SYSTEMATIC REVIEW PENGGUNAAN *ISOLATION FOREST* DAN *SUPPORT VECTOR MACHINE* UNTUK *ANOMALY DETECTION*

M. Raihan Tri Wardhana, Aditya Septiandy, Ivana Sintani,
Veronika Saprilliana Anugrah, Satria Dwi Cahya, Widiatry
Program Studi Teknik Informatika S1, Fakultas Teknik, Universitas Palangka Raya
Jl. Yos Sudarso, Palangka, Kec. Jekan Raya, Kota Palangka Raya, Kalimantan Tengah 74874, Indonesia
raihanwardhana24@gmail.com

ABSTRAK

Deteksi anomali penting untuk menjaga keamanan dan konsistensi data pada berbagai sistem. Namun, penelitian terkait metode yang digunakan masih tersebar dan belum terpetakan secara sistematis. Permasalahan utama dalam kajian ini adalah ketiadaan pemetaan komprehensif mengenai penggunaan *Isolation Forest* (IF) dan *Support Vector Machine* (SVM) dalam penelitian anomaly detection, terutama terkait domain penelitian, metrik evaluasi yang digunakan, serta kualitas pelaporan eksperimen. Penelitian ini bertujuan menyusun *systematic review* terhadap studi-studi yang memanfaatkan IF dan SVM untuk deteksi anomali setelah tahun 2019, sehingga diperoleh gambaran menyeluruh mengenai tren penggunaannya. Metode yang digunakan adalah *systematic literature review* melalui pencarian artikel di Google Scholar, seleksi berbasis kriteria inklusi dan eksklusi, ekstraksi data terstruktur, serta sintesis deskriptif untuk menjawab tiga pertanyaan penelitian terkait domain dan dataset, metrik evaluasi, dan aspek reprodusibilitas. Hasil penelitian menunjukkan bahwa IF banyak digunakan pada domain keamanan jaringan, *log sistem*, keuangan, dan evaluasi kinerja, sedangkan SVM lebih dominan pada domain keamanan dan sistem seperti *intrusion detection*, IoT, dan *malware*. SVM juga lebih konsisten dalam pelaporan metrik evaluasi. Namun, pelaporan *hyperparameter*, *preprocessing*, *runtime*, dan ketersediaan kode sumber pada kedua kelompok studi masih terbatas, sehingga reprodusibilitas penelitian belum optimal.

Kata Kunci : *Deteksi Anomali, Isolation Forest, Support Vector Machine, Systematic Literature Review, Machine Learning.*

1. PENDAHULUAN

Pemanfaatan *machine learning* tidak hanya digunakan untuk deteksi anomali, tetapi juga telah banyak diterapkan untuk menganalisis opini dan perilaku pengguna pada data tekstual. Angel et al. (2024), misalnya, menganalisis sentimen dan emosi dari ulasan Google Maps terhadap layanan rumah sakit di Palangka Raya menggunakan beberapa algoritma klasifikasi seperti K-Nearest Neighbors, Logistic Regression, dan Decision Tree, dengan akurasi terbaik mencapai 92% pada model Decision Tree. Hasil tersebut memperlihatkan bahwa pemilihan algoritma dan desain proses pra-pemrosesan data berpengaruh signifikan terhadap kinerja model *machine learning* dalam memahami pola laten pada data, sehingga pendekatan serupa juga relevan ketika merancang sistem deteksi anomali yang bergantung pada kualitas representasi data dan metrik evaluasi yang digunakan[1].

Deteksi anomali (*anomaly detection*) merupakan salah satu topik penting dalam analisis data *modern*, terutama karena kemampuannya mengidentifikasi pola yang menyimpang dari perilaku normal dan berpotensi mengindikasikan masalah seperti serangan *cyber*, *fraud*, kerusakan sistem, maupun penyalahgunaan aset. Anomali didefinisikan sebagai nilai atau perilaku data yang berbeda secara signifikan dari pola umum [2], dan dalam banyak konteks menjadi indikator awal terjadinya kegagalan proses, ancaman keamanan, atau tindakan kecurangan [3]. Seiring meningkatnya volume data dalam berbagai sektor, proses identifikasi anomali secara manual menjadi tidak efisien dan membutuhkan teknik otomatis berbasis *machine learning* yang mampu mendeteksi pola-pola tidak wajar secara cepat, skalabel, dan akurat [4].

Dua metode yang banyak digunakan dalam penelitian *anomaly detection* adalah *Isolation Forest* (IF) dan *Support Vector Machine* (SVM). *Isolation Forest* merupakan algoritma *unsupervised* yang bekerja dengan mengisolasi data melalui pemisahan acak menggunakan struktur pohon, sehingga instance yang lebih mudah terisolasi dianggap sebagai anomali [5]. IF dikenal efisien untuk data berdimensi tinggi dan berskala besar, serta banyak digunakan pada konteks yang tidak memiliki label anomali, seperti analisis konsumsi energi, deteksi pola belanja tidak wajar, dan pemantauan trafik jaringan [2]. Pengembangan *Extended Isolation Forest* (EIF) juga meningkatkan performa IF dengan mengurangi bias pemisahan menggunakan *hyperplane* berkemiringan acak [6]. Sementara itu, SVM merupakan metode supervised maupun *semi-supervised* yang efektif dalam memisahkan kelas normal dan tidak normal melalui *hyperplane* optimal, serta banyak digunakan pada deteksi intrusi jaringan, sistem kelistrikan industri, komunikasi IoT, *malware*, dan *fraud* transaksi *online* [7].

Walaupun penelitian terkait IF dan SVM berkembang pesat, terdapat sejumlah kesenjangan yang mencerminkan perlunya kajian sistematis. Pertama, domain penggunaan IF dan SVM sangat

bervariasi, namun belum ada pemetaan komprehensif mengenai area mana yang paling dominan dilakukan dan jenis dataset apa yang digunakan (misalnya *benchmark* publik, data simulasi, atau data nyata). Kedua, praktik pelaporan metrik evaluasi tidak konsisten. Penelitian berbasis SVM umumnya melaporkan metrik seperti *accuracy, precision, recall, F1-score,* dan ROC-AUC, sedangkan sebagian penelitian IF hanya menampilkan skor anomali atau visualisasi tanpa metrik evaluasi standar [8]. Ketiga, aspek reprodusibilitas penelitian masih rendah. Sebagian besar studi tidak menjelaskan *hyperparameter* secara detail, tidak menyertakan langkah *preprocessing*, tidak melaporkan *runtime* pelatihan maupun prediksi, serta hampir tidak menyediakan *repository* kode. Hanya sedikit penelitian, seperti implementasi *Extended Isolation Forest* di H2O-3, yang memberikan informasi teknis dan kode sumber secara terbuka [6].

Ketiadaan *systematic review* yang secara khusus membahas penggunaan *Isolation Forest* dan SVM untuk *anomaly detection pasca-2019*, khususnya dalam konteks penelitian Indonesia, menunjukkan adanya kebutuhan kajian menyeluruh. Untuk menjawab kebutuhan tersebut, penelitian ini merumuskan tiga pertanyaan penelitian sebagai berikut:

RQ1: Pada domain apa saja *Isolation Forest* dan *Support Vector Machine* digunakan untuk deteksi anomali, dan jenis dataset seperti apa yang dipakai (misalnya dataset *benchmark*, data simulasi, atau data nyata)?

RQ2: Metrik evaluasi apa saja yang digunakan dalam menilai kinerja model IF dan SVM pada studi-studi tersebut, dan bagaimana praktik pelaporan hasil evaluasinya? Apakah pelaporannya hanya terbatas pada akurasi, atau sudah mencakup *precision, recall, F1-score, dan ROC-AUC*?

RQ3: Bagaimana praktik pelaporan aspek reprodusibilitas dalam penelitian-penelitian tersebut, khususnya terkait penyajian *hyperparameter*, *skema preprocessing data*, *runtime* pelatihan/prediksi, serta ketersediaan kode atau repository publik yang memungkinkan eksperimen direplikasi?

Penelitian ini bertujuan untuk memetakan domain dan jenis dataset yang digunakan dalam studi-studi *anomaly detection* berbasis IF dan SVM, mengkaji pola pelaporan metrik evaluasi yang digunakan, serta menganalisis sejauh mana aspek reprodusibilitas telah diperhatikan dalam literatur terkini. Kontribusi utama penelitian ini adalah menyediakan gambaran komprehensif mengenai pola penerapan IF dan SVM, sekaligus memberikan dasar bagi penelitian selanjutnya untuk meningkatkan kualitas eksperimen, pelaporan, dan transparansi dalam riset deteksi anomali.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian-penelitian terdahulu menunjukkan bahwa *Isolation Forest* (IF) telah banyak dimanfaatkan untuk deteksi anomali pada berbagai domain. Al-Akbar mengusulkan model deteksi trafik anomali berbasis IF menggunakan *dataset* LUFlow yang merepresentasikan trafik jaringan nyata. Penelitian tersebut menerapkan serangkaian tahapan mulai dari *preprocessing*, standarisasi fitur, pemodelan IF, hingga evaluasi menggunakan *confusion matrix, precision, recall*, dan *F1-score*. Hasilnya menunjukkan bahwa model mampu mengidentifikasi outlier dalam trafik jaringan, meskipun performa deteksi terhadap serangan bot masih belum optimal dan akurasi deteksi *outlier* baru mencapai sekitar 49%, sehingga membuka ruang pengembangan lebih lanjut dalam pemilihan fitur dan konfigurasi *hyperparameter* model [2].

Pada konteks yang berbeda, Novaldi menerapkan *Isolation Forest* untuk mendeteksi anomali pada permintaan kebutuhan farmasi di Rumah Sakit Mitra Sejati Medan. Dengan menggunakan pendekatan *Cross-Industry Standard Process for Data Mining* (CRISP-DM), penelitian ini mengolah lebih dari 2,1 juta transaksi permintaan obat, barang medis habis pakai, dan alat kesehatan. Model IF yang dikembangkan berhasil mengidentifikasi sekitar 13.503 transaksi (0,62%) sebagai anomali statistik, yang kemudian dimanfaatkan sebagai dasar pengambilan keputusan dalam pengelolaan stok dan verifikasi permintaan oleh bagian farmasi. Implementasi sistem berbasis *web* menunjukkan bahwa IF dapat menjadi alat bantu keputusan yang efektif untuk meningkatkan efisiensi operasional dan akurasi stok farmasi [8].

Di sisi lain, metode *Support Vector Machine* (SVM) banyak digunakan pada permasalahan klasifikasi, termasuk untuk deteksi anomali dan penilaian kualitas layanan. Putri menerapkan SVM untuk klasifikasi kualitas air berdasarkan *Water Quality Index* (WQI) dengan sembilan parameter utama seperti pH, *hardness, solids, chloramines, sulfate, conductivity, organic carbon, trihalomethanes*, dan *turbidity*. Penelitian tersebut membandingkan beberapa jenis kernel, yaitu *linear, RBF, polynomial*, dan *sigmoid*, serta mengevaluasi performanya menggunakan metrik akurasi dan *F1-score*. Hasilnya menunjukkan bahwa kernel RBF dengan parameter optimum $C = 1000$ dan $\gamma = 4$ memberikan performa terbaik dengan akurasi mencapai 100%, sehingga direkomendasikan sebagai konfigurasi SVM yang paling efektif untuk klasifikasi kualitas air dalam studi tersebut [9].

Penerapan SVM pada domain keamanan jaringan IoT dilakukan oleh Luthfiawan yang mengembangkan sistem deteksi intrusi untuk mitigasi anomali komunikasi pada sistem IoT menggunakan *dataset* CIC-ToN-IoT. Penelitian ini melakukan pra-pemrosesan data berupa normalisasi, ekstraksi fitur,

dan *Principal Component Analysis* (PCA), kemudian melatih model SVM dengan tiga jenis kernel: linear, *polynomial*, dan *Radial Basis Function* (RBF). Evaluasi dilakukan menggunakan metrik akurasi, *precision, recall, F1-score*, serta *Area Under Curve* (AUC). Hasil eksperimen menunjukkan bahwa kernel polynomial orde dua memberikan performa paling seimbang dengan *recall* yang sangat tinggi, sementara kernel RBF unggul dari sisi nilai AUC, dan kernel linear relatif lebih efisien untuk data yang hampir linier terpisah [10]. Temuan ini menegaskan bahwa pemilihan kernel dan optimasi hyperparameter sangat menentukan keberhasilan SVM dalam mendeteksi anomali komunikasi pada perangkat IoT yang memiliki keterbatasan sumber daya komputasi.

Penelitian-penelitian terdahulu tersebut memperlihatkan bahwa baik *Isolation Forest* maupun *Support Vector Machine* telah berhasil diterapkan pada berbagai domain seperti trafik jaringan, manajemen permintaan farmasi, kualitas air, dan keamanan komunikasi IoT dengan konfigurasi dan metrik evaluasi yang beragam. Namun, masing-masing studi umumnya berfokus pada satu kasus atau satu domain tertentu sehingga belum tersedia pemetaan komprehensif lintas studi mengenai domain aplikasi, jenis *dataset*, metrik evaluasi, serta aspek reprodusibilitas eksperimen kedua metode tersebut. Kekosongan inilah yang menjadi dasar perlunya dilakukan *systematic review* untuk merangkum, membandingkan, dan mengkritisi penggunaan IF dan SVM dalam penelitian *anomaly detection* terkini.

2.2. Anomali dan Deteksi Anomali

Anomali merujuk pada data, pola, atau kejadian yang menyimpang dari perilaku normal atau ekspektasi umum [11]. Keberadaan anomali sering kali menjadi indikator awal adanya masalah seperti *fraud*, intrusi jaringan, *error* sistem, atau penyalahgunaan aset [12]. Dalam konteks data operasional, anomali dapat muncul akibat aktivitas tidak wajar, perubahan pola perilaku, *noise*, atau kesalahan input. Oleh karena itu, deteksi anomali menjadi proses penting untuk menjaga keamanan, integritas data, dan efisiensi operasional

Teknik deteksi anomali berkembang seiring bertambahnya volume dan kompleksitas data. Pendekatan klasik seperti *thresholding* dan statistika memiliki keterbatasan pada data yang bersifat non-linear atau berdimensi tinggi. Untuk menjawab tantangan tersebut, *machine learning* mulai digunakan untuk mengidentifikasi pola anomali secara otomatis dan skalabel [9].

2.3. *Machine Learning* untuk *Anomaly Detection*

Deteksi anomali dapat dilakukan dengan *supervised* maupun *unsupervised learning*. Pada *supervised learning*, model belajar memisahkan data normal dan anomali berdasarkan label [10]. Namun, banyak kasus dunia nyata tidak memiliki label anomali yang lengkap, sehingga pendekatan *unsupervised* lebih banyak digunakan.

Unsupervised learning bekerja dengan mencari pola intrinsik dalam data tanpa mengandalkan label [13]. Teknik yang umum digunakan meliputi *clustering*, *dimensionality reduction*, dan *anomaly scoring*. Karena sifatnya yang fleksibel dan tidak membutuhkan pelabelan data, metode *unsupervised* menjadi pilihan ketika jumlah anomali sangat sedikit, tidak terstruktur, atau tidak terdokumentasi.

2.4. *Isolation Forest* (IF)

Isolation Forest diperkenalkan oleh Kardiyasa & Santika (2025) sebagai algoritma *unsupervised* yang dirancang khusus untuk deteksi *anomaly* [14]. Konsep utamanya adalah bahwa instance anomali lebih mudah diisolasi melalui pemisahan acak menggunakan struktur pohon dibandingkan *instance* normal. Semakin pendek *path length* yang dibutuhkan untuk mengisolasi suatu data, semakin tinggi kemungkinan bahwa data tersebut merupakan anomali.

IF memiliki keunggulan pada efisiensi komputasi, skalabilitas, serta kemampuannya bekerja pada dataset berdimensi tinggi tanpa asumsi distribusi data [15]. Studi-studi terbaru menggunakan IF pada berbagai domain seperti deteksi anomali trafik jaringan [2], log aktivitas *website* [4], belanja persediaan konsumsi pemerintah [3], permintaan farmasi [8], hingga evaluasi kinerja dosen berbasis data survei [7].

EIF dikembangkan untuk mengatasi bias pemisahan pada IF klasik yang hanya menggunakan pemotongan *axis-aligned*. Dengan memperkenalkan *hyperplane* berkemiringan acak, EIF menghasilkan pemisahan yang lebih adaptif terhadap bentuk distribusi data [6]. EIF terbukti lebih akurat pada data berdimensi tinggi dan menyediakan *anomaly score* yang lebih stabil. Penelitian implementatif pada platform H2O-3 menunjukkan bahwa EIF memiliki performa tinggi, meskipun waktu pelatihan dapat meningkat seiring kompleksitas *hyperplane*.

2.5. *Support Vector Machine* (SVM)

SVM adalah algoritma *supervised learning* yang bekerja dengan mencari *hyperplane* terbaik untuk memisahkan kelas data [15]. Untuk deteksi anomali, SVM dapat digunakan sebagai *binary classifier* atau *One-Class SVM*. *Kernel trick* memungkinkan SVM menangani data *non-linear* secara efektif.

SVM telah diterapkan pada berbagai studi *anomaly detection* seperti fraud transaksi *online* [11], sistem kelistrikan industri [12], kualitas air [9], intrusi jaringan dan IoT [14], serta deteksi *malware* [15]. Keunggulan SVM terletak pada kemampuannya memberikan margin pemisahan yang optimal dan kinerja tinggi pada dataset dengan fitur kompleks.

2.6. Metrik Evaluasi pada *Anomaly Detection*

Evaluasi *anomaly detection* sangat bergantung pada ketersediaan label. Pada *supervised* SVM, metrik

umum yang digunakan meliputi *accuracy, precision, recall, F1-score,* dan *ROC-AUC* [11]. Studi berbasis SVM biasanya melaporkan kombinasi metrik tersebut secara konsisten.

Sebaliknya, penelitian yang menggunakan *Isolation Forest* sering kali tidak menyertakan metrik evaluasi formal karena tidak tersedia label anomali. Beberapa studi hanya menggunakan *anomaly score*, visualisasi *scatter plot*, atau pendekatan statistik seperti *silhouette score* atau *Rand Index* [7]. Hal ini menyebabkan perbandingan kinerja antarstudi menjadi kurang terstandarisasi.

2.7. *Reprodusibilitas* Penelitian *Machine Learning*

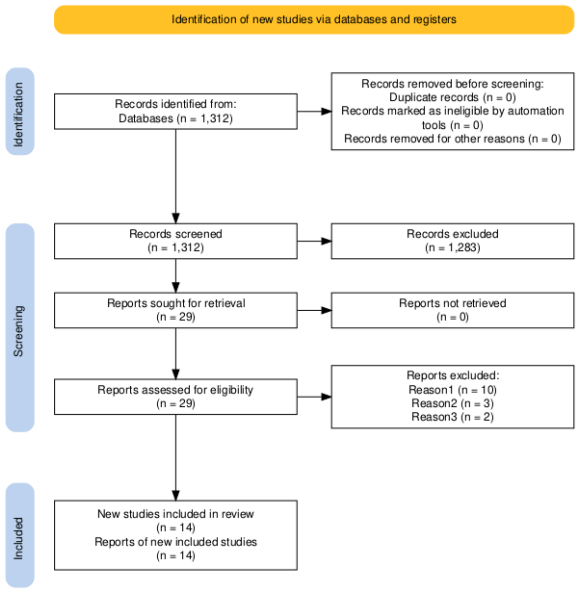
Reprodusibilitas mencakup pelaporan *hyperparameter, preprocessing* data, arsitektur model, *runtime*, serta ketersediaan kode sumber. Namun, banyak penelitian *anomaly detection* yang tidak menyertakan detail teknis tersebut. Penelitian IF untuk belanja pemerintah, misalnya, tidak menyebutkan *hyperparameter* selain *contamination* dan n_*estimators* [3]. Beberapa penelitian SVM juga hanya menyebut penggunaan kernel tanpa menyajikan nilai parameter seperti C, gamma, atau nu secara lengkap.

Penelitian *Extended Isolation Forest* menjadi salah satu contoh yang baik dalam reprodusibilitas karena menyediakan dokumentasi implementasi, analisis skalabilitas, serta kode publik di *platform* H2O-3 [6]. Kurangnya pelaporan ini menjadi salah satu alasan perlunya *systematic review* untuk menilai kualitas pelaporan eksperimen pada studi-studi terkait IF dan SVM.

3. METODE PENELITIAN

Penelitian ini menggunakan metode *Systematic Literature Review* (SLR) untuk mengidentifikasi, mengevaluasi, dan mensintesis penelitian-penelitian yang membahas penerapan *Isolation Forest* (IF) dan *Support Vector Machine* (SVM) dalam deteksi anomali. Metode SLR dipilih karena memberikan prosedur pengumpulan literatur yang terstruktur, objektif, dan dapat direplikasi dibandingkan metode tinjauan naratif biasa [2]. Pendekatan ini sangat relevan mengingat penelitian mengenai *anomaly detection* menggunakan IF dan SVM berkembang pesat di berbagai domain, namun masih bersifat terfragmentasi dan belum dipetakan secara komprehensif.

Fokus *systematic review* ini diarahkan pada tiga aspek utama sesuai rumusan masalah, yaitu: (1) pemetaan domain dan jenis dataset yang digunakan pada studi-studi IF dan SVM; (2) identifikasi metrik evaluasi yang dilaporkan peneliti; dan (3) analisis kualitas pelaporan eksperimen, termasuk *hyperparameter, preprocessing, runtime*, serta ketersediaan kode sumber. Tahapan *systematic review* mengikuti siklus umum SLR, yaitu identifikasi studi, seleksi studi berdasarkan kriteria inklusi–eksklusi, ekstraksi data, dan sintesis temuan secara deskriptif untuk menjawab RQ1–RQ3.



Gambar 1. Diagram Prisma

Pencarian literatur dilakukan menggunakan satu basis data daring, yaitu Google Scholar, karena menyediakan cakupan luas terhadap artikel jurnal dan prosiding akademik berbahasa Indonesia maupun Inggris [14]. Penelusuran dilakukan dengan menggunakan kombinasi kata kunci "anomali deteksi", "*isolation forest*", dan "svm", yang dipilih untuk menangkap artikel yang secara eksplisit mengimplementasikan IF atau SVM sebagai metode deteksi anomali. Rentang publikasi dibatasi pada tahun 2019 ke atas untuk memastikan bahwa literatur yang direview merupakan penelitian terbaru sesuai perkembangan mutakhir teknik *anomaly detection*. Selain itu, Google Scholar secara default mengurutkan hasil pencarian berdasarkan relevansi, sehingga penyaringan awal difokuskan pada hasil teratas karena artikel pada halaman yang semakin jauh umumnya memiliki relevansi yang lebih rendah terhadap topik penelitian.

Kriteria inklusi–eksklusi ditetapkan untuk menjaga validitas dan relevansi studi yang disertakan dalam review. Artikel dimasukkan apabila memenuhi syarat sebagai berikut: (1) merupakan jurnal atau prosiding ilmiah berformat full-text; (2) terbit pada tahun 2019 atau lebih baru; (3) menjadikan *anomaly detection* sebagai fokus utama penelitian; (4) menggunakan *Isolation Forest* dan/atau SVM sebagai metode inti atau metode pembanding; dan (5) menyajikan hasil pengujian berbasis *dataset* nyata ataupun *benchmark* publik. Sebaliknya, artikel dikeluarkan apabila hanya tersedia dalam bentuk abstrak atau poster, bersifat *non-peer-reviewed*, tidak relevan dengan *anomaly detection*, hanya membahas IF atau SVM pada tataran teori tanpa implementasi, merupakan duplikasi, atau berfokus pada metode lain selain IF dan SVM. Penyusunan kriteria ini mengacu pada prinsip seleksi literatur yang lazim dalam SLR untuk meminimalisasi bias seleksi [8].

Proses seleksi studi dilakukan dalam tiga tahap: identifikasi, *screening*, dan evaluasi *full-text*. Pada tahap identifikasi, pencarian kata kunci "anomali deteksi" AND "isolation forest" menghasilkan sekitar 202 artikel, dan pencarian "anomali deteksi" AND "svm" menghasilkan sekitar 1.312 artikel. Seluruh hasil pencarian kemudian diseleksi berdasarkan judul dan abstrak untuk menghilangkan artikel yang tidak relevan atau tidak memenuhi format ilmiah. Artikel yang lolos pada tahap *screening* kemudian dianalisis secara full-text untuk memastikan keterlibatan IF dan/atau SVM secara eksplisit dalam eksperimen *anomaly detection* serta ketersediaan data hasil evaluasi. Setelah seluruh tahap seleksi dilalui, diperoleh 14 artikel yang memenuhi kriteria, terdiri atas 7 studi bertema *Isolation Forest* dan 7 studi bertema *Support Vector Machine*. Artikel-artikel ini menjadi dasar analisis untuk menjawab ketiga pertanyaan penelitian.

Artikel yang memenuhi kriteria selanjutnya diekstraksi menggunakan formulir ekstraksi data terstruktur. Informasi yang diambil mencakup domain dan jenis *dataset*, metrik evaluasi, teknik *preprocessing*, konfigurasi model (*hyperparameter*), detail implementasi, runtime komputasi, serta ketersediaan kode sumber. Ekstraksi data ini mengikuti praktik umum SLR yang menekankan pentingnya pencatatan variabel-variabel metodologis yang berpengaruh terhadap kualitas dan interpretasi hasil penelitian [9]. Pada studi IF, *hyperparameter* yang dicari meliputi_*estimators*, *max_samples*, dan *contamination*; sedangkan pada studi SVM meliputi jenis *kernel*, parameter C, *gamma*, *nu*, dan *degree*. Data *preprocessing* yang dicatat mencakup pembersihan data, normalisasi atau standardisasi, *encoding* variabel kategorikal, *feature selection*, serta penerapan PCA.

Setelah seluruh data terekstraksi, sintesis dilakukan secara deskriptif dengan mengelompokkan temuan dari studi IF dan SVM berdasarkan tema analisis. Sintesis ini mencakup pembuatan tabel perbandingan, penghitungan frekuensi karakteristik tertentu (misalnya jumlah studi yang melaporkan *accuracy* atau menyertakan *hyperparameter* lengkap), serta penyusunan narasi interpretatif untuk menjelaskan pola-pola yang ditemukan. Mengingat *heterogenitas dataset*, domain, dan metrik antarstudi yang cukup besar, penelitian ini tidak melakukan meta-analisis kuantitatif, melainkan menghasilkan sintesis naratif untuk menjawab RQ1–RQ3. Pendekatan ini sesuai dengan rekomendasi SLR ketika literatur bersifat variatif dan tidak memungkinkan integrasi numerik yang seragam [10].

Dengan demikian, metode penelitian ini memastikan bahwa proses peninjauan literatur dilakukan secara sistematis, transparan, dan dapat direplikasi, sekaligus menghasilkan gambaran menyeluruh mengenai pola penggunaan *Isolation Forest* dan *Support Vector Machine* dalam penelitian *anomaly detection* terkini.

4. HASIL DAN PEMBAHASAN

Berdasarkan proses seleksi literatur yang telah dijelaskan pada metode penelitian, diperoleh empat belas artikel yang memenuhi kriteria inklusi dan dianalisis lebih lanjut dalam *systematic review* ini. Empat belas studi tersebut terdiri dari tujuh penelitian yang menggunakan *Isolation Forest* (IF) dan tujuh penelitian yang menggunakan *Support Vector Machine* (SVM) dalam konteks deteksi anomali. Seluruh artikel terbit setelah tahun 2019 sehingga sesuai dengan batasan temporal yang ditetapkan. Secara umum, penelitian-penelitian tersebut mencerminkan dua karakter utama, yakni studi terapan yang mengimplementasikan IF dan SVM pada berbagai kasus nyata serta satu studi algoritmik yang mengembangkan metode *Extended Isolation Forest* (EIF) pada *platform* H2O-3 [6].

Hasil ekstraksi awal menunjukkan bahwa IF dan SVM diterapkan pada domain yang beragam dengan karakteristik dataset yang berbeda-beda. *Isolation Forest* digunakan pada keamanan jaringan [2], analisis transaksi dan belanja pemerintah [7], analisis log aktivitas *website* [4], evaluasi kinerja dosen berbasis data *Likert* [7], hingga eksperimen *algoritmik* menggunakan data *Gaussian* sintetis dalam studi EIF [6]. Sebaliknya, SVM cenderung digunakan pada *domain* yang lebih fokus pada keamanan dan sistem, seperti *fraud* transaksi *online* [11], deteksi gangguan sistem kelistrikan industri [12], kualitas air [9], deteksi intrusi jaringan menggunakan *dataset* Kaggle dan CIC-ToN-IoT [10], gangguan jaringan FTTH [14], serta analisis perilaku proses *malware* [15]. Temuan ini mengindikasikan bahwa IF lebih fleksibel untuk masalah tanpa label (*unsupervised*), sedangkan SVM lebih banyak digunakan pada *dataset* dengan label normal/anomali yang jelas.

Perbedaan domain dan sumber dataset turut berdampak pada teknik evaluasi yang digunakan masing-masing metode. Studi-studi berbasis SVM secara konsisten melaporkan metrik performa klasifikasi seperti *Accuracy*, *Precision*, *Recall*, dan *F1-score*, serta beberapa studi juga menambahkan ROC-AUC [9]; [15]; [13]. Konsistensi pelaporan metrik ini memudahkan pembanding antarmodel serta memungkinkan interpretasi performa yang lebih objektif. Hal yang kontras tampak pada penelitian IF, karena hanya sebagian kecil studi yang melaporkan metrik klasifikasi lengkap. Beberapa penelitian IF hanya menampilkan visualisasi skor anomali, statistik dasar, atau uji statistik tanpa metrik klasifikasi formal [3]; [7]; [4]. Satu studi menggunakan *Rand Index* dan *Silhouette Score* dalam konteks *clustering* [7], menunjukkan bahwa studi IF cenderung lebih heterogen dalam pelaporannya dibandingkan studi SVM.

Temuan lain menunjukkan variasi yang signifikan dalam pelaporan konfigurasi model dan *preprocessing*. Pada IF, beberapa artikel menjelaskan *hyperparameter* seperti *estimators*, *max_samples*, dan *contamination*, terutama pada penelitian keuangan

publik yang menekankan sensitivitas parameter *threshold* anomali [3]. Studi EIF memberikan informasi teknis yang sangat rinci mengenai struktur *hyperparameter*, *extension level*, serta pengaruh konfigurasi terhadap waktu komputasi [6]. Namun demikian, terdapat juga studi IF yang tidak menjelaskan konfigurasi model sehingga sulit direplikasi. Pada SVM, sebagian besar studi melaporkan jenis kernel yang digunakan serta nilai *hyperparameter* seperti C, *gamma*, *degree*, dan variasi proses *grid search* yang dilakukan untuk *tuning* parameter [10]; [9]; [13]. Studi SVM umumnya menerapkan langkah *preprocessing* yang lebih komprehensif, seperti normalisasi fitur, *encoding* variabel kategorikal, reduksi dimensi menggunakan PCA, serta penanganan class *imbalance* seperti *oversampling* [11]; [10]. Hal ini menunjukkan bahwa pipeline SVM cenderung lebih terdokumentasi secara sistematis dibandingkan pipeline IF

Dari perspektif reprodusibilitas, hasil *review* menunjukkan bahwa aspek ini merupakan salah satu kelemahan terbesar, baik pada studi IF maupun SVM. Hanya satu studi, yaitu penelitian *Extended Isolation Forest* [6], yang menyediakan kode sumber terbuka serta menyajikan analisis *runtime* secara kuantitatif. Enam penelitian IF lainnya tidak memberikan kode maupun runtime, dan sebagian hanya menampilkan potongan kode tanpa informasi lingkungan komputasi. Pada SVM, tidak ada satu pun studi yang menyediakan repository publik, dan informasi runtime hampir tidak pernah dilaporkan secara numerik kecuali dalam bentuk pernyataan kualitatif tanpa angka konkret [10]. Minimnya laporan reprodusibilitas ini berdampak signifikan pada kemampuan peneliti lain untuk memverifikasi hasil, terutama pada konteks *anomaly detection* yang sangat sensitif terhadap perubahan *hyperparameter* dan kondisi komputasi.

Hasil temuan dari seluruh studi memperlihatkan bahwa penggunaan IF dan SVM menunjukkan karakter tren yang berbeda. *Isolation Forest* fleksibel digunakan untuk domain yang benar-benar unsupervised dan bervariasi, menjadikannya pilihan yang tepat untuk skenario keterbatasan label, misalnya pada data transaksi pemerintah atau *log* aktivitas sistem. Sebaliknya, SVM menonjol pada domain dengan struktur label yang jelas, terutama pada keamanan jaringan, IoT, *malware*, dan *fraud*, karena kemampuannya mempelajari *decision boundary* dengan baik. Konsistensi pelaporan SVM menjadikannya lebih mudah direplikasi, meskipun ketiadaan kode publik masih menjadi kekurangan besar. Di sisi lain, IF membutuhkan peningkatan standar pelaporan terutama dalam hal *hyperparameter* dan metrik evaluasi.

Hasil *review* ini juga memberikan implikasi praktis. Pertama, ketika data tidak memiliki label atau memiliki tingkat ketidakseimbangan yang tinggi, IF dapat menjadi metode yang efisien dan sederhana untuk mendeteksi anomali awal. Kedua, ketika data memiliki label yang cukup lengkap, SVM menjadi pilihan lebih kuat karena dapat dipertajam dengan tuning kernel dan optimasi parameter. Ketiga, preprocessing terbukti merupakan komponen penting dalam *pipeline anomaly detection*, terutama pada SVM yang sangat sensitif terhadap skala fitur. Penggunaan normalisasi, *encoding*, *feature selection*, dan PCA terbukti berkontribusi pada tingginya performa model pada beberapa studi. Keempat, minimnya pelaporan runtime dan ketersediaan kode menunjukkan perlunya peningkatan standar dokumentasi di penelitian lokal agar penelitian *anomaly detection* menjadi lebih transparan, objektif, dan mudah direplikasi.

Adapun keterbatasan *systematic review* ini perlu dicatat. Literasi pencarian hanya dilakukan pada Google Scholar sehingga tidak menutup kemungkinan terdapat penelitian relevan di basis data lain yang terlewat. Jumlah artikel yang terbatas dan domain yang sangat heterogen juga membatasi kemampuan *review* ini untuk melakukan meta-analisis kuantitatif. Selain itu, keterbatasan pelaporan informasi pada banyak studi menyebabkan beberapa bagian analisis harus ditandai sebagai "tidak dilaporkan," sehingga kedalaman interpretasi menjadi terbatas. Meski demikian, temuan-temuan dalam *review* ini tetap memberikan gambaran yang jelas mengenai kondisi riset *anomaly detection* berbasis IF dan SVM setelah tahun 2019, serta membuka arah penelitian lanjutan yang lebih kuat dan komparatif.

5. KESIMPULAN DAN SARAN

Berdasarkan *systematic review* terhadap 14 artikel yang menggunakan *Isolation Forest* (IF) dan *Support Vector Machine* (SVM) untuk *anomaly detection* sejak tahun 2019, dapat disimpulkan bahwa IF cenderung dimanfaatkan pada domain yang lebih beragam (keamanan jaringan, keuangan publik dan pembayaran tunjangan, log aktivitas *website*, evaluasi kinerja dosen, serta analisis konsumsi atau realisasi keuangan), sedangkan SVM lebih dominan diaplikasikan pada domain yang berhubungan dengan keamanan dan sistem (*intrusion detection*, komunikasi IoT, gangguan jaringan FTTH, deteksi *malware*, anomali transaksi, dan sistem kelistrikan), dengan keduanya sama-sama memanfaatkan kombinasi *dataset* publik dan *dataset* internal. Dari sisi metrik evaluasi, studi-studi SVM jauh lebih konsisten melaporkan *Accuracy*, *Precision*, *Recall/Sensitivity*, *F1-score*, dan kadang ROC-AUC, sementara sebagian studi IF masih terbatas pada visualisasi skor anomali atau analisis statistik sehingga sulit dibandingkan secara numerik. Pelaporan *hyperparameter* dan *preprocessing* pada SVM juga lebih rinci dan terstruktur dibandingkan IF, namun dari sisi reprodusibilitas kedua kelompok masih lemah karena hampir semua studi tidak melaporkan *runtime* secara kuantitatif dan tidak menyediakan kode sumber, kecuali satu studi *Extended Isolation Forest* yang membagikan implementasi dan analisis skalabilitas. Berdasarkan temuan tersebut, disarankan agar

penelitian selanjutnya meningkatkan kualitas pelaporan eksperimen dengan selalu menyajikan metrik evaluasi yang lengkap, mendokumentasikan *hyperparameter* beserta prosedur *tuning*, serta menyertakan informasi *runtime* dan, bila memungkinkan, kode sumber atau konfigurasi eksperimen dalam *repository* publik; selain itu, diperlukan studi komparatif yang lebih sistematis antara IF, SVM (termasuk *One-Class* SVM), dan metode *anomaly detection* lainnya pada beberapa *dataset benchmark* yang sama dengan protokol evaluasi yang seragam, serta perluasan cakupan systematic review menggunakan lebih banyak basis data dan melibatkan lebih dari satu peninjau agar peta penggunaan IF dan SVM dalam *anomaly detection* menjadi lebih komprehensif, kuat, dan dapat dijadikan landasan bagi penelitian lanjutan maupun implementasi praktis di berbagai domain.

DAFTAR PUSTAKA

[1] A. C. T. Angel, V. H. Pranatawijaya, and Widiatry, "Analisis Sentimen dan Emosi dari Ulasan Google Maps untuk Layanan Rumah Sakit di Palangka Raya Menggunakan Machine Learning," *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 4, no. 1, pp. 35–49, 2024.

[2] M. 'Azam Al-Akbar, A. P. Y, and A. N. A. H. Gurning, "Deteksi Trafik Anomali Berdasarkan Pola Trafik Menggunakan Isolation Forest," *Cosm. J. Tek.*, vol. 2, no. 3, pp. 88–95, 2025.

[3] A. Zulfikar, F. A. Rahmani, and N. Azizah, "DETEKSI ANOMALI MENGGUNAKAN ISOLATION FOREST BELANJA BARANG PERSEDIAAN KONSUMSI PADA SATUAN KERJA KEPOLISIAN REPUBLIK INDONESIA," *J. Manaj. Perbendaharaan*, vol. 4, no. 1, pp. 1–15, 2023.

[4] D. A. Artika, D. Rumahorbo, M. H. Al-Majid, and D. Kiswanto, "IMPLEMENTASI SISTEM KEAMANAN WEBSITE DENGAN ANALISIS LOG DAN DETEKSI AKTIVITAS ANOMALI MENGGUNAKAN ISOLATION FOREST," *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 13, no. 3, pp. 1968–1877, 2025.

[5] C. L. S. Putri and R. Rachman, "Deteksi Anomali Pembayaran TPD dan TKGB dengan Isolation Forest dan Evaluasi Risiko Berbasis COSO ERM," *J. Ilm. Inform. Glob.*, vol. 16, no. 2, pp. 307–312, 2025.

[6] M. W. Sunarto, D. Kurniawan, E. Siswanto, and H. I. Huda, "Deteksi Anomali Menggunakan Extended Isolation Forest (Eif)," *J. ILMU Tek. DAN Inform.*, vol. 1, no. 2, pp. 96–111, 2021.

[7] Mutmainah and W. Yustanti, "Studi Komparasi Local Outlier Factor (LOF) dan Isolation Forest (IF) pada Analisis Anomali Kinerja Dosen," *JINACS (Journal Informatics Comput. Sci.*, vol. 06, no. 02, pp. 532–540, 2024.

[8] F. Novaldi, Syafrijon, Y. Hendriyani, and M. Anwar, "Deteksi Anomali menggunakan Isolation Forest pada Permintaan Kebutuhan Farmasi Pasien di Rumah Sakit Mitra Sejati Medan," *J. Pendidik. Tambusai*, vol. 9, no. 2, pp. 25367–25374, 2025.

[9] S. S. M. Putri, M. Arhami, and Hendrawaty, "Penerapan Metode SVM pada Klasifikasi Kualitas Air," *JAISE J. Artif. Intell. Softw. Eng.*, vol. 3, no. 2, pp. 94–101, 2023.

[10] F. R. Luthfiawan, "PENERAPAN SVM UNTUK MITIGASI ANOMALI KOMUNIKASI PADA SISTEM IoT," *IKRAITH-INFORMATIKA*, vol. 9, no. 2, pp. 175–186, 2025.

[11] H. Eldo, Ayuliana, D. Suryadi, G. Chrisnawati, and L. Judijanto, "Penggunaan Algoritma Support Vector Machine (SVM) Untuk Deteksi Penipuan pada Transaksi Online," *J. Minfo Polgan*, vol. 13, no. 2, pp. 1627–1632, 2024.

[12] B. M. Basuki and D. Cahyono, "Model Machine Learning SVM (Support Vector Machine) untuk Deteksi Anomali pada Sistem Kelistrikan Perusahaan Kerajinan Kayu GS4," *J. Tek. Mesin, Ind. Elektro dan Inform.*, vol. 4, no. 1, pp. 310–320, 2025.

[13] R. P. S. Pohan, A. P. Juledi, and I. R. Munthe, "Penerapan Algoritma Support Vector Machine untuk Mendeteksi Anomali pada Jaringan Komputer," *J. Ilmu Komput. dan Sist. Inf.*, vol. 00, no. 00, pp. 174–182, 2019.

[14] K. A. P. Kardiyasa and P. A. Santika, "Teknik Machine Learning dengan Metode Support Vector Machine (SVM) untuk Deteksi Anomali pada Kendala Jaringan Fiber to The Home (FTTH)," *Maj. Ilm. Teknol. Elektro*, vol. 24, no. 1, pp. 71–78, 2025.

[15] H. S. Simbolon and A. Maslan, "DETEKSI SERANGAN MALWARE MENGGUNAKAN METODE SUPPORT VECTOR MACHINE," *J. Comasie*, vol. 13, no. 02, pp. 119–130, 2025.