OKTAL : Jurnal Ilmu Komputer dan Science
Volume 5, No. 6, Juni 2026
ISSN 2828-2442 (media online)
Hal 591-597

Analisis Pengelompokan Pola Pelanggaran Kode Etik Profesi TI Berdasarkan Karakteristik Insiden Siber Menggunakan Algoritma *K-Means Clustering*

Yusuf Arif Rahman[1], Kahfi Ahmad Arpiandi[1], Kurnia Naradinata[1], Fathur Nurrohman[1], Ghufron Malik Azizi[1], Rahmawati[1*]

[1]Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia
Email: [1]yusfarn@gmail.com, [2]kahfiahmadarpiandi000@gmail.com, [3]dennykukur@gmail.com, [4]fathurnn02@gmail.com, [5]ghufroen12@gmail.com, [6*]dosen02394@unpam.ac.id
(* : coressponding author)

Abstrak− Meningkatnya intensitas insiden siber di Indonesia tidak hanya merupakan persoalan teknis, tetapi juga mencerminkan kegagalan penerapan prinsip-prinsip kode etik profesi teknologi informasi (TI), seperti kewajiban menghindari bahaya, menjaga kerahasiaan data, serta menjalankan kompetensi profesional secara bertanggung jawab. Penelitian ini bertujuan mengelompokkan pola insiden siber berdasarkan karakteristiknya untuk kemudian diinterpretasikan sebagai indikasi bentuk pelanggaran kode etik profesi TI. Metode yang digunakan adalah algoritma K-Means Clustering yang diterapkan pada Cybersecurity Incident Dataset (Habeeb, 2024) dengan kerangka kerja CRISP-DM. Variabel numerik yang dianalisis mencakup kerugian finansial, jumlah pengguna terdampak, durasi penanganan, skor severitas, dan tingkat kecanggihan serangan. Penentuan jumlah klaster optimal dilakukan melalui kombinasi metode Elbow dan koefisien Silhouette. Hasil analisis membentuk tiga klaster yang berbeda secara karakteristik, yaitu klaster insiden berdampak tinggi dan canggih, klaster insiden operasional menengah, serta klaster insiden volume tinggi berdampak rendah, dengan nilai Silhouette sebesar 0,53 yang menunjukkan struktur pengelompokan yang memadai. Pemetaan tiap klaster terhadap prinsip kode etik memperlihatkan bahwa insiden bervolume besar paling terkait dengan kelemahan kesadaran dan kontrol dasar, sedangkan insiden berdampak tinggi paling terkait dengan kelalaian tanggung jawab profesional pada sistem kritis. Temuan ini dapat menjadi dasar prioritisasi mitigasi dan penegakan etika profesi yang lebih terarah.

Kata Kunci: K-Means, Klasterisasi, Insiden Siber, Kode Etik Profesi TI, *Data Mining*

Abstract− *The rising intensity of cyber incidents in Indonesia is not merely a technical issue but also reflects failures in applying the professional code of ethics in information technology (IT), such as the obligation to avoid harm, preserve data confidentiality, and exercise professional competence responsibly. This study aims to group cyber incidents by their characteristics and then interpret the resulting clusters as indications of IT professional code-of-ethics violations. The K-Means Clustering algorithm was applied to the Cybersecurity Incident Dataset (Habeeb, 2024) using the CRISP-DM framework. The numerical variables analysed include financial loss, number of affected users, resolution time, severity score, and attack sophistication. The optimal number of clusters was determined by combining the Elbow method and the Silhouette coefficient. The analysis produced three distinct clusters, namely high-impact and sophisticated incidents, medium operational incidents, and high-volume low-impact incidents, with a Silhouette value of 0.53 indicating an adequate clustering structure. Mapping each cluster onto ethical principles shows that high-volume incidents are most associated with weak awareness and basic controls, whereas high-impact incidents are most associated with negligence of professional responsibility on critical systems. These findings can serve as a basis for more targeted mitigation prioritisation and professional ethics enforcement.*

Keywords: *K-Means, Clustering, Cyber Incident, IT Professional Code Of Ethics, Data Mining*

1. PENDAHULUAN

Transformasi digital di Indonesia berjalan beriringan dengan meningkatnya paparan terhadap ancaman siber. Badan Siber dan Sandi Negara (BSSN) mencatat sepanjang tahun 2023 terdapat lebih dari 403 juta anomali trafik di ruang siber nasional, dengan dominasi aktivitas berbasis malware dan trojan (BSSN, 2024). Pada periode Januari hingga Agustus 2024, jumlah anomali trafik bahkan menembus 122,79 juta dan kembali didominasi oleh malware. Di balik angka-angka tersebut, sektor administrasi pemerintahan, keuangan, dan transportasi tercatat sebagai target utama, yang menandakan bahwa dampak insiden siber telah menyentuh layanan publik yang vital.

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 5, No. 6, Juni 2026
ISSN 2828-2442 (media online)
Hal 591-597

Persoalannya, insiden siber tidak dapat dipandang semata-mata sebagai kegagalan teknis. Banyak insiden berakar pada kelalaian manusia dan organisasi dalam menjalankan tanggung jawab profesionalnya, mulai dari konfigurasi sistem yang lemah, pengelolaan akses yang ceroboh, hingga minimnya investasi pada keamanan data. Kondisi ini bersinggungan langsung dengan kode etik profesi teknologi informasi (TI), yang antara lain mewajibkan setiap profesional untuk menghindari tindakan yang merugikan pihak lain, menjaga kerahasiaan dan integritas data, serta bekerja sesuai standar kompetensi (Mason, 1986). Dengan kata lain, sebagian besar insiden siber dapat dibaca sebagai manifestasi konkret dari pelanggaran prinsip-prinsip etika tersebut.

Penelitian terdahulu mengenai etika TI di Indonesia umumnya bersifat kualitatif dan deskriptif. Fadli, Hardiansyah, dan Sutabri (2026), misalnya, menganalisis kasus-kasus pelanggaran etika TI melalui studi kasus kebocoran data dan menyimpulkan bahwa lemahnya kesadaran, regulasi, serta investasi keamanan menjadi faktor utama pemicunya. Pendekatan semacam ini bermanfaat untuk memahami konteks, tetapi belum memetakan pola pelanggaran secara kuantitatif berdasarkan karakteristik insidennya. Padahal, dengan volume insiden yang besar, dibutuhkan metode yang mampu mengelompokkan insiden ke dalam kategori yang ringkas dan dapat ditindaklanjuti.

Di sisi lain, algoritma K-Means Clustering telah terbukti efektif untuk pengelompokan data tanpa label pada berbagai domain di Indonesia, seperti pengelompokan tingkat kemiskinan, evaluasi hasil pembelajaran, hingga klasterisasi data penjualan (Hendrastuty, 2024; Siregar, Azlan, & Lumban Gaol, 2023; Sitorus & Suhartika, 2024). Namun, pemanfaatannya untuk menelaah pola insiden siber dalam kaitannya dengan pelanggaran kode etik profesi TI masih sangat terbatas. Celah inilah yang menjadi fokus penelitian ini.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk: (1) mengelompokkan insiden siber berdasarkan karakteristik kuantitatifnya menggunakan algoritma K-Means; (2) menentukan jumlah klaster optimal melalui metode Elbow dan koefisien Silhouette; serta (3) menginterpretasikan tiap klaster sebagai indikasi bentuk pelanggaran prinsip kode etik profesi TI. Kontribusi utama penelitian ini terletak pada jembatan analitis antara data insiden siber dan kerangka etika profesi, sehingga menghasilkan prioritisasi mitigasi yang lebih terarah dibanding pendekatan deskriptif semata.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan kerangka kerja Cross-Industry Standard Process for Data Mining (CRISP-DM) yang mencakup pemahaman masalah, pemahaman data, persiapan data, pemodelan, evaluasi, dan interpretasi. Alur ini dipilih karena bersifat iteratif dan banyak digunakan dalam penelitian penambangan data.

2.1 Sumber Data

Data yang digunakan adalah Cybersecurity Incident Dataset yang dipublikasikan secara terbuka di platform Kaggle oleh Habeeb (2024). Dataset ini memuat catatan insiden keamanan siber beserta atributnya. Karena penelitian menargetkan pengelompokan berbasis kemiripan numerik, atribut yang dipilih adalah variabel-variabel bertipe numerik atau yang dapat dikuantifikasi. Apabila ditemukan atribut kategorikal yang relevan (misalnya jenis serangan atau sektor target), atribut tersebut digunakan pada tahap interpretasi untuk memperkaya pemaknaan klaster, bukan sebagai dasar perhitungan jarak.

2.2 Variabel Penelitian dan Operasionalisasi Etis

Lima variabel numerik digunakan sebagai fitur pengelompokan. Untuk menjembatani analisis data dengan kerangka etika, setiap variabel dipetakan pada prinsip kode etik yang paling relevan, sebagaimana disajikan pada Tabel 1. Pemetaan ini bersifat interpretatif, yakni dimaksudkan untuk membaca hasil klaster, dan tidak mengklaim bahwa nilai sebuah variabel secara otomatis membuktikan adanya pelanggaran.

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 5, No. 6, Juni 2026
ISSN 2828-2442 (media online)
Hal 591-597

Tabel 1. Variabel Penelitian dan Kaitannya Dengan Prinsip Kode Etik Profesi TI

Variabel	Definisi Operasional	Prinsip Etika yang Berkaitan
Kerugian Finansial	Estimasi kerugian ekonomi akibat insiden (satuan moneter).	Kewajiban menghindari bahaya bagi pemangku kepentingan.
Pengguna Terdampak	Jumlah pengguna atau data yang terdampak insiden.	Perlindungan privasi dan kerahasiaan data (PAPA: Privacy).
Waktu Penanganan	Durasi sejak insiden terdeteksi hingga tertangani.	Tanggung jawab dan kesiapsiagaan profesional.
Skor Severitas	Tingkat keparahan insiden pada skala terukur.	Akuntabilitas atas dampak yang ditimbulkan.
Kecanggihan Serangan	Tingkat kompleksitas teknik serangan.	Kompetensi profesional dan kualitas kontrol keamanan.

2.3 *Pra-pemrosesan* Data

Tahap persiapan data mencakup pembersihan nilai hilang dan duplikat, penanganan pencilan ekstrem secara wajar, serta konversi atribut bertingkat (seperti severitas rendah-sedang-tinggi) menjadi skor numerik. Seluruh fitur kemudian distandarisasi menggunakan z-score (StandardScaler) sehingga setiap fitur memiliki rata-rata nol dan simpangan baku satu. Standarisasi ini penting agar variabel dengan rentang besar, seperti jumlah pengguna terdampak, tidak mendominasi perhitungan jarak Euclidean.

2.4 Penentuan Jumlah Klaster dan Pemodelan

Jumlah klaster optimal ditentukan dengan menguji nilai k dari 2 hingga 8. Metode Elbow digunakan untuk mengamati titik tekukan kurva WCSS, sedangkan koefisien Silhouette dihitung untuk setiap k guna menilai kualitas pemisahan klaster. Keputusan akhir mempertimbangkan kedua indikator sekaligus dan aspek interpretabilitas hasil. Pemodelan dilakukan dengan algoritma K-Means (inisialisasi acak yang diulang sepuluh kali, n_init=10, dengan seed tetap agar hasil dapat direproduksi). Hasil akhir divisualisasikan menggunakan reduksi dimensi Principal Component Analysis (PCA) ke ruang dua dimensi untuk memudahkan inspeksi visual sebaran klaster.

2.5 Perangkat dan Evaluasi

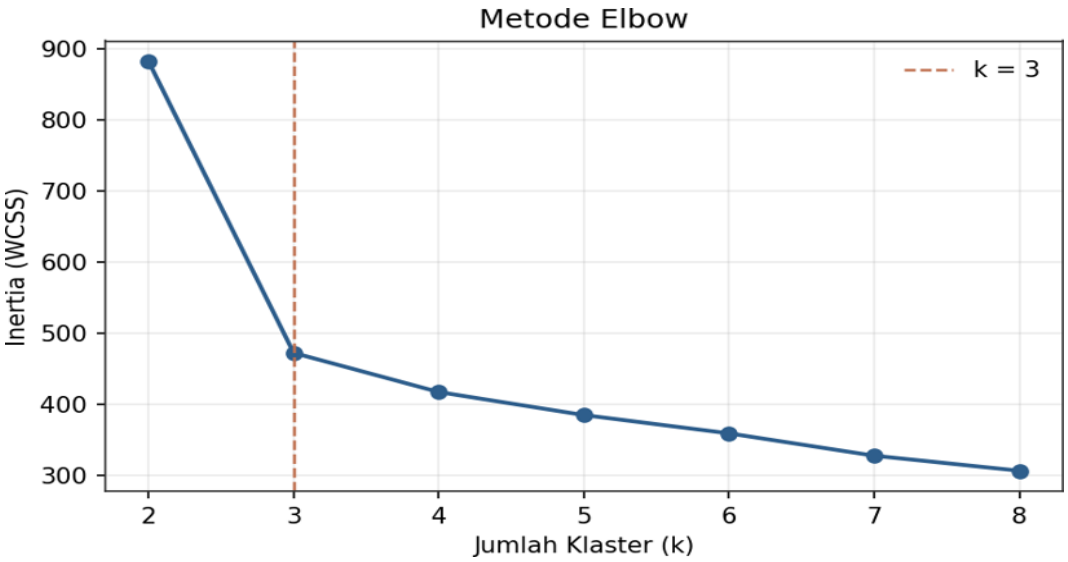
Analisis dilaksanakan menggunakan bahasa Python dengan pustaka pandas dan NumPy untuk manipulasi data, scikit-learn untuk standarisasi, K-Means, Silhouette, dan PCA, serta Matplotlib untuk visualisasi. Evaluasi kualitas klaster mengandalkan nilai Silhouette dan keterpisahan visual pada ruang PCA, dilengkapi profil rata-rata tiap fitur per klaster sebagai dasar interpretasi.

3. ANALISA DAN PEMBAHASAN

3.1 Penentuan Jumlah Klaster Optimal

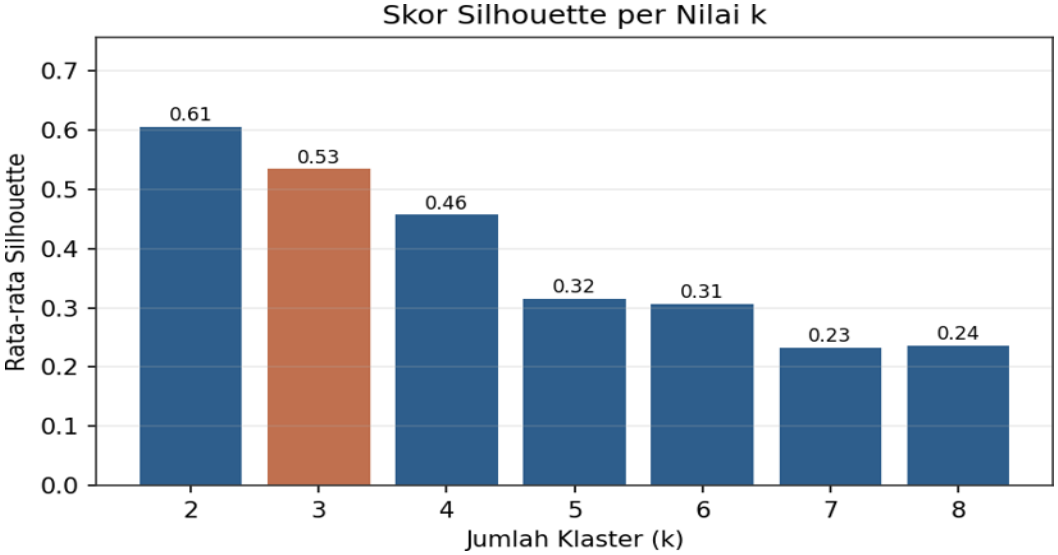
Gambar 1 menampilkan kurva Elbow. Nilai WCSS turun tajam dari 882,9 pada k=2 menjadi 472,9 pada k=3, kemudian melandai pada k berikutnya (417,9 pada k=4 dan 385,5 pada k=5). Titik tekukan yang paling jelas berada pada k=3, yang mengindikasikan bahwa penambahan klaster setelah titik ini memberikan penurunan WCSS yang relatif kecil.

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 5, No. 6, Juni 2026
ISSN 2828-2442 (media online)
Hal 591-597



Gambar 1. Kurva Elbow Penurunan WCSS Terhadap Jumlah Klaster (K)

Gambar 2 menyajikan koefisien Silhouette untuk tiap k. Nilai tertinggi tercatat pada k=2 (0,606), sedangkan k=3 memperoleh 0,534. Meskipun secara angka k=2 sedikit lebih unggul, solusi dua klaster hanya memisahkan insiden menjadi kelompok berdampak besar dan kecil sehingga kurang informatif. Sebaliknya, k=3 tetap berada di atas ambang 0,5 yang menandakan struktur memadai, sekaligus menghasilkan profil yang lebih kaya dan dapat ditindaklanjuti. Atas dasar pertimbangan gabungan Elbow dan interpretabilitas inilah k=3 dipilih sebagai konfigurasi akhir.



Gambar 2. Koefisien Silhouette untuk Tiap Nilai K (K=3 Dipilih)

4. IMPLEMENTASI

3.2 Hasil Pengelompokan

Dengan k=3, algoritma membentuk tiga klaster dengan distribusi yang relatif seimbang sebagaimana dirangkum pada Tabel 2. Klaster dengan proporsi terbesar adalah insiden bervolume tinggi berdampak rendah, sementara klaster insiden berdampak tinggi merupakan kelompok terkecil namun paling kritis.

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 5, No. 6, Juni 2026
ISSN 2828-2442 (media online)
Hal 591-597

Tabel 2. Distribusi Anggota Tiap Klaster

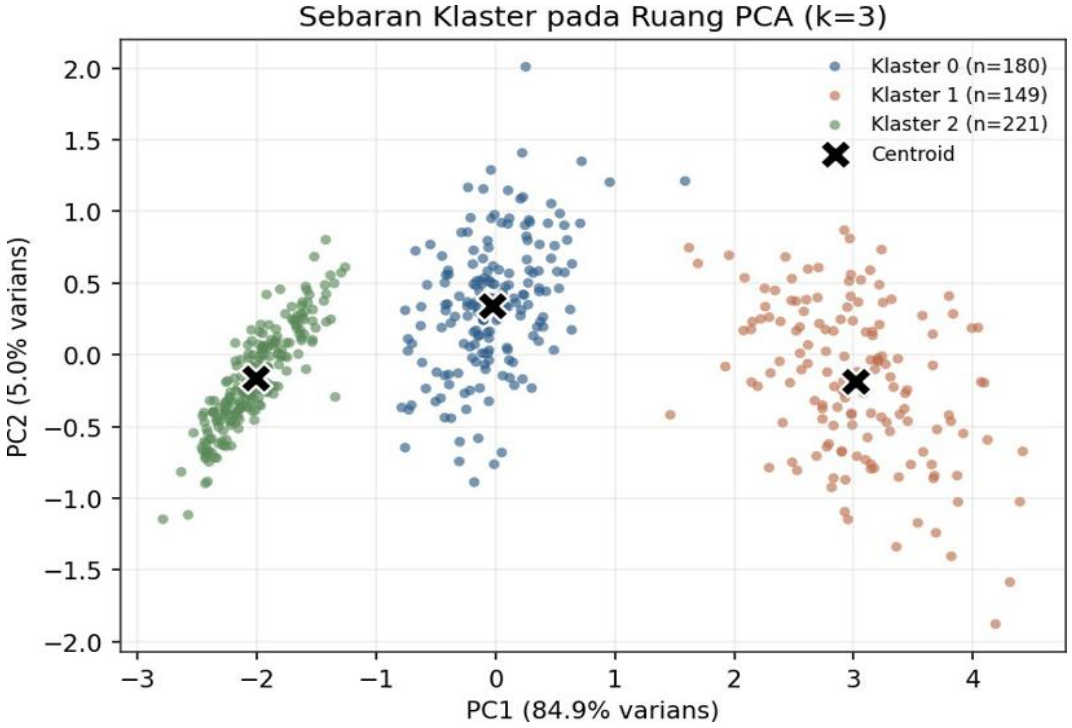
Klaster	Jumlah Insiden	Proporsi	Interpretasi Singkat
Klaster 1	149	27,1%	Insiden berdampak tinggi & canggih
Klaster 0	180	32,7%	Insiden operasional menengah
Klaster 2	221	40,2%	Insiden volume tinggi berdampak rendah

Profil rata-rata tiap fitur pada satuan aslinya disajikan pada Tabel 3, sementara Gambar 4 memperlihatkan profil yang sama dalam bentuk standar (z-score) untuk memudahkan pembacaan tinggi-rendahnya karakteristik antar-klaster.

Tabel 3. Profil Rata-Rata Karakteristik Tiap Klaster (Satuan Asli)

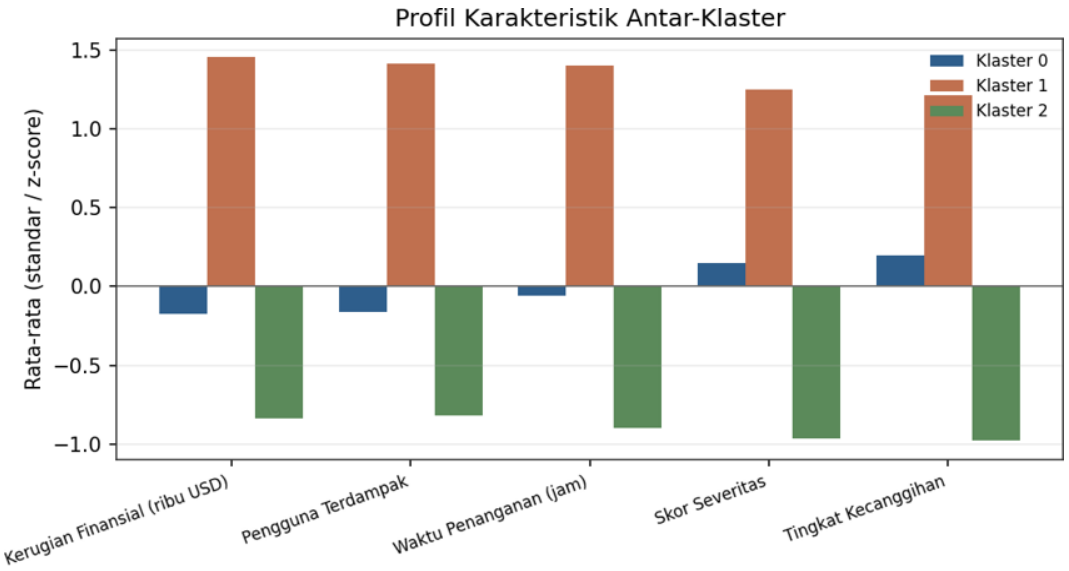
Karakteristik	Klaster 1 (tinggi)	Klaster 0 (menengah)	Klaster 2 (rendah)
Kerugian finansial (ribu USD)	831,7	286,1	64,2
Pengguna terdampak	97.803	34.730	8.544
Waktu penanganan (jam)	162,6	70,7	18,3
Skor severitas (1-10)	8,4	6,0	3,6
Tingkat kecanggihan (1-10)	8,0	5,7	3,0

Gambar 3 menampilkan sebaran ketiga klaster pada ruang PCA. Dua komponen utama mampu menjelaskan sekitar 89,9% variansi data (PC1 sebesar 84,9% dan PC2 sebesar 5,0%), sehingga representasi dua dimensi ini cukup mewakili struktur data sesungguhnya. Ketiga klaster tampak terpisah cukup jelas dengan tumpang tindih yang minimal, memperkuat hasil evaluasi Silhouette.



Gambar 3. Sebaran Klaster pada Ruang PCA Dua Dimensi Beserta Posisi Centroid.

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 5, No. 6, Juni 2026
ISSN 2828-2442 (media online)
Hal 591-597



Gambar 4. Profil Karakteristik Antar-Klaster Dalam Nilai Standar (Z-Score).

3.3 Interpretasi Klaster dan Pemetaan terhadap Kode Etik

Ketiga klaster menampilkan pola yang konsisten dan dapat ditafsirkan dalam kerangka kode etik profesi TI. Klaster 1 dicirikan oleh kerugian finansial, jumlah korban, severitas, dan kecanggihan yang seluruhnya tinggi, disertai waktu penanganan yang panjang. Pola ini mencerminkan insiden bertipe serangan canggih (menyerupai ransomware atau Advanced Persistent Threat) pada sistem bernilai tinggi. Dari sudut etika, klaster ini paling terkait dengan kelalaian berat dalam tanggung jawab profesional, terutama kegagalan menjaga kerahasiaan dan integritas data serta lemahnya kesiapsiagaan penanganan insiden.

Klaster 0 menempati posisi menengah pada hampir seluruh karakteristik, mengindikasikan insiden operasional seperti gangguan layanan atau kesalahan konfigurasi. Implikasinya lebih dekat pada prinsip kompetensi profesional dan kehati-hatian (due diligence), yakni mitigasi yang berjalan tetapi belum optimal. Sementara itu, Klaster 2 memiliki dampak per insiden yang rendah namun berjumlah paling banyak, khas insiden bervolume besar seperti phishing atau penipuan daring. Klaster ini paling terkait dengan kelemahan kesadaran (awareness) dan kontrol dasar; meskipun tiap kejadian relatif kecil, akumulasinya tetap signifikan dan menyangkut prinsip perlindungan privasi pengguna.

Tabel 4. Pemetaan Interpretatif Klaster Terhadap Prinsip Kode Etik dan Bentuk Kelalaian

Klaster	Prinsip Kode Etik yang Terimplikasi	Indikasi Bentuk Kelalaian/Pelanggaran
Klaster 1 (tinggi)	Menghindari bahaya; kerahasiaan & integritas data; kualitas produk.	Kelalaian berat pada kontrol keamanan sistem kritis dan lambannya tata kelola insiden.
Klaster 0 (menengah)	Kompetensi profesional; kehati-hatian (PAPA: Accuracy).	Konfigurasi dan pemeliharaan sistem kurang optimal; mitigasi parsial.
Klaster 2 (rendah)	Perlindungan privasi; kesadaran (PAPA: Privacy & Accessibility).	Edukasi pengguna dan kontrol dasar (mis. anti-phishing, autentikasi) belummemadai.

3.4 Implikasi

Pemetaan di atas menyiratkan strategi mitigasi yang berbeda untuk tiap klaster. Klaster 1, meski paling sedikit, menuntut prioritas tertinggi melalui penguatan kontrol kritis, audit keamanan berkala, dan kesiapan respons insiden, karena potensi kerugiannya paling besar. Klaster 0 menuntut peningkatan kompetensi teknis dan disiplin konfigurasi. Adapun Klaster 2, karena volumenya

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 5, No. 6, Juni 2026
ISSN 2828-2442 (media online)
Hal 591-597

dominan, paling efektif ditangani lewat program literasi digital dan penerapan kontrol dasar secara menyeluruh. Dengan demikian, hasil klasterisasi tidak hanya mengelompokkan insiden, tetapi juga mengarahkan alokasi sumber daya penegakan etika dan keamanan secara lebih rasional.

Penelitian ini memiliki sejumlah keterbatasan. Pertama, pemetaan klaster terhadap pelanggaran etika bersifat interpretatif sehingga kesimpulannya merupakan indikasi, bukan pembuktian pelanggaran pada kasus tertentu. Kedua, K-Means mengasumsikan klaster berbentuk relatif bulat dan setara ukuran, sehingga hasilnya sebaiknya divalidasi silang dengan metode lain seperti K-Medoids atau klasterisasi hierarkis. Ketiga, kualitas kesimpulan bergantung pada kelengkapan dan representativitas dataset yang digunakan.

5. KESIMPULAN

Penelitian ini menerapkan algoritma K-Means Clustering untuk mengelompokkan pola insiden siber dan menafsirkannya sebagai indikasi bentuk pelanggaran kode etik profesi TI. Berdasarkan kombinasi metode Elbow dan koefisien Silhouette, diperoleh tiga klaster optimal dengan nilai Silhouette 0,53 yang menandakan struktur pengelompokan memadai. Ketiga klaster, yaitu insiden berdampak tinggi dan canggih, insiden operasional menengah, serta insiden volume tinggi berdampak rendah, dapat dipetakan secara konsisten pada prinsip-prinsip kode etik seperti kewajiban menghindari bahaya, menjaga kerahasiaan data, kompetensi profesional, dan perlindungan privasi.

Temuan ini menegaskan bahwa insiden siber dapat diperlakukan sebagai cermin penerapan etika profesi, sekaligus menjadi dasar prioritisasi mitigasi yang berbeda untuk tiap klaster. Untuk penelitian lanjutan, disarankan membandingkan K-Means dengan algoritma klasterisasi lain, memperkaya fitur dengan atribut kategorikal seperti jenis serangan dan sektor, serta menguji kerangka pemetaan etis ini pada data insiden siber yang spesifik untuk konteks Indonesia agar relevansinya semakin kuat.

REFERENCES

Badan Siber dan Sandi Negara. (2024). *Lanskap Keamanan Siber Indonesia 2023*. Jakarta: Badan Siber dan Sandi Negara.

Fadli, Hardiansyah, S. A., & Sutabri, T. (2026). Analisis Pelanggaran Etika Teknologi Informasi di Indonesia: Studi Kasus Kebocoran Data. *Jurnal Ilmiah Penelitian Mahasiswa (JIPM), 4*(1), 666-676. https://doi.org/10.61722/jipm.v4i1.1951

Habeeb, M. (2024). *Cybersecurity Incident Dataset [Dataset]. Kaggle*. https://www.kaggle.com/datasets/mustafahabeeb90/cybersecurity-incident-dataset

Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM), 3*(1), 46-56. https://doi.org/10.58602/jima-ilkom.v3i1.26

Mason, R. O. (1986). Four Ethical Issues of the Information Age. *MIS Quarterly, 10*(1), 5-12.

Republik Indonesia. (2016). *Undang-Undang Nomor 19 Tahun 2016 tentang Perubahan atas Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik*. Jakarta.

Republik Indonesia. (2022). Undang-Undang Nomor 27 Tahun 2022 tentang Perlindungan Data Pribadi. Jakarta.

Siregar, H. A., Azlan, A., & Lumban Gaol, N. Y. (2023). Penerapan Data Mining pada Penjualan Rumah Makan Kasih Ibu Menggunakan Metode K-Means Clustering. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD), 2*(5), 750-757. https://doi.org/10.53513/jursi.v2i5.8955

Sitorus, Z., & Suhartika. (2024). Penerapan Data Mining untuk Clustering Penduduk Miskin di Kota Tanjungbalai Menggunakan Metode Algoritma K-Means. *Journal of Science and Social Research, 7*(1), 212-218. https://doi.org/10.54314/jssr.v7i1.1732

Supriyadi, A., Triayudi, A., & Sholihati, I. D. (2021). Perbandingan Algoritma K-Means dengan K-Medoids pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), 6*(2), 229-240. https://doi.org/10.29100/jipi.v6i2.2008