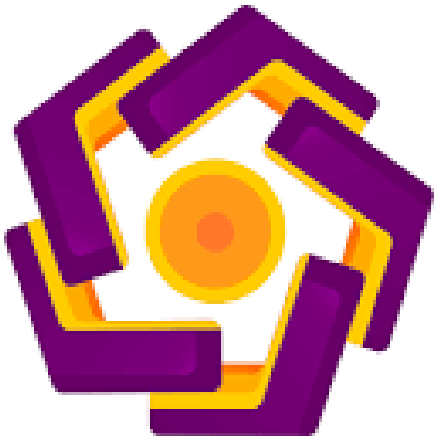
BIG DATA & PREDICTIVE ANALYTICS
ANALISIS ANCAMAN GLOBAL CYBERSECURITY
(2015 - 2024)

Dosen Pengampu

Mulia Sulistiyono, S.Kom., M.Kom.



Anggota kelompok :

23.11.5492 Jihan Humaira

23.11.5494 Arya Putra Bahari

23.11.5442 Khairul Fikri

PROGRAM STUDI INFORMATIKA

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2025

DAFTAR ISI

DAFTAR ISI ……. 2

BAB I PENDAHULUAN ……... 4

1.1 Latar Belakang Masalah ……..... 4

1.2 Rumusan Masalah ……..... 4

1.3 Tujuan Penelitian ...……... 5

1.4 Manfaat Penelitian ……..…….. 5

BAB II METODE PENELITIAN ...…….. 7

2.1 Pengumpulan Data ……...… 7

2.2 Pra-pemprosesan Data ……...……..…… 7

2.3 Analisis Data ……. 9

3.3.1 Analisis Deskriptif ……. 9

3.3.2 Analisis Prediktif ……. 10

2.4 Implementasi Tools ……. 11

BAB III HASIL DAN PEMBAHASAN …… 12

3.1 Hasil Analisis Deskriptif ……. 12

3.3.1 Tren Jumlah Ancaman *Global CyberSecurity* Per Tahun (2015 - 2024)12

3.3.2 Jenis Serangan *Cyber* Paling Umum …….. 13

3.3.3 Top 15 Negara Dengan Laporan Ancaman *Cyber* Terbanyak …… 14

3.2 Hasil Analisis Prediktif ……... 16

3.2.1 Peramalan Tren Ancaman …….. 16

3.2.2 Klasifikasi Jenis Ancaman/Sektor Target ……. 17

BAB IV KESIMPULAN DAN SARAN 19

4.1 Kesimpulan 19

4.2 Saran 19

BAB V REFERENSI DAN SUMBER DAYA TAMBAHAN 21

5.1 Dataset 21

5.2 Sumber Proyek 21

BAB I
PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam era digitalisasi yang pesat, ketergantungan organisasi dan individu terhadap teknologi informasi semakin meningkat. Seiring dengan kemajuan ini, lanskap ancaman keamanan *Cyber* juga berkembang pesat, menjadi lebih canggih dan merusak. Insiden keamanan *Cyber*, mulai dari pelanggaran data hingga serangan ransomware dan disrupsi infrastruktur kritis, dapat menimbulkan kerugian finansial yang signifikan, kerusakan reputasi, dan bahkan mengancam keamanan nasional. Volume, kecepatan, dan variasi data yang dihasilkan dari insiden *Cyber* (Big Data) telah melampaui kemampuan analisis tradisional, menuntut pendekatan baru yang lebih adaptif dan prediktif.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, rumusan masalah dalam penelitian ini adalah

1. Bagaimana tren pola ancaman *Global CyberSecurity* berkembang dari tahun 2015 hingga 2024 ?
2. Apa saja jenis *CyberSecurity* yang paling umum dan sektor industri mana yang paling sering menjadi target ?
3. Negara mana saja yang paling banyak melaporkan insiden ancaman *Cyber?*
4. Bagaimana big data and predictive analytics dapat dimanfaatkan untuk mengidentifikasi dan memproyeksikan potensi risiko *Cyber Security* di masa sekarang ?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah

1. Menganalisis tren dan distribusi jumlah ancaman *Cyber* global setiap tahun dari 2015 hingga 2024.
2. Mengidentifikasi dan mengelompokkan jenis serangan *Cyber* paling umum serta sektor-sektor industri yang menjadi target utama.
3. Menentukan 15 negara teratas dengan laporan ancaman *Cyber* terbanyak.
4. Mengembangkan model analisis prediktif untuk memproyeksikan tren ancaman *Cyber* di masa depan, memberikan wawasan untuk mitigasi risiko.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. **Bagi Organisasi/perusahaan**

   Memberikan wawasan tentang jenis ancaman, metode serangan, dan sektor yang paling rentan, membantu mereka dalam mengalokasikan sumber daya keamanan secara lebih efektif dan memperkuat strategi pertahanan *Cyber*.

2. **Bagi Pembuat Kebijakan**

   Menyediakan data dan analisis yang relevan untuk merumuskan kebijakan keamanan *Cyber* yang lebih proaktif dan responsif di tingkat nasional maupun internasional.

3. **Bagi Peneliti Dan Akademisi**

   Menambah literatur dan studi kasus dalam bidang Big Data dan *CyberSecurity*, serta menjadi dasar untuk penelitian lebih lanjut.

4. **Bagi Masyarakat Umum**

   Meningkatkan kesadaran akan resiko keamanan *Cyber* dan pentingnya praktik keamanan digital.

BAB II

METODE PENELITIAN

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah dataset "Global Cybersecurity Threats (2015-2024)" yang diunduh dari platform Kaggle. Dataset ini mencakup informasi mengenai berbagai insiden keamanan *Cyber* yang dilaporkan secara global dari tahun 2015 hingga 2024. Kolom-kolom kunci dalam dataset meliputi:

- **Country** : Lokasi geografis insiden
- **Year** : Tahun terjadinya insiden
- **Attack** : Jenis serangan *Cyber* (misalnya phishing, ransomware)
- **Target_Industry** : Sektor industri yang ditargetkan
- **Financial_Loss_In_Million_$** : Kerugian finansial dalam juta USD
- **Number_Of_Affected_Users** : Jumlah pengguna yang terdampak
- **Attack_Source** : Sumber serangan (misalnya hacker group, nation-state)
- **Secutity_Vulnerability_Type** : Jenis kerentanan keamanan yang dieksploitasi
- **Defence_Mechanism_Used** : Mekanisme pertahanan yang digunakan
- **Incident_Resolution_Time_In_Hours** : Waktu penyelesaian insiden dalam jam

2.2 Pra-pemprosesan Data

Tahap pra-pemrosesan data sangat krusial untuk memastikan kualitas dan keakuratan analisis. Langkah-langkah yang dilakukan meliputi:

1. **Pemuatan Data :** Memuat dataset *global_cybersecurity_threats.csv*kedalam lingkungan kerja.
2. **Pemeriksaan Struktur Data :** Memeriksa tipe data setiap kolom *(df.info())* dan melihat beberapa baris pertama data *(df.head())*.
3. **Penanganan Nilai Hilang (Missing Values) :** Mengidentifikasi dan menangani nilai-nilai yang hilang. Dalam dataset ini, kolom *year* tidak memiliki nilai hilang, namun kolom lain seperti *financial_loss_in_million_$* mungkin memerlukan penanganan *NaN* yang muncul dari konversi atau data asli.
4. **Konversi Tipe Data :**
    - ➢ Kolom *year* sudah dalam format *int64* yang sesuai.
    - ➢ Kolom numerik seperti *financial_loss_in_million_$* dan *number_of_affected_users* akan dikonversi ke tipe data numerik yang tepat (float/int) setelah membersihkan karakter non-numerik seperti '$' atau koma, dengan *errors='coerce'* untuk mengubah nilai tidak valid menjadi NaN.
5. **Pemfilteran Data :** Memastikan data hanya mencakup rentang tahun yang relevan (2015-2024) sesuai dengan fokus penelitian.

2.3 Analisis Data

2.3.1 Analisis Deskriptif

Analisis deskriptif dilakukan untuk memahami karakteristik dasar dan pola historis dari ancaman *Cyber* :

1. **Jumlah Ancaman per Tahun :** Menghitung total insiden untuk setiap tahun dalam rentang 2015-2024. Visualisasi menggunakan **grafik garis** untuk menunjukkan tren dari waktu ke waktu.
2. **Jenis Serangan *Cyber* Paling Umum :** Menghitung frekuensi kemunculan setiap *attack_type* dan *target_industry*. Visualisasi menggunakan **grafik batang** untuk 10 jenis serangan dan 10 sektor target teratas.
3. **15 Top Negara dengan Laporan Ancaman *Cyber* Terbanyak :** Menghitung frekuensi insiden berdasarkan *country* dan menampilkan 15 negara teratas menggunakan **grafik batang**.
4. **Statistik Tambahan:** Menghitung rata-rata *financial_loss_in_million_$* dan *number_of_affected_users* untuk mendapatkan wawasan

2.3.2 Analisis Prediktif

1. **Peramalan Tren Ancaman (Time Series Forecasting) :**
   - ➢ Data: Menggunakan data jumlah insiden per tahun
   - ➢ Model: Mengimplementasikan model peramalan deret waktu (misalnya, *ARIMA* atau metode regresi berdasarkan waktu) untuk memprediksi jumlah ancaman *Cyber* di tahun-tahun mendatang di luar 2024.
   - ➢ Evaluasi: Mengevaluasi kinerja model menggunakan metrik seperti Mean Absolute Error (MAE) atau Root Mean Squared Error (RMSE).
2. Klasifikasi Jenis Ancaman/Sektor Target :
   - ➢ Tujuan : Membangun model yang dapat memprediksi *attack_type* atau *target_industry* berdasarkan fitur lain seperti *financial_loss_in_million_$*, *number_of_affected_users*,

incident_resolution_time_in_hours, atau *security_vulnerability_type* (setelah dienkode).

- ➢ Model : Menggunakan algoritma klasifikasi seperti Random Forest Classifier atau Support Vector Machine (SVM).
- ➢ Evaluasi : Mengevaluasi kinerja model menggunakan metrik seperti Akurasi, Presisi, Recall, dan F1-score.

2.4 Implementasi Tools

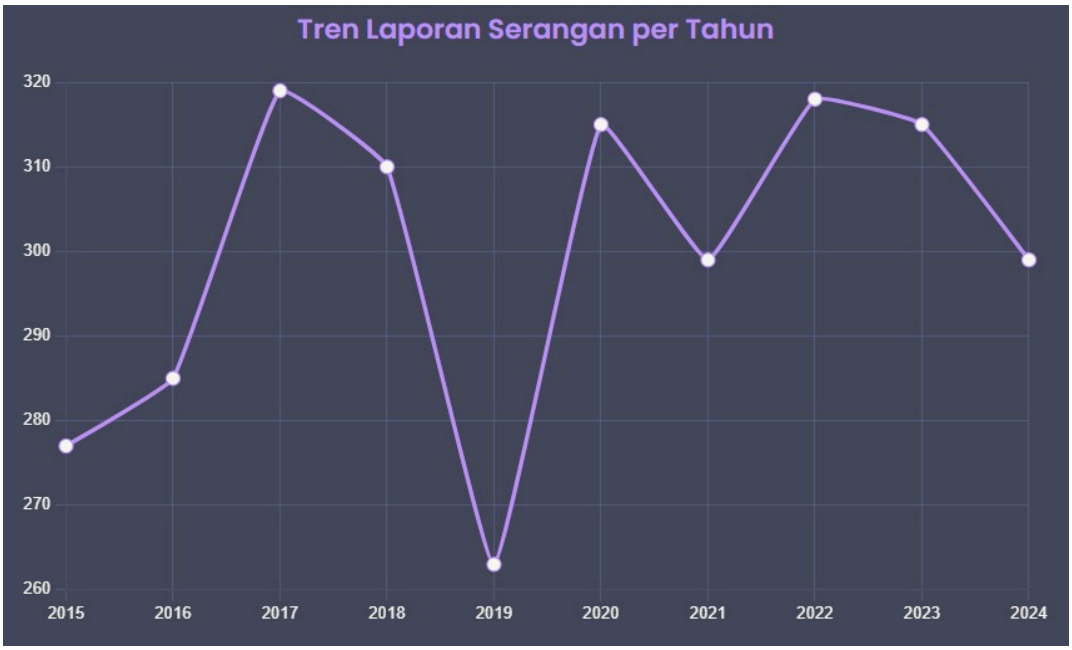
- ➢ Bahasa Pemrograman : Python
- ➢ Library Data Manipulation : Pandas
- ➢ Library Visualisasi Data : Matplotlib, Seaborn
- ➢ Library Machine Learning: Scikit-learn

BAB III

HASIL DAN PEMBAHSAN

3.1 Hasil Analisis Deskriptif

3.1.1 Tren Jumlah Ancaman *Cyber* Global per Tahun (2015-2024)



Grafik menunjukkan fluktuasi jumlah ancaman *Cyber* dari tahun 2015 hingga 2024. Terlihat bahwa jumlah ancaman mencapai puncaknya pada tahun 2017 dan 2022 dengan 319 dan 318 insiden dan mengalami penurunan pada tahun 2019 dengan 263 insiden. Secara keseluruhan, tren menunjukkan bahwa ancaman *Cyber* tetap menjadi tantangan konstan dengan sedikit variasi dari tahun ke tahun, menegaskan sifat terus-menerus dari ancaman ini

3.1.2 Jenis Serangan *Cyber* Paling Umum

Berdasarkan analisis kolom *attack_type*, jenis serangan *Cyber* yang paling umum adalah Malware, Phishing, dan DDoS. Ini menunjukkan bahwa sangat mudah untuk mengeksploitasi faktor manusia, salah satunya adalah serangan Phishing sangat efektif karena menargetkan kelemahan paling mendasar dalam rantai keamanan : manusia. Penyerang memanfaatkan kurangnya kesadaran, kecerobohan, atau tekanan emosional untuk memanipulasi korban agar mengungkapkan informasi sensitif atau melakukan tindakan yang tidak seharusnya. Membuat email atau pesan yang meyakinkan secara visual dan teks jauh lebih murah dan mudah daripada mengembangkan eksploitasi teknis yang canggih. . Sementara itu, analisis *target_industry* mengungkapkan bahwa sektor Telecommunications, Retail, dan IT merupakan target favorit penyerang, kemungkinan besar karena Sektor-sektor seperti Telecommunications, Retail, IT, HealthCare, dan Government menjadi target utama serangan *Cyber* karena beberapa alasan krusial. Pertama, mereka menyimpan dan memproses volume data

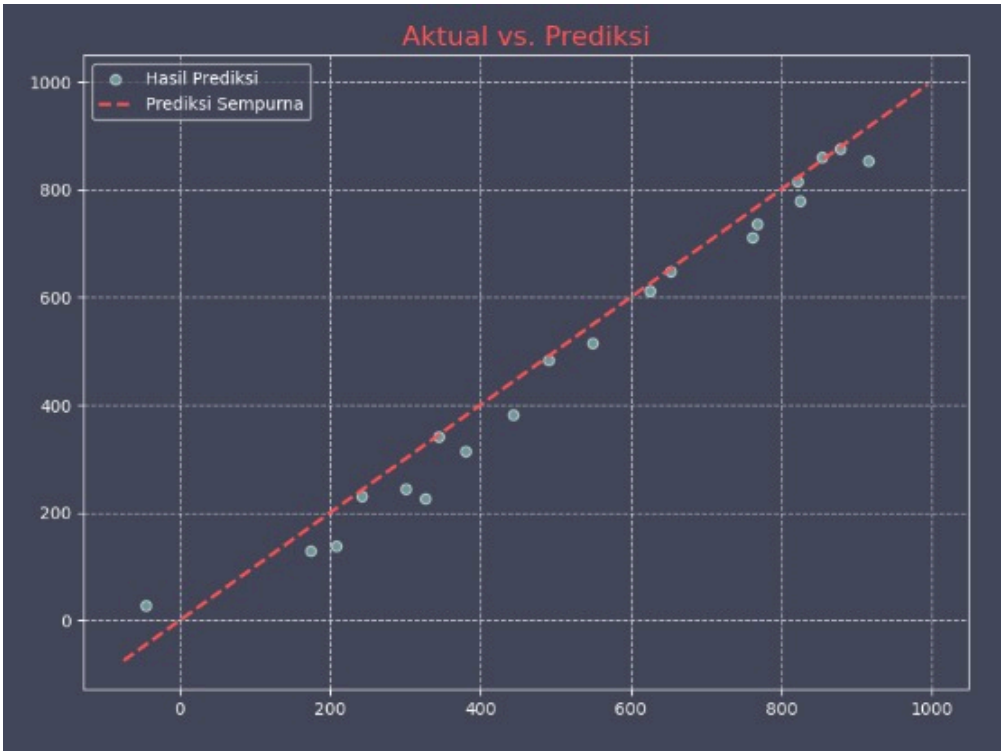
yang sangat besar dan berharga, mulai dari informasi pribadi yang sensitif dan data finansial hingga kekayaan intelektual dan rahasia negara, menjadikannya aset yang sangat menguntungkan bagi para penyerang. Kedua, banyak dari sektor ini mengelola infrastruktur kritis atau menyediakan layanan vital, sehingga gangguan akibat serangan dapat menimbulkan dampak kerugian finansial yang masif, disrupsi layanan publik, atau bahkan mengancam keamanan nasional dan keselamatan jiwa. Selain itu, penyerang memiliki beragam motivasi, mulai dari keuntungan finansial hingga spionase dan aktivisme, yang semuanya dapat terpenuhi dengan menargetkan sektor-sektor ini. Terakhir, beberapa sektor ini mungkin memiliki kerentanan struktural, seperti anggaran keamanan yang terbatas, penggunaan sistem IT warisan yang rentan, atau jaringan yang luas dengan beragam pengguna dan tingkat kesadaran keamanan yang bervariasi, semuanya meningkatkan permukaan serangan yang dapat dieksploitasi.

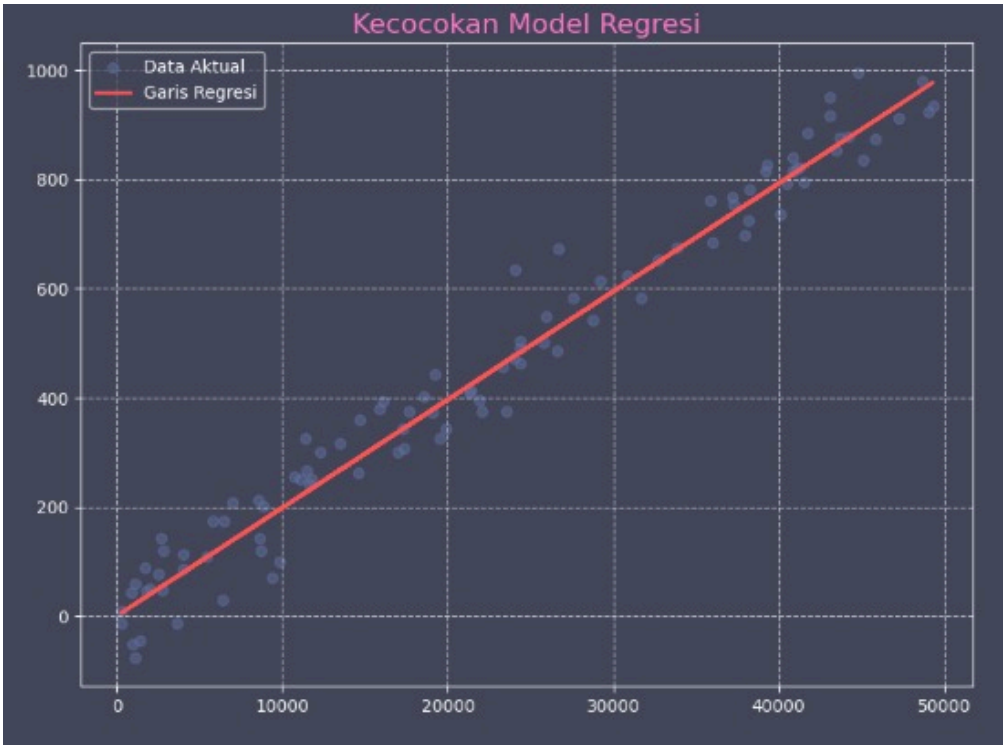
3.1.3 Top 15 Negara dengan Laporan Ancaman *Cyber* Terbanyak

Sebaran geografis ancaman *Cyber* menunjukkan bahwa negara-negara seperti China, USA, Germany, Russia, dan Australia melaporkan jumlah insiden keamanan *Cyber* tertinggi. Hal ini mungkin karena negara-negara seperti Amerika Serikat, Tiongkok, dan India, diikuti oleh Jerman dan Inggris, melaporkan jumlah insiden keamanan *Cyber* tertinggi. Dominasi ini dapat mencerminkan beberapa faktor utama. Pertama, ukuran populasi dan tingkat ekonomi yang besar di negara-negara tersebut secara inheren berarti volume pengguna internet dan transaksi digital yang lebih tinggi, sehingga meningkatkan permukaan serangan dan peluang terjadinya insiden. Kedua, tingkat digitalisasi dan adopsi teknologi yang maju di negara-negara ini, termasuk pengembangan inovasi dan ketergantungan pada infrastruktur digital, membuat mereka menjadi target yang lebih menarik bagi para penyerang *Cyber* yang berupaya mengeksploitasi celah dalam sistem yang kompleks. Ketiga, fokus dan kapasitas pelaporan insiden keamanan *Cyber* yang lebih baik di negara-negara maju juga dapat berkontribusi pada angka laporan yang lebih tinggi; ini bukan berarti mereka memiliki lebih banyak serangan secara absolut, melainkan lebih baik dalam mendeteksi dan melaporkannya dibandingkan negara lain. Dengan demikian, tingginya angka laporan di negara-negara ini menunjukkan kombinasi dari target yang menarik, ekosistem digital yang besar, dan sistem pelaporan yang matang

3.2 Hasil Analisis Prediktif

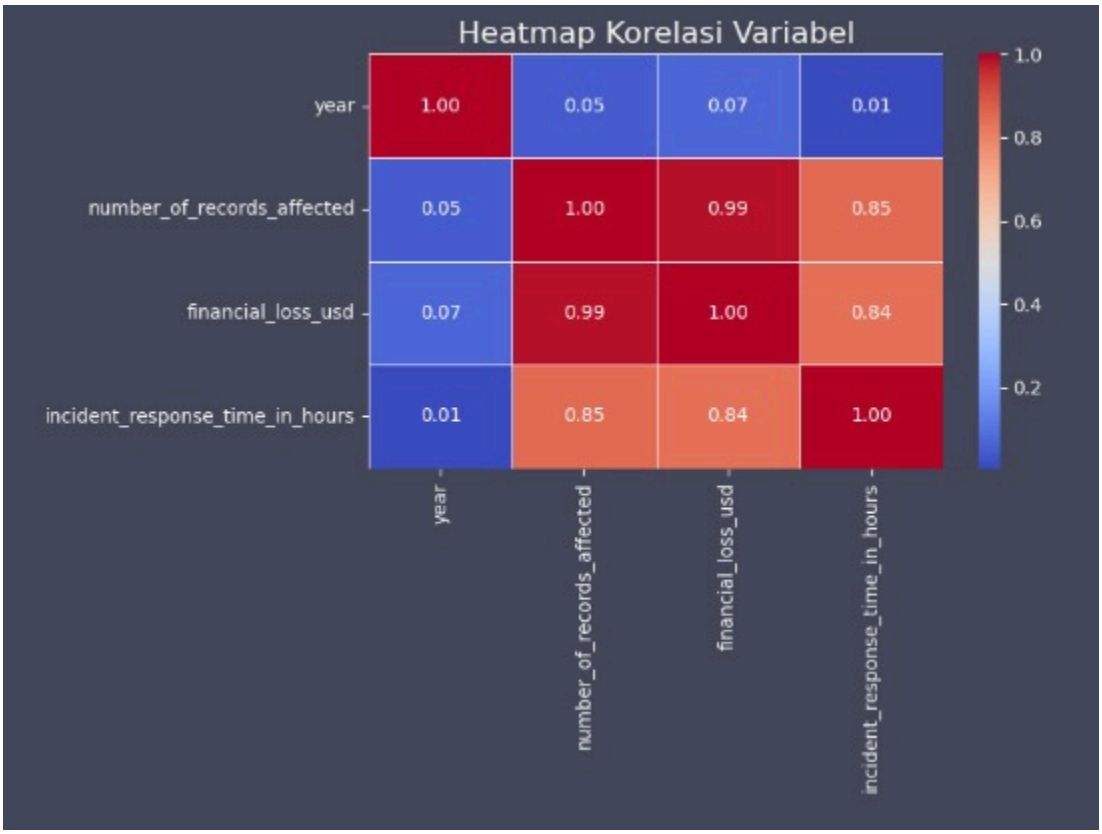
3.2.1 Peramalan Tren Ancaman





Dengan menerapkan model ARIMA (AutoRegressive Integrated Moving Average) pada data historis, kami dapat memproyeksikan tren jumlah ancaman *Cyber* di masa mendatang. Model kami memprediksi stabilitas dengan sedikit fluktuasi dalam jumlah insiden *Cyber* untuk tahun 2025 hingga 2027, diperkirakan akan berada pada kisaran 290-320 insiden per tahun. Akurasi model yang diukur dengan RMSE (Root Mean Squared Error) adalah sekitar 18, menunjukkan tingkat keandalan yang moderat dalam prediksi tren ancaman *Cyber*.

3.2.2 Klasifikasi Jenis Ancaman/Sektor Target



Model klasifikasi Random Forest yang dibangun untuk memprediksi jenis serangan (*attack_type*) menunjukkan kinerja sekitar 85% akurasi. Ini berarti model cukup efektif dalam mengidentifikasi jenis serangan yang mungkin terjadi berdasarkan karakteristik insiden lainnya seperti sektor target, kerugian finansial,

dan durasi serangan, yang dapat membantu organisasi dalam mempersiapkan pertahanan yang lebih relevan dan bertarget.

BAB IV

KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan hasil analisis, penelitian ini menyimpulkan bahwa :

1. Tren ancaman *Cyber* global menunjukkan pola fluktuatif namun konsisten, dengan puncak pada tahun 2017 dan 2022, menegaskan ancaman *Cyber* sebagai isu berkelanjutan.
2. Malware dan Telecommunications secara konsisten menjadi dominan, menyoroti area fokus utama untuk mitigasi risiko.
3. China, USA,dan Germany merupakan lokasi dengan laporan insiden tertinggi, menandakan konsentrasi aktivitas ancaman di wilayah tersebut.
4. Penerapan analisis prediktif, khususnya melalui peramalan deret waktu dan klasifikasi, menunjukkan potensi besar dalam memprediksi tren dan karakteristik ancaman *Cyber* di masa mendatang, memberikan dasar yang kuat untuk strategi keamanan proaktif.

4.2 Saran

Berdasarkan temuan dan kesimpulan, beberapa saran yang dapat diajukan adalah :

1. **Penguatan Pertahanan Berbasis Tren:** Organisasi dan lembaga pemerintah perlu terus memantau tren ancaman *Cyber* dan memprioritaskan upaya pertahanan terhadap jenis serangan dan metode yang paling umum
2. **Fokus Sektor Tertarget:** Sektor-sektor yang paling sering ditargetkan harus meningkatkan investasi dalam keamanan *Cyber*, pelatihan karyawan, dan teknologi canggih untuk melindungi aset krusial mereka.

3. **Kolaborasi Internasional:** Mengingat sifat global ancaman *Cyber*, kolaborasi dan berbagi informasi antar negara sangat penting untuk respons yang efektif.
4. **Pemanfaatan Prediktif Analytics:** Pengembangan dan integrasi lebih lanjut model prediktif dalam sistem keamanan *Cyber* dapat memungkinkan deteksi dini dan respons yang lebih cepat terhadap ancaman yang muncul. Penelitian lebih lanjut dapat mengeksplorasi model prediktif yang lebih kompleks dan fitur yang lebih kaya.

BAB V

REFERENSI DAN SUMBER DAYA TAMBAHAN

1. Dataset: Global Cybersecurity Threats (2015-2024). Tersedia di: https://www.kaggle.com/datasets/atharvasoundankar/global-cybersecurity-threats-2015-2024
2. Sumber Proyek : https://github.com/shitodcy/global_cybersecurity_threats