INJOPHARS (Innovation Journal of Pharmaceutical Science)
Vol. 2, No. 2, November 2026, pp. 73-80
DOI: 10.68300/injophars.v2i2.119
E-ISSN 3108-9542

Research Article Open Access (CC–BY-SA)

Klasifikasi Tingkat Keparahan Kerentanan Keamanan Siber pada Sistem Kesehatan Menggunakan Algoritma Random Forest

Classification of Cybersecurity Vulnerability Severity Levels in Healthcare Systems Using the Random Forest Algorithm

Firman Aziz[1,*]; Irmawati[2]; Ayu Lestari Azis[3]; Wahyuni Azis[4]
[1] *Universitas Pancasakti, Makassar 90121, Indonesia*
[2, 3] *IRMEX Digital Akademika, Makassar 90155, Indonesia*
[4] *Branch Educatian and Training Center, Makassar 90155, Indonesia*
[1] *firman.aziz@unpacti.ac.id*; [2] *irmawati@irmexdigika.com*; [3] *ayulestariazis@irmexdigika.com*; [4] *wahyuni.betc@gmail.com*
** Corresponding author*

Abstrak

Keamanan siber merupakan aspek penting dalam sistem kesehatan karena kerentanan pada perangkat, aplikasi, dan infrastruktur teknologi informasi dapat mengancam kerahasiaan, integritas, serta ketersediaan informasi dan layanan kesehatan. Penelitian ini bertujuan mengevaluasi algoritma Random Forest untuk mengklasifikasikan tingkat keparahan kerentanan keamanan siber ke dalam empat kelas, yaitu LOW, MEDIUM, HIGH, dan CRITICAL. Dataset yang digunakan adalah Healthcare Cybersecurity Vulnerabilities Dataset yang terdiri atas 1.515 baris dan 11 atribut. Sebanyak 18 data tanpa label Severity dikeluarkan sehingga diperoleh 1.497 data berlabel. Fitur teks diproses menggunakan Term Frequency–Inverse Document Frequency (TF-IDF), sedangkan fitur kategorikal diproses menggunakan One-Hot Encoding. Data dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan stratifikasi. Model Random Forest dibangun menggunakan 300 pohon keputusan dan pembobotan kelas untuk menangani ketidakseimbangan data. Pada pengujian terhadap 300 data, model memperoleh akurasi 71,00%, balanced accuracy 62,80%, macro F1-score 66,02%, dan weighted F1-score 70,47%. Recall tertinggi diperoleh pada kelas MEDIUM sebesar 81,94%, sedangkan recall kelas CRITICAL hanya mencapai 43,33%. Hasil tersebut menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar, tetapi performanya belum merata pada seluruh kelas. Evaluasi lanjutan melalui validasi silang, optimasi hiperparameter, dan perbandingan dengan algoritma lain diperlukan sebelum model dipertimbangkan untuk mendukung prioritas penanganan kerentanan pada sistem kesehatan.

Kata Kunci: keamanan siber; sistem kesehatan; kerentanan; Random Forest; klasifikasi; machine learning.

Abstract

Cybersecurity is an important aspect of healthcare systems because vulnerabilities in devices, applications, and information technology infrastructure can threaten the confidentiality, integrity, and availability of healthcare information and services. This study evaluates the Random Forest algorithm for classifying cybersecurity vulnerability severity into four classes: LOW, MEDIUM, HIGH, and CRITICAL. The dataset used was the Healthcare Cybersecurity Vulnerabilities Dataset, containing 1,515 rows and 11 attributes. Eighteen records without a Severity label were excluded, leaving 1,497 labeled records. Text features were processed using Term Frequency–Inverse Document Frequency (TF-IDF), while categorical features were processed using One-Hot Encoding. The data were divided into training and testing sets using a stratified 80:20 split. The Random Forest model used 300 decision trees and class weighting to address class imbalance. On the 300-record test set, the model achieved 71.00% accuracy, 62.80% balanced accuracy, a 66.02% macro F1-score, and a 70.47% weighted F1-score. The highest recall was obtained for the MEDIUM class at 81.94%, whereas recall for the CRITICAL class reached only 43.33%. These results indicate that the model can correctly classify a substantial proportion of the records, but its performance is uneven across classes. Further evaluation through cross-validation, hyperparameter optimization, and comparison with other algorithms is required before the model can be considered for supporting vulnerability prioritization in healthcare systems.

Keywords: *cybersecurity; health systems; vulnerability; Random Forest; classification; machine learning.*

Pendahuluan

Perkembangan teknologi informasi dan komunikasi telah mendorong transformasi digital pada sektor kesehatan melalui pemanfaatan rekam medis elektronik, sistem informasi rumah sakit, perangkat medis yang terhubung ke jaringan, dan berbagai aplikasi pelayanan kesehatan. Penerapan teknologi tersebut mendukung efisiensi operasional,

pertukaran informasi, dan peningkatan kualitas pelayanan. Namun, meningkatnya ketergantungan terhadap sistem digital juga memperluas potensi risiko keamanan siber yang dapat mengganggu kerahasiaan, integritas, dan ketersediaan informasi maupun layanan kesehatan. Ewoh dan Vartiainen mengidentifikasi sejumlah faktor yang berkontribusi terhadap kerentanan sistem kesehatan terhadap serangan siber, termasuk kompleksitas perangkat yang terhubung ke jaringan, penggunaan sistem lama (*legacy systems*), keterbatasan investasi, serta perkembangan teknologi digital [1]. Oleh karena itu, keamanan siber menjadi salah satu aspek penting dalam menjaga keberlangsungan operasional sistem kesehatan.

Salah satu permasalahan utama dalam keamanan siber adalah keberadaan kerentanan pada perangkat keras, perangkat lunak, aplikasi, dan konfigurasi sistem. Kerentanan tersebut dapat dimanfaatkan oleh pihak yang tidak berwenang untuk mengakses sistem, memanipulasi informasi, atau mengganggu ketersediaan layanan. Dalam lingkungan kesehatan, dampak gangguan keamanan tidak hanya berkaitan dengan kebocoran informasi, tetapi juga berpotensi menghambat akses terhadap sistem pendukung pelayanan medis. Tinjauan sistematis mengenai keamanan siber pada sektor kesehatan juga menekankan pentingnya penerapan kontrol keamanan yang sesuai untuk mengurangi risiko terhadap infrastruktur dan perangkat medis [2]. Dengan demikian, identifikasi serta penilaian tingkat keparahan kerentanan diperlukan untuk membantu organisasi kesehatan menentukan prioritas penanganan berdasarkan karakteristik risiko yang dimiliki setiap kerentanan.

Penilaian tingkat keparahan kerentanan dapat mengacu pada *Common Vulnerability Scoring System* (CVSS), yaitu kerangka penilaian yang menyediakan pendekatan terstandar untuk menggambarkan karakteristik dan dampak kerentanan keamanan teknologi informasi [3]. Informasi kerentanan juga dapat diidentifikasi melalui *Common Vulnerabilities and Exposures* (CVE), yang menyediakan pengenal untuk membedakan entri kerentanan yang diketahui. Dalam praktiknya, informasi kerentanan dapat memiliki beragam atribut, seperti deskripsi teknis, jalur serangan (*attack vector*), jenis kelemahan (*weakness*), tanggal publikasi, dan skor keparahan. Keragaman atribut tersebut memberikan peluang untuk menganalisis pola kerentanan menggunakan pendekatan berbasis data. Meskipun demikian, jumlah informasi yang perlu dianalisis dan variasi karakteristik antarkerentanan dapat menyulitkan proses klasifikasi apabila seluruh analisis dilakukan secara manual.

Machine learning merupakan salah satu pendekatan yang dapat digunakan untuk mempelajari pola dari data historis dan menghasilkan prediksi berdasarkan karakteristik yang tersedia. Salah satu algoritma yang relevan untuk permasalahan klasifikasi adalah Random Forest. Algoritma ini merupakan metode *ensemble learning* yang membangun sejumlah pohon keputusan dan menggabungkan hasil prediksi pohon-pohon tersebut untuk menentukan kelas akhir. Breiman menjelaskan bahwa Random Forest mengombinasikan prediktor berbasis pohon dengan pemilihan fitur secara acak, sehingga dapat menghasilkan kinerja klasifikasi yang kompetitif dan relatif tangguh terhadap gangguan pada data [4]. Karakteristik tersebut menjadikan Random Forest sebagai salah satu kandidat metode untuk mengklasifikasikan tingkat keparahan kerentanan berdasarkan kombinasi atribut numerik, kategorikal, dan tekstual yang telah melalui tahap preprocessing.

Penelitian ini menggunakan *Healthcare Cybersecurity Vulnerabilities Dataset* yang tersedia dalam format CSV. Berdasarkan pemeriksaan awal terhadap file yang digunakan, dataset terdiri atas 1.515 baris dan 11 atribut, yang mencakup identitas kerentanan, kata kunci, tanggal publikasi, status, deskripsi, tingkat keparahan, skor CVSS, jalur serangan, kelemahan keamanan, dan referensi. Variabel target penelitian adalah atribut *Severity*, yang dikelompokkan menjadi empat kelas, yaitu LOW, MEDIUM, HIGH, dan CRITICAL. Sebanyak 18 baris tidak memiliki label *Severity*, sehingga data tersebut tidak digunakan dalam klasifikasi terawasi. Dengan karakteristik tersebut, penelitian dapat diarahkan untuk mempelajari hubungan antara atribut yang tersedia dan kategori tingkat keparahan kerentanan.

Tantangan penting dalam penelitian ini adalah menentukan representasi fitur yang sesuai serta mengevaluasi kemampuan model pada setiap kelas. Deskripsi kerentanan mengandung informasi tekstual yang dapat membantu membedakan karakteristik kelemahan, sedangkan atribut seperti *Attack_Vector* dan *Weakness* memberikan informasi kategorikal mengenai jalur serangan dan jenis kelemahan. Oleh karena itu, penelitian ini mengintegrasikan representasi teks menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF) dengan pengolahan atribut kategorikal sebelum data digunakan untuk melatih Random Forest. Selain itu, evaluasi tidak hanya didasarkan pada akurasi keseluruhan, tetapi juga menggunakan *precision*, *recall*, *F1-score*, *balanced accuracy*, dan *confusion matrix*. Penggunaan beberapa metrik diperlukan karena model dapat menghasilkan akurasi yang relatif tinggi, tetapi tetap memiliki kelemahan dalam mengenali kelas tertentu, terutama apabila distribusi data antarkelas tidak seimbang.

Meskipun penelitian mengenai keamanan siber di sektor kesehatan telah membahas faktor risiko, kerentanan infrastruktur, dan kontrol mitigasi [1], [2], masih diperlukan evaluasi empiris terhadap penerapan model klasifikasi pada dataset kerentanan yang memiliki karakteristik khusus sektor kesehatan. Penelitian ini berfokus pada penerapan Random Forest untuk mengklasifikasikan empat kategori tingkat keparahan menggunakan kombinasi informasi tekstual dan kategorikal. Fokus evaluasi per kelas juga digunakan untuk mengidentifikasi kategori yang masih sulit dikenali model. Kesenjangan penelitian yang hendak dikaji dalam studi ini adalah bagaimana pendekatan tersebut bekerja pada dataset yang digunakan, seberapa baik model membedakan setiap kategori keparahan, dan sejauh mana kesalahan klasifikasi terjadi pada kategori CRITICAL. Kesenjangan tersebut diposisikan sebagai fokus evaluasi penelitian ini, bukan sebagai klaim bahwa belum pernah ada penelitian lain yang menggunakan Random Forest untuk klasifikasi kerentanan.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan algoritma Random Forest dalam klasifikasi tingkat keparahan kerentanan keamanan siber pada sistem kesehatan, mengevaluasi kinerja model menggunakan metrik klasifikasi, serta menganalisis pola kesalahan melalui *confusion matrix*. Kontribusi penelitian ini meliputi penerapan alur preprocessing dan ekstraksi fitur yang menggabungkan data tekstual dan kategorikal, evaluasi performa klasifikasi pada kelas LOW, MEDIUM, HIGH, dan CRITICAL, serta identifikasi keterbatasan model dalam mendeteksi kerentanan dengan tingkat keparahan tinggi. Hasil penelitian diharapkan dapat memberikan gambaran empiris mengenai potensi dan keterbatasan Random Forest sebagai pendekatan pendukung analisis kerentanan. Model yang dihasilkan tidak dimaksudkan untuk menggantikan penilaian risiko oleh tenaga keamanan siber, melainkan sebagai pendekatan analitis yang masih memerlukan validasi lebih lanjut sebelum diterapkan dalam lingkungan operasional.

Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk mengklasifikasikan tingkat keparahan kerentanan keamanan siber pada sistem kesehatan menggunakan algoritma Random Forest. Tahapan penelitian meliputi pengumpulan data, preprocessing, pembentukan fitur, pembagian data, pelatihan model, prediksi, dan evaluasi kinerja. Proses pengolahan data dan pemodelan dilakukan menggunakan Python dengan pustaka Scikit-learn. Alur penelitian dirancang agar proses klasifikasi dan evaluasi dapat dilakukan secara sistematis dan terukur.

A. Dataset

Dataset yang digunakan adalah *Healthcare Cybersecurity Vulnerabilities Dataset* yang diperoleh dari Kaggle dalam format CSV [5]. Berdasarkan pemeriksaan file, dataset memiliki 1.515 baris dan 11 atribut, yaitu CVE_ID, Keyword, Published, Last_Modified, Status, Description, Severity, CVSS_Score, Attack_Vector, Weakness, dan Reference.

Atribut Severity digunakan sebagai variabel target yang terdiri atas empat kelas, yaitu LOW, MEDIUM, HIGH, dan CRITICAL. Distribusi label awal terdiri atas 134 data LOW, 720 data MEDIUM, 492 data HIGH, dan 151 data CRITICAL. Selain itu, terdapat 18 baris yang tidak memiliki label Severity. Karena penelitian ini menggunakan pendekatan klasifikasi terawasi, baris tanpa label target dikeluarkan dari proses pemodelan. Dengan demikian, jumlah data yang digunakan adalah 1.497 baris berlabel.

Tabel 1. Distribusi Kelas Target

Kelas	*Jumlah data*	*Persentase*
LOW	134	8,95%
MEDIUM	720	48,10%
HIGH	492	32,87%
CRITICAL	151	10,09%
Total	**1.497**	**100,00%**

Distribusi tersebut menunjukkan bahwa kelas MEDIUM memiliki jumlah sampel terbesar, sedangkan kelas LOW memiliki jumlah sampel terkecil. Ketidakseimbangan ini perlu diperhatikan karena dapat memengaruhi kemampuan model dalam mengenali setiap kategori. Oleh sebab itu, evaluasi dilakukan menggunakan metrik keseluruhan dan metrik per kelas.

B. Preprocessing Data

Preprocessing dilakukan untuk memastikan bahwa data dapat diproses oleh algoritma klasifikasi. Tahapan preprocessing yang diterapkan adalah sebagai berikut.

1. Pembersihan label target. Baris yang tidak memiliki nilai Severity dihapus. Nilai label yang tersisa diseragamkan dengan menghilangkan spasi di awal dan akhir serta mengubah seluruh karakter menjadi huruf kapital.
2. Penanganan nilai kosong. Nilai kosong pada atribut Description diganti dengan string kosong. Nilai kosong pada atribut kategorikal Keyword, Attack_Vector, Weakness, dan Status diganti dengan kategori Unknown.
3. Ekstraksi tahun publikasi. Atribut Published dikonversi menjadi format tanggal, kemudian tahun publikasi diekstraksi menjadi fitur Published_Year. Tanggal yang tidak dapat dikonversi diberi kategori Unknown.
4. Pemilihan fitur. Fitur prediktor dipilih dari Description, Keyword, Attack_Vector, Weakness, Status, dan Published_Year. Atribut CVE_ID dan Reference tidak digunakan sebagai fitur karena berfungsi sebagai pengenal dan sumber referensi. Atribut CVSS_Score tidak digunakan untuk mengurangi risiko kebocoran target apabila kategori Severity ditentukan berdasarkan skor CVSS.
5. Transformasi fitur. Fitur teks diubah menjadi representasi numerik dengan TF-IDF, sedangkan fitur kategorikal diubah menggunakan One-Hot Encoding.

Tahapan transformasi disusun dalam sebuah *pipeline* agar proses preprocessing pada data latih dan data uji mengikuti aturan yang sama. Parameter transformasi yang memerlukan pembelajaran dari data, seperti kosakata TF-IDF dan kategori One-Hot Encoding, dipelajari hanya dari data latih.

C. Ekstraksi Fitur Teks Menggunakan TF-IDF

Atribut Description berisi uraian tekstual mengenai karakteristik kerentanan, mekanisme serangan, dan potensi dampaknya. Karena algoritma Random Forest membutuhkan representasi numerik, deskripsi tersebut ditransformasikan menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF).

TF-IDF memberikan bobot berdasarkan frekuensi suatu istilah dalam dokumen dan tingkat kekhususan istilah tersebut terhadap keseluruhan dokumen. Bobot TF-IDF dapat dirumuskan sebagai berikut:

$$w_{\{t,d\}} = \backslash operatorname\{tf\}(t,d) \times \backslash log(\backslash frac\{N\}\{\backslash operatorname\{df\}(t)\}) \quad (1)$$

dengan $w_{\{t,d\}}$ menyatakan bobot istilah *(t)* pada dokumen *(d)*, (\operatorname{tf}(t,d) merupakan frekuensi istilah dalam dokumen, *(N)* adalah jumlah dokumen, dan ({$operatorname\{df\}(t)$}) adalah jumlah dokumen yang memuat istilah tersebut. Implementasi Scikit-learn menggunakan varian TF-IDF dengan pengaturan normalisasi dan smoothing sesuai konfigurasi pustaka [6].

Konfigurasi TF-IDF dalam penelitian ini menggunakan parameter max_features=5000, ngram_range= (1,2), min_df= 2, dan stop_words= 'english'. Parameter tersebut membatasi jumlah fitur maksimum hingga 5.000, mempertimbangkan unigram dan bigram, mengabaikan istilah yang muncul pada kurang dari dua dokumen, serta menghapus stop words bahasa Inggris.

Penggunaan unigram dan bigram memungkinkan representasi istilah tunggal maupun pasangan istilah. Dengan demikian, kata seperti *vulnerability* dan frasa seperti *remote code* dapat menjadi fitur yang dipertimbangkan model dalam mempelajari pola klasifikasi.

D. Encoding Fitur Kategorikal

Selain fitur teks, penelitian ini menggunakan fitur kategorikal yang meliputi Keyword, Attack_Vector, Weakness, Status, dan Published_Year. Fitur tersebut ditransformasikan menggunakan One-Hot Encoding, yaitu metode yang merepresentasikan setiap kategori sebagai fitur biner.

One-Hot Encoding dipilih agar kategori yang berbeda tidak dianggap memiliki urutan numerik tertentu. Parameter handle_unknown='ignore' digunakan untuk menangani kategori yang muncul pada data uji tetapi tidak ditemukan ketika encoder dilatih. Dengan pendekatan ini, fitur tekstual dan kategorikal dapat digabungkan menjadi satu representasi fitur untuk proses klasifikasi.

E. Pembagian Data

Data berlabel dibagi menjadi data latih (*training set*) dan data uji (*testing set*) menggunakan rasio 80:20. Pembagian dilakukan dengan train_test_split dari Scikit-learn, menggunakan random_state=42 dan stratify=y.

Parameter stratifikasi digunakan untuk mempertahankan proporsi relatif kelas target pada kedua subset. Dari 1.497 data berlabel, sekitar 1.197 data digunakan untuk melatih model dan 300 data digunakan untuk menguji model.

Data uji tidak digunakan dalam proses pelatihan model. Seluruh transformasi fitur yang mempelajari karakteristik data dilakukan melalui pipeline yang dilatih menggunakan data latih, kemudian diterapkan pada data uji. Pengaturan ini dimaksudkan untuk mengurangi risiko kebocoran informasi dari data uji ke proses pelatihan.

F. Klasifikasi Menggunakan Random Forest

Random Forest merupakan algoritma *ensemble learning* yang menggabungkan prediksi dari sejumlah pohon keputusan. Algoritma ini membangun pohon berdasarkan sampel dan pemilihan fitur secara acak, kemudian menggabungkan hasil prediksi pohon untuk menentukan kelas akhir [4].

Dalam penelitian ini, model Random Forest dikonfigurasi dengan parameter n_estimators= 300, random_state= 42, class_weight ='balanced_subsample', dan n_jobs= -1. Parameter n_estimators= 300 menentukan jumlah pohon keputusan yang digunakan. Parameter random_state= 42 membantu menjaga keterulangan eksperimen, sedangkan n_jobs= -1 memungkinkan pemanfaatan seluruh prosesor yang tersedia.

Parameter class_weight= 'balanced_subsample' digunakan untuk memberikan bobot kelas berdasarkan distribusi kelas pada sampel bootstrap setiap pohon. Pengaturan ini dimaksudkan untuk membantu mengurangi pengaruh ketidakseimbangan kelas terhadap pembelajaran model, tetapi tidak menjamin bahwa seluruh kelas akan memperoleh performa yang sama baiknya.

Model diimplementasikan dalam sebuah pipeline yang menggabungkan preprocessing, ekstraksi fitur, encoding, dan klasifikasi. Setelah dilatih menggunakan data latih, model menghasilkan prediksi empat kategori Severity pada data uji. Penelitian ini menggunakan konfigurasi awal tersebut dan belum menerapkan optimasi hyperparameter secara sistematis.

G. Evaluasi Kinerja Model

Evaluasi dilakukan dengan membandingkan label aktual dan label hasil prediksi pada data uji. Metrik yang digunakan meliputi *confusion matrix*, accuracy, precision, recall, F1-score, dan balanced accuracy. Pemilihan beberapa metrik bertujuan memberikan gambaran yang lebih menyeluruh daripada hanya menggunakan accuracy.

Hasil dan Diskusi

A. Hasil Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan data dapat digunakan dalam proses klasifikasi. Dataset kerentanan keamanan siber pada sistem kesehatan memiliki empat kelas tingkat keparahan, yaitu LOW, MEDIUM, HIGH, dan CRITICAL. Berdasarkan pengolahan data yang telah dilaporkan sebelumnya, dataset terdiri atas 1.515 baris dengan 18 data yang tidak memiliki label Severity. Oleh karena itu, 1.497 data berlabel digunakan untuk proses klasifikasi.

Tabel 1. Distribusi label

Kelas	*Jumlah data*	*Persentase*
LOW	134	8,95%
MEDIUM	720	48,10%
HIGH	492	32,87%
CRITICAL	151	10,09%
Total	1.497	100%*

Berdasarkan Tabel I, kelas MEDIUM memiliki jumlah data terbesar, yaitu 720 data atau sekitar 48,10% dari keseluruhan data berlabel. Sebaliknya, kelas LOW memiliki jumlah data paling sedikit, yaitu 134 data atau sekitar 8,95%. Distribusi tersebut menunjukkan adanya ketidakseimbangan kelas (class imbalance) yang berpotensi memengaruhi kinerja model. Untuk mengurangi pengaruh ketidakseimbangan kelas, pendekatan class_weight='balanced_subsample' digunakan pada algoritma Random Forest, apabila konfigurasi eksperimen yang diterapkan sesuai dengan pengaturan tersebut. Selain itu, data dibagi menjadi data latih dan data uji dengan proporsi 80:20 menggunakan stratifikasi agar proporsi kelas relatif terjaga pada kedua bagian data.

B. Hasil Klasifikasi Random Forest

Evaluasi model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score. Selain itu, confusion matrix digunakan untuk mengamati kesesuaian antara kelas aktual dan kelas hasil prediksi. Berdasarkan hasil evaluasi yang telah dilaporkan sebelumnya, model memperoleh akurasi sebesar 71% pada 300 data uji. Artinya, sebanyak 213 data diprediksi dengan benar, sedangkan 87 data lainnya salah diklasifikasikan.

Tabel 2. Hasil Evaluasi Model Random Forest

Metrik	*Nilai*
Accuracy	71,00%

Balanced accuracy	62,80%
Macro precision	72,43%
Macro recall	62,80%
Macro F1-score	66,02%
Weighted F1-score	70,47%

Akurasi sebesar 71% menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data uji dengan benar. Namun, nilai balanced accuracy sebesar 62,80% lebih rendah daripada akurasi keseluruhan. Perbedaan ini menunjukkan bahwa performa model belum merata pada seluruh kelas tingkat keparahan.

Nilai macro F1-score sebesar 66,02% juga memperlihatkan bahwa kinerja antarkelas masih bervariasi. Metrik macro memberikan bobot yang sama kepada setiap kelas, sehingga kelemahan dalam mengklasifikasikan kelas yang memiliki jumlah data lebih sedikit tetap terlihat dalam evaluasi.

C. Analisis Confusion Matrix

Confusion matrix digunakan untuk melihat jumlah prediksi benar dan kesalahan klasifikasi pada setiap kelas. Berdasarkan matriks yang sebelumnya dilaporkan, diperoleh hasil berikut.

Tabel 3. Confusion Matrix Model Random Forest

Kelas aktual	*Prediksi LOW*	*Prediksi MEDIUM*	*Prediksi HIGH*	*Prediksi CRITICAL*
LOW	16	7	4	0
MEDIUM	6	118	19	1
HIGH	0	31	66	2
CRITICAL	1	3	13	13

Angka pada diagonal utama menunjukkan prediksi yang benar. Model berhasil mengklasifikasikan 16 data LOW, 118 data MEDIUM, 66 data HIGH, dan 13 data CRITICAL dengan benar.

Kelas MEDIUM memiliki jumlah prediksi benar paling tinggi, yaitu 118 dari 144 data atau sekitar 81,94%. Hal ini menunjukkan bahwa model relatif baik dalam mengenali kelas tersebut.

Sebaliknya, kelas CRITICAL hanya berhasil dikenali pada 13 dari 30 data, sehingga nilai recall-nya sebesar 43,33%. Sebanyak 13 data CRITICAL diprediksi sebagai HIGH. Kesalahan ini penting untuk diperhatikan karena kerentanan yang tergolong kritis dapat memerlukan penanganan lebih cepat dan prioritas mitigasi yang lebih tinggi.

D. Diskusi

Hasil evaluasi menunjukkan bahwa Random Forest berpotensi digunakan untuk mengklasifikasikan tingkat keparahan kerentanan keamanan siber pada sistem kesehatan. Meskipun demikian, akurasi keseluruhan belum cukup untuk menyatakan bahwa model telah bekerja secara optimal. Pertama, ketidakseimbangan distribusi kelas dapat memengaruhi proses pembelajaran. Kelas MEDIUM mendominasi dataset, sedangkan kelas LOW dan CRITICAL memiliki jumlah sampel yang lebih sedikit. Penggunaan pembobotan kelas dapat membantu mengurangi pengaruh ketidakseimbangan tersebut, tetapi tidak menjamin bahwa seluruh kelas akan memiliki performa yang sama baiknya. Kedua, kesalahan klasifikasi antara kelas HIGH dan CRITICAL perlu menjadi perhatian. Kedua kelas tersebut memiliki tingkat keparahan yang berdekatan, sehingga informasi dalam deskripsi kerentanan mungkin memiliki karakteristik yang serupa. Apabila model mengandalkan teks deskripsi dan informasi kategoris, kualitas informasi tersebut dapat memengaruhi kemampuannya membedakan tingkat keparahan. Ketiga, penelitian ini perlu mempertimbangkan validitas fitur yang digunakan. Jika nilai CVSS_Score dikeluarkan karena berpotensi menyebabkan kebocoran target (target leakage), keputusan tersebut perlu dijelaskan secara konsisten dalam metodologi. Fitur lain juga perlu diperiksa agar tidak memuat informasi yang secara langsung mengungkap label target.

Berdasarkan confusion matrix yang dilaporkan, sebagian kesalahan klasifikasi terjadi antara kelas HIGH dan CRITICAL. Dari 30 data aktual CRITICAL, sebanyak 13 data berhasil dikenali, 13 data diprediksi sebagai HIGH, tiga sebagai MEDIUM, dan satu sebagai LOW. Pola tersebut menunjukkan bahwa model masih kesulitan membedakan sebagian kerentanan berkategori HIGH dan CRITICAL berdasarkan fitur yang tersedia. Temuan ini tidak dengan sendirinya membuktikan bahwa kedua kelas memiliki deskripsi yang serupa. Untuk memastikan penyebabnya, perlu dilakukan pemeriksaan terhadap deskripsi data yang salah diklasifikasikan, distribusi fitur, kualitas label, serta kemungkinan duplikasi atau kemiripan data antara data latih dan data uji. Pemeriksaan tersebut juga diperlukan untuk memastikan bahwa kinerja model bukan dipengaruhi kebocoran informasi.

Secara keseluruhan, hasil ini menunjukkan bahwa Random Forest dapat menjadi pendekatan awal untuk klasifikasi kerentanan keamanan siber. Namun, penerapan pada sistem kesehatan nyata memerlukan evaluasi lebih lanjut, terutama untuk meningkatkan deteksi kelas CRITICAL, menguji kestabilan model melalui validasi silang, dan membandingkan hasilnya dengan algoritma klasifikasi lain. Penelitian ini juga memiliki beberapa keterbatasan. Pertama, evaluasi menggunakan satu pembagian data latih dan data uji sehingga belum menggambarkan variasi kinerja pada pembagian data yang berbeda. Kedua, model belum dibandingkan dengan baseline atau algoritma alternatif menggunakan protokol evaluasi yang sama. Ketiga, optimasi hiperparameter dan validasi silang belum dilakukan. Keempat, recall kelas CRITICAL masih rendah sehingga model belum dapat dianggap memadai untuk menentukan prioritas mitigasi secara otomatis. Terakhir, penggunaan dataset yang tersedia tidak secara langsung membuktikan bahwa hasil dapat digeneralisasikan ke seluruh sistem kesehatan nyata.

Untuk mengatasi keterbatasan tersebut, eksperimen selanjutnya perlu mencakup pemeriksaan duplikasi, audit target leakage, validasi silang terstratifikasi, optimasi hiperparameter, serta perbandingan dengan algoritma alternatif. Seluruh hasil tambahan harus diperoleh melalui pengujian aktual sebelum dimasukkan ke dalam artikel.

Kesimpulan

Penelitian ini mengevaluasi algoritma Random Forest untuk mengklasifikasikan tingkat keparahan kerentanan keamanan siber pada sistem kesehatan ke dalam empat kelas, yaitu LOW, MEDIUM, HIGH, dan CRITICAL. Dari 1.515 baris dataset awal, sebanyak 18 data tanpa label Severity dikeluarkan sehingga diperoleh 1.497 data berlabel. Pemrosesan fitur dilakukan menggunakan TF-IDF untuk teks dan One-Hot Encoding untuk fitur kategorikal. Pada satu pembagian data latih dan data uji dengan rasio 80:20, model memperoleh akurasi 71,00%, balanced accuracy 62,80%, macro F1-score 66,02%, dan weighted F1-score 70,47%. Hasil evaluasi menunjukkan bahwa model memiliki kinerja terbaik pada kelas MEDIUM dengan recall sebesar 81,94%, sedangkan recall kelas CRITICAL hanya mencapai 43,33%. Temuan tersebut menunjukkan bahwa kemampuan model dalam mengenali tingkat keparahan belum merata dan masih terdapat risiko kesalahan klasifikasi pada kelas kritis. Oleh karena itu, model ini lebih tepat dipandang sebagai pendekatan awal untuk klasifikasi kerentanan, bukan sebagai sistem yang siap digunakan secara mandiri dalam menentukan prioritas mitigasi. Penelitian selanjutnya perlu melakukan validasi silang, optimasi hiperparameter, pemeriksaan kebocoran target, dan perbandingan dengan algoritma lain. Evaluasi lanjutan juga perlu memberi perhatian pada recall kelas CRITICAL serta kestabilan performa antarkelas. Langkah-langkah tersebut diperlukan untuk mengetahui apakah model dapat memberikan hasil yang lebih konsisten sebelum dipertimbangkan dalam lingkungan operasional sistem kesehatan.

Deklarasi Kepentingan

Penulis menyatakan bahwa tidak terdapat konflik kepentingan yang dapat memengaruhi pelaksanaan penelitian, analisis data, maupun penyusunan artikel ini.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada pihak-pihak yang telah mendukung penyusunan penelitian ini. Penulis juga menyampaikan apresiasi kepada penyedia dataset *Healthcare Cybersecurity Vulnerabilities Dataset* pada platform Kaggle yang telah menyediakan data untuk mendukung pelaksanaan penelitian ini.

Ketersediaan Data

Data yang digunakan dalam penelitian ini berasal dari *Healthcare Cybersecurity Vulnerabilities Dataset* yang tersedia pada platform Kaggle dan dapat diakses melalui tautan berikut:

https://www.kaggle.com/datasets/chuneeb/healthcare-cybersecurity-vulnerabilities-dataset

Dataset tersebut digunakan sebagai sumber data untuk proses preprocessing, pelatihan model, dan evaluasi klasifikasi tingkat keparahan kerentanan keamanan siber menggunakan algoritma Random Forest.

Penggunaan Ai Dan Deklarasi Penggunaan Ai Generatif

Dalam penyusunan artikel ini, penulis menggunakan kecerdasan buatan generatif sebagai alat bantu untuk mendukung penyusunan dan penyuntingan teks, perbaikan tata bahasa, serta penjelasan konsep yang berkaitan dengan penelitian. Penggunaan AI generatif tidak dimaksudkan untuk menggantikan proses analisis ilmiah maupun tanggung jawab penulis terhadap keakuratan isi artikel. Penulis bertanggung jawab untuk memeriksa, memverifikasi, dan memastikan bahwa seluruh data, hasil eksperimen, interpretasi, serta referensi yang disajikan dalam artikel sesuai dengan sumber dan hasil penelitian yang sebenarnya.

Daftar Pustaka

[1] P. Ewoh and T. Vartiainen, “Vulnerability to cyberattacks and sociotechnical solutions for health care systems: Systematic review,” *Journal of Medical Internet Research*, vol. 26, Art. no. e46904, May 2024, doi: 10.2196/46904.

[2] M. Auz-Torres and S. G. Yoo, “Key factors for reducing cybersecurity risks in healthcare: A systematic literature review,” *IEEE Access*, vol. 14, pp. 33002–33033, 2026, doi: 10.1109/ACCESS.2026.3667073.

[3] P. M. Mell, K. Scarfone, and S. Romanosky, “Common vulnerability scoring system,” *IEEE Security & Privacy*, 2006. [Online]. Available: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=50899.

[4] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[5] Chuneeb, “Healthcare Cybersecurity Vulnerabilities Dataset,” *Kaggle*. [Online]. Available: https://www.kaggle.com/datasets/chuneeb/healthcare-cybersecurity-vulnerabilities-dataset. [Accessed: Oct. 24, 2026].

[6] F. Pedregosa *et al*., “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.