

KEAMANAN SIBER DI ERA DIGITAL: TINJAUAN PENERAPAN DATA MINING UNTUK DETEKSI ANCAMAN PADA PELAKU USAHA KECIL

Faizah Amirah Harahap¹, Taufiq², Yanita³

^{1,2,3}Universitas Malikussaleh, Indonesia

¹faizah.240170228@mhs.unimal.ac.id, ²taufiq.te@unimal.ac.id, ³yanita@unimal.ac.id

ABSTRACT

The rapid digital transformation of Indonesian MSMEs has increased exposure to cyber threats such as phishing, malware, ransomware, and DDoS. This study aims to review the application of data mining and machine learning for cyber threat detection, especially in the context of small businesses. A systematic literature review (SLR) with the PRISMA approach was used. Articles were collected from Google Scholar and Semantic Scholar (2018–2026), screened using inclusion and exclusion criteria, and analyzed descriptively. From 19 collected articles, 11 were reviewed: seven on data mining/machine learning for cyber threat detection and four on MSME cybersecurity. The results show that Random Forest, C4.5/J48, SVM/SMO, Naive Bayes, and K-Means are commonly used. Feature selection (Information Gain, PCA) and class balancing (SMOTE) improve accuracy and training efficiency, with reported accuracy generally above 90%. However, most studies use public datasets such as KDD Cup'99, AWID, CICIDS2017, and CIC-MalMem-2022, not data from MSME environments. Key challenges include class imbalance, false positives, limited resources, data quality, scalability, and low digital literacy. Non-technical factors—management commitment, governance, training, and cost—are also critical. Data mining has potential for MSME threat detection, but implementation requires lightweight, context-specific models and organizational readiness.

Keywords:

Cybersecurity, Data Mining, Machine Learning, Threat Detection, Intrusion Detection System, SMEs.

PENDAHULUAN

Transformasi digital di Indonesia berkembang pesat, termasuk di kalangan Usaha Mikro, Kecil, dan Menengah (UMKM) yang menyumbang lebih dari 97% tenaga kerja dan hampir 62% produk domestik bruto (PDB) nasional. Namun, pertumbuhan ini juga diiringi lonjakan ancaman siber. Badan Siber dan Sandi Negara (BSSN) mencatat rata-rata 182 potensi serangan siber per detik di Indonesia sepanjang 2025, atau setara lebih dari 5,1 miliar anomali trafik internet dalam setahun (Sriyono, 2026). Kondisi ini diperparah oleh temuan Kaspersky yang menyebutkan bahwa 92% UMKM di Indonesia mengalami insiden siber dalam satu tahun terakhir, menjadikan Indonesia negara ketiga dengan tingkat insiden siber terhadap UMKM tertinggi di Asia Tenggara, dengan phishing dan malware sebagai jenis ancaman paling umum (Hestin Untari, 2026b).

Ironisnya, sejauh yang dapat dilihat, kesiapan pelaku UMKM Indonesia dalam menghadapi berbagai ancaman tersebut masih tertinggal jika dibandingkan dengan besarnya risiko yang sebenarnya dihadapi. Masih banyak UMKM yang mengandalkan sistem keamanan digital yang minim dan rendahnya literasi digital juga turut menjadi faktor utama yang membuat mereka rentan terhadap serangan. Serangan ransomware yang menargetkan UMKM di kawasan Asia Tenggara terus meningkat. Ini termasuk Indonesia, yang juga mengalami peningkatan dari waktu ke waktu (Hestin Untari, 2026a). Kondisi ini menunjukkan bahwa deteksi ancaman secara manual sudah tidak lagi memadai untuk mengimbangi volume serta kompleksitas serangan siber yang terus berkembang setiap tahunnya.

Dalam konteks ini, pendekatan berbasis data mining serta machine learning menjadi relevan untuk dikaji, karena mampu mengenali pola, mengklasifikasikan aktivitas, serta mendeteksi anomali dari data secara otomatis, sehingga ancaman dapat diidentifikasi lebih cepat sebelum menimbulkan kerugian yang lebih besar bagi pelaku usaha kecil. Namun, kajian yang secara spesifik membahas mengenai penerapan data mining untuk mendeteksi ancaman siber pada konteks UMKM di Indonesia masih terbatas. Oleh karena itu, penelitian ini memiliki tujuan untuk melakukan tinjauan literatur sistematis terhadap penerapan metode data mining dalam deteksi ancaman siber. Hal ini bertujuan guna merangkum metode yang paling relevan serta untuk mengidentifikasi peluang dan tantangan penerapannya pada konteks UMKM di Indonesia.

KAJIAN LITERATUR

Penelitian Terdahulu

Beberapa penelitian sebelumnya telah mengkaji penerapan machine learning dan data mining untuk keamanan siber pada Usaha Mikro, Kecil, dan Menengah (UMKM). Sebuah penelitian menguji empat algoritma pembelajaran mesin terlabeli, yaitu *Naive Bayes*, *Sequential Minimal Optimaization (SMO)*, *C4.5*, dan *Random Forest*, pada sistem

deteksi intrusi untuk usaha kecil untuk usaha kecil dan menengah menggunakan dataset AWID. Berdasarkan hasil pengujian, penerapan seleksi fitur Information Gain mengurangi jumlah fitur dari 130 menjadi 39, membuat akurasi seluruh algoritma berada di atas 95%, serta memangkas waktu pelatihan model secara drastis (Baci et al., 2022).

Studi lain membandingkan delapan artikel jurnal dan menyimpulkan bahwa pengelolaan dan pengendalian keamanan sistem informasi sangat perlu diterapkan pada UMKM karena ancaman terhadap sistem informasi merupakan kondisi yang tidak dapat dihindari. Temuan utama menunjukkan bahwa sebagian besar UMKM hanya menerapkan langkah keamanan yang paling dasar seperti kata sandi (Nugroho et al., 2024).

Keamanan Siber

Keamanan siber merujuk pada praktik melindungi sistem, jaringan, dan data digital dari akses yang tidak sah, kerusakan, atau pencurian. Ancaman siber yang umum dihadapi usaha kecil yaitu, phishing (upaya penipuan untuk mencuri data melalui email/pesan palsu), malware (perangkat lunak yang berbahaya), ransomware (malware yang mengenkripsi dan meminta tebusan), serta serangan DDoS (upaya membanjiri sistem dengan trafik palsu hingga lumpuh) (Jonker et al., 2025).

Data Mining

Data mining merupakan proses mengekstraksi pola, korelasi, atau informasi tersembunyi dari kumpulan data berskala besar. Beberapa teknik utamanya meliputi klasifikasi (mengelompokkan data hingga menemukan pola yang mana merupakan *phishing* dan yang mana yang bukan merupakan *phishing*), clustering (mengelompokkan data tanpa label berdasarkan kemiripan), dan deteksi anomali (mengenal data yang menyimpang dari pola normal). Clustering merupakan metode unsupervised yang digunakan untuk mengelompokkan data berdasarkan kemiripan atau kesamaan tanpa perlu dilabeli sebelumnya. Algoritma seperti ini dapat digunakan untuk mengidentifikasi pola yang tidak biasa (Maulid, 2024).

Machine Learning

Machine Learning merupakan cabang dari kecerdasan buatan yang memanfaatkan data atau algoritma untuk memungkinkan sistem belajar dan berkembang seperti manusia, tanpa diberikan instruksi secara gamblang. Dalam konteks keamanan siber, machine learning merupakan salah satu pendekatan yang dipakai dalam data mining dan keduanya beririsan.

METODE PENELITIAN

Penelitian ini menggunakan metode *Systematic Literature Review* (SLR) dengan pendekatan kualitatif deskriptif, yaitu pendekatan yang menggambarkan dan merangkum temuan secara apa adanya tanpa pengujian statistik. Teknik *Systematic Literature Review* atau SLR merupakan cara sistematis untuk mengidentifikasi, mengumpulkan, menilai secara kritis, dan menyajikan temuan dari berbagai studi penelitian (Tumurang, 2024). Pendekatan ini dipilih karena penelitian bertujuan meninjau bagaimana data mining diterapkan untuk mendeteksi ancaman siber, khususnya yang relevan bagi pelaku usaha kecil, dengan merangkum dan membandingkan temuan penelitian terdahulu, bukan mengumpulkan data yang langsung berasal dari responden. Tinjauan ini dilakukan untuk menjawab tiga pertanyaan:

1. Metode data mining dan *machine learning* apa yang digunakan untuk mendeteksi ancaman siber,
2. Bagaimana efektivitas metode tersebut berdasarkan hasil penelitian yang dikaji,
3. Apa tantangan penerapannya pada usaha kecil/UMKM.

Pencarian mencakup dua kelompok topik, yaitu penerapan data mining/*machine learning* untuk mendeteksi ancaman siber (misalnya sistem deteksi intrusi dan malware) serta kondisi keamanan siber pada UMKM sebagai konteks kajian. Tahapan seleksi artikel dilakukan bertahap mengikuti alur PRISMA, yaitu mengidentifikasi, melakukan penyaringan berdasarkan judul dan abstrak, melakukan penilaian kelayakan melalui teks lengkap, lalu menentukan artikel akhir. Artikel dimasukkan apabila membahas penerapan data mining/*machine learning* untuk deteksi ancaman siber atau membahas keamanan siber pada UMKM/usaha kecil, berupa artikel jurnal atau prosiding berbahasa Indonesia atau Inggris, serta teks lengkapnya dapat diakses. Artikel dikeluarkan apabila tidak membahas keamanan siber atau deteksi ancaman, bukan artikel ilmiah (berita, blog, opini), atau klaim dan daftar pustakanya tidak dapat ditelusuri.

Pencarian artikel dilakukan melalui Google Scholar dan Semantic Scholar, pada Oktober 2026. Pencarian dibatasi pada artikel yang terbit di rentang tahun 2018-2026 agar temuan mencerminkan perkembangan metode dan kondisi ancaman siber terkini. Pencarian menggunakan kata kunci dalam bahasa Indonesia dan Inggris, yaitu “keamanan siber”, “cyber security”, “data mining”, “machine learning”, “intrusion detection”, “malware detection”, “phishing detection”, “UMKM”, “SME” dan “small business” yang dikombinasikan dengan operator Boolean AND dan OR.

Setiap artikel terpilih diekstraksi ke dalam tabel yang memuat penulis dan tahun, konteks penelitian, dataset, algoritma, metrik evaluasi, hasil utama, serta tantangan yang dilaporkan. Data kemudian dianalisis secara deskriptif melalui sintesis tematik yang dikelompokkan sesuai tiga pertanyaan penelitian. Karena dataset dan metrik evaluasi tiap artikel berbeda, perbandingan hasil dilakukan secara naratif bukan statistik. Pengelolaan referensi penelitian ini menggunakan Mendeley Web Importer.

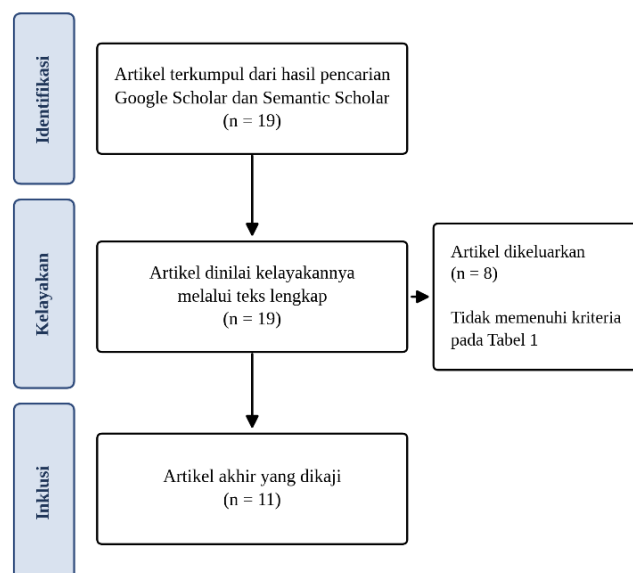
Tabel 1. Kriteria Inklusi dan Eksklusi

Kriteria Inklusi (Yang Dimasukkan)	Kriteria Eksklusi (Yang Dibuang)
Artikel jurnal atau prosiding ilmiah	Artikel berupa berita, blog, atau opini
Membahas penerapan Data Mining/ <i>Machine Learning</i> untuk deteksi ancaman siber atau keamanan siber pada UMKM/usaha kecil	Tidak membahas keamanan siber atau deteksi ancaman secara spesifik
Daftar pustaka dapat ditelusuri	Daftar pustaka tidak dapat ditelusuri
Berbahasa Indonesia atau Inggris	Diluar dari bahasa Indonesia atau Inggris
Jurnal yang teks lengkapnya dapat diakses	Jurnal yang teks lengkapnya (<i>full-text</i>) tidak dapat diakses
Terbit pada rentang tahun 2018-2026	Terbit diluar rentang tahun 2018-2026

HASIL DAN PEMBAHASAN

Hasil Seleksi Artikel

Proses seleksi mengikuti alur PRISMA sebagaimana ditunjukkan pada Gambar 1. Sebanyak 19 artikel terkumpul dari hasil pencarian di Google Scholar dan Semantic Scholar, kemudian seluruhnya dinilai kelayakannya melalui teks lengkap berdasarkan kriteria pada Tabel 1. Lalu sebanyak 8 artikel dikeluarkan, yaitu 1 artikel daftar pustakanya tidak sesuai dengan judul sehingga tidak layak dijadikan referensi, 2 artikel yang terbit diluar rentang tahun 2018-2026, 2 artikel tidak membahas keamanan siber atau deteksi ancaman karena berfokus pada web dan 3 artikel karena tidak dapat ditelusuri daftar pustakanya dan . Dengan demikian, 11 artikel akhir dikaji yang terdiri dari 7 artikel yang membahas penerapan *data mining* dan *machine learning* untuk deteksi ancaman siber serta 4 artikel yang membahas keamanan siber pada UMKM.



Gambar 1. Diagram Seleksi Artikel Berdasarkan Alur PRISMA

Gambar 1 menunjukkan bahwa proses seleksi berlangsung dalam 3 tahap. Pada tahap identifikasi terkumpul 19 artikel dari dua basis data. Pada tahap kelayakan, seluruh artikel dibaca teks lengkapnya dan 8 artikel dikeluarkan, sehingga 11 artikel masuk ke tahap inklusi. Artikel yang dikeluarkan karena topik di luar keamanan siber ($n = 2$), karena daftar pustaka bermasalah ($n = 4$) dan karena terbit diluar rentang tahun 2018-2026 ($n = 2$) menunjukkan bahwa pencarian berbasis kata kunci masih menjaring artikel yang kurang relevan atau kurang dapat dipertanggungjawabkan, sehingga penilaian teks lengkap diperlukan. Sebelas artikel akhir ini kemudian diekstraksi pada Tabel 2 (7 artikel) dan Tabel 3 (4 artikel) dan menjadi dasar pembahasan untuk ketiga pertanyaan penelitian.

Metode Data Mining dan Machine Learning untuk Mendeteksi Ancaman Siber

Tabel 2. Ekstraksi Data Artikel Penerapan Data Mining/Machine Learning untuk Deteksi Ancaman Siber

Penulis (tahun)	Dataset	Algoritma	Hasil utama	Tantangan
(Navirgo & Habibbullaah, 2019)	KDD Cup'99; 247.010 dari 494.021 record (DoS 220.145; R2L hanya 1)	Hoeffding Tree, J48, Random Forest, Random Tree, REP Tree (Weka)	RF tertinggi 99,9891% (10-fold; build 91,64 dtk). Random Tree tercepat 1,03 dtk (99,9798%). Hoeffding terendah 99,9571%	Build RF sampai 100,16 dtk
(Elezaj et al., 2019)	AWID-CLS-R	C4.5, Bayesian Network, RF, SVM, ANN; seleksi fitur (16 dan 34 fitur); SMOTE	34 fitur: C4.5 98,4508% (FP 0,029), SVM terendah 88,2149%. Setelah SMOTE 50:50: C4.5 99,9872%, TPR >99%	Dataset tidak seimbang; false positive membebani SME; skalabilitas
(Desai et al., 2021)	Tanpa eksperimen	Ringkasan: klasifikasi, SVM, KNN, K-means, Naïve Bayes, ANN, GA, graph mining	Deteksi signature vs behavior (statis, dinamis, hybrid); tanpa metrik	Signature gagal pada malware polimorfik; behavior butuh waktu dan storage lebih
(Baci et al., 2022)	AWID-CLS-R (latih 1.795.575; uji 530.643), 130 fitur	Naïve Bayes, SMO, C4.5, RF + Information Gain (→39 fitur)	Akurasi sebelum→sesudah IG: NB 83,02→95,02; SMO 95,21→96,69; C4.5 93,99→99,76; RF 95,98→97,98. Waktu latih SMO 4j19m→4m22d; C4.5 1j20m→5m29d	Dataset tidak seimbang; FP tinggi di UKM; skalabilitas
(Sitorus et al., 2025)	Kaggle <i>Cybersecurity Intrusion Detection Dataset</i> , 9.537 record, 11 atribut (RapidMiner)	Random Forest (5 pohon); split 90:10, 80:20, 70:30	Terbaik 90:10: akurasi 88,16%, precision 100%, recall 72,77%, F-measure 84,24%	Recall turun jadi 60% (80:20) dan 52,45% (70:30)
(Putri & Mulyana, 2025)	CIC-MalMem-2022 (58.596 data), split 70:30	Random Forest (Grid Search; n_estimators=100, max_depth=10) + TOPSIS	Akurasi 92,3%, F1 0,91; ranking + pop-up ±9 dtk	Belum pakai log real-time; biaya dan pembaruan sistem
(wafdan et al., 2026)	CICIDS2017: 2,64 juta record bersih, 78→61 fitur, PCA 18 komponen (95% variansi)	K-Means (<i>unsupervised</i>), k=6 (Elbow + Silhouette)	Silhouette 0,512; Davies-Bouldin 0,891; purity global 0,874; NMI 0,691; ARI 0,623. Purity tertinggi: Normal 0,967, DDoS 0,941; terendah Web Attack 0,621	Kluster Web Attack paling heterogen; balanced accuracy 0,71 (RF 0,94, angka dari studi lain)

Lima artikel menerapkan klasifikasi terlabeli (*supervised*) secara eksperimen, satu artikel menerapkan pengelompokkan tanpa label (*unsupervised*) menggunakan K-MEANS (wafdan et al., 2026) dan satu artikel berupa tinjauan teknik deteksi malware tanpa eksperimen (Desai et al., 2021). Random Forest dipakai dalam kelima eksperimen terlabeli ((Navirgo & Habibbullaah, 2019), (Elezaj et al., 2019), (Baci et al., 2022), (Sitorus et al., 2025), (Putri & Mulyana, 2025), disusul C4.5 atau J48 pada tiga artikel serta SVM atau variannya (SMO) pada dua artikel. (Desai et al., 2021) membedakan deteksi berbasis *signature*, yang tidak dapat mengenali malware polimorfik, dari deteksi berbasis perilaku yang membutuhkan waktu dan ruang penyimpanan lebih besar.

Dataset yang digunakan adalah dataset publik, yaitu KDD Cup'99, AWID, CIC-MalMem-2022, CICIDS2017, dan dataset *intrusion detection detection* dari Kaggle. Tidak ada artikel yang memakai data yang dikumpulkan langsung dari lingkungan UMKM. Hanya dua artikel yang secara eksplisit menempatkan UMKM sebagai konteks pengembangan sistem deteksi intrusi (Elezaj et al., 2019), (Baci et al., 2022), dan keduanya memakai dataset AWID. Tiga artikel juga

melaporkan upaya pra-pemrosesan untuk mengurangi dimensi fitur atau menyeimbangkan data, yaitu seleksi fitur dan SMOTE (Elezaj et al., 2019), *Information Gain* (Baci et al., 2022), serta seleksi fitur berbasis varians, korelasi, dan PCA (wafdan et al., 2026).

Efektivitas Metode Berdasarkan Hasil Penelitian yang Dikaji

Kinerja yang dilaporkan umumnya tinggi, tetapi dihitung pada dataset, skema pembagian data, dan metrik yang berbeda sehingga dibandingkan secara naratif. (Navirgo & Habibbullaah, 2019) melaporkan akurasi tertinggi 99,9891% pada Random Forest (10-fold cross validation), sedangkan Random Tree memerlukan waktu pembentukan model tersingkat, yaitu 1,03 detik dengan akurasi 99,9798% pada pembagian data 66%. (Elezaj et al., 2019) melaporkan C4.5 mencapai akurasi 98,4508% dengan 34 fitur dan meningkat menjadi 99,9872% setelah data diseimbangkan dengan SMOTE. (Baci et al., 2022) melaporkan bahwa *Information Gain* mengurangi fitur dari 130 menjadi 39 dan membuat akurasi keempat algoritma berada di atas 95%, dengan waktu pelatihan SMO turun dari 4 jam 19 menit menjadi 4 menit 22 detik dan C4.5 dari 1 jam 20 menit menjadi 5 menit 29 detik. (Putri & Mulyana, 2025) melaporkan akurasi Random Forest 92,3% dengan F1-score 0,91 pada klasifikasi tiga jenis malware, sedangkan (Sitorus et al., 2025) memperoleh akurasi tertinggi 88,16% dengan *precision* 100% dan *recall* 72,77% pada pembagian data 90:10. Pada pendekatan tanpa label, (wafdan et al., 2026) memperoleh *Silhouette Coefficient* 0,512 dan *Purity* global 0,874 dengan enam kluster, serta *balanced accuracy* 0,71 terhadap label. Artikel tersebut membandingkannya dengan Random Forest (0,94), tetapi angka Random Forest itu diambil dari studi lain.

Dua temuan penting dari pembahasan ini. Pertama, seleksi fitur dan penyeimbangan data memperbaiki kinerja sekaligus efisiensi waktu ((Elezaj et al., 2019), (Baci et al., 2022)), hal yang relevan bagi UMKM yang memiliki keterbatasan sumber daya (Baci et al., 2022). Kedua, akurasi yang sangat tinggi perlu dibaca hati-hati. Pada (Navirgo & Habibbullaah, 2019), sekitar 89% data yang digunakan berasal dari kelas DoS (220.145 dari 247.010 *record*), sedangkan kelas R2L hanya satu *record*. Pada (Sitorus et al., 2025), *recall* turun menjadi 60% (80:20) dan 52,45% (70:30) ketika porsi data uji diperbesar, sementara *precision* tetap 100%, artinya model tidak menghasilkan *false positive* tetapi melewatkan sebagian serangan. Perlu dicatat pula bahwa abstrak (Putri & Mulyana, 2025) melaporkan akurasi 95,87%, sedangkan tabel evaluasi dan kesimpulannya 92,3% kajian ini menggunakan angka pada tabel evaluasi.

Tantangan Penerapan pada Usaha Kecil/UMKM

Tabel 3. Ekstraksi Data Artikel Keamanan Siber pada UMKM

Penulis (tahun)	Metode	Temuan utama	Tantangan
(Benjamin et al., 2024)	SLR 2015–2024 (IEEE Xplore, ScienceDirect, JSTOR, Google Scholar)	UKM rentan phishing, malware, data breach, ransomware, mitigasi: kebijakan, pelatihan, teknologi; AI/ML diprediksi makin dipakai	Keterbatasan sumber daya, kurang kesadaran
(Amanda et al., 2026)	SLR Scopus, thematic synthesis (4 tema)	Risiko fraud UMKM digital bersifat socio-technical; teknologi anti-fraud (analytics, ML) berpotensi efektif	Governance, explainability, kualitas data, biaya, kesiapan organisasi
(Rahmawati, 2025)	Survei 100 UKM Surabaya, regresi berganda	Literasi digital (β 0,752; p 0,000) dan komitmen manajemen (β 0,362; p 0,001) signifikan; R^2 0,896	Pengendalian internal tidak signifikan (β -0,108; p 0,233)
(Sari et al., 2025)	Pelatihan + pre/post-test, 34 UMKM Semarang, Wilcoxon	Skor 5,64→10,18 (Z -4,476; p 0,0001)	Pemahaman “jenis ancaman data pribadi” nyaris tidak naik (11,8%→12,1%)

(Benjamin et al., 2024) menemukan bahwa UKM rentan terhadap *phishing*, malware, kebocoran data, dan *ransomware* akibat keterbatasan sumber daya dan rendahnya kesadaran, serta memprediksi penggunaan *AI/machine learning* untuk deteksi ancaman akan meningkat. (Amanda et al., 2026) menyimpulkan bahwa risiko *fraud* pada UMKM digital bersifat *socio-technical* dan bahwa efektivitas teknologi anti-fraud dipengaruhi oleh tata kelola, *explainability*, kualitas data, biaya, dan kesiapan organisasi. (Rahmawati, 2025) menemukan bahwa literasi digital (β = 0,752; p = 0,000) dan komitmen manajemen (β = 0,362; p = 0,001) berpengaruh signifikan terhadap kesiapan UKM menghadapi ancaman siber, sedangkan pengendalian internal tidak signifikan pada tabel koefisiennya (p = 0,233). (Sari et al., 2025) melaporkan pelatihan meningkatkan skor rata-rata dari 5,64 menjadi 10,18 (p = 0,0001), tetapi pemahaman mengenai jenis ancaman terhadap data pribadi nyaris tidak berubah (11,8% menjadi 12,1%).

Di sisi teknis, artikel kelompok pertama melaporkan tantangan berupa dataset yang tidak seimbang ((Elezaj et al., 2019), (Baci et al., 2022)), *false positive* yang membebani analisis pada UKM ((Elezaj et al., 2019), (Baci et al., 2022)), waktu pelatihan model yang panjang (Baci et al., 2022), serta biaya dan pembaruan sistem yang berkelanjutan (Putri &

Mulyana, 2025). (wafdan et al., 2026) menyatakan bahwa pendekatan tanpa label relevan bagi UMKM yang tidak memiliki tim keamanan khusus.

Jika kedua kelompok dibandingkan, tampak kesenjangan: studi teknis belum menguji metodenya pada data yang berasal dari lingkungan UMKM, sedangkan studi tentang UMKM berfokus pada literasi, kebijakan, dan kesiapan organisasi, dengan *machine learning* hanya dibahas sebagai potensi. Tantangan penerapan *data mining* di UMKM dengan demikian bersifat teknis (data tidak seimbang, *false positive*, waktu komputasi) sekaligus organisasional (literasi, komitmen manajemen, tata kelola, biaya).

KESIMPULAN

Berdasarkan tinjauan sistematis terhadap 11 artikel, penerapan data mining dan machine learning untuk deteksi ancaman siber pada pelaku usaha kecil menunjukkan potensi yang besar, terutama melalui algoritma klasifikasi seperti Random Forest, C4.5/J48, SVM/SMO, dan Naive Bayes, serta pendekatan clustering K-Means. Random Forest paling dominan dan konsisten memberikan akurasi tinggi, sementara seleksi fitur seperti Information Gain dan PCA serta penyeimbangan data dengan SMOTE terbukti meningkatkan kinerja dan efisiensi pelatihan model. Namun, sebagian besar penelitian masih menggunakan dataset publik seperti KDD Cup'99, AWID, CICIDS2017, dan CIC-MalMem-2022, bukan data yang dikumpulkan langsung dari lingkungan UMKM. Keterbatasan ini membuat hasil belum sepenuhnya merepresentasikan kondisi nyata UMKM, terutama terkait keterbatasan sumber daya, kualitas data, ketidakseimbangan kelas, *false positive*, dan skalabilitas.

Di sisi lain, kajian tentang keamanan siber UMKM lebih banyak menyoroti aspek non-teknis, seperti literasi digital, komitmen manajemen, pengendalian internal, tata kelola, pelatihan, dan biaya. Machine learning dalam kajian tersebut masih dibahas sebagai potensi, belum diuji secara teknis pada data UMKM. Oleh karena itu, penerapan data mining untuk deteksi ancaman pada usaha kecil tidak dapat hanya mengandalkan model berakurasi tinggi, tetapi perlu mempertimbangkan konteks organisasi, kesiapan SDM, dan keberlanjutan biaya.

Penelitian ini memiliki keterbatasan karena hanya mengkaji 11 artikel dari dua basis data dan tidak melakukan pengujian eksperimen. Penelitian selanjutnya disarankan untuk mengumpulkan dataset spesifik UMKM, menguji model ringan yang sesuai dengan keterbatasan perangkat, mengevaluasi biaya *false positive*, serta mengintegrasikan solusi teknis dengan pelatihan, kebijakan, dan tata kelola keamanan informasi. Dengan demikian, data mining dapat menjadi bagian dari strategi deteksi ancaman siber yang lebih adaptif dan terjangkau bagi pelaku usaha kecil di Indonesia.

REFERENSI

- Amanda, Izza Muthmainnah, B., & Rabilah Affandy, N. (2026). MANAJEMEN RISIKO KECURANGAN PADA UMKM DIGITAL: TINJAUAN SISTEMATIS KEAMANAN SIBER, PENGENDALIAN, DAN TEKNOLOGI ANTIFRAUD. *E-Prosiding Akuntansi*, 6(1). <https://ojs.trilogi.ac.id/index.php/EPAKT/article/view/2937>
- Baci, N., Vukatana, K., & Baci, M. (2022). Machine Learning Approach for Intrusion Detection Systems as a Cyber Security Strategy for Small and Medium Enterprises. *WSEAS Transactions on Business and Economics*, 19, 474–480. <https://doi.org/10.37394/23207.2022.19.43>
- Benjamin, L. B., Adegbola, A. E., Amajuoyi, P., Adegbola, M. D., Adeusi, K. B., Benjamin, L. B., Adegbola, A. E., Amajuoyi, P., Adegbola, M. D., & Adeusi, K. B. (2024). Digital transformation in SMEs: Identifying cybersecurity risks and developing effective mitigation strategies. *Global Journal of Engineering and Technology Advances*, 19(2), 134–153. <https://doi.org/10.30574/GJETA.2024.19.2.0084>
- Desai, V. P., Oza, Dr. K. S., & Naik, Dr. P. G. (2021). Data Mining Approach for Cyber Security. *International Journal of Computer Applications Technology and Research*, 10(01), 35–41. <https://doi.org/10.7753/IJCATR1001.1007>
- Elezaj, O., Yayilgan, S. Y., Abomhara, M., Yeng, P., & Ahmed, J. (2019). Data-driven Intrusion Detection System for Small and Medium Enterprises. *IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks, CAMAD, 2019-September*. <https://doi.org/10.1109/CAMAD.2019.8858166>
- Hestin Untari, P. (2026a, July 13). *Serangan Ransomware ke UMKM Indonesia Lebih Tinggi Dibandingkan Negara Kawasan*. teknologi bisnis. <https://teknologi.bisnis.com/read/20260713/84/1987638/serangan-ransomware-ke-umkm-indonesia-lebih-tinggi-dibandingkan-negara-kawasan>
- Hestin Untari, P. (2026b, September 14). *Kaspersky: 92% UMKM Indonesia Alami Insiden Siber, Phishing-Malware Paling Sering*. teknologi bisnis. <https://teknologi.bisnis.com/read/20260914/84/2003898/kaspersky-92-umkm-indonesia-alami-insiden-siber-phishing-malware-paling-sering>
- Jonker, A., Lindemulder, G., & Kosinski, M. (2025, September 30). *Apa itu Keamanan Siber?* | IBM. IBM. <https://www.ibm.com/id-id/think/topics/cybersecurity>
- Maulid, R. (2024, October 22). *Algoritma Data Science untuk Perkuat Keamanan Siber*. DQLab. <https://dqlab.id/algoritma-data-science-untuk-perkuat-keamanan-siber>
- Navirgo, A., & Habibullaah, A. (2019). Implementasi Data Mining Dengan Algoritma Berbasis Tree Untuk Klasifikasi Serangan Pada Intrusion Detection System (Ids). *Prosiding Seminar Nasional Darmajaya*, 1(0), 120–126.

- <https://jurnal.darmajaya.ac.id/index.php/PSND/article/view/1708>
- Nugroho, M. A., Yerika Septininditya, A., Andro, R., & Nugraha, Z. (2024). Urgensi Manajemen Keamanan dan Pengendalian Sistem Informasi bagi UMKM di Indonesia. *Indo-Fintech Intellectuals: Journal of Economics and Business*, 4(3), 734–744. <https://doi.org/10.54373/IFIJEB.V4I3.1373>
- Putri, N. H., & Mulyana, D. I. (2025). DETEKSI DAN PRIORITAS MITIGASI MALWARE MENGGUNAKAN RANDOM FOREST DAN METODE MCDM. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(3S1), 2830–7062. <https://doi.org/10.23960/JITET.V13I3S1.7762>
- Rahmawati, S. P. (2025). Analisis Pengaruh Literasi Digital, Komitmen Manajemen, Dan Pengendalian Internal Terhadap Ukm Cyber Preparedness. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(3), 1192–1203. <https://doi.org/10.31004/RIGGS.V4I3.2010>
- Sari, P. K., Candiwan, C., Trianasari, N., & Prasetyo, A. (2025). PENINGKATAN KESADARAN KEAMANAN SIBER MELALUI PELATIHAN KEPADA PELAKU UMKM BINAAN YAYASAN PURBA DANARTA SEMARANG. *Jurnal Pengabdian Kolaborasi Dan Inovasi IPTEKS*, 3(4), 849–856. <https://doi.org/10.59407/JPKI2.V3I4.2574>
- Sitorus, Z., Pratama, A., Hidayatullah, O. A., & Widarma, A. (2025). Model Data Mining Klasifikasi Serangan Siber untuk Deteksi Dini Serangan Menggunakan Algoritma Random Forest. *CESS (Journal of Computer Engineering, System and Science)*, 10(1), 335–346. <https://doi.org/10.24114/CESS.V10I1.65940>
- Sriyono. (2026, August 6). *BSSN: Ada 182 Potensi Serangan Siber Tiap Detik Lewat Internet di Indonesia*. koran jakarta. https://koran-jakarta.com/2026-08-06/bssn-ada-182-potensi-serangan-siber-tiap-detik-lewat-internet-di-indonesia#google_vignette
- Tumurang, M. (Marjes). (2024). METODOLOGI PENELITIAN. *PT Media Pustaka Indo*, 1–376. <https://www.neliti.com/publications/617648/>
- wafdan, juwardi, suratman, muhammad, mujiono, kawet, & ridha, muh. rasyid. (2026). PENERAPAN K-MEANS CLUSTERING UNTUK KLASIFIKASI POLA SERANGAN SIBERPADA INTRUSION DETECTION SYSTEM (IDS) BERBASIS DATA LOG JARINGAN | TEKNOFILE : Jurnal Sistem Informasi. *Teknofile - Jurnal Sistem Informasi*, 4, 520–531. <https://jurnal.nawansa.com/index.php/teknofile/article/view/986>