(Hermawan and Hanif 2025) . Hasil penelitian diharapkan dapat memberikan kontribusi sebagai pendekatan pendukung dalam upaya peningkatan kesadaran dan mitigasi risiko keamanan siber pada ekosistem game online.

2. TINJAUAN PUSTAKA

2.1 Keamanan Siber pada Game Online

Game online merupakan bagian dari ekosistem layanan digital yang memiliki karakteristik khusus, seperti interaksi pengguna secara masif, konektivitas jaringan yang tinggi, serta integrasi dengan sistem pembayaran digital. Karakteristik tersebut menjadikan game online rentan terhadap berbagai ancaman keamanan siber. Beberapa ancaman yang umum terjadi pada platform game online meliputi pencurian dan pengambilalihan akun (account takeover), kebocoran data pribadi, phishing melalui pesan dalam game atau tautan eksternal, penipuan transaksi in-app purchase, serta penyebaran malware yang menyamar sebagai aplikasi atau pembaruan game (Riviani and Santoso 2025).

Ancaman keamanan siber pada game online tidak hanya berdampak pada kerugian finansial, tetapi juga dapat mengganggu kepercayaan pengguna terhadap platform (Praja, Nuruzzaman, and Sugiantoro 2025). Oleh karena itu, pengelola game online perlu memiliki mekanisme yang tidak hanya bersifat reaktif, tetapi juga mampu memberikan indikasi awal terhadap potensi risiko keamanan siber yang sedang berkembang.

2.2 Ulasan Pengguna sebagai Indikator Risiko Keamanan Siber

Ulasan pengguna (user reviews) merupakan salah satu bentuk user-generated content yang mencerminkan pengalaman langsung pengguna dalam menggunakan suatu aplikasi digital. Dalam konteks keamanan siber, ulasan pengguna dapat berfungsi sebagai sinyal awal terhadap adanya ancaman keamanan, terutama pada kasus-kasus yang melibatkan interaksi sosial dan manipulasi pengguna, seperti phishing dan penipuan digital (Iranda and Huda 2025).

Beberapa penelitian menunjukkan bahwa laporan pengguna sering kali menjadi sumber awal terungkapnya insiden keamanan siber, sebelum dilakukan investigasi teknis lebih lanjut. Oleh karena itu, analisis terhadap ulasan pengguna dapat dimanfaatkan sebagai pendekatan pendukung dalam proses identifikasi dan pemantauan risiko keamanan siber, khususnya pada platform dengan jumlah pengguna yang besar seperti game (Dwijaya and Laksito 2023).

2.3 Kerangka Kerja Keamanan Siber sebagai Dasar Kategori Risiko

Penentuan kategori risiko keamanan siber dalam penelitian ini didasarkan pada kerangka kerja keamanan siber yang telah diakui secara luas, khususnya NIST Cybersecurity Framework dan ISO/IEC 27001. Kedua kerangka kerja tersebut menyediakan pendekatan sistematis dalam mengidentifikasi, melindungi, mendeteksi, dan merespons ancaman keamanan siber terhadap aset informasi, termasuk data pengguna dan sistem digital (Prastya et al. 2024). Dengan mengacu pada framework standar ini, kategori risiko yang digunakan dalam penelitian tidak hanya bersifat konseptual, tetapi memiliki dasar teoritis dan praktis yang relevan dengan praktik keamanan siber modern.

Kategori Account Security Risk dalam penelitian ini selaras dengan domain Protect pada NIST Cybersecurity Framework serta kontrol keamanan akses pada ISO/IEC 27001, yang menekankan pentingnya pengelolaan identitas, autentikasi, dan kontrol akses untuk mencegah pengambilalihan akun dan akses tidak sah. Sementara itu, kategori Data Privacy Risk berkaitan dengan perlindungan informasi pribadi dan data sensitif pengguna, yang merupakan bagian inti dari pengelolaan aset informasi dalam ISO/IEC 27001 serta aspek Identify dan Protect pada NIST Cybersecurity Framework (Sheren et al. 2023).

Kategori Phishing & Fraud Risk dan Malware Risk dipetakan ke dalam domain Detect dan Respond pada NIST Cybersecurity Framework, yang menekankan kemampuan organisasi dalam mengidentifikasi ancaman siber dan merespons insiden secara tepat. Ancaman phishing, penipuan digital, serta malware yang menyamar sebagai aplikasi game atau konten terkait merupakan bentuk serangan siber yang umum terjadi pada platform digital dan memiliki potensi dampak signifikan terhadap keamanan akun dan data pengguna (Riadi et al. 2023). Oleh karena itu, pengelompokan risiko ini dipandang relevan dan konsisten dengan literatur keamanan siber yang kredibel.

Dengan mengadopsi kerangka kerja keamanan siber standar sebagai dasar penentuan kategori risiko, penelitian ini memastikan bahwa proses klasifikasi ulasan pengguna tidak dilakukan secara arbitrer. Pendekatan ini menegaskan bahwa kategori risiko keamanan siber yang digunakan memiliki keterkaitan langsung dengan konsep dan praktik keamanan siber yang diakui secara internasional, sekaligus memperkuat validitas konseptual penelitian dalam konteks deteksi dini indikasi risiko keamanan siber berbasis ulasan pengguna.

2.4 Text Mining dalam Analisis Keamanan Siber

Text mining merupakan teknik yang digunakan untuk mengekstraksi informasi dan pola dari data tekstual dalam jumlah besar. Dalam bidang keamanan siber, text mining telah banyak dimanfaatkan untuk analisis laporan insiden, deteksi phishing, analisis malware berbasis teks, serta pemantauan ancaman siber melalui media sosial dan forum daring. Penerapan text mining pada ulasan pengguna memungkinkan identifikasi kata kunci dan pola bahasa yang berkaitan dengan indikasi ancaman keamanan siber. Meskipun pendekatan ini tidak dapat menggantikan analisis teknis berbasis sistem, text mining berperan sebagai mekanisme pendukung deteksi dini yang dapat membantu meningkatkan kewaspadaan terhadap risiko keamanan siber (Reandito et al. 2024).

2.5 Klasifikasi Teks Menggunakan Naïve Bayes

Naïve Bayes merupakan salah satu algoritma klasifikasi probabilistik yang banyak digunakan dalam pengolahan teks karena kesederhanaan, efisiensi, dan performanya yang baik pada dataset berukuran besar. Algoritma ini bekerja berdasarkan Teorema Bayes dengan asumsi independensi antar fitur, sehingga cocok diterapkan pada representasi teks seperti bag-of-words atau TF-IDF. Dalam konteks penelitian ini, algoritma Naïve Bayes digunakan untuk mengklasifikasikan ulasan pengguna ke dalam kategori indikasi risiko keamanan siber dan non-keamanan. Pemilihan Naïve Bayes didasarkan pada kemampuannya dalam menangani data teks yang tidak terstruktur serta kebutuhan penelitian untuk memperoleh model klasifikasi yang interpretatif dan efisien sebagai bagian dari pendekatan deteksi dini (Pamuji 2022).

2.6 Penelitian Terkait

Beberapa penelitian sebelumnya telah memanfaatkan pendekatan text mining dan teknik klasifikasi untuk menganalisis ulasan pengguna serta opini publik pada berbagai platform digital. Pendekatan ini umumnya digunakan dalam konteks analisis sentimen dan evaluasi persepsi pengguna terhadap layanan digital, termasuk aspek keamanan secara umum.

Tabel 1 menyajikan ringkasan penelitian terkait yang menggunakan metode text mining dan algoritma Naïve Bayes pada berbagai objek penelitian. (Galena et al. 2024) mengkaji persepsi keamanan pengguna pada platform ecommerce dan menunjukkan bahwa pendekatan berbasis text mining efektif dalam mengolah data persepsi keamanan. Namun, penelitian tersebut masih terbatas pada klasifikasi sentimen dan belum memetakan tingkat risiko keamanan siber secara spesifik. Penelitian lain oleh (Kariman et al. 2025), (Mas'ud et al. 2024), dan (Pratama et al. 2024)juga menunjukkan efektivitas algoritma Naïve Bayes dalam mengklasifikasikan opini dan ulasan pengguna, tetapi fokus penelitian masih berada pada aspek sentimen dan kualitas layanan, tanpa mengaitkannya secara langsung dengan ancaman keamanan siber.

Tabel 1. Ringkasan penelitian terdahulu

Peneliti & Tahun	Judul & Metode Penelitian	Kelebihan Penelitian	Kekurangan Penelitian
Galena *et al.* (2024)	*Analisis Sentimen Masyarakat terhadap Keamanan Penggunaan E-Commerce B2C Menggunakan Naïve Bayes Berbasis Text Mining*	Penelitian mampu mengkaji persepsi keamanan pengguna secara sistematis serta membuktikan efektivitas Naïve Bayes dalam analisis teks terkait keamanan	Klasifikasi masih terbatas pada sentimen umum dan belum mengelompokkan tingkat risiko keamanan siber secara spesifik

Kariman *et al.* (2025)	*Analisis Sentimen TikTok Shop pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes*	Menghasilkan akurasi yang baik dan menunjukkan bahwa Naïve Bayes efektif digunakan pada data media sosial berbahasa Indonesia	Fokus penelitian hanya pada sentimen pengguna dan belum membahas risiko keamanan siber
Mas'ud *et al.* (2024)	*Analisis Sentimen Deepseek Berdasarkan Ulasan Google Play Store Menggunakan Metode Naïve Bayes*	Menggunakan dataset ulasan yang besar serta evaluasi model yang sistematis	Penelitian belum memetakan tingkat risiko keamanan siber dan masih berfokus pada klasifikasi sentimen
Pratama *et al.* (2024)	*Analisis Sentimen Kebijakan Pembelian Gas 3 Kg Menggunakan Metode Naïve Bayes*	Menunjukkan kemampuan Naïve Bayes dalam menganalisis opini publik pada isu kebijakan	Objek penelitian bukan aplikasi digital dan tidak membahas aspek keamanan siber

Berdasarkan ringkasan penelitian pada Tabel 1, dapat disimpulkan bahwa sebagian besar penelitian terdahulu masih memposisikan text mining sebagai alat untuk analisis sentimen dan evaluasi persepsi pengguna, bukan sebagai sarana untuk mengidentifikasi indikasi risiko keamanan siber secara lebih spesifik. Kategori keamanan yang digunakan umumnya bersifat umum dan belum mencakup klasifikasi ancaman seperti pencurian akun, kebocoran da ta pribadi, phishing, penipuan transaksi digital, atau malware.

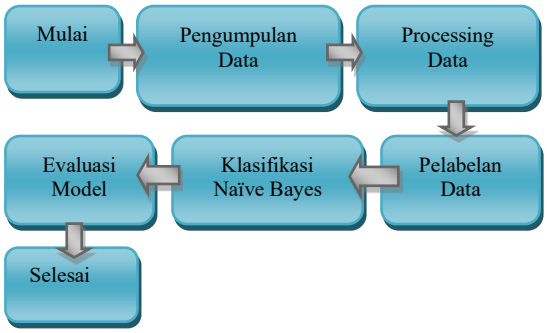
Dengan demikian, terdapat celah penelitian dalam pemanfaatan ulasan pengguna sebagai sumber data untuk deteksi dini indikasi risiko keamanan siber, khususnya pada konteks game online. Penelitian ini berupaya mengisi celah tersebut dengan mengombinasikan pendekatan text mining dan klasifikasi Naïve Bayes untuk memetakan indikasi risiko keamanan siber secara lebih terstruktur dan terfokus, sehingga dapat mendukung upaya peningkatan kewaspadaan dan mitigasi risiko keamanan siber pada ekosistem game online.

3. METODE PENELITIAN

3.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode text mining dan klasifikasi untuk melakukan deteksi dini indikasi risiko keamanan siber berdasarkan ulasan pengguna game online. Tahapan penelitian secara umum meliputi pengumpulan data, praproses data teks, pelabelan data berdasarkan kategori risiko keamanan siber, ekstraksi fitur, proses klasifikasi menggunakan algoritma Naïve Bayes, serta evaluasi kinerja model (Darmawan, Alam, and Sulistyo 2023).

Pendekatan ini dirancang sebagai mekanisme pendukung deteksi dini yang memanfaatkan laporan pengguna sebagai sumber sinyal awal terhadap potensi ancaman keamanan siber, dan tidak dimaksudkan untuk menggantikan analisis teknis atau forensik digital berbasis sistem.



Gambar 1. Alur Penelitian

3.2 Pengumpulan Data

Data penelitian berupa ulasan pengguna aplikasi Mobile Legends yang diperoleh dari platform Google Play Store. Proses pengumpulan data dilakukan dengan teknik web scraping menggunakan kata kunci yang relevan dengan pengalaman penggunaan aplikasi (Prasetyo, Ridwan, and Voutama 2024). Total data yang berhasil dikumpulkan sebanyak 3.069 ulasan pengguna, yang mencerminkan berbagai pengalaman dan laporan pengguna terkait penggunaan game online. Data ulasan yang dikumpulkan mencakup teks ulasan tanpa menyertakan informasi identitas pribadi pengguna, sehingga penelitian ini tidak melanggar aspek privasi data.

3.3 Praproses Data Teks

Tahap praproses data dilakukan untuk meningkatkan kualitas data teks sebelum dilakukan proses klasifikasi.

Tahapan praproses yang diterapkan meliputi:

1. Case folding, untuk mengubah seluruh teks menjadi huruf kecil.
2. Tokenizing, untuk memecah teks menjadi kata-kata.
3. Stopword removal, untuk menghapus kata-kata umum yang tidak memiliki makna signifikan.
4. Stemming, untuk mengubah kata ke bentuk dasarnya.

Praproses data ini bertujuan untuk mengurangi noise dan meningkatkan representasi fitur teks yang relevan dengan indikasi risiko keamanan siber.

3.4 Proses Pelabelan dan Relabeling Data

Proses pelabelan data dalam penelitian ini dilakukan melalui dua tahapan utama, yaitu

pelabelan awal berdasarkan tingkat keparahan risiko dan relabeling ke dalam kategori risiko keamanan siber yang lebih spesifik. Pendekatan dua tahap ini diterapkan untuk menjaga konsistensi dengan naskah sebelumnya sekaligus meningkatkan relevansi kategorisasi terhadap konteks keamanan siber.

Pada tahap awal, ulasan pengguna dilabeli secara manual ke dalam empat kategori umum, yaitu High Risk, Medium Risk, Low Risk, dan No Risk. Pelabelan awal ini didasarkan pada tingkat keparahan permasalahan yang diindikasikan dalam teks ulasan, tanpa memisahkan secara rinci jenis ancaman keamanan siber yang terkandung di dalamnya. Kategori High Risk dan Medium Risk mencerminkan ulasan yang menunjukkan potensi dampak signifikan terhadap keamanan akun, data, atau transaksi digital, sedangkan Low Risk dan No Risk digunakan untuk ulasan dengan dampak yang terbatas atau tidak berkaitan dengan keamanan siber.

Selanjutnya, dilakukan proses relabeling untuk memetakan kategori risiko umum tersebut ke dalam lima kategori risiko keamanan siber yang lebih spesifik, yaitu Account Security Risk, Data Privacy Risk, Phishing & Fraud Risk, Malware Risk, dan Non-Security Issue. Proses relabeling dilakukan dengan menelaah kembali konteks teks ulasan, kata kunci, serta indikasi ancaman yang dilaporkan oleh pengguna. Ulasan yang pada pelabelan awal termasuk dalam kategori High Risk dan Medium Risk dipetakan ke dalam kategori Account Security Risk, Data Privacy Risk, Phishing & Fraud Risk, atau Malware Risk, tergantung pada jenis ancaman keamanan siber yang diindikasikan. Sementara itu, ulasan yang sebelumnya dilabeli sebagai Low Risk dan No Risk dievaluasi kembali dan dipetakan ke dalam kategori Non-Security Issue apabila tidak ditemukan indikasi risiko keamanan siber yang relevan.

Secara operasional, ulasan yang mengindikasikan pengambilalihan akun, kehilangan akses, atau aktivitas login yang mencurigakan diklasifikasikan sebagai Account Security Risk. Ulasan yang menyebutkan dugaan kebocoran, penyalahgunaan, atau eksposur data pribadi pengguna dipetakan ke dalam Data Privacy Risk. Ulasan yang melaporkan adanya tautan palsu, penipuan transaksi digital, atau manipulasi pembayaran in-app purchase dikategorikan sebagai Phishing & Fraud Risk. Sementara itu, ulasan yang mengindikasikan keberadaan aplikasi palsu, file berbahaya, atau gangguan sistem akibat instalasi dari sumber tidak resmi diklasifikasikan sebagai Malware Risk. Ulasan yang hanya berkaitan dengan performa permainan, gangguan jaringan, bug, atau sistem matchmaking tetap dikategorikan sebagai Non-Security Issue dan digunakan sebagai kelas pembanding.

Dengan adanya proses relabeling ini, kategori risiko keamanan siber yang digunakan dalam penelitian tidak hanya merepresentasikan tingkat keparahan, tetapi juga jenis ancaman yang lebih spesifik dan relevan dengan konteks keamanan siber. Pendekatan ini memastikan bahwa proses klasifikasi dilakukan secara sistematis dan transparan, sekaligus memperkuat keterkaitan antara tujuan penelitian, metodologi, dan hasil yang diperoleh.

3.5 Ekstraksi Fitur

Setelah proses pelabelan, data teks diubah ke dalam bentuk numerik menggunakan metode Term Frequency– Inverse Document Frequency (TF-IDF). Metode ini digunakan untuk merepresentasikan tingkat kepentingan suatu kata dalam dokumen relatif terhadap keseluruhan dataset. Representasi TF-IDF dipilih karena efektif dalam menangkap kata-kata kunci yang berkaitan dengan indikasi risiko keamanan siber pada data teks yang tidak terstruktur.

3.6 Klasifikasi Menggunakan Naïve Bayes

Proses klasifikasi dilakukan menggunakan algoritma Naïve Bayes, yang merupakan algoritma probabilistik berbasis Teorema Bayes dengan asumsi independensi antar fitur. Algoritma ini dipilih karena memiliki performa yang baik pada klasifikasi teks, efisiensi komputasi yang tinggi, serta kemampuannya dalam menangani dataset berukuran besar. Dalam penelitian ini, Naïve Bayes digunakan untuk mengklasifikasikan ulasan pengguna ke dalam kategori indikasi risiko keamanan siber dan non-keamanan berdasarkan fitur TF-IDF yang telah diekstraksi.

3.7 Pembagian Data dan Evaluasi Model

Dataset penelitian berjumlah 3.069 ulasan dengan distribusi kelas yang tidak seimbang (class imbalance), di mana kategori Non-Security Issue mendominasi sebesar 79,8% dari keseluruhan data. Kondisi ini berpotensi menyebabkan bias klasifikasi apabila evaluasi dilakukan secara langsung menggunakan pembagian data konvensional, karena model dapat mencapai akurasi tinggi hanya dengan memprediksi kelas mayoritas. Pada tahap awal, pembagian data dilakukan menggunakan skenario 80:20 sebagai praktik umum dalam penelitian text mining. Namun, untuk memperoleh evaluasi performa yang lebih adil dan representatif terhadap seluruh kategori risiko keamanan siber, pengujian akhir dilakukan menggunakan balanced test set. Data uji disusun dengan mengambil 55 ulasan dari masing-masing kategori risiko keamanan siber (Account Security Risk, Data Privacy Risk, Phishing & Fraud Risk, dan Malware Risk) serta 150 ulasan dari kategori Non-Security Issue sebagai kelas mayoritas, sehingga total data uji yang digunakan dalam evaluasi berjumlah 370 ulasan. Pendekatan ini dipilih untuk mengurangi bias akibat dominasi kelas mayoritas dan memungkinkan analisis performa model pada kategori risiko minoritas secara lebih proporsional. Evaluasi kinerja model dilakukan

menggunakan metrik accuracy, precision, recall, dan F1-score untuk memberikan pengukuran yang komprehensif terhadap kemampuan klasifikasi. Penggunaan balanced evaluation pada dataset yang imbalanced ini merupakan praktik metodologis yang umum dalam penelitian klasifikasi teks guna mencegah terjadinya inflated accuracy dan memastikan bahwa performa model terhadap kelas minoritas tetap terukur secara objektif.

4. HASIL DAN PEMBAHASAN

4.1 Pengambilan Data

Data yang digunakan dalam penelitian ini diperoleh dari dataset sekunder berupa ulasan pengguna aplikasi Mobile Legends yang diambil dari platform Google Play Store. Dataset tersebut berisi komentar pengguna yang merepresentasikan pengalaman, keluhan, serta tanggapan terhadap layanan aplikasi digital. Dalam konteks keamanan siber, ulasan pengguna tidak hanya mencerminkan tingkat kepuasan layanan, tetapi juga dapat berfungsi sebagai laporan awal berbasis pengguna (user-reported incidents) yang mengindikasikan potensi risiko keamanan siber, seperti pencurian akun, pemblokiran akun, manipulasi sistem, maupun gangguan layanan digital.

Data dikumpulkan dalam format berkas CSV dan digunakan sebagai sumber utama dalam proses analisis risiko keamanan siber berbasis text mining. Jumlah keseluruhan data yang diperoleh sebanyak 3.069 ulasan pengguna dan seluruh data tersebut dipertahankan tanpa dilakukan penghapusan pada tahap awal untuk menjaga representasi alami distribusi opini pengguna. Dataset ini menjadi dasar dalam proses pelabelan, pemodelan, serta evaluasi klasifikasi yang dilakukan pada tahapan selanjutnya.

Tabel 2. Contoh data mentah ulasan pengguna

No	User Name	Review Teks	Rating	Date
1	Pengguna Google	DRAK SISTEM SEMUA ISI SATU TEAM SINYAL NGELAG NGELEG PADAHAL BUAT BUKA APK LAIN LANCAR JAYA EMANG GAME SESAT INI LAMA LAMA	1	23/08/2025 12:39
2	Pengguna Google	Makin kesini makin rusak ni game loading stuck ga masuk2 game sinyal di game jadi jelek padahal sinyal bagus, padahal sebelumnya gk prnh kyk gini makin di update makin rusak sinyal di ingame	2	23/08/2025 11:34
3	Pengguna Google	dpt tim yg ga bener mulu ,sistem ny ga adil pdhl satu bintang lagi naik glory -,-	3	23/08/2025 10:58
4	Pengguna Google	MOHON DIPERBAIKI SAYA UDAH 14 PERTANDINGAN SINYAL HIJAU TAPI NGE LAG DAN ITU KALAH SEMUA MOHON DIPERBAIKI DARI DULU KAYA GINI , TAPI LAMA KELAMAAN JADI KESEL. MOHON DIPERBAIKI DAN KASIH SAYA MENANG	5	23/08/2025 07:28
5	Pengguna Google	Game ga guna Masa ngelag padahal pake wifi hadeh	1	23/08/2025 05:39
6	Pengguna Google	woi monton babi minimal kasih tim dan lawan yang sepadan su	1	23/08/2025 05:20
7	Pengguna Google	saya kecewa terhadap system moonton karna akun saya di ban padahal saya tidak melakukan kesalahan apa pun. kalo tau gini mending main honor of king	1	23/08/2025 04:22
8	Pengguna Google	gak adil ngasih tim nya	1	23/08/2025 03:32
9	Pengguna Google	montoon baik kasih lah tim setara Jan tim bot Mulu	1	23/08/2025 03:21
10	Pengguna Google	Nemu Tim Bot Mulu Sampek Lostrike 10 Kali Ini Sih Bukan Games Hiburan Games Penyiksa Batin	1	23/08/2025 02:10
11	Pengguna Google	sya ksih 4 bintang dulu ya kalau saya Uda naik ke level atas sya kasih lgiðŸ™□	4	23/08/2025 01:33
12	Pengguna Google	asli game apa ini Cok gw udh Lose 12 kali di kasih tim gak bener Mulu sumpah tim ngetrol lah nge feed lah.atau gw nih kan udh glory tapi satu tim Ama legend 2 nah logikanya dimana plis lah sistem pemilihan tim di benerin lah atau kalo gak ban aja tuh orgÂ² yg suka ngetrol ngefeed capek Cok dari glory bintang 55 ke honor bintang 44.	1	23/08/2025 00:17
13	Pengguna Google	udahlah tim nya dongo terus ini tier atas udah ga ada harga dirinya asli	5	23/08/2025 00:12
14	Pengguna Google	GEm jelek apa bagusan nya bgusss honor of king lagi	1	22/08/2025 22:22
15	Pengguna Google	ini game sangat bagus, saya bisa winstreak 13x dan sekarang bintang saya sudah 43, semoga dalam 3 hari kedepan saya bisa mencapai 50 bintang terima kasih moontoon	5	22/08/2025 22:14

Seluruh data tersaji dalam format terstruktur yang mencakup nama pengguna, teks ulasan, rating, dan tanggal unggahan, sehingga memudahkan proses pra-pemrosesan dan ekstraksi fitur sebelum dilakukan klasifikasi risiko keamanan siber.

4.2 Pra-pemrosesan

Tahap pra-pemrosesan dilakukan untuk membersihkan dan menyiapkan data teks agar dapat dianalisis secara optimal menggunakan algoritma Naïve Bayes. Tahapan ini penting untuk mengurangi noise linguistik, menstandarkan representasi teks, serta meningkatkan kualitas fitur yang digunakan dalam proses klasifikasi risiko keamanan siber.

4.2.1 Pembersihan

Pembersihan data dilakukan dengan menghilangkan bagian teks yang tidak diperlukan, seperti URL, penanda akun pengguna (@username), tanda kutip, karakter tidak terbaca, serta karakter selain huruf yang tidak memiliki pengaruh terhadap proses analisis

Tabel 3. Hasil dari pembersihan data

Text_Lengkap	Pembersihan
@agus03 matchmaking jadi lebih lama dari sebelum ada pra pilih Lane, tolong sistem pilih Lane nya seperti yang lama😭😭😭	matching jadi lebih lama dari sebelumnya ada pra pilih lane tolong sistem pilih lane nya seperti yang lama
@dwi7890 Jan kecanduan ni game kalo kalah mulu😂😂😂😂😂😂😂😂😂	jan kecanduan ni game kalo kalah mulu
@ikbal771game nya seru bagus 😀👍	game nya seru bagus
@marko95Saya sangat menyukai game mobile legend ini😍😍	saya sangat menyukai game mobile legend ini
@lance11game suka ngelag² gak jelasssss	game suka ngelag gak jelas

4.2.2 Penyeragaman huruf

Penyeragaman huruf dilakukan dengan mengonversi seluruh teks ulasan ke dalam bentuk huruf kecil (lowercase). Tahap ini bertujuan untuk menghindari perbedaan makna akibat variasi penggunaan huruf besar dan kecil sehingga proses analisis dapat dilakukan secara konsisten. Penyeragaman huruf penting dalam pendekatan berbasis frekuensi kata seperti Naïve Bayes, karena variasi kapitalisasi dapat menghasilkan token yang berbeda meskipun memiliki makna yang sama.

Tabel 4. Hasil dari penyeragaman huruf

Pembersihan	Penyeragaman huruf
Ngasih tim yg GK adil dan suka lag	ngasih tim yg gk adil dan suka lag
masih oke cuman suka NGELAGGGGGG!	ngasih tim yg gk adil dan suka lag masih oke cuman suka ngelaggggggg!
woi montoon kenapa lukasih tim busuk Mulu	woi montoon kenapa lu kasih tim busuk. mulu woi rank
woi Rank Saya naik turun terus Temen temen saya pada naik honor KLO kamu masih kasih sistem naik turun terus. saya kasih bintang 1	saya naik turun terus temen temen saya pada naik honor klo kamu masih kasih sistem naik turun terus saya kasih bintang 1
bintang SATU SETENGAH SAJA semakin sulit menikmati game ini	bintang satu setengah saja semakin sulit menikmati game ini
GAME SETTINGAN MASA KALO KITA LAGI WINN TRUS DI KASIH TIM YANG BENER KALO LAGI LOSE TRUS DI KETEMUIN AMA LEGEND LAWAN NYA GLORY KAN GILLA YAA	game settingan masa kalo kita lagi winn trus di kasih tim yang bener kalo lagi lose trus di ketemuin ama legend lawan nya glory kan gilla yaa

4.2.3 Ulasan Berdasarkan Kategori Risiko Keamanan Siber

Sebagai upaya untuk memperjelas perbedaan antara ulasan yang mengandung indikasi risiko keamanan siber dan keluhan teknis umum, penelitian ini menyajikan contoh ulasan pengguna untuk setiap kategori risiko keamanan siber yang digunakan dalam proses klasifikasi. Penyajian contoh ini bertujuan untuk memberikan gambaran konkret mengenai karakteristik teks yang merepresentasikan masing-masing kategori risiko.

Tabel 5 menampilkan contoh ulasan pengguna yang diklasifikasikan ke dalam kategori Account Security Risk, Data Privacy Risk, Phishing & Fraud Risk, Malware Risk, serta Non-Security Issue. Setiap contoh dipilih berdasarkan adanya kata kunci, konteks, dan pola bahasa yang secara eksplisit mengindikasikan jenis risiko keamanan siber tertentu, seperti kehilangan akses akun, dugaan kebocoran data pribadi, tautan mencurigakan yang mengarah pada penipuan digital, serta gangguan sistem akibat aplikasi pihak ketiga. Sementara itu, ulasan yang hanya berkaitan dengan performa permainan, jaringan, atau sistem matchmaking diklasifikasikan sebagai Non-Security Issue. Proses pembersihan dilakukan secara konsisten terhadap seluruh dataset sebelum tahap pembagian data pelatihan dan pengujian untuk menjaga keseragaman transformasi teks.

Perlu ditegaskan bahwa seluruh contoh ulasan pada Tabel 5 merupakan kutipan langsung dari dataset asli tanpa dilakukan parafrase, penyuntingan struktur kalimat, maupun perapihan bahasa. Perbedaan tingkat kerapian bahasa dibandingkan sebagian data mentah pada Tabel 2 terjadi secara alami karena Tabel 2 menampilkan sampel acak ulasan pengguna dengan variasi kesalahan penulisan yang tinggi, sedangkan Tabel 5 menampilkan contoh representatif dari masing-masing kategori risiko yang secara kontekstual lebih eksplisit menggambarkan indikasi keamanan siber.

Tabel 5. Contoh Ulasan Pengguna untuk Setiap Kategori Risiko Keamanan Siber

Kategori Risiko	Contoh Ulasan Pengguna
Account Security Risk	"Akun saya tiba-tiba tidak bisa login, padahal email dan password tidak pernah saya bagikan. Mohon dicek, akun saya seperti diretas."
Data	"Setelah install game ini, banyak spam SMS

Privacy Risk	dan email aneh masuk. Sepertinya data saya bocor."
Phishing & Fraud Risk	"Ada link event gratis diamond yang mengarah ke website lain dan akun saya malah diambil."
Malware Risk	"Setelah download APK Mobile Legends dari luar Play Store, HP saya jadi sering error dan muncul aplikasi aneh."
Non-Security Issue	"Game sering lag dan matchmaking tidak adil walaupun sinyal bagus."

Penyajian contoh ulasan ini menegaskan bahwa kategori risiko keamanan siber dalam penelitian ini tidak ditentukan secara abstrak, melainkan didasarkan pada indikasi nyata yang disampaikan oleh pengguna. Dengan demikian, ulasan pengguna dapat dipahami sebagai sumber informasi awal yang relevan untuk mendukung deteksi dini indikasi risiko keamanan siber pada ekosistem game online, sekaligus membedakannya secara tegas dari keluhan teknis non-keamanan. Dataset yang telah melalui tahap pra-pemrosesan selanjutnya digunakan dalam proses pelabelan dan pembagian data untuk tahap pelatihan serta evaluasi model sebagaimana dijelaskan pada bagian berikutnya.

4.3 Pelabelan data keamanan siber

Setelah melalui tahap pra-pemrosesan, data ulasan diberi label berdasarkan indikasi risiko keamanan siber yang terkandung dalam teks. Proses pelabelan dilakukan secara manual berbasis aturan (rule-based labeling) dengan mengacu pada definisi operasional masing-masing kategori risiko yang disusun berdasarkan literatur keamanan siber dan konteks insiden digital pada platform game daring. Setiap ulasan dianalisis dengan mempertimbangkan konteks kalimat, kata kunci, serta indikasi eksplisit yang mengarah pada potensi risiko keamanan, sehingga proses pelabelan tidak hanya bergantung pada kemunculan kata tertentu, tetapi juga pada makna semantik yang terkandung dalam teks.
Kategori risiko yang digunakan meliputi:

1. *Account Security Risk* - indikasi pencurian akun, akun diblokir, atau penyalahgunaan akses.
2. *Data Privacy Risk* - indikasi kebocoran atau penyalahgunaan data pribadi.
3. *Phishing & Fraud Risk* - indikasi penipuan, manipulasi transaksi, atau rekayasa sosial.
4. *Malware Risk* - indikasi gangguan sistem atau aplikasi mencurigakan.
5. *Non-Security Issue* - keluhan nonkeamanan seperti performa dan matchmaking.

Distribusi hasil pelabelan menunjukkan bahwa kategori Non-Security Issue mendominasi dataset, sementara kategori risiko keamanan siber memiliki proporsi yang relatif lebih kecil. Kondisi ini mencerminkan karakteristik umum fenomena keamanan siber, di mana insiden bersifat jarang namun berdampak tinggi (rare but high-impact events), serta menunjukkan adanya ketidakseimbangan distribusi kelas dalam dataset.

Tabel 6. Distribusi Ulasan Berdasarkan Kategori Risiko Keamanan Siber

Kategori	Jumlah Ulasan	Persentase (%)
Account Security Risk	215	7,0
Data Privacy Risk	148	4,8
Phishing & Fraud Risk	159	5,2
Malware Risk	97	3,2
Non-Security Issue	2.450	79,8
Total	**3.069**	**100**

Distribusi alami yang tidak seimbang ini menjadi pertimbangan metodologis dalam strategi evaluasi model untuk memastikan bahwa kinerja klasifikasi tidak terdistorsi oleh dominasi kelas mayoritas, sebagaimana dijelaskan pada bagian evaluasi model.

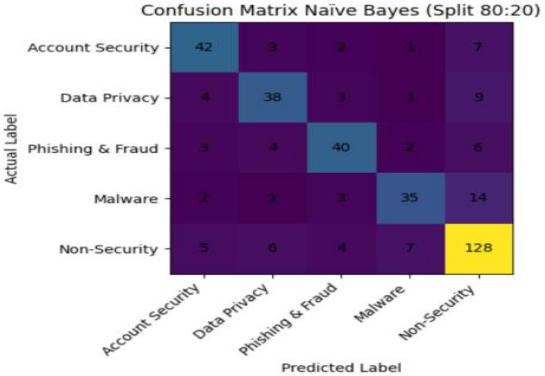
4.4 Klasifikasi Risiko Keamanan Siber Menggunakan Naïve Bayes

Proses klasifikasi dilakukan menggunakan algoritma Naïve Bayes. Model dilatih menggunakan skema pembagian data 80% sebagai data latih, sedangkan evaluasi kinerja dilakukan menggunakan balanced test set yang disusun secara proporsional dari masing-masing kategori risiko keamanan siber. Pendekatan ini diterapkan untuk memastikan bahwa pengukuran performa model tidak terdistorsi oleh dominasi kelas mayoritas dan mampu merepresentasikan kemampuan klasifikasi secara lebih adil pada seluruh kategori risiko.

4.4.1 Confusion Matrix Model Naïve Bayes

Tabel 7. Confusion Matrix Klasifikasi Risiko Keamanan Siber

Aktual \ Prediksi	Account Security	Data Privacy	Phishing & Fraud	Malware	Non-Security
Account Security	42	3	2	1	7
Data Privacy	4	38	3	1	9
Phishing & Fraud	3	4	40	2	6
Malware	2	1	3	35	14
Non-Security	5	6	4	7	128



Gambar 2. Confusion Matrix Naïve Bayes

Berdasarkan Tabel 7 dan Gambar 2, hasil confusion matrix menunjukkan bahwa model Naïve Bayes memiliki kemampuan klasifikasi yang relatif konsisten antar kategori risiko keamanan siber. Jika dihitung berdasarkan total prediksi benar dibagi jumlah data uji pada balanced test set, yaitu (42 + 38 + 40 + 35 + 128) / 370, maka diperoleh akurasi keseluruhan model sebesar 76,5%. Nilai ini merepresentasikan performa agregat model dalam mengklasifikasikan seluruh kategori risiko secara proporsional.

Pada kategori Non-Security Issue, model berhasil mengklasifikasikan 128 data secara benar, sementara kesalahan klasifikasi tersebar dalam jumlah yang relatif terbatas ke kategori lainnya. Hal ini menunjukkan bahwa karakteristik linguistik pada ulasan non-keamanan cukup stabil dan dapat dikenali dengan baik oleh model.

Pada kategori Account Security Risk, sebanyak 42 data berhasil diprediksi dengan benar, sedangkan sebagian lainnya mengalami salah klasifikasi, dengan kesalahan terbesar mengarah ke kategori Non-Security Issue. Temuan ini menunjukkan bahwa sebagian laporan terkait keamanan akun menggunakan pola bahasa yang menyerupai keluhan teknis umum, sehingga menimbulkan ambiguitas dalam proses klasifikasi berbasis teks.

Kategori Phishing & Fraud Risk menunjukkan hasil yang relatif stabil dengan 40 data terklasifikasi secara tepat. Kesalahan klasifikasi yang terjadi cenderung tersebar dan tidak terpusat pada satu kategori tertentu, yang mengindikasikan bahwa indikasi penipuan dan rekayasa sosial dalam ulasan pengguna memiliki ciri linguistik yang cukup eksplisit dan dapat dikenali oleh model.

Sebaliknya, kategori Malware Risk masih menunjukkan tantangan dalam proses klasifikasi. Dari total data pada kategori ini, 35 data berhasil diklasifikasikan secara benar, sementara sebagian lainnya salah diprediksi sebagai Non-Security Issue. Hal ini mengindikasikan adanya kemiripan pola bahasa antara laporan gangguan teknis dan indikasi malware, sehingga menyulitkan model dalam membedakan kedua kategori tersebut secara tegas.

Secara keseluruhan, hasil confusion matrix menegaskan bahwa variasi performa antar kategori lebih dipengaruhi oleh tingkat eksplisitnya indikasi risiko dalam teks ulasan serta kemiripan semantik antar kategori, dibandingkan oleh faktor distribusi jumlah data. Dengan demikian, tantangan utama dalam klasifikasi risiko keamanan siber berbasis ulasan pengguna terletak pada kompleksitas representasi bahasa alami. Oleh karena itu, hasil ini memperkuat temuan bahwa ulasan pengguna dapat dimanfaatkan sebagai indikator awal risiko keamanan siber, meskipun masih diperlukan pengembangan lanjutan untuk meningkatkan sensitivitas model terhadap kategori risiko yang memiliki tingkat ambiguitas linguistik tinggi.

4.4.2 Evaluasi Kinerja Model Berdasarkan Presisi, Recall, dan F1-Score

Untuk memperoleh gambaran yang lebih komprehensif mengenai kinerja model Naïve Bayes, evaluasi tidak hanya dilakukan berdasarkan tingkat akurasi, tetapi juga menggunakan metrik precision, recall, dan F1-score pada setiap kategori risiko keamanan siber. Pendekatan ini memungkinkan analisis performa model secara lebih rinci pada masing-masing kategori, sehingga tidak hanya bergantung pada nilai akurasi agregat.

Tabel 8. Hasil Evaluasi Presisi, Recall, dan F1-Score per Kategori Risiko Keamanan Siber

Kategori Risiko	Precision	Recall	F1-Score
Account Security Risk	0,80	0,76	0,78
Data Privacy Risk	0,75	0,69	0,72
Phishing & Fraud Risk	0,83	0,73	0,78
Malware Risk	0,70	0,64	0,67
Non-Security Issue	0,90	0,95	0,92

Berdasarkan Tabel 8, kategori Non-Security Issue menunjukkan performa terbaik dengan nilai precision sebesar 0,90, recall 0,95, dan F1-score 0,92. Nilai recall yang tinggi menunjukkan bahwa karakteristik linguistik pada kategori ini relatif lebih mudah dikenali oleh model dibandingkan kategori lainnya. Perlu ditegaskan bahwa nilai 0,92 tersebut merupakan F1-score untuk kategori Non-Security Issue, bukan nilai akurasi keseluruhan model. Berdasarkan perhitungan confusion matrix pada balanced test set, akurasi total model adalah sebesar 76,5%. Dengan demikian, angka 92% merepresentasikan performa spesifik pada satu kategori, sedangkan akurasi agregat model secara keseluruhan adalah 76,5%.

Pada kategori Account Security Risk, nilai precision sebesar 0,80 dan recall 0,76 menghasilkan F1-score sebesar 0,78. Nilai ini menunjukkan keseimbangan yang cukup baik antara ketepatan prediksi dan kemampuan menjangkau data pada kategori tersebut, meskipun masih terdapat kesalahan klasifikasi akibat ambiguitas konteks bahasa.

Kategori Phishing & Fraud Risk menunjukkan precision sebesar 0,83 dan recall 0,73 dengan F1-score 0,78. Nilai precision yang relatif tinggi mengindikasikan bahwa ketika model memprediksi adanya indikasi phishing atau penipuan, prediksi tersebut umumnya benar. Namun demikian, sebagian indikasi masih belum sepenuhnya teridentifikasi sebagaimana tercermin pada nilai recall.

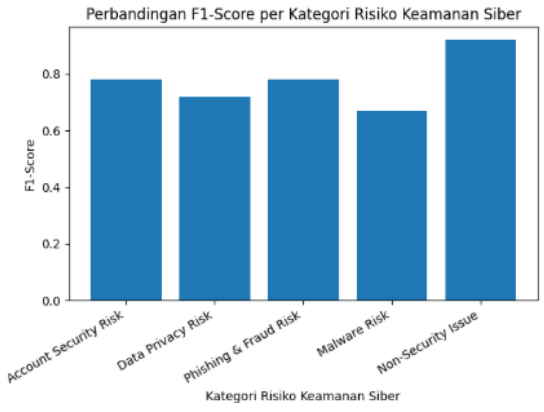
Kategori Data Privacy Risk memiliki F1-score sebesar 0,72, sedangkan Malware Risk menunjukkan nilai terendah sebesar 0,67. Rendahnya performa pada kategori Malware Risk menunjukkan adanya tantangan dalam membedakan indikasi malware

dengan keluhan teknis non-keamanan akibat kemiripan pola bahasa yang digunakan oleh pengguna.

Secara keseluruhan, evaluasi berbasis presisi, recall, dan F1-score menegaskan bahwa interpretasi performa model perlu mempertimbangkan analisis per kategori, bukan hanya nilai agregat. Variasi performa antar kategori lebih mencerminkan kompleksitas linguistik dalam ulasan pengguna dibandingkan faktor distribusi kelas pada tahap evaluasi. Fenomena ini sejalan dengan karakteristik umum dalam deteksi ancaman keamanan siber, di mana indikasi risiko yang berdampak tinggi sering kali muncul dalam bentuk ekspresi yang implisit dan kontekstual. Oleh karena itu, hasil ini menguatkan posisi penelitian sebagai pendekatan deteksi dini berbasis persepsi pengguna terhadap potensi risiko keamanan siber, bukan sebagai sistem deteksi teknis insiden keamanan secara langsung.

4.4.3 Analisis F1-Score per Kategori Risiko Keamanan Siber

Gambar 3 menampilkan perbandingan nilai F1-score untuk setiap kategori risiko keamanan siber hasil klasifikasi menggunakan algoritma *Naïve Bayes*. F1-score digunakan sebagai metrik utama karena merepresentasikan keseimbangan antara presisi dan recall pada masing-masing kategori, sehingga mampu memberikan gambaran performa model secara lebih komprehensif dibandingkan hanya mengandalkan akurasi agregat. Dalam konteks klasifikasi multi-kategori, metrik ini penting untuk memastikan bahwa kemampuan model tidak hanya baik dalam mengidentifikasi prediksi yang benar, tetapi juga konsisten dalam meminimalkan kesalahan klasifikasi pada setiap jenis risiko.



Gambar 3. Perbandingan F1-score per kategori risiko keamanan siber

Berdasarkan grafik tersebut, kategori Non-Security Issue memiliki nilai F1-score tertinggi sebesar 0,92, sedangkan kategori Malware Risk memiliki nilai terendah sebesar 0,67. Nilai F1-score pada kategori Account Security Risk dan Phishing & Fraud Risk masing-masing sebesar 0,78, menunjukkan bahwa indikasi risiko yang bersifat eksplisit relatif lebih mudah dikenali oleh model dibandingkan risiko yang memiliki ambiguitas linguistik tinggi. Tingginya performa pada kategori Non-Security Issue juga mengindikasikan bahwa model mampu membedakan dengan cukup baik antara ulasan yang tidak mengandung indikasi risiko dan ulasan yang mengandung potensi ancaman keamanan.

Sebaliknya, performa yang lebih rendah pada kategori Malware Risk mengindikasikan adanya tantangan dalam mengidentifikasi istilah atau frasa yang bersifat implisit, teknis, atau kontekstual. Ulasan pengguna sering kali tidak menggunakan terminologi keamanan siber secara langsung, melainkan mengekspresikan pengalaman dalam bentuk keluhan umum atau deskripsi gejala aplikasi, sehingga meningkatkan kompleksitas klasifikasi.

Perbedaan performa antar kategori lebih dipengaruhi oleh karakteristik semantik dan tingkat eksplisitnya indikasi risiko dalam teks ulasan. Temuan ini menegaskan bahwa pendekatan klasifikasi berbasis ulasan pengguna memiliki potensi sebagai mekanisme deteksi dini (early warning system) risiko keamanan siber, dengan ruang pengembangan lanjutan pada peningkatan sensitivitas terhadap kategori yang memiliki tingkat ambiguitas tinggi serta optimalisasi representasi fitur linguistik.

4.5 Pembahasan

Hasil penelitian ini menunjukkan bahwa pendekatan text mining terhadap ulasan pengguna game online dapat dimanfaatkan sebagai mekanisme deteksi dini indikasi risiko keamanan siber dalam ekosistem aplikasi digital. Berbeda dengan penelitian terdahulu yang umumnya memanfaatkan ulasan pengguna untuk analisis sentimen atau evaluasi kualitas layanan, penelitian ini memperluas fungsi ulasan sebagai sumber informasi risiko keamanan siber berbasis pengalaman komunitas. Temuan ini memperkaya literatur dengan menempatkan ulasan pengguna tidak hanya sebagai refleksi kepuasan layanan, tetapi juga sebagai early risk signal dalam konteks keamanan digital.

Evaluasi berbasis confusion matrix, presisi, recall, dan F1-score menunjukkan bahwa performa model bervariasi antar kategori risiko. Klarifikasi terhadap hasil pengujian menunjukkan bahwa akurasi keseluruhan model berdasarkan confusion matrix adalah 76,5%, sedangkan analisis F1-score per kategori memberikan gambaran yang lebih representatif mengenai sensitivitas model terhadap masing-masing jenis risiko. Variasi performa antar kategori lebih dipengaruhi oleh kompleksitas linguistik dan tingkat eksplisitnya indikasi risiko dalam teks dibandingkan oleh distribusi kelas pada tahap evaluasi. Hal ini menegaskan bahwa interpretasi performa model dalam klasifikasi risiko keamanan siber perlu dilakukan secara granular, bukan hanya berdasarkan metrik agregat.

Secara konseptual, temuan ini sejalan dengan karakteristik domain keamanan siber, di mana indikasi ancaman tidak selalu muncul dalam bentuk terminologi teknis yang eksplisit, melainkan dalam narasi pengalaman pengguna yang kontekstual dan implisit. Kondisi ini menjelaskan mengapa kategori dengan ambiguitas linguistik tinggi menunjukkan performa yang relatif lebih rendah. Dengan demikian, tantangan utama dalam pendekatan ini terletak pada representasi semantik teks dan kemampuan model dalam menangkap makna kontekstual, bukan semata-mata pada aspek kuantitatif data.

Kontribusi teoretis penelitian ini terletak pada integrasi perspektif cyber threat intelligence dengan analisis ulasan pengguna berbasis machine learning. Penelitian ini memperkenalkan pendekatan community-based cyber risk identification, di mana persepsi dan pengalaman pengguna diposisikan sebagai sumber intelijen awal yang bersifat komplementer terhadap mekanisme deteksi teknis seperti analisis log sistem atau intrusion detection systems. Dengan demikian, model yang dikembangkan tidak dimaksudkan untuk menggantikan sistem deteksi teknis, melainkan untuk melengkapi strategi manajemen risiko keamanan siber pada level aplikasi.

Dari sisi praktis, temuan ini memberikan implikasi bagi pengembang dan pengelola platform digital untuk mengintegrasikan analisis ulasan pengguna sebagai bagian dari proses monitoring risiko keamanan secara proaktif. Pendekatan ini berpotensi membantu organisasi dalam mengidentifikasi pola risiko sejak tahap awal sebelum berkembang menjadi insiden keamanan yang berdampak signifikan. Untuk penelitian selanjutnya, disarankan pengembangan representasi teks yang lebih kontekstual, seperti pendekatan berbasis word embedding atau model bahasa yang lebih canggih, serta integrasi dengan data teknis keamanan guna menghasilkan sistem deteksi risiko yang lebih komprehensif dan adaptif.

5. Kesimpulan dan Saran

Penelitian ini menunjukkan bahwa pendekatan text mining terhadap ulasan pengguna game online dapat dimanfaatkan sebagai mekanisme deteksi dini terhadap indikasi risiko keamanan siber dalam ekosistem aplikasi digital. Melalui tahapan pra-pemrosesan, pelabelan risiko berbasis kategori keamanan siber, serta klasifikasi menggunakan algoritma Naïve Bayes, penelitian ini berhasil mengidentifikasi keberadaan ulasan yang mengandung indikasi risiko seperti pencurian akun, penyalahgunaan akses, pelanggaran privasi data, malware risk, dan penipuan digital berbasis rekayasa sosial. Temuan ini memperlihatkan bahwa ulasan pengguna tidak hanya merefleksikan pengalaman layanan, tetapi juga memuat sinyal awal yang relevan terhadap potensi ancaman keamanan.

Evaluasi model berdasarkan confusion matrix menunjukkan akurasi keseluruhan sebesar 76,5%, sementara analisis presisi, recall, dan F1-score per kategori memberikan gambaran yang lebih komprehensif mengenai variasi performa model. Hasil ini menegaskan bahwa interpretasi kinerja klasifikasi risiko keamanan siber perlu mempertimbangkan analisis per kategori, bukan hanya metrik agregat. Variasi performa yang ditemukan lebih berkaitan dengan kompleksitas linguistik dan tingkat eksplisitnya indikasi risiko dalam teks ulasan dibandingkan faktor distribusi kelas pada tahap evaluasi. Dengan demikian, tantangan utama dalam pendekatan ini terletak pada representasi semantik dan pemaknaan kontekstual bahasa alami.

Secara teoretis, penelitian ini berkontribusi dengan memperkenalkan konsep community-based cyber risk identification, yaitu pemanfaatan ulasan pengguna sebagai bentuk community-driven cyber threat intelligence yang bersifat komplementer terhadap mekanisme deteksi teknis. Secara praktis, pendekatan ini berpotensi membantu pengembang dan pengelola platform digital dalam melakukan monitoring risiko secara lebih proaktif pada tahap awal.

Namun demikian, penelitian ini memiliki keterbatasan pada penggunaan algoritma klasifikasi berbasis representasi fitur konvensional serta belum mengintegrasikan data teknis keamanan seperti log sistem atau laporan insiden aktual. Oleh karena itu, penelitian selanjutnya disarankan untuk mengembangkan pendekatan berbasis representasi teks yang lebih kontekstual, mengombinasikan metode klasifikasi yang lebih adaptif, serta mengintegrasikan analisis ulasan dengan sumber data keamanan lainnya guna menghasilkan sistem deteksi risiko yang lebih komprehensif dan akurat.

DAFTAR PUSTAKA

Arie, Yoyon, Budi Suprio, and Moch Najib. 2023. "Implementasi Naïve Bayes Terhadap Kesadaran Keamanan Informasi Dengan Infeksi Virus Pada Computer Implementation of Naïve Bayes Against Information Security Awareness with Virus Infections On Computers." 13(2): 162–71.

Darmawan, Gilbert, Syariful Alam, and M. Imam Sulistyo. 2023. "Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi Mypertamina Pada Google Playstore Menggunakan Metode Naïve Bayes 1,2,3)." 2(3): 100–108.

Dwijaya, Ahmad Rais, and Arif Dwi Laksito. 2023. "Sentiment Analysis of Pedulilindungi Application Reviews Using Machine Learning and Deep Learning." 5(2).

Galena, Marcelena Vicky, Adnan Syawal Adilaha Sadikin, Aprilia Prastyaningrum, Reza Febrian

Nugroho, and M. Fariz Fadillah Mardianto. 2024. "Analisis Sentimen Masyarakat Terhadap Keamanan Penggunaan E-Commerce B2C Menggunakan Pendekatan Naïve Bayes Berbasis Text Mining Untuk Mencegah Penipuan." 8(3): 2003–12.

Hermawan, Ahmad Rijal, and Isa Faqihuddin Hanif. 2025. "Sentiment Classification of Public Perception on LHKPN Using SVM and Naive Bayes." 14: 148–55.

Iranda, Muhammad, and Nurul Huda. 2025. "Analisis Kinerja Algoritma SVM Dan Naïve Bayes Untuk Klasifikasi Sentimen Program Makan Gratis." 14(3): 1452–64.

Kariman, Delsi, Ainil Mardiyah, Junios, Lisma Sari, Muthia Ananda, and Jasril. 2025. "Analisis Sentimen TikTok Shop Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes." 14(3): 1881–96.

Mas'ud, Khalid Al, Muhammad Izzan Fieldi, M. Hadi Al-Farisy, Fathoni M. Alfarizi4, and Ali Ibrahim. 2024. "Analisis Sentimen Deepseek Berdasarkan Ulasan Google Play Store Menggunakan Metode Naïve Bayes." *JUTISI* 14(2): 1020–30.

Pamuji, Agus. 2022. "PREDIKSI OTORISASI PENGGUNA SISTEM BERKAS PADA ALGORITMA." 24: 35–44. doi:10.23969/infomatek.v24i1.4604.

Pongoh, Arthur Gregorius, Rizqy Achmad Fahreza, Bilal Al Kindi, Feddy Setio Pribadi, and Rizky Ajie. 2024. "Systematic Literature Review (SLR): Dampak Pemanfaatan Artificial Intelligence Untuk Meningkatkan Cyber Security Systematic Literature Review (SLR): The Impact of Utilizing Artificial Intelligence to Enhance Cyber Security." 7(1): 34–41.

Praja, Panggih Gumelaring, Muhammad Taufiq Nuruzzaman, and Bambang Sugiantoro. 2025. "Tingkat Keamanan Jaringan Home Wi-Fi Di Kota Yogyakarta Terhadap Password Attack The Security Level Of Home Wi-Fi Networks Against Password Attacks." 8(2): 80–88.

Prasetyo, Andika, Taufik Ridwan, and Apriade Voutama. 2024. "GBWHATSAPP MENGGUNAKAN NAIVE BAYES CLASSIFIER DAN RANDOM FOREST." 11(1): 1–9. doi:10.30656/jsii.v11i1.6936.

Prastya, M Wildan Alvian, Muhlis Tahir, Ayu Agustyas Ningrum, and Aqiqul Putra Zaibintoro. 2024. "Analisis Ancaman Pishing Melalui Aplikasi WhatsApp : Review Metode Studi Literatur." 7(3): 190–97.

Pratama, Muhammad Ridwan, Ahmad Fauzi, Deden Wahiddin, and Adi Rizky Pratama. 2024. "Analisis Sentimen Kebijakan Pembelian Gas 3 Kg Dengan KTP Menggunakan Naïve Bayes."

Reandito, Alingga, Ikhwan Sumantri, Muhamad Fatchan, and Tri Ngudi Wiyatno. 2024. "Analisis Sentimen Produk Makanan Jepang Di Indonesia Pada Twitter Menggunakan Naïve Bayes." 13(2): 1635–45.

Riadi, Imam, Takdir Ruslan, Falkutas Teknologi Industri, Universitas Ahmad Dahlan, Falkutas Teknologi Industri, Universitas Ahmad Dahlan, Teknik Informatika, Falkutas Teknologi Industri, and Universitas Ahmad Dahlan. 2023. "Analisis Forensik Digital Pada Whatsapp Dan Facebook Menggunakan Metode NIST." 13(2): 286–92.

Riviani, Aulia, and Cahyono Budy Santoso. 2025. "Perbandingan Naïve Bayes Dan SVM Untuk Analisis Sentimen Pengguna Aplikasi X Terhadap Danantara."

Saputra, Adika Kaka, Maya Rini Handayani, Nur Cahyo, Hendro Wibowo, and Khothibul Umam. 2025. "Sentiment Analysis of User Reviews on the Game GTA V Using Support Vector Machine." 14: 284–90.

Sheren, Angellica, Romauli Sirait, Adinda Syifa Kamalia, Askia Diska, and Mercurius Broto Legowo. 2023. "MANAJEMEN RESIKO KEBOCORAN DATA NASABAH DAN SERANGAN SIBER MENGGUNAKAN NIST-RISK." : 144–51.

Triana, Risma Faris, Ade Irma Purnama Sari, Agus Bahtiar, and Edi Wahyudin. 2023. "Implementasi Algoritma Naïve Bayes Untuk Klasifikasi Sentimen Ulasan Pengguna KAI Access." : 12–21.

Zy, Ahmad Turmudi, and Wahyu Hadikristanto. 2023. "Implementasi Algoritma Metode Naive Bayes Dan Support Vector Machine Tentang Pembobolan Dan Kebocoran Data Di Twitter." 4(1): 49–56.