IDENTIFIKASI ANOMALI KEAMANAN SERVER NGINX MENGGUNAKAN ALGORITMA *ISOLATION FOREST*

Abdullah Ragil Saputro, Nurchim, Vihi Atina
Program S1-Teknik Informatika, Universitas Duta Bangsa Surakarta
Jl. Bhayangkara No.55, Tipes, Serengan, Surakarta, Jawa Tengah
230103287@mhs.udb.ac.id

ABSTRAK

Marak terjadinya serangan siber yang mengancam keamanan data dan sistem informasi menjadikan keamanan informasi merupakan hal yang perlu diperhatikan. Web server merupakan salah satu komponen dalam infrastruktur teknologi informasi, dimana web server berfungsi sebagai pintu gerbang bagi pengguna untuk mengakses berbagai layanan dan aplikasi secara daring. Dalam upaya peningkatan keamanan server Nginx, salah satu pendekatan yang dapat digunakan adalah dengan melakukan analisis terhadap log akses server. Penelitian ini bertujuan untuk mengaplikasikan algoritma *Isolation Forest* dalam analisis log akses server Nginx untuk mengidentifikasi anomali yang berpotensi sebagai ancaman keamanan. Dengan mengoptimalkan deteksi anomali, diharapkan keamanan server Nginx dapat ditingkatkan, sehingga dapat meminimalisir risiko serangan dan kerugian yang diakibatkannya. Dalam penelitian ini menggunakan metode *Isolation Forest* untuk melakukan pendeteksian anomali pada log server nginx. Data yang digunakan adalah data log server Nginx yang memiliki kode respon selain HTTP 301, kemudian dari data log tersebut dilakukan proses pendeteksian anomali berdasarkan data *country, path* dan *size*. Hasil pengujian menunjukkan bahwa algoritma *Isolation Forest* berhasil mendeteksi 218 anomali dari 1529 data dan waktu pemrosesan 1 menit 14 detik dan menunjukkan 97,8% dari deteksi anomali adalah benar. Implementasi sistem ini dapat meningkatkan efisiensi dan efektivitas dalam melakukan deteksi terhadap ancaman siber. Pengembangan lebih lanjut dari penelitian ini diantaranya adalah dapat memanfaatkan sitemap pada sistem informasi untuk melakukan transformasi data *path*, selain itu dapat dikembangkan aplikasi monitoring yang dapat memudahkan pengelola server dalam melakukan monitoring keamanan server.

Kata kunci : *keamanan server,server nginx, isolation forest*

1. PENDAHULUAN

Pada era digital saat ini keamanan informasi merupakan sesuatu hal yang penting untuk diperhatikan, seperti diketahui akhir-akhir ini marak terjadi serangan siber yang mengancam keamanan data dan sistem informasi. Salah satu hal yang penting dalam pengelolaan sistem informasi adalah keamanan server. Web server merupakan salah satu komponen dalam infrastruktur teknologi informasi, dimana web server berfungsi sebagai pintu gerbang bagi pengguna untuk mengakses berbagai layanan dan aplikasi secara daring. Berbagai macam web server yang sering digunakan antara lain Apache, Nginx, IIS(*Internet Information Services*) dan masih banyak lagi. Berdasarkan survei Netcraft pada Januari 2024 Nginx merupakan salah satu web server yang paling populer dan banyak digunakan dimana pangsa pasarnya mencapai 23,21%, hal ini dikarenakan Nginx menawarkan performa tinggi dan keandalan yang sangat baik [1].

Namun, seiring dengan meningkatnya penggunaan, risiko keamanan yang mengancam Nginx juga semakin kompleks dan beragam.

Dalam upaya meningkatkan keamanan server Nginx, diperlukan metode untuk membantu mendeteksi dan mencegah aktivitas mencurigakan atau serangan siber. Salah satu pendekatan yang dapat digunakan adalah dengan melakukan analisis terhadap log akses server. Log akses server berisi catatan aktivitas yang terjadi pada server, seperti permintaan yang diterima, respon yang diberikan, dan informasi lainnya yang dapat digunakan untuk mengidentifikasi pola-pola tidak biasa atau indikasi serangan.

Teknik *machine learning* salah satunya adalah *Isolation Forest*, merupakan algoritma yang dirancang untuk mendeteksi anomali dengan cara mengisolasi data poin yang berbeda secara signifikan dari mayoritas data lainnya. Penelitian ini bertujuan untuk mengaplikasikan algoritma *Isolation Forest* dalam analisis log akses server Nginx untuk mengidentifikasi anomali yang berpotensi sebagai ancaman keamanan. Dengan mengoptimalkan deteksi anomali, diharapkan keamanan server Nginx dapat ditingkatkan, sehingga dapat meminimalisir risiko serangan dan kerugian yang diakibatkannya.

2. TINJAUAN PUSTAKA

Dalam pembuatan penelitian ini ada beberapa tinjauan pustaka yang dijadikan sebagai bahan perbandingan untuk menganalisis sistem yang ada, antara lain :

2.1. Penelitian Terkait

Penelitian milik I Made Sudarsana Taksa Wibawaa, Anak Agung Istri Ngurah Eka Karyawati yang berjudul "*Isolation Forest* dengan *Exploratory Data Analysis* pada *Anomaly Detection* untuk Data Transaksi", menghasilkan bahwa metode *Isolation*

Forest memiliki kinerja yang sangat baik. Penelitian ini berhasil mendeteksi 1.646 data anomali dan mencatat *anomaly score* sebesar 0,47%. Selain itu, penggunaan *Exploratory Data Analysis* (EDA) membantu dalam memahami dataset serta menghilangkan fitur-fitur yang kurang relevan, sehingga proses perhitungan model menjadi lebih efisien [2].

Penelitian milik Oktavia Triana yang berjudul "Deteksi Anomali Jaringan Menggunakan Algoritma *Isolation Forest*'' menunjukkan bahwa algoritma *Isolation Forest* berhasil mendeteksi 950 anomali data dari 10.000 data, serta menunjukkan 94% dari deteksi anomali adalah benar [3].

Penelitian milik Kurnia Setiawan, Arief Wibowo yang berjudul "*Data Mining Implementation For Detection Of Anomalies In Network Traffic Packets Using Outlier Detection Approach*'', hasil pengujian menunjukkan algoritma *Isolation Forest* mengidentifikasi 1.643 record (4,86%) sebagai outlier dan 32.098 record (95,13%) sebagai inlier [4].

2.2. Nginx

Nginx adalah perangkat lunak open-source yang dikenal memiliki kinerja tinggi sebagai server HTTP dan *reverse* proxy. Nginx mampu menyajikan konten statis dengan cepat dan efisien dalam penggunaan sumber daya sistem. Selain itu, Nginx juga mampu mendistribusikan HTTP konten dinamis melalui jaringan menggunakan *FastCGI handler* untuk *script*, serta berfungsi sebagai perangkat lunak penyeimbang beban yang andal. Nginx memiliki arsitektur modular yang mampu mendukung berbagai fitur seperti *Load Balancing dan Reverse Proxying*, *Virtual hosts* berbasis nama dan IP, *Fast CGI*, akses langsung ke *cache, SSL, Flash Video Streaming* dan fitur standar lainnya. Nginx kompatibel untuk berbagai platform termasuk Unix, Linux, varian dari BSD, MacOS X, Solaris, dan Microsoft Windows [5].

2.3. Log Akses Server

File log akses adalah catatan yang mencatat aktivitas yang terjadi dalam suatu sistem, seperti akses pengguna terhadap situs web, permintaan server, atau kejadian keamanan. Data log dapat digunakan untuk melacak dan menganalisis aktivitas pengguna yang mengunjungi situs web. Akses log menyimpan informasi tentang Admin alamat IP administrator, nama administrator, URL, permintaan akses, penanda waktu (*timestamp*), kode kesalahan, dan informasi lainnya [6].

Format Log akses Nginx mencakup informasi sebagai berikut:

a. IP Address: Alamat IP dari klien yang membuat permintaan.
b. Tanggal dan Waktu: Waktu ketika permintaan dibuat.
c. Request: URL yang diminta oleh klien.
d. Status HTTP: Kode status HTTP yang dikembalikan oleh server.
e. User Agent: Informasi tentang browser atau perangkat yang digunakan.
f. Referer: URL halaman sebelumnya yang mengarahkan pengguna ke halaman yang diminta.

   Contoh

   192.168.1.1 - - [12/Dec/2024:14:12:09 +0000] "GET /index.html HTTP/1.1" 200 3422 "https://example.com" "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/58.0.3029.110 Safari/537.36"

2.4. Deteksi Anomali

Menurut Anwar deteksi anomali adalah salah satu metode dalam *Intrusion Detection Sistem* (IDS) yang bertujuan untuk mengidentifikasi serangan yang menyimpang dari pola normal berdasarkan probabilitas statistik. Dengan semakin tingginya penggunaan internet, risiko serangan dari intruder atau cracker yang mengeksploitasi kelemahan protokol internet dan perangkat lunak juga meningkat. Deteksi anomali menjadi sangat penting untuk mengamankan sistem dari ancaman tersebut[7]. Deteksi anomali dalam analisis data adalah proses penting untuk mengidentifikasi pengamatan atau pola yang berbeda dari mayoritas data. Prediksi berbasis model, seperti Isolation Forest atau Gaussian Mixture Models, membangun model dengan menggunakan data normal dan mengidentifikasi data yang menyimpang dari model sebagai anomali[8].

2.5. Isolation Forest

Isolation forest adalah salah satu metode detelsi anomali berbasis *unsupervised machine learning* yang pertama kali diperkenalkan oleh Liu et al.(2008) [9].

Metode ini merupakan algoritma *unsupervised learning* dan *nonparametric* yang menggunakan pendekatan berbasis pohon keputusan atau *trees*. Proses deteksi anomali menggunakan *Isolation Forest* melewati dua tahap, yaitu tahap pelatihan (*training*) dan tahap evaluasi (*evaluation*). Pada tahap pelatihan, model *Isolation Forest* membuat pohon isolasi menggunakan sampel dari seluruh set data. Selanjutnya pada tahap evaluasi, setiap individu/instance dalam dataset melalui pohon isolasi untuk menghitung skor anomali setiap instance[10].

Prinsip dasar dalam pendeteksian anomali menggunakan *Isolation Forest* adalah dengan membagi ruang data menjadi beberapa subruang secara acak. Hal ini dilakukan karena data anomali cenderung berada pada titik yang jarang muncul dalam dataset. Pendekatan ini memungkinkan algoritma untuk mengidentifikasi anomali dengan cepat. *Isolation Forest* dikenal karena kemampuannya mendeteksi anomali dengan efisiensi tinggi, dimana algoritma ini dapat mengisolasi outlier lebih cepat dibandingkan data normal. Skor anomali (*anomaly score*) adalah nilai yang menunjukkan sejauh mana suatu data dianggap sebagai anomali. Secara

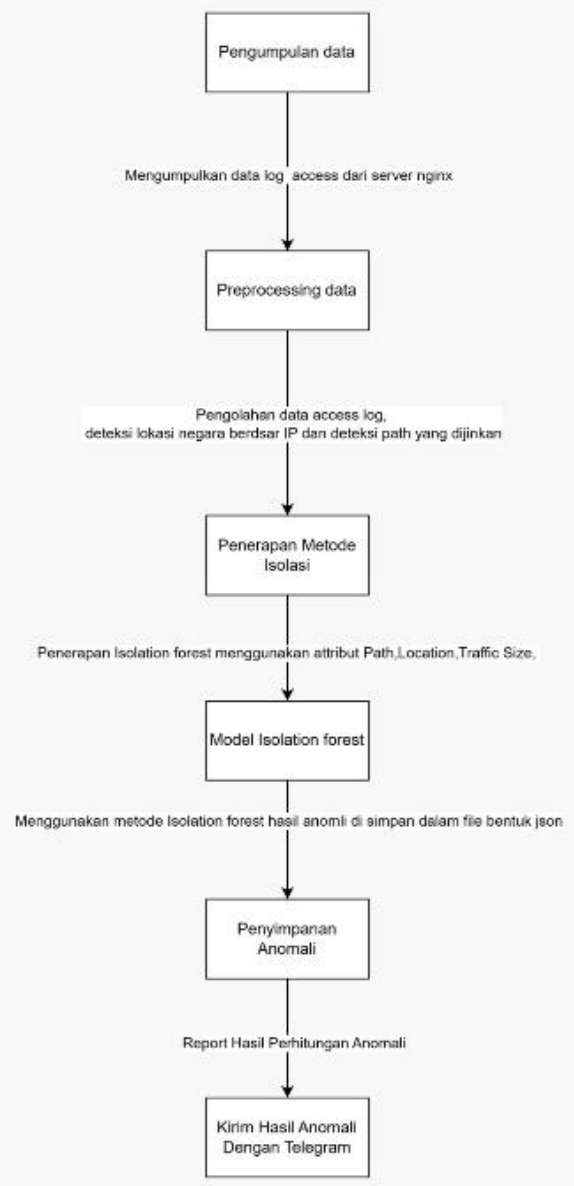
matematis, skor anomali dihitung menggunakan rumus:

$$s(\mathrm{x}+\mathrm{n}) = 2^{-\frac{E(h(x))}{c(n)}} \quad (1)$$

Ket.
h(x) : path length
E(h(x)) : rata-rata path length setiap isolation tree
C(n) : rata-dara h(x) dari nilai variabel ke-n

Path length adalah ukuran jarak yang menunjukkan titik di mana data tidak dapat dibagi lebih lanjut dengan root node dalam struktur algoritma pohon keputusan. Skor anomali (anomaly score) memiliki rentang nilai antara -1 hingga 1. Semakin mendekati -1, nilai tersebut dianggap sebagai anomali, sedangkan semakin mendekati 1, nilai tersebut dianggap normal.

3. METODE PENELITIAN



Gambar 1. Alur proses deteksi anomali

Metode pengumpulan data yang digunakan dalam penelitian ini adalah metode observasi dan studi literatur. Pengumpulan data dilakukan dengan cara melakukan pengamatan langsung terhadap objek penelitian, yaitu data log akses server Nginx yang diambil dari server Nginx. Tujuan dari observasi ini adalah untuk mengidentifikasi karateristik anomali sehingga menjadi kunci dalam pengidentifikasian anomali pola akses. Sedangkan untuk pengumpulan data melalui studi pustaka, penulis mencari jurnal, laporan, artikel ilmiah, dan referensi lain yang berkaitan dengan penelitian.

Setelah data terkumpul, tahap selanjutnya adalah melakukan pembersihan dan persiapan data untuk menghilangkan data yang tidak berhubungan langsung dengan tujuan akhir dari penelitian ini. Selain itu data yang masih memiliki nilai yang berbentuk string diubah menjadi bentuk numerik.

Setelah data melalui proses *prepocessing*, selanjutnya algoritma *Isolation Forest* diterapkan untuk mendeteksi anomali. *Isolation forest* bekerja dengan membangun sejumlah pohon keputusan yang secara efektif memisahkan data normal dari anomali, sehingga memungkinkan identifikasi pola yang tidak biasa dalam dataset. Selanjutnya setelah mendapatkan data anomaly, data tersebut disimpan dalam file log akses nginx isolation forest.xlsx, yang selanjutnya akan mengirimkan notifikasi memalui telegram kepada pengelola server.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa log yang dihasilkan langsung oleh server Nginx. Pada penelitian ini terdapat 1530 sampel data log yang akan digunakan, data tersebut tersimpan dalam format JSON. File tersebut berisikan informasi mengenai siapa yang mengakses server, kapan, dan apa yang diakses. Informasi tersebut mencakup hal-hal seperti alamat IP pengunjung, metode HTTP yang digunakan (GET, POST, dll.), URL yang diminta, kode status HTTP yang diberikan, dan informasi lain terkait permintaan. Pada penelitian ini parameter yang akan digunakan meliputi *ip, path, size, status*. Adapun contoh data yang digunakan terdapat pada gambar dibawah ini :

Gambar 2. Contoh data akses log

4.2. Processing Data

Berdasarkan hasil pengumpulan data terdapat 1530 sampel data log yang akan digunakan dalam penelitian, adapun untuk *processing* data yang digunakan dalam penelitian ini terdiri dari :

a. Pembersihan data, dari data tersebut diambil data log akses server Nginx yang hanya menyertakan entri dengan kode respon HTTP selain 301. Hal ini dilakukan dengan asumsi bahwa jika kode respon HTTP 301 adalah permintaan yang hanya sebagai pengalihan ke laman lain sehingga tidak diproses didalam server. Sehingga diperoleh data sejumlah 1529 data yang dapat digunakan.
b. Transformasi Data, Log Nginx berisikan data dalam bentuk text berupa alamat IP, metode HTTP, path URL. Dalam melakukan analisis menggunakan Isolation Forest perlu mengubah data text menjadi format numerik. Pada tahapan reduksi data ini, data-data yang masih dalam bentuk text diubah menjadi data numerik. Adapun data yang ditransformasi sebagai berikut:

Tabel 1. Data yang ditransformasi

No	Nama Data	Jenis Data Awal	Jenis Data Akhir
1	Country	String	Numerik
2	Path	String	Numerik

Pada data country untuk mengubah string menjadi numerik dengan memeriksa alamat IP berdasarkan basis data lokasi geografis (geolocation database) untuk dapat menentukan negara asal. Jika negara asal IP tersebut terdaftar dalam daftar negara yang telah ditentukan, maka nilainya diberikan 1 dan jika tidak terdaftar maka nilainya diberikan -1.

Sedangkan pada data path dilakukan dengan mendeteksi apakah path yang diminta merupakan path yang terdaftar dalam path yang ditentukan. Jika path tersebut terdaftar maka diberikan nilai 1 dan jika tidak maka diberikan nilai -1.

Contoh hasil dari proses *processing* data dapat dilihat pada tabel berikut:

Tabel 2. Hasil transformasi data

Index	Country	Path	Size	Status	IP
1	1	1	111	200	108.137.128.70
2	1	1	111	200	108.137.128.70
3	1	1	111	200	108.137.128.70
4	-1	-1	128	200	111.119.220.219
5	1	1	111	200	108.137.128.70
6	1	1	111	200	108.137.128.70
7	1	-1	353	200	110.238.111.223
8	1	1	111	200	108.137.128.70
9	1	1	111	200	108.137.128.70
10	-1	1	128	200	111.119.220.21 9

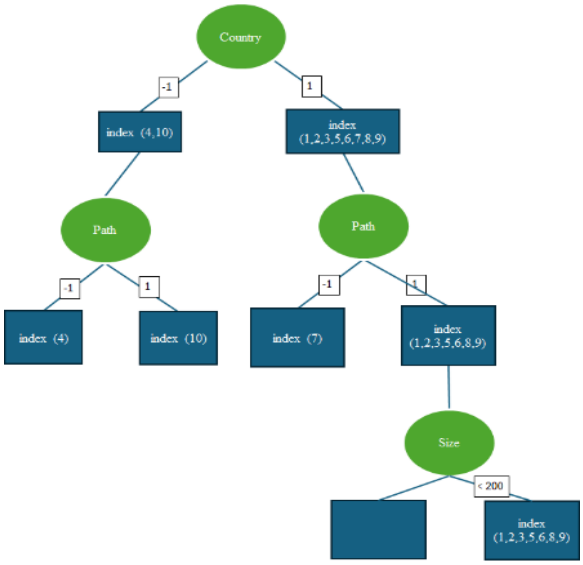
4.3. Metode Isolation Forest

Metode *Isolation Forest* adalah algoritma yang dirancang khusus untuk mendeteksi anomali (*outlier*) dalam dataset. Inti dari metode ini adalah memanfaatkan sifat unik *outlier* yang biasanya lebih mudah "terisolasi" dibandingkan data normal. Dalam penelitian ini tahapan metode isolation forest sebagai berikut

4.4. Pembangunan Pohon Isolasi

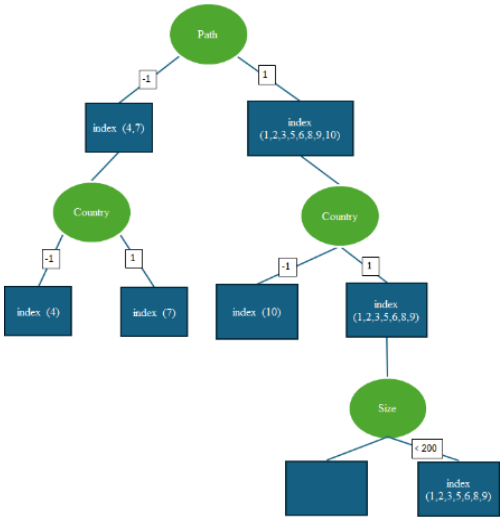
Pada penelitian ini membangun 3 pohon isolasi, setiap pohon dibangun dengan membagi data hingga kondisi terminasi tercapai (simpul homogen atau kedalaman maksimum).

a. Pohon 1 :
Pada pohon 1 dataset awal dibagi berdasarkan fitur *country*. Pada level 2 untuk cabang kiri pembagian lebih lanjut dilakukan pada fitur *path* dan cabang kanan fitur *path*. Selanjutnya data di sub cabang kanan pembagian dilakukan pada fitur *size* sehingga dapat dilihat pada gambar hasil dari pohon 1.



Gambar 3. Pohon isolasi 1

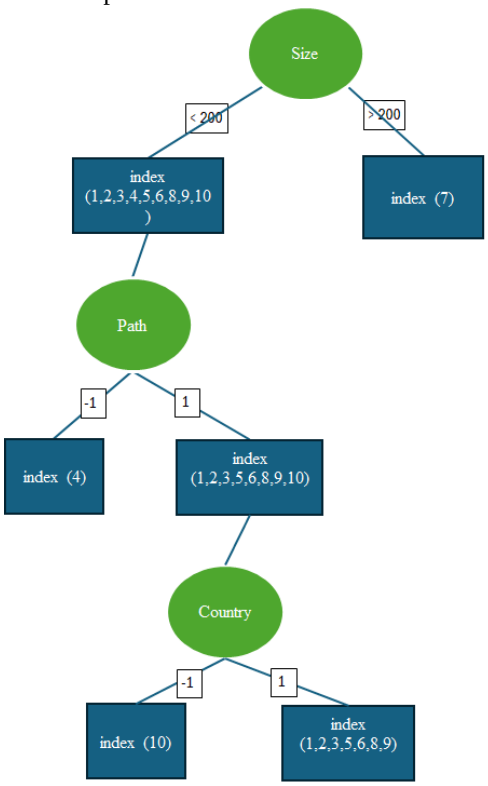
b. Pohon 2 :



Gambar 4. Pohon isolasi 2

Pada pohon 2 dataset awal dibagi berdasarkan fitur *path*. Pada level 2 untuk cabang kiri pembagian lebih lanjut dilakukan pada fitur *country* dan cabang kanan fitur *country*. Selanjutnya data di sub cabang kanan pembagian dilakukan pada fitur *size* sehingga dapat dilihat pada gambar hasil dari pohon 2.

c. Pohon 3 :
Pada pohon 3 dataset awal dibagi berdasarkan fitur *size*. Pada level 2 untuk cabang kiri pembagian lebih lanjut dilakukan pada fitur *path*. Selanjutnya data di sub cabang kiri pembagian dilakukan pada fitur *country* sehingga dapat dilihat pada gambar hasil dari pohon 3.



Gambar 5. Pohon isolasi 3

4.5. Perhitungan Kedalaman Simpul

Kedalaman simpul mengacu pada jumlah langkah yang diperlukan untuk mencapai suatu simpul dari akar pohon (root node). Langkah-langkah perhitungan:

a. Awal dari root node:
Kedalaman dimulai dari 0.
b. Setiap kali pembagian terjadi:
Jika data masuk ke simpul kiri atau kanan, kedalaman bertambah +1.
c. Pohon berhenti membagi:
Ketika data menjadi homogen (satu data di simpul) atau mencapai kedalaman maksimum.
Contoh hasil dari perhitungan kedalaman simpul adalah sebagai berikut:

Tabel 3. Hasil perhitungan kedalaman simpul

Index	E1	E2	E3	E(avg)
1	6,9	6,9	6,9	6,9000
2	6,9	6,9	6,9	6,9000
3	6,9	6,9	6,9	6,9000
4	2	2	2	2,0000
5	3	6,9	6,9	5,6000
6	2	2	3	2,3333
7	6,9	6,9	1	4,9333
8	6,9	6,9	6,9	6,9000
9	6,9	6,9	6,9	6,9000
10	2	2	3	2,3333

Keterangan:
E1 : merupakan nilai kedalaman simpul dari pohon 1
E2 : merupakan nilai kedalaman simpul dari pohon 2
E3 : merupakan nilai kedalaman simpul dari pohon 3
E(avg) : merupakan rata-rata kedalaman dari tiga pohon untuk setiap data.

4.6. Perhitungan Skor Anomali

Skor anomali adalah nilai yang menunjukkan seberapa besar kemungkinan suatu data adalah anomali. Dalam Isolation Forest, skor ini didasarkan pada kedalaman simpul rata-rata dari pohon isolasi. Contoh hasil perhitungan skor anomali sebagai berikut:

Tabel 4. Hasil perhitungan skor anomali

Index	E(avg)	Score Anomali
1	6,9000	0,289
2	6,9000	0,289
3	6,9000	0,289
4	2,0000	0,698
5	5,6000	0,366
6	2,3333	0,658
7	4,9333	0,412
8	6,9000	0,289
9	6,9000	0,289
10	2,3333	0,658

4.7. Keputusan Outlier

Dari data hasil perhitungan skor anomali yang memiliki nilai lebih dari 0,5 dianggap data tersebut anomali. Sehingga dari 10 data diatas ada 3 data yang dianggap anomali yaitu data(4), data(6), dan data (10).

4.8. Penerapan Metode Isolation Forest

Setelah melakukan perhitungan secara manual menggunakan metode *Isolation Forest*, langkah selanjutnya adalah mengimplementasikan metode ini ke dalam bahasa pemrograman agar proses identifikasi anomali dapat diimplementasikan ke dalam server. Pada penelitian ini metode *Isolation Forest* diimplementasikan dalam bahasa pemrograman python dan menyimpan hasil dalam file xlxs serta mengirimkan notifikasi via telegram.

Pada tahapan ini, melakukan pengujian dengan sejumlah 1530 data, dimana data yang memiliki respon HTTP kode selain 301 sejumlah 1529 data, proses ini memakan waktu 1 menit 5 detik. Dataset kemudian dibagi menjadi dua bagian, yaitu 80% untuk pelatihan

(*training*) dan 20% untuk pengujian (*testing*). Selanjutnya data tersebut diproses menggunakan model *Isolation Forest* dengan waktu pemrosesan sebesar 0,13 detik. Hasil yang diperoleh dari pengujian data tersebut terdapat 1311 data yang normal dan 218 data yang dianggap anomali. Hasil pengujian tersebut dilakukan evaluasi menggunakan metode *K-Fold Cross-Validation* dan menghasilkan tingkat akurasi 99,4% yang berarti model berhasil mengklasifikasi 99,4% dari semua *fold*, nilai *precision* 97.3% yang berarti model berhasil memprediksi data anomali dari semua fold sebesar 97,3%, nilai recall 98.4% yang menunjukkan bahwa model berhasil mengidentifikasi anomali data sebesar 98,4% disemua *fold*, dan F1 Score 97,8% yang berarti keseimbangan antara *precision* dan *recall* sebesar 97,8%. Hasil pengujian dapat dilihat pada gambar berikut:

```
K-Fold Cross-Validation Results:
Total Data: 1530
Total Data Digunakan: 1529
Jumlah Data Training: 1224
Jumlah Data Testing: 305
Total Normal: 1311
Total Anomaly: 218
Rata-rata Accuracy: 0.99411550412514 74
Rata-rata Precision: 0.97387714274506 72
Rata-rata Recall: 0.98431114040870 15
Rata-rata F1 Score: 0.97887469742255 22
```

. Gambar 6. Hasil pengujian

4.9. Penyimpanan Anomali

Data yang telah selesai diproses dalam model *Isolation Forest* selanjutnya disimpan dalam file dengan format .xlsx. Terdapat 218 data yang terdeteksi anomali yang tersimpan dalam file Log Access Nginx log akses nginx isolation forest.xlsx yang nantinya akan dikirimkan kepada pengelola server melalui telegram. Berikut contoh data anomali yang berhasil tersimpan:

	country	path	size	status	url	ip	waktu	anomaly
2	702	3356189	364	200	/payment	110.238.11	1,73E+09	Anomaly
3	702	3356189	362	200	/payment	110.238.11	1,73E+09	Anomaly
4	702	3356189	363	200	/payment	110.238.11	1,73E+09	Anomaly
5	702	3356189	352	200	/payment	111.119.22	1,73E+09	Anomaly
6	840	84704107	11971	200	/login	147.185.15	1,73E+09	Anomaly
7	702	3356189	366	200	/payment	110.238.11	1,73E+09	Anomaly
8	702	3356189	352	200	/payment	111.119.22	1,73E+09	Anomaly
9	840	84704107	11977	200	/login	198.235.24	1,73E+09	Anomaly
10	702	3356189	363	200	/payment	110.238.11	1,73E+09	Anomaly
11	702	3356189	352	200	/payment	111.119.22	1,73E+09	Anomaly
12	702	3356189	365	200	/payment	110.238.11	1,73E+09	Anomaly
13	702	3356189	364	200	/payment	110.238.11	1,73E+09	Anomaly
14	702	3356189	362	200	/payment	111.119.22	1,73E+09	Anomaly
15	840	84704107	264	400	/login	198.235.24	1,73E+09	Anomaly
16	702	3356189	363	200	/payment	110.238.11	1,73E+09	Anomaly
17	-1	96698807	6628	404	/login.asp	194.50.16.	1,73E+09	Anomaly
18	840	84704107	11977	200	/login	198.235.24	1,73E+09	Anomaly
19	840	84704107	11977	200	/login	198.235.24	1,73E+09	Anomaly
20	702	3356189	365	200	/payment	111.119.22	1,73E+09	Anomaly
21	360	18983344	6252	200	/sales-inv	182.253.18	1,73E+09	Anomaly
22	360	69854112	5263	200	/canvas-re	182.253.18	1,73E+09	Anomaly
23	360	99841737	3097	200	/canvas-re	182.253.18	1,73E+09	Anomaly
24	360	20127755	4346	200	/canvas-re	182.253.18	1,73E+09	Anomaly
25	360	25196592	10686	200	/canvas-re	182.253.18	1,73E+09	Anomaly
26	840	84704107	11977	200	/login	205.210.31	1,73E+09	Anomaly
27	360	7129801	6941	200	/canvas-re	182.253.18	1,73E+09	Anomaly
28	360	81181867	630408	500	/canvas-re	182.253.18	1,73E+09	Anomaly

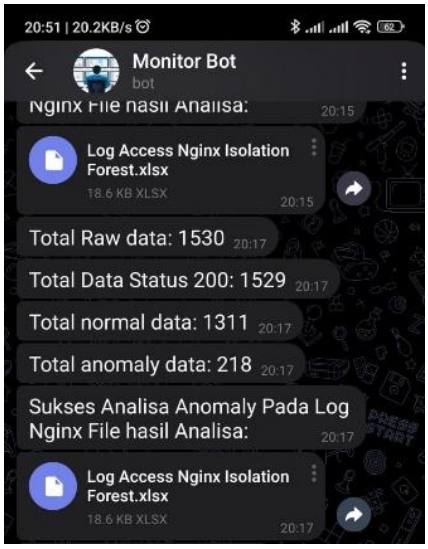
Gambar 7. Hasil penyimpanan data anomali

4.10. Notifikasi Telegram

Setelah seluruh proses klasifikasi dan penyimpanan selesai, selanjutnya terdapat notifikasi yang dikirimkan melalui telegram ke pengelola server. Notifikasi tersebut memberikan informasi mengenai hasil klasifikasi dari log server menggunakan model *Isolation Forest*. Berikut proses pengiriman data telegram memerlukan waktu 8 detik untuk dapat terkirim.

```
Proses Preprocessing data Dimulai
Selesai Melakukan Preprocessing data: 65.9269769191742 detik ---
Proses cleaning data berhasil
Proses Perhitungan Anomali Dimulai
Selesai Melakukan Perhitungan anomaly data: 0.13524532318115234 detik ---
Proses Penyimpanan Data Anomali Dimulai
Selesai Melakukan Export anomaly data: 0.14121556282043457 detik ---
K-Fold Cross-Validation Results:
Total Data: 1530
Total Data Digunakan: 1529
Jumlah Data Training: 1224
Jumlah Data Testing: 305
Total Normal: 1311
Total Anomaly: 218
Rata-rata Accuracy: 0.99411550412514 74
Rata-rata Precision: 0.97387714274506 72
Rata-rata Recall: 0.98431114040870 15
Rata-rata F1 Score: 0.97887469742255 22
Jumlah data training: 1224
Jumlah data testing: 305
Confusion Matrix:
[[398   0]
 [  0  61]]
Proses Pengiriman Notifikasi Data Anomali Dimulai
Message successfully sent to Telegram.
Message successfully sent to Telegram.
Message successfully sent to Telegram.
Message successfully sent to Telegram.
Message successfully sent to Telegram.
File successfully sent to Telegram.
Selesai Melakukan Pengiriman Notifikasi anomaly data: 8.118016242980957 detik ---
```

Gambar 8. Proses pengiriman notifikasi



Gambar 9. Hasil notifikasi terkirim ke telegram

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilaksanakan menunjukkan bahwa *Isolation Forest* berhasil dalam melakukan pendeteksian anomali pada data log server Nginx. Telah dilakukan ujicoba pada model *Isolation Forest* dengan 1530 data log akses server Nginx, dari data tersebut dilakukan processing data dan terdapat 1529 data yang di proses pada model *isolation forest*. Proses mendeteksi data anomali memakan waktu sebesar 1 menit 6 detik untuk memperoleh hasil. Hasil yang diperoleh dari pengujian data tersebut terdapat 1311 data yang normal dan 218 data yang dianggap anomali. Hasil pengujian tersebut dilakukan evaluasi menggunakan metode *K-Fold Cross-Validation* dan menghasilkan tingkat akurasi 99,4% yang berarti

model berhasil mengklasifikasi 99,4% dari semua *fold*, nilai *precision* 97.3% yang berarti model berhasil memprediksi data anomali dari semua fold sebesar 97,3%, nilai recall 98.4% yang menunjukkan bahwa model berhasil mengidentifikasi anomali data sebesar 98,4% disemua *fold*, dan F1 Score 97,8% yang berarti keseimbangan antara *precision* dan *recall* sebesar 97,8%. Sistem ini juga mampu mengirimkan notifikasi melalui pesan telegram dengan waktu pengiriman 8 detik.

Penelitian ini membuktikan bahwa model *Isolation Forest* mampu memberikan solusi untuk mendukung dalam pengelolaan keamanan server Nginx, dimana dapat menggantikan metode manual dengan pendekatan otomatisasi. Meskipun demikian, penelitian ini masih memiliki kekurangan yang dapat dikembangkan kedepannya guna menyempurnakan sistem ini diantaranya adalah dapat memanfaatkan sitemap pada sistem informasi untuk melakukan transformasi data *path*, selain itu dapat dikembangkan aplikasi monitoring yang dapat memudahkan pengelola server dalam melakukan monitoring keamanan server.

DAFTAR PUSTAKA

[1] Netcraft. (2024). January 2024 Web Server Survey. Netcraft. Diakses pada https://news.netcraft.com/archives/category/web-server-survey/.

[2] Wibawaa, I. M. S. T., and Karyawati, A. A. I. N. E., 2023. Isolation Forest dengan Exploratory Data Analysis pada Anomaly Detection untuk Data Transaksi.

[3] Triana, O., 2024. Deteksi Anomali Jaringan Menggunakan Algoritma Isolation Forest.

[4] Setiawan, K., and Wibowo, A., 2023. Data Mining Implementation for Detection of Anomalies in Network Traffic Packets Using Outlier Detection Approach.

[5] Clément Nedelcu, *Nginx HTTP Server*, Packt Publishing, Birmingham, UK, ISBN 978-1-849510-86-8, 2010.

[6] Nguyen, M.T., Diep, T.D., Vinh, T. H., Nakajima, T., & Thoai, N. (2018). Analyzing dan visualizing web server access log file. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11251 LNCS, 349-367. https://doi.org/10.1007/978-3-030-03192-3_27.

[7] S. Anwar, F. Septian, and R. D. Septiana, "Klasifikasi Anomali Intrusion Detection System (IDS) Menggunakan Algoritma Naïve Bayes Classifier dan Correlation-Based Feature Selection," *J. Teknol. Sist. Inf. dan Apl.*, vol. 2, no. 4, pp. 135–140, 2019.

[8] R. R. Widalala, et al, "Dampak Penggunaan Artificial Intelligence pada Keamanan Siber: Sebuah Kajian Terhadap Potensi Keuntungan dan Ancaman," *J. Pembelajaran dan Pengemb. Diri*, vol. 4, no. 8, pp. 1541–1552, 2024.

[9] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," in *2008 Eighth IEEE International Conference on Data Mining*, 2008, pp. 413–422.

[10] 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.