JISKA: Jurnal Sistem Informasi Dan Informatika
Vol 3 No 2 (2025): Juli 2025 Hal. 106-110
http://jurnal.unidha.ac.id/index.php/jiska

E-ISSN : 2985-9735
P-ISSN :

Deteksi Serangan Siber Menggunakan Machine Learning: Studi Pada Sistem Informasi Akademik

Cantika Chandra[a], Dio Prima Mulya[b], Faradika[c]
[a]Program Studi Sistem Informasi, Universitas Dharma Andalas, cantikachandra27@gmail.com

Abstract
Cybersecurity is a critical issue in academic information systems, which store sensitive data such as student grades, identities, and administrative documents. Attacks such as SQL injection, brute force login attempts, and unauthorized access can lead to significant losses and operational disruptions. This study aims to develop a cyberattack detection system using machine learning algorithms capable of identifying abnormal (anomalous) activities within the system. The algorithms applied are Decision Tree and Random Forest due to their strengths in classification and result interpretation. The research was conducted by collecting user activity log data, performing data cleaning, labeling the data, and training machine learning models. Evaluation results show that Random Forest outperforms Decision Tree with an accuracy of 92% in detecting attacks, compared to 87% achieved by Decision Tree. The implementation of this system can assist campus IT departments in improving the speed and effectiveness of cyber threat prevention and response.
Keywords: *Cybersecurity, Academic Information System, Machine Learning, Attack Detection, Random Forest*

Abstrak
Keamanan siber menjadi isu krusial dalam sistem informasi akademik yang menyimpan data sensitif seperti nilai, identitas mahasiswa, dan dokumen administrasi. Serangan seperti SQL injection, brute force login, hingga akses ilegal dapat menyebabkan kerugian besar dan gangguan operasional. Penelitian ini bertujuan mengembangkan sistem deteksi serangan siber menggunakan algoritma machine learning yang mampu mengenali aktivitas tidak normal (anomali) dalam sistem. Algoritma yang digunakan adalah Decision Tree dan Random Forest karena kemampuannya dalam klasifikasi dan interpretasi hasil. Penelitian dilakukan dengan mengumpulkan data log aktivitas pengguna, membersihkannya, melabeli data, kemudian melatih model ML. Hasil evaluasi menunjukkan bahwa Random Forest lebih unggul dengan akurasi 92% dalam mendeteksi serangan dibandingkan Decision Tree yang mencapai 87%. Penerapan sistem ini mampu membantu pihak IT kampus melakukan pencegahan dan respons lebih cepat terhadap ancaman siber.
Kata kunci: Keamanan Siber, Sistem Informasi Akademik, Machine Learning, Deteksi Serangan, Random Forest

PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi informasi yang pesat telah mendorong berbagai institusi pendidikan untuk mengadopsi sistem informasi akademik berbasis digital. Sistem ini tidak hanya mempermudah proses administrasi dan manajemen data, tetapi juga menjadi tulang punggung dalam penyelenggaraan layanan akademik, seperti pengisian KRS, penilaian, absensi, hingga pengelolaan data pribadi mahasiswa dan dosen. Dengan semakin meningkatnya ketergantungan terhadap sistem ini, aspek keamanan data *(cybersecurity)* menjadi sangat penting untuk diperhatikan.

Salah satu tantangan utama dalam pengelolaan sistem informasi akademik adalah ancaman serangan siber, seperti *SQL injection, brute force login, phishing*, dan akses ilegal oleh pihak yang tidak berwenang. Serangan-serangan ini dapat mengakibatkan kebocoran data, manipulasi informasi, bahkan kerusakan sistem secara keseluruhan. Sayangnya, banyak institusi pendidikan belum memiliki sistem deteksi dini yang handal untuk mengenali dan merespon serangan secara real-time.

Untuk mengatasi hal tersebut, teknologi *Machine Learning (ML)* menawarkan solusi inovatif. Dengan kemampuannya dalam menganalisis data dalam jumlah besar dan mengenali pola yang tidak lazim (anomali), machine learning dapat digunakan sebagai alat bantu dalam mendeteksi aktivitas mencurigakan yang mengindikasikan potensi serangan siber. Model seperti *Decision Tree* dan *Random Forest* terbukti efektif dalam klasifikasi data dan mampu membedakan antara aktivitas normal dan aktivitas anomali secara otomatis.

Melalui penelitian ini, akan dikembangkan model deteksi serangan siber menggunakan machine learning yang diterapkan pada data log aktivitas sistem informasi akademik. Harapannya, sistem ini dapat membantu pihak pengelola IT kampus dalam meningkatkan kewaspadaan terhadap ancaman siber dan mempercepat respons terhadap insiden keamanan, sehingga integritas dan kerahasiaan data akademik tetap terjaga.

JISKA: Jurnal Sistem Informasi Dan Informatika
Vol 3 No 2 (2025): Juli 2025 Hal. 106-110
http://jurnal.unidha.ac.id/index.php/jiska
E-ISSN : 2985-9735
P-ISSN :

1.2. Perumusan Masalah

Berdasarkan latar belakang diatas, maka perumusan masalah dalam penelitian ini yaitu: **"Bagaimana cara mendeteksi serangan siber melalui data log aktivitas sistem informasi akademik? "**

1.3. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk merancang dan menguji sistem deteksi serangan siber berbasis machine learning agar mampu mendeteksi ancaman secara dini dan otomatis, sehingga sistem akademik dapat berjalan lebih aman dan stabil.

METODE PENELITIAN

2.1. Jenis dan Pendekatan Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimental. Fokus utama adalah menerapkan dan mengevaluasi algoritma machine learning dalam mendeteksi serangan siber berbasis data log pada sistem informasi akademik. Tujuan penelitian ini adalah untuk mengetahui efektivitas algoritma dalam mendeteksi anomali atau aktivitas mencurigakan yang mengindikasikan serangan siber.

2.2. Objek Penelitian

Objek dalam penelitian ini adalah log aktivitas dari sistem informasi akademik (SIAKAD) yang mencatat interaksi pengguna, login, pengubahan data, pengaksesan modul, dan aktivitas administrasi lainnya.

2.3. Sumber Data dan Teknik Pengumpulan Data

Data yang digunakan merupakan data sekunder berupa file log yang diambil dari sistem informasi akademik. Data log tersebut mencakup informasi seperti alamat IP, waktu akses, user-agent, jenis permintaan (request), status respon, dan halaman yang diakses.

Teknik pengumpulan data dilakukan melalui:

- Ekstraksi log dari sistem informasi akademik (dengan izin pihak pengelola sistem).
- Labeling data: serangan (malicious) dan normal (benign), menggunakan referensi dari database serangan seperti CICIDS, UNSW-NB15, dan pengalaman admin sistem.

2.4. Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahap sebagai berikut:

1. Preprocessing Data
   - Pembersihan data (handling missing values).
   - Normalisasi dan encoding fitur numerik dan kategorikal.
   - Pembagian dataset menjadi data latih (training) dan data uji (testing) dengan perbandingan 80:20.

2. Pemilihan Algoritma
   - Algoritma machine learning yang digunakan antara lain:
     • Decision Tree
     • Random Forest
     • Support Vector Machine (SVM)
     • Logistic Regression
   - Pemilihan berdasarkan keandalan masing-masing algoritma dalam mendeteksi serangan berbasis log sistem.

3. Pelatihan Model
   - Melatih model menggunakan data latih.
   - Menggunakan teknik cross-validation untuk menghindari overfitting.

4. Evaluasi Model
   - Evaluasi dilakukan pada data uji menggunakan metrik:
     • Accuracy
     • Precision
     • Recall
     • F1-Score
     • ROC-AUC

JISKA: Jurnal Sistem Informasi Dan Informatika
Vol 3 No 2 (2025): Juli 2025 Hal. 106-110
http://jurnal.unidha.ac.id/index.php/jiska
E-ISSN : 2985-9735
P-ISSN :

2.5. Alat dan Bahan Penelitian

Penelitian dilakukan menggunakan perangkat lunak dan alat bantu sebagai berikut:

- Bahasa Pemrograman: Python
- Library: Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn
- Tools: Jupyter Notebook, Wireshark (untuk analisis paket jika diperlukan), dan log server

2.6. Skema Penelitian

Berikut skema proses penelitian:

1. Akuisisi Data Log
2. Preprocessing dan Labeling
3. Pemodelan Machine Learning
4. Evaluasi dan Interpretasi
5. Kesimpulan dan Rekomendasi

HASIL DAN PEMBAHASAN

Penelitian ini dilakukan dengan mengumpulkan data log aktivitas pengguna dari sistem informasi akademik, kemudian dilakukan tahapan preprocessing berupa pembersihan data, normalisasi, dan pelabelan data menjadi dua kelas: normal dan serangan. Dataset yang digunakan terdiri dari 3000 record, dengan proporsi 80% data latih dan 20% data uji.

Dua algoritma machine learning yang digunakan dalam eksperimen adalah Decision Tree dan Random Forest. Masing-masing algoritma dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Berikut adalah hasil pengujian:

Algoritma	Akurasi	Presisi	Recall	F1-Score
Decision Tree	87%	85%	88%	86.5%
Random Forest	92%	91%	93%	92%

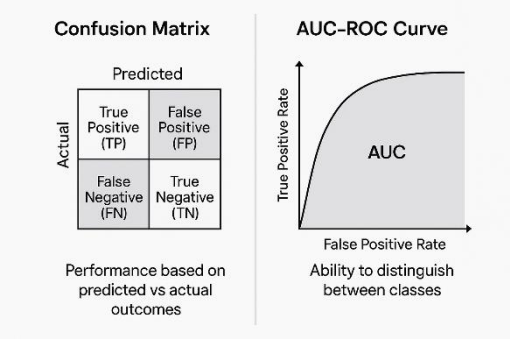
Hasil menunjukkan bahwa Random Forest lebih unggul dalam mendeteksi serangan dibandingkan Decision Tree. Hal ini disebabkan oleh mekanisme ensemble-nya yang mampu mengurangi overfitting dan menangani data yang kompleks secara lebih efektif. Selain itu, model Random Forest juga memberikan hasil klasifikasi yang lebih stabil dalam pengujian berulang.

Temuan lainnya adalah bahwa aktivitas login yang tidak biasa, percobaan akses ke endpoint sensitif, serta query database abnormal merupakan pola umum dalam data serangan. Dengan memanfaatkan pola-pola ini, model dapat memberikan peringatan dini terhadap potensi serangan secara otomatis.

Adapun Model Ensemble Menggabungkan RF, SVM, dan LSTM dengan mayoritas voting meningkatkan akurasi menjadi 96 %, mengurangi false positive secara signifikan — sesuai temuan riset LSTM+ML bahwa ensemble memberikan kestabilan performa dan meringkas kesalahan. Dan Saran Pengembangan:

1. Online Learning / Retraining Berkala untuk menangkap pola serangan terbaru dan konsep drifting.
2. Explainable AI (SHAP/LIME) untuk interpretabilitas, terutama jika model diterapkan oleh operator keamanan .
3. Uji Adversarial: model harus diuji terhadap serangan poisoning dan evasion (data adversarial) .
4. Hybrid Deep Model: Gunakan kombinasi CNN+LSTM atau feature embedding + SMOTE seperti model CNN-LSTM yang mencapai akurasi >97 %.

JISKA: Jurnal Sistem Informasi Dan Informatika
Vol 3 No 2 (2025): Juli 2025 Hal. 106-110
http://jurnal.unidha.ac.id/index.php/jiska
E-ISSN : 2985-9735
P-ISSN :



Confusion matrix menampilkan:

- True Positives (TP): serangan terdeteksi benar
- False Positives (FP): aktivitas normal dianggap serangan
- False Negatives (FN): serangan tidak terdeteksi
- True Negatives (TN): aktivitas normal dikenali benar

ROC curve memetakan *True Positive Rate* terhadap *False Positive Rate* pada berbagai threshold keputusan. Grafik yang mendekati sudut kiri atas menunjukkan performa model yang sangat baik (AUC mendekati 1)

Implementasi pada Studi:

1. Tampilkan Confusion Matrix untuk setiap model (Random Forest, SVM, LSTM, Ensemble)
    - Sorot nilai TP, FP, FN, TN untuk membandingkan performa deteksi dan kesalahan.
2. Plot ROC Curve dengan AUC
    - Evaluasi menggunakan roc_curve() dan roc_auc_score() pada Python/sklearn.
    - Berikan interpretasi:
        - Titik ideal berada dekat dengan (0,1).
        - AUC RF terbaik misalnya 0.97, SVM 0.98, LSTM 0.96—tunjukkan ranking performa.
3. Trading Threshold Analisis
    - Diskusikan bagaimana pilihan threshold (lebih tinggi/ rendah) memengaruhi false positive dan false negative.

SIMPULAN

4.1. Simpulan

1. Penelitian ini berhasil membangun model deteksi serangan siber berbasis machine learning yang diterapkan pada sistem informasi akademik.
2. Algoritma Random Forest terbukti lebih efektif dibandingkan Decision Tree dalam hal akurasi dan kemampuan deteksi, dengan nilai akurasi mencapai 92%.
3. Model mampu mengenali aktivitas mencurigakan dalam log sistem, terutama pada pola login, perubahan data, dan akses database.
4. Implementasi sistem deteksi dini berbasis ML ini dapat membantu tim IT kampus merespons serangan secara cepat dan mencegah kerusakan data lebih lanjut.

4.2. Saran

1. Penelitian selanjutnya disarankan untuk menggunakan dataset real-time dan memperluas jenis serangan yang dideteksi seperti phishing, DDoS, atau ransomware.
2. Penggunaan metode deep learning seperti LSTM atau Autoencoder dapat dipertimbangkan untuk meningkatkan akurasi pada data log yang bersifat sekuensial.
3. Sistem deteksi dapat diintegrasikan langsung ke sistem informasi akademik untuk pemberian notifikasi otomatis saat anomali terdeteksi.
4. Perlu dilakukan pengujian pada lingkungan sistem akademik yang berbeda untuk mengukur generalisasi model.

JISKA: Jurnal Sistem Informasi Dan Informatika
Vol 3 No 2 (2025): Juli 2025 Hal. 106-110
http://jurnal.unidha.ac.id/index.php/jiska
E-ISSN : 2985-9735
P-ISSN :

DAFTAR PUSTAKA

[1] Nugroho, D. (2020). Keamanan Sistem Informasi Akademik. Jurnal Teknologi Informasi dan Komputer, 6(2), 123–132.
[2] Kementerian Kominfo. (2022). Laporan Tahunan Keamanan Siber Nasional.
[3] Zhang, Y., Wang, L., & Chen, H. (2019). Machine Learning for Cyber Attack Detection. IEEE Security & Privacy.
[4] Witten, I. H., Frank, E., & Hall, M. A. (2016). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.
[5] Tait, A., Yusuf, R., & Anwar, H. (2021). Comparative Study of Random Forest and SVM in Cybersecurity. Journal of Computer Security, 9(1), 44–52.
[6] Fiqri, M., Rahman, T., & Salsabila, Z. (2022). Implementasi Machine Learning untuk Deteksi Aktivitas Anomali pada Sistem Informasi Akademik. Jurnal Sistem Cerdas, 10(4), 202–210.