Analisis Forensik Dokumen Hasil *Web Scraping* pada *Carding Forum* dan *Carding Shop:* Pendekatan *Profiling* Forensik dan *Latent Dirichlet Allocation* untuk Investigasi *Cybercrime*

Laporan Akhir Tesis

Fikri Irfan Adristi
23917020

Tesis diajukan sebagai syarat untuk meraih gelar Magister Komputer
Konsentrasi Forensika Digital
Program Studi Informatika Program Magister
Fakultas Teknologi Industri
Universitas Islam Indonesia
2025

Lembar Pengesahan Pembimbing

Analisis Forensik Dokumen Hasil *Web Scraping* pada *Carding Forum* dan *Carding Shop*: Pendekatan *Profiling* dan *Latent Dirichlet Allocation* untuk Investigasi *Cybercrime*

Fikri Irfan Adristi
23917020

Yogyakarta, September, 2025

Pembimbing

Dr. Yudi Prayudi, S.Si., M.Kom.

Lembar Pengesahan Penguji

Analisis Forensik Dokumen Hasil *Web Scraping* pada *Carding Forum* dan *Carding Shop*: Pendekatan *Profiling* dan *Latent Dirichlet Allocation* untuk Investigasi *Cybercrime*

Fikri Irfan Adristi
23917020

Yogyakarta, September, 2025

Tim Penguji,

Dr. Yudi Prayudi, S.Si., M.Kom.
Ketua

Dr. Ahmad Luthfi, S.Kom., M.Kom.
Anggota I

Ir. Irving Vitra Paputungan, S.T., M.Sc., Ph.D
Anggota II

10/09/2025

Mengetahui,
Ketua Program Studi Informatika Program Magister
Universitas Islam Indonesia

Ir. Irving Vitra Paputungan, S.T., M.Sc., Ph.D

Daftar Publikasi

Publikasi yang menjadi bagian dari tesis ini adalah:

Adristi, F. I., & Prayudi, Y. (2025). Forensic Analysis of Web Scraping Documents on Carding Forums and Shops using Latent Dirichlet Allocation. *Jurnal Online Informatika, 10*(2), 323–339. https://join.if.uinsgd.ac.id/index.php/join/article/view/1603

JOIN (Jurnal Online Informatika)
https://join.if.uinsgd.ac.id/index.php/join/article/view/1603

Kontributor	Jenis Kontribusi
Author Fikri Irfan Adristi	1. Menyusun struktur dan isi artikel (100%). 2. Melakukan pengolahan, analisis, dan interpretasi data (100%). 3. Menyempurnakan naskah melalui proses revisi substansial dan teknis (100%).
Author Yudi Prayudi	1. Pengembangan ide dan arah penelitian (70%). 2. Memberi masukan kritis terhadap metodologi dan kerangka analisis (70%). 3. Melakukan tinjauan menyeluruh terhadap draf artikel dan memberi umpan balik (30%). 4. Mendukung penyelarasan naskah dengan kaidah ilmiah dan kebijakan jurnal (30%).

Halaman Kontribusi

Publikasi ini dalam tesis ini merupakan bagian dari proyek penelitian yang didanai oleh Direktorat Riset, Teknologi, dan Pengabdian kepada Masyarakat – Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi (Diktiristek), Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, melalui skema Hibah Penelitian Tesis Tahun Anggaran 2025, dengan Nomor Kontrak: 0498.01/LL5-INT/AL.04/2025. Penulis mengucapkan terima kasih atas dukungan dana yang telah diberikan, yang berperan penting dalam keberhasilan pelaksanaan penelitian ini.

Pernyataan Keaslian Tesis

Dengan ini saya menyatakan bahwa tesis ini merupakan tulisan asli dari penulis, dan tidak berisi material yang telah diterbitkan sebelumnya atau tulisan dari penulis lain terkecuali referensi atas material tersebut telah disebutkan dalam tesis. Apabila ada kontribusi dari penulis lain dalam tesis ini, maka penulis lain tersebut secara eksplisit telah disebutkan dalam tesis ini.

Dengan ini saya juga menyatakan bahwa segala kontribusi dari pihak lain terhadap tesis ini, termasuk bantuan analisis statistik, desain survei, analisis data, prosedur teknis yang bersifat signifikan, dan segala bentuk aktivitas penelitian yang dipergunakan atau dilaporkan dalam tesis ini telah secara eksplisit disebutkan dalam tesis ini.

Segala bentuk hak cipta yang terdapat dalam material dokumen tesis ini berada dalam kepemilikan pemilik hak cipta masing-masing. Apabila dibutuhkan, penulis juga telah mendapatkan izin dari pemilik hak cipta untuk menggunakan ulang materialnya dalam tesis ini.

Yogyakarta, September 2025

Fikri Irfan Adristi

Halaman Persembahan

I dedicate this thesis to

Father: Taufik Joko Susilo

Mother: Imelda Razak

Brother: Rayhan Dwi Fajar & Norman Sariningrat

Kata Pengantar

بِسْمِ اللهِ الرَّحْمنِ الرَّحِيمِ

Assalamu'alaikum Warahmatullahi Wabarakaatuuh

Alhamdulillahirabbil'alamin, segala puji dan syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan tesis dengan judul,„"Analisis Forensik Dokumen Hasil *Web Scraping* pada *Carding Forum* dan *Carding Shop:* Pendekatan *Profiling* Forensik dan *Natural Language Processing* untuk Investigasi *Cybercrime*".

Penulisan skripsi ini dilakukan dengan maksud memenuhi salah satu persyaratan guna *meraih gelar Magister Komputer* di Program Studi Informatika Program Magister, Fakultas Teknologi Industri, Universitas Islam Indonesia. Pada kesempatan ini, penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Prof. Fathul Wahid, ST., M.Sc., Ph.D. selaku Rektor Universitas Islam Indonesia
2. Dr. Raden Teduh Dirgahayu, S.T., M.Sc. selaku Ketua Jurusan Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia
3. Ir. Irving Vitra Paputungan, S.T., M.Sc., Ph.D selaku Ketua Program Studi Informatika Program Magister, Fakultas Teknologi Industri, Universitas Islam Indonesia
4. Dr. Ahmad Luthfi, S.Kom., M.Kom. selaku Manajer Akademik Program Studi Informatika Program Magister, Fakultas Teknologi Industri, Universitas Islam Indonesia
5. Dr. Yudi Prayudi, S.Si., M.Kom. selaku Kepala Pusat Studi Forensika Digital, Universitas Islam Indonesia dan Dosen Pembimbing yang telah memberikan arahan, nasihat, dan dukungan selama penyusunan tesis
6. Arif Hartono, S.E., M.Ec., Ph.D. selaku Ketua Jurusan Manajemen, Fakultas Bisnis dan Ekonomika, Universitas Islam Indonesia
7. Raden Roro Ratna Roostika, S.E., MAC, Ph.D. selaku Kepala Center for International Language and Cultural Studies – Universitas Islam Indonesia (Cilacs UII)
8. Dr.rer.soc.oec Jaya Addin Linando S.E., M.B.A. selaku Editor in Chief Jurnal Siasat Bisnis
9. Abdur Rafik, SE., M.Sc., CSA., ASPM. selaku Ketua Program Studi Manajemen Program Sarjana, Fakultas Bisnis dan Ekonomika, Universitas Islam Indonesia

10. Seluruh jajaran Dosen dan Staff Fakultas Teknologi Industri, Universitas Islam Indonesia
11. Toto Raharjo selaku teman setia perkuliahan di Program Studi Informatika Program Magister, Fakultas Teknologi Industri, Universitas Islam Indonesia
12. Aji Andrianto selaku Chief Business Development Officer PT Kamadeva Indonesia Mandiri
13. Bambang Wirawan selaku Aktivis Nasional yang memberikan inspirasi dan semangat perjuangan kepada penulis
14. Seluruh teman dan kolega terkait teman-teman dan kolega yang telah memberikan doa dan dukungannya baik moril dan materil untuk penulis
15. Seluruh keluarga besar penulis yang senantiasa memberikan doa dan dukungannya baik moril dan materil untuk penulis

Penulis menyadari bahwa tesis ini masih banyak kekurangan dan jauh dari kesempurnaan. Oleh sebab itu, segala saran serta kritik yang konstruktif dari seluruh pihak sangat diharapkan penulis demi penyempurnaan berikutnya. Penulis berharap semoga tesis ini dapat memberikan sumbangsih serta manfaat bagi semua pihak yang membutuhkan.

Wabillahi taufiq wal hidayah
Wassalamu'alaikum Warahmatullahi Wabarakaatuuh

Yogyakarta, September 2025
Penulis,

Fikri Irfan Adristi

Abstrak

Analisis Forensik Dokumen Hasil *Web Scraping* pada *Carding Forum* dan *Carding Shop*: Pendekatan *Profiling* dan *Latent Dirichlet Allocation* untuk Investigasi *Cybercrime*

Penelitian ini berangkat dari masifnya aktivitas *cybercrime* di *carding forum* dan *carding shop*. Berdasarkan banyaknya jumlah korban dan kerugian dari aktivitas tersebut, tentunya diperlukan suatu tindakan investigasi *cybercrime* oleh investigator digital forensik. Tujuan penelitian ini adalah untuk mengembangkan *framework* investigasi forensik *carding* berbasis analisis dokumen hasil *web scraping* pada *carding forum* dan *carding shop*, yang menerapkan metode analisis *profiling* forensik dan *natural language processing* berbasis algoritma *latent dirichlet allocation*. *Tools* yang digunakan untuk *web scraping* pada penelitian ini WebHarvy Version 7.3.0.222. Adapun *tools* yang digunakan untuk pengolahan data pada penelitian ini adalah Microsoft Excel; *packages* Python yaitu Pandas; serta Orange Data Mining. Kesimpulan dari penelitian ini menunjukkan bahwa penerapan teknik investigasi *web scraping* pada *carding forum* dan *carding shop* berbasis *framework* investigasi telah efektif dalam mengumpulkan data yang relevan dan menganalisis aktivitas pelaku *cybercrime* dengan tepat. Secara keseluruhan, penelitian ini berhasil mengembangkan sebuah pendekatan yang lebih terorganisir dan berbasis data dalam menangani kejahatan di *carding forum dan carding shop*, yang dapat menjadi acuan bagi penelitian dan penerapan selanjutnya di bidang investigasi forensik digital.

Kata kunci

web scraping, *latent dirichlet allocation*, kejahatan siber*, carding forum, carding shop*

Abstract

Forensic Analysis of Web Scraping Documents from Carding Forums and Carding Shops: Profiling and Latent Dirichlet Allocation Approaches for Cybercrime Investigation

This research is based on the massive cybercrime activity in carding forums and carding shops. Based on the many victims and losses from these activities a cybercrime investigation action is needed by a digital forensic investigator. The purpose of this research is to develop a forensic carding investigation framework based on document analysis of web scraping results in carding forums and carding shops, which applies forensic profiling analysis methods and natural language processing based on the latent dirichlet allocation algorithm. The tools used for web scraping in this research are WebHarvy Version 7.3.0.222. The tools used for data processing in this research are Microsoft Excel; Python packages namely Pandas; and Orange Data Mining. The conclusion of this research shows that the application of web scraping investigation techniques in carding forums and carding shops based on an investigative framework has been effective in collecting relevant data and analyzing the activities of cybercrime perpetrators appropriately. Overall, this research has succeeded in developing a more organized and data-driven approach in handling crimes in carding forums and carding shops, which can be a reference for further research and application in the field of digital forensic investigation.

Keywords

web scraping, latent dirichlet allocation, cybercrime, carding forum, carding shop

Daftar Isi

Daftar Tabel

Daftar Gambar

Glosarium

AI	- *Artificial intelligence* atau kecerdasan buatan adalah kemampuan sistem komputer untuk meniru kecerdasan manusia, termasuk belajar, memahami, memecahkan masalah, dan membuat keputusan.
BAP	- Berita Acara Pemeriksaan
BAPSA	- Berita Acara Pemeriksaan Saksi Ahli
Carder	- Pelaku kegiatan kriminal *carding*
Carding	- Penipuan daring menggunakan data kartu kredit atau debit
Carding Forum	- *Platform* daring untuk berbagi informasi *carding* ilegal
Carding Shop	- *Platform* daring yang menjual data kartu kredit dan/atau akses ilegal
Cybercrime	- Kejahatan yang dilakukan dengan menggunakan teknologi *cyber*
Cybercriminal	- Kriminal yang melakukan aksinya menggunakan teknologi *cyber*
Framework	- Struktur tahapan sistematis dalam investigasi digital forensik untuk analisis dan pemecahan masalah
LDA	- *Latent dirichlet allocation* yaitu model statistik untuk mengidentifikasi topik tersembunyi dalam kumpulan teks.
ML	- *Machine learning* adalah cabang dari kecerdasan buatan yang memungkinkan sistem komputer untuk belajar dari data dan membuat keputusan tanpa diprogram secara eksplisit.
NLP	- *Natural language processing* yaitu komputasi untuk memahami dan mengolah bahasa manusia secara komputeris.
Profiling Forensik	- Pendekatan investigasi digital untuk mengidentifikasi pola, karakteristik, dan perilaku pelaku kejahatan siber berdasarkan bukti digital yang diperoleh dari data terkait aktivitas mereka.
URL	- *Uniform Resource Locator* adalah alamat unik yang digunakan untuk menemukan sumber daya di internet, seperti halaman web, gambar, video, atau dokumen lainnya.
Web Scraping	- Ekstraksi otomatis informasi dari *web page*

BAB 1
Pendahuluan

1.1. Latar Belakang

Cybercrime semakin menjadi perhatian di era digital saat ini, dengan individu dan organisasi menjadi korban berbagai jenis serangan dunia maya (Buil-Gil et al., 2021; Koops, 2011; M. S. Malik & Islam, 2019; Teodoro et al., 2015). Saat ini diperkirakan bahwa *cybercrime* menjadi lebih terorganisir; berskala besar; terdiversifikasi dengan meningkatnya pembagian kerja; dan diperkirakan akan mengembangkan hubungan yang semakin erat dengan kejahatan terorganisir *offline* (Ahmad, 2008; Alsaraireh, 2025; Buil-Gil et al., 2021; H. Chen et al., 2025; Grabosky, 2014; Jirovský et al., 2018; Koops, 2011; Wang et al., 2025; Whittaker et al., 2025; Yip et al., 2013).

Lebih lanjut, para *cybercriminal* juga senantiasa mengembangkan teknik dan strateginya, menggunakan kemajuan teknologi untuk tujuan ilegal, khususnya untuk keuntungan finansial. Bidang perilaku kriminal yang sedang berkembang ini, dikenal sebagai *carding*, yang mana melibatkan penggunaan komputer dan internet untuk melakukan kejahatan, khususnya menargetkan informasi keuangan seperti nomor kartu kredit, rincian rekening bank, dan *personal identification information* (Kshetri, 2006; Yip et al., 2013).

Dampak kejahatan siber *carding* sangat signifikan, berdampak pada perusahaan kartu kredit, *merchant*, dan konsumen (A. Agarwal et al., 2020; Azhan & Meraj, 2020; Smadi & Min, 2020). *Carding* menimbulkan ancaman signifikan terhadap keamanan dan integritas sistem keuangan dan informasi pribadi individu lainnya (Zahra & Urumsah, 2025). Sudah menjadi fakta umum bahwa penipuan kartu kredit merupakan masalah yang terus berkembang. Berbagai skema *skimming*, pemalsuan, dan *phishing* terjadi setiap tahunnya, menyebabkan kerugian miliaran US Dollar bagi perusahaan dan korban. Meskipun perusahaan dan *merchant* kartu kredit telah menerapkan berbagai cara untuk membantu mencegah penipuan kartu kredit, hal ini masih menjadi kekhawatiran (Barker et al., 2008).

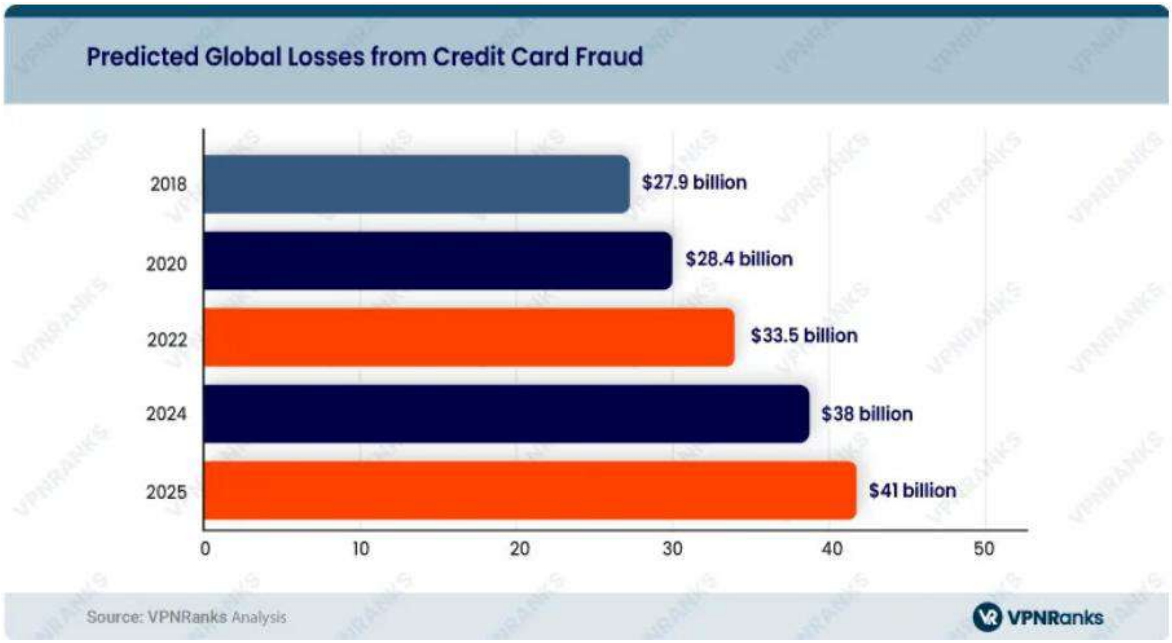
Kekhawatiran ini tentunya didukung dengan data laporan pencurian identitas sejak kuartal 1 tahun 2019 hingga kuartal 1 tahun 2025 yang dirilis oleh Federal Trade Commission (2025) dan disajikan pada tabel 1.1 dibawah berikut:

Tabel 1.1 Perbandingan Jenis Laporan Pencurian Identitas
Sejak Kuartal 1 Tahun 2019 hingga Kuartal 1 Tahun 2025

Jenis Identitas yang Dicuri	Total Laporan
Kartu Kredit	2.512.995
Pencurian Identitas Lainnya	2.013.941
Dokumen atau Manfaat Pemerintah	1.070.103
Pinjaman atau Sewa	1.044.290
Akun Bank	711.724
Terkait Ketenagakerjaan atau Pajak	581.293
Telepon atau Utilitas	534.332

Sumber: Federal Trade Commission (2025)

Data laporan di atas telah menunjukkan bahwa total laporan pencurian identitas yang paling tinggi terdapat pada tipe kartu kredit dengan jumlah laporan sebanyak 2.512.995 (Federal Trade Commission, 2025). Selanjutnya, berikut dibawah ini pada gambar 1.1 memprediksi kerugian global akibat penipuan kartu kredit (Ashraf & Tilawat, 2024):



Gambar 1.1 Kerugian Global yang Diprediksi Akibat Penipuan Kartu Kredit
Sumber: Ashraf & Tilawat (2024)

Berdasarkan gambar 1.1. jika dilihat dari tahun 2018 hingga 2025, tren kerugian global akibat penipuan kartu kredit diprediksi terus meningkat. Pada tahun 2024, kerugian diperkirakan mencapai sekitar $38 miliar (berdasarkan tren dan kelanjutan teknik penipuan). Pada tahun 2025, kerugian diperkirakan dapat melebihi $41 miliar (Ashraf & Tilawat, 2024).

Besarnya jumlah laporan pencurian identitas dan dampak kerugian dari penipuan kartu finansial tersebut, tentunya merupakan akibat aktivitas *cybercrime* yang dilakukan para *cybercriminal* di *carding forum* dan *carding shop*. Berdasarkan hal tersebut, tentunya perlu

menjadi atensi bagi investigator digital forensik untuk melakukan investigasi dan analisis forensik pada *carding forum* dan *carding shop* (Ashraf & Tilawat, 2024; Federal Trade Commission, 2025).

Berdasarkan penelitian terdahulu terdapat sejumlah solusi yang dapat diimplementasikan guna menginvestigasi kejahatan siber. Diantaranya seperti *footprinting* (Agrawal et al., 2024; Levy & Gafni, 2021; Sharma et al., 2023; Sharmila & Aparna, 2024); *reconnaissance* (Bollikonda & Kiran, 2024; Mazurczyk & Caviglione, 2021; Pringle et al., 2024); *undercover operations* (Ndubuisi et al., 2024; Steinmetz et al., 2023; Valiño Ces, 2024; Vasoya et al., 2024); dan kerjasama dengan penyedia layanan *hosting* (Benhamou, 2017; Frosio & Bulayenko, 2021; Kalacska & Bouchard, 2011; Kopel, 2013) untuk memperoleh informasi tentang pengguna dan aktivitas yang terjadi di *carding forum* dan *carding shop*. Teknik investigasi *footprinting, reconnaissance, undercover operations*, dan kerjasama dengan penyedia layanan *hosting* memiliki kelemahan seperti risiko terdeteksi oleh pelaku, proses yang memakan waktu dan sumber daya, kehilangan identitas, keterbatasan kerjasama yang terjadi, dan perbedaan yurisdiksi.

Meskipun teknik-teknik investigasi tersebut *applicable* dan bermanfaat namun, dalam beberapa kasus, penerapan *web scraping* terbukti lebih efektif dalam investigasi kejahatan siber. Alasannya, penerapan *web scraping* memungkinkan akses *real-time* ke informasi *carding*, identifikasi pelaku, dan jejak digital. *Web scraping* memiliki istilah yang mirip, yaitu *web crawling*, namun keduanya memiliki perbedaan dalam tujuan dan metode yang digunakan. *Web scraping* adalah teknik untuk mengubah data web yang tidak terstruktur menjadi data terstruktur yang dapat disimpan dan dianalisis dalam c*entral database* atau *spreadsheet* (Sirisuriya, 2015). Di sisi lain, *web crawling* adalah proses yang lebih luas yang melibatkan penggunaan *bot* untuk menjelajahi seluruh situs web atau bahkan seluruh internet. Tujuan utama dari *web crawling* adalah untuk mengindeks konten web sehingga dapat diakses dan dicari dengan mudah oleh mesin pencari seperti Google atau Bing. *Web crawler* akan mengikuti tautan dari satu halaman ke halaman lain, mengumpulkan dan menyimpan informasi tentang setiap halaman yang dikunjungi (ScrapeOps, 2024).

Perbedaan antara *"web scraping"* dan *"web crawling"* relatif samar, karena banyak penulis dan programmer akan menggunakan kedua istilah tersebut secara bergantian. Secara umum, istilah *"crawler"* menunjukkan kemampuan program untuk menavigasi halaman web sendiri, mungkin bahkan tanpa tujuan akhir atau maksud yang jelas, menjelajahi situs atau web tanpa henti. *Web crawler* banyak digunakan oleh mesin pencari seperti Google untuk mengambil konten untuk URL, memeriksa halaman tersebut untuk tautan lain,

mengambil URL untuk tautan tersebut, dan sebagainya (Broucke & Baesens, 2018). Namun secara praktis, dapat diidentifikasi bahwa fungsi *web crawling* memiliki cakupan yang lebih luas dan umumnya digunakan untuk mengindeks konten secara menyeluruh, sedangkan *web scraping* berfokus pada ekstraksi data tertentu yang kemudian dianalisis dalam format yang terstruktur (Khder, 2021; Nigam & Biswas, 2021).

Selanjutnya, terdapat juga beberapa penelitian terdahulu yang menerapkan *web scraping* sebagai teknik dalam investigasi forensik, seperti pada penelitian (Maybir & Chapman, 2021; Muehlethaler & Albert, 2021). Pada penelitian Maybir & Chapman (2021) hasilnya menunjukkan bahwa *open source investigation* menggunakan teknik *web scraping* terhadap data laporan ekstasi *online* terbukti paling efektif untuk memperoleh data ringkasan umum yang sesuai dibandingkan dengan pendekatan populasi lainnya yang lebih mahal dan memberatkan seperti *wastewater analyses* dan *population surveys*.

Pada penelitian Muehlethaler & Albert (2021) telah ditekankan bahwa survei populasi fiber merupakan bagian penting dari bidang pemeriksaan serat forensik. Penelitian tersebut menunjukkan bahwa teknik *web scraping* memiliki potensi untuk memberikan penelitian *near real-time population* yang dapat memberikan manfaat besar bagi praktisi forensik. Meskipun peneltian (Maybir & Chapman, 2021; Muehlethaler & Albert, 2021) telah menerapkan *web scraping* sebagai teknik dalam investigasi forensik, namun penelitian tersebut memiliki kelemahan dalam aspek teknik analisis datanya yang masih sederhana berupa analisis deskriptif *profiling* dalam penyajiannya.

Solusi teknik investigasi kejahatan siber seperti *footprinting*, dan *undercover operations* penting; namun penerapan *web scraping*, analisis *profiling* forensik, dan analisis *natural language processing* dapat mengakomodir kebutuhan untuk respons cepat dan fokus pada sumber masalah. Kombinasi strategi ini penting dalam memerangi *cybercrime*. Contohnya pada penelitian (Basheer & Alkhatib, 2024; Li et al., 2016; Sonmez & Codal, 2024; Szigeti et al., 2024) telah menggunakan pendekatan *natural language processing* pada teks dokumen dari *dark web* dan *dark forum* untuk mengidentifikasi pola komunikasi, aktivitas ilegal, serta profil penjual, yang mendukung investigasi forensik *cybercrime* dalam konteks perdagangan narkoba, terorisme, dan *credit card fraud*.

Penelitian ini memiliki urgensi yang tinggi dalam ranah forensik digital, mengingat aktivitas *carding* yang mencakup pencurian dan perdagangan data kartu kredit secara daring, semakin marak dilakukan melalui *carding forum* dan *carding shop*. *Platform* ini digunakan oleh pelaku kejahatan siber untuk berbagi informasi, menjual data curian, dan melakukan transaksi secara anonim, sehingga menyulitkan proses pelacakan dan investigasi oleh aparat

penegak hukum. Forum-forum ini umumnya memuat konten teks tidak terstruktur, bersifat terselubung, dan sulit diproses dengan metode analisis konvensional.

Guna mengatasi tantangan tersebut, penelitian ini memanfaatkan kombinasi teknik *web scraping,* analisis *profiling* forensik, dan pendekatan *natural language processing* untuk mengakses dan menganalisis konten yang tersembunyi dalam ekosistem *carding*. Pendekatan ini mengacu pada temuan Wiratmoko et al. (2025), yang menunjukkan efektivitas model *natural language processing* dalam mengekstrak informasi secara otomatis dari kumpulan data besar, serta diperkuat oleh perspektif keamanan digital dari Alam & Gupta (2024) yang menekankan pentingnya integrasi teknologi seperti *blockchain* dan otomasi untuk deteksi dan verifikasi konten daring secara efisien. Pendekatan dalam penelitian ini juga mengacu pada penelitian (Jin et al., 2024; Sonmez & Codal, 2024). Sonmez & Codal (2024) mengeksplorasi aktivitas kriminal *dark web*, mengidentifikasi topik terkait terorisme menggunakan pemodelan topik LDA. Sementara itu, penelitian Jin et al. (2024) membahas tantangan pelacakan penjahat di *dark web*, menggunakan *advanced crawlers* dan *machine learning* untuk mengatasinya.

Pemilihan objek *carding forum* dan *carding shop* dalam penelitian ini didasarkan pada pertimbangan strategis, relevansi konteks, serta keterkaitan langsung kedua *platform* tersebut dengan aktivitas kejahatan pencurian dan perdagangan data kartu kredit secara daring. Tidak seperti *darkweb marketplaces* yang bersifat lebih umum dan menjual berbagai jenis barang atau jasa ilegal, *carding forum* dan *carding shop* secara khusus menjadi wadah utama pertukaran informasi, transaksi, serta interaksi antar pelaku *carding*. Di dalamnya, ditemukan diskusi teknis, *tutorial*, penawaran data curian, serta reputasi anggota yang menunjukkan struktur sosial yang khas dalam ekosistem kejahatan ini.

Sementara itu, objek lain seperti *paste sites*, *leak databases*, dan *ransomware leak sites* cenderung lebih bersifat satu arah dan tidak memiliki dinamika percakapan atau interaksi antar pengguna yang cukup untuk dianalisis secara mendalam melalui pendekatan berbasis *forensic profiling* dan *natural language processing-latent dirichlet allocation*. *Carding forum* menyediakan jejak digital yang lebih kaya dan kontekstual seperti nama pengguna, *thread*, komentar, dan pola interaksi, yang sangat relevan untuk dianalisis guna mengidentifikasi peran, aktivitas, dan struktur komunikasi pelaku. Dengan demikian, pemilihan *carding forum* dan *carding shop* sebagai fokus utama tidak hanya memberikan sumber data yang spesifik dan relevan dengan tujuan investigasi, tetapi juga memungkinkan pengembangan *framework* analisis yang lebih terarah dan aplikatif dalam konteks penegakan hukum serta penelitian forensik digital secara empiris dan terfokus.

Secara ilmiah, penelitian ini memberikan kontribusi dalam pengembangan pendekatan baru berbasis *data science* untuk investigasi forensik digital, dengan kebaruan utama sebagai berikut: (1) Integrasi teknik *web scraping* dengan *natural language processing* dan *descriptive forensic profiling* secara terpadu; (2) pemanfaatan algoritma *latent dirichlet allocation* (LDA) untuk mengungkap topik-topik tersembunyi dalam komunikasi *carding*; (3) penyajian hasil analisis dalam bentuk visualisasi data forensik yang komunikatif dan informatif. Selain itu, data yang digunakan berasal dari *platform* internasional yang belum banyak diteliti secara spesifik dalam konteks *carding ecosystem*, sehingga memperkaya literatur dan praktik forensik digital global.

Kontribusi sosial dari penelitian ini terletak pada efisiensi proses investigasi yang tidak memerlukan keterlibatan langsung dalam ekosistem ilegal sehingga, mengurangi risiko hukum dan etika bagi penyidik atau peneliti. Lebih jauh, *framework* investigasi forensik *carding* dirancang agar kontekstual dan relevan untuk mendukung aparat penegak hukum di Indonesia dalam memanfaatkan data global guna memperkuat upaya pencegahan dan penindakan kejahatan siber secara lebih adaptif dan berbasis bukti. Pengembangan *framework* investigasi ini juga mendukung penegakan hukum yang lebih sistematis serta mengarahkan pada laporan yang dapat dipertanggungjawabkan di pengadilan.

Dalam penelitian ini, peneliti mengusulkan solusi berupa pengembangan *framework* investigasi forensik *carding* yang didasarkan pada analisis dokumen hasil *web scraping* dari *carding forum* dan *carding shop*. *Framework* ini melibatkan penerapan metode analisis deskriptif *profiling* forensik dan teknik *natural language processing* menggunakan pendekatan *latent dirichlet allocation* untuk menganalisis dokumen hasil *web scraping* yang diperoleh dari *carding forum* dan *carding shop*. Peneliti menggunakan Microsoft Excel serta *packages* Python yaitu Pandas agar dapat membuat visualiasi data deskriptif *profiling* forensik yang baik (Microsoft, 2024; The pandas development team, 2020). Dalam analisis *profiling* forensik yang baik tentunya juga harus menyajikan infografis yang akurat, relevan, dan komunikatif. Informasi deskriptif *profiling* forensik dengan penyajian infografis yang baik tentunya akan memudahkan *stakeholder* penegakan hukum dan pembuat kebijakan dalam memahami tren *cybercrime* dan pengambilan keputusan atas tren tersebut seperti halnya pada penelitian (Böhm et al., 2020; Cullen et al., 2024; Siricharoen & Siricharoen, 2018) yang memanfaatkan infografis untuk *stakeholder*.

Selanjutnya, analisis *natural language processing* dengan pendekatan *latent dirichlet allocation* atas dokumen hasil *web scraping* pada *carding forum* dan *carding shop* diperlukan untuk mengidentifikasi dan memahami topik tersembunyi, pola komunikasi, dan

profil pelaku kejahatan, guna membantu dalam investigasi dan pencegahan *cybercrime* secara efektif. Investigasi dalam penelitian ini difokuskan pada konteks analisis informasi yang tersebar di *platform* ilegal seperti *carding forum* dan *carding shop*, dimana pelaku kejahatan siber sering berdiskusi, menawarkan, atau memperjualbelikan data hasil kejahatan.

Framework yang diusulkan memanfaatkan dokumen hasil *web scraping* dari situs-situs tersebut sebagai sumber data utama, yang kemudian dianalisis menggunakan metode deskriptif *profiling* forensik dan teknik *natural language processing* (NLP) dengan pendekatan *latent dirichlet allocation* (LDA). Pendekatan ini diharapkan dapat mengidentifikasi pola komunikasi, peran pengguna, serta topik-topik dominan yang berkaitan dengan aktivitas *carding*, sehingga mendukung upaya penegakan hukum dan pencegahan kejahatan siber.

Peneliti menggunakan *software* Orange Data Mining untuk melakukan analisis *natural language processing* dengan pendekatan *latent dirichlet allocation* (Demšar et al., 2013). Berikut adalah tabel perbandingan mengapa solusi *web-scraping*, analisis *profiling*, dan LDA cocok diterapkan dalam investigasi *cybercrime*, sedangkan *footprinting, reconnaissance, undercover operations,* dan kerjasama dengan penyedia layanan *hosting* tidak cocok:

Tabel 1.2 Perbandingan Metode Investigasi *Cybercrime*

Metode Investigasi	**Kelebihan Metode**	**Kekurangan Metode**	**Alasan Cocok/Tidak Cocok**	**Referensi Terkait**
Web Scraping	Dapat mengumpulkan data dalam jumlah besar secara cepat dan otomatis.	Memerlukan pemahaman teknis tinggi dan bisa terblokir oleh situs target; rentan terhadap perubahan struktur situs web.	Cocok: Dapat mengumpulkan data dari *carding forum* secara sistematis tanpa interaksi langsung dengan pelaku.	(Maybir & Chapman, 2021; Muehlethaler & Albert, 2021; Nurseno et al., 2024)
Analisis *Profiling* Forensik	Dapat mengidentifikasi pola perilaku pelaku berdasarkan data digital.	Membutuhkan data historis yang cukup lengkap untuk analisis.	Cocok: Membantu memahami modus operandi pelaku *carding* melalui analisis pola perilaku dan karakteristik mereka.	(Agrawal et al., 2024; Christian et al., 2022; Li et al., 2016)

Metode Investigasi	Kelebihan Metode	Kekurangan Metode	Alasan Cocok/Tidak Cocok	Referensi Terkait
Latent Dirichlet Allocation (LDA)	Mengungkap topik tersembunyi dalam teks besar, memberikan *insight* penting.	Hasil analisis bisa sulit diinterpretasikan tanpa pemahaman mendalam tentang algoritma dan topik yang dihasilkan; membutuhkan pemahaman statistik yang baik; Tidak cocok untuk data non-teks.	Cocok: Dapat mengidentifikasi topik terkait *carding* yang tidak terlihat secara langsung, sehingga memperkaya investigasi.	(Basheer & Alkhatib, 2024; Sonmez & Codal, 2024; Szigeti et al., 2024)
Footprinting	Efektif untuk mengidentifikasi infrastruktur dan komponen jaringan yang digunakan oleh pelaku.	Hanya mengumpulkan informasi di permukaan, kurang relevan untuk menganalisis konten teks atau transaksi *carding*.	Tidak Cocok: Fokus pada informasi jaringan, bukan pada analisis konten atau dokumen *carding* yang menjadi target investigasi.	(Agrawal et al., 2024; Levy & Gafni, 2021; Sharma et al., 2023; Sharmila & Aparna, 2024)
Reconnaissance	Berguna untuk mendapatkan informasi awal sebelum serangan.	Informasi yang diperoleh umumnya bersifat umum dan kurang spesifik untuk investigasi mendalam.	Tidak Cocok: Tidak dirancang untuk analisis mendalam pada data digital yang dibutuhkan dalam investigasi *carding forum* atau *carding shop*.	(Bollikonda & Kiran, 2024; Mazurczyk & Caviglione, 2021; Pringle et al., 2024)
Undercover Operations	Memungkinkan interaksi langsung dengan pelaku dan pengumpulan bukti dari dalam jaringan kriminal.	Sangat berisiko, memakan waktu lama, dan sulit diterapkan dalam skala besar.	Tidak Cocok: Berisiko tinggi dan membutuhkan waktu lama, sementara data *carding* bisa diakses lebih efisien melalui metode otomatis	(Ndubuisi et al., 2024; Steinmetz et al., 2023; Valiño Ces, 2024; Vasoya et al., 2024)

Metode Investigasi	Kelebihan Metode	Kekurangan Metode	Alasan Cocok/Tidak Cocok	Referensi Terkait
			seperti *web-scraping*.	
Kerjasama dengan Penyedia *Hosting*	Dapat memberikan akses langsung ke *server* tempat aktivitas *carding* berlangsung.	Bergantung pada izin pihak ketiga dan bisa memakan waktu lama, akses ke data historis mungkin terbatas.	Tidak Cocok: Tergantung pada pihak ketiga dan akses bisa terbatas, sementara investigasi membutuhkan data segera dan seringkali dalam jumlah besar.	(Benhamou, 2017; Hidayat, 2020; Kalacska & Bouchard, 2011; Kopel, 2013)

1.2. Rumusan Masalah

Berdasarkan pada latar belakang yang telah dipaparkan maka, rumusan masalah pada penelitian ini adalah adalah masifnya aktivitas *cybercrime* pada *carding forum* dan *carding shop*. *Cybercrime* tersebut dapat ditangani secara pro-aktif melalui pengembangan *framework* investigasi forensik berbasis analisis dokumen hasil *web scraping* pada *carding forum* dan *carding shop*, sehingga *framework* ini dapat memberikan *insight* yang mendalam untuk tindakan pencegahan dan penanggulangan *cybercrime*.

1.3. Pertanyaan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan di atas, maka pertanyaan dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan teknik investigasi *web scraping* pada *carding forum* dan *carding shop*?
2. Bagaimana penerapan analisis deskriptif *profiling* forensik berdasarkan dokumen hasil *web scraping* pada *carding forum* dan *carding shop*?
3. Bagaimana penerapan analisis *natural language processing* berdasarkan dokumen hasil *web scraping* pada *carding forum* dan *carding shop*?
4. Bagaimana pengembangan dan penerapan *framework* investigasi forensik *carding* berdasarkan analisis dokumen hasil *web scraping* pada *carding forum* dan *carding shop*?

1.4. Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Objek analisis pada penelitian ini spesifik berfokus pada situs web *carding forum* dan *carding shop*.
2. Analisis dokumen hasil *web scraping* dilakukan menggunakan secara analisis deskriptif *profiling* forensik dan *natural language processing*.
3. Analisis *natural language processing* pada penelitian ini dilakukan menggunakan pendekatan *topic modelling:* algoritma *latent dirichlet allocation.*
4. Pengembangan *framework* investigasi forensik *carding* disusun berdasarkan sudut pandang ahli digital forensik.

1.5. Tujuan Penelitian

Penelitian ini bertujuan untuk mengembangkan *framework* investigasi forensik *carding* berdasarkan analisis dokumen hasil *web scraping* pada *carding forum* dan *carding shop*, yang mana menerapkan metode analisis *profiling* forensik dan *natural language processing* berbasis algoritma *latent dirichlet allocation*, sesuai *Alexiou Principle* untuk efektivitas dan efisiensi investigasi *cybercrime*.

1.6. Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah sebagai berikut:

1. Manfaat Teoritis: Penelitian ini diharapkan memperkaya referensi pada topik *machine learning for cyber forensic* khususnya pada aspek pengembangan *framework* investigasi forensik berbasis analisis dokumen hasil *web scraping* pada *carding forum* dan *carding shop*, yang mana analisisnya menerapkan metode analisis deskriptif *profiling* forensik dan *natural language processing* dengan algoritma *latent dirichlet allocation*.
2. Manfaat Praktis: Hasil penelitian ini diharapkan dapat membantu *stakeholder* penegakan hukum dan pembuat kebijakan dalam memahami tren *cybercrime;* serta mengambil keputusan yang tepat untuk tindakan pencegahan dan penanggulangannya.

1.7. Metodologi Penelitian Secara Umum

Metodologi penelitian ini disusun secara sistematis melalui beberapa tahapan, yaitu: perumusan tahapan penelitian, identifikasi masalah, kajian pustaka, serta penentuan objek berupa *carding forum* dan *carding shop*. Selanjutnya dilakukan analisis kebutuhan

informasi, pengembangan *framework* investigasi, pemahaman struktur web, dan pengumpulan data menggunakan *web scraping*. Data yang diperoleh diolah dengan analisis deskriptif profiling forensik dan LDA, diinterpretasikan, serta dibahas melalui validasi kualitatif triangulasi sebelum ditarik kesimpulan dan saran.

Analisis deskriptif *profiling* forensik adalah teknik untuk mengidentifikasi pola dan karakteristik penting dari data forensik melalui penggunaan basis kata kunci dan visualisasi data secara detail dalam investigasi kriminal atau keamanan digital (Maybir & Chapman, 2021; Muehlethaler & Albert, 2021). Sementara itu, *latent dirichlet allocation* (LDA) merupakan model probabilistik yang merepresentasikan setiap dokumen sebagai campuran beberapa topik tersembunyi, di mana tiap topik dibentuk oleh kumpulan kata tertentu (Blei et al., 2003; Zulhanif, 2016).

Penelitian ini menggunakan analisis deskriptif *profiling* forensik dengan bantuan *tool* Microsoft Excel dan Pandas untuk menyajikan visualisasi data serta infografis yang akurat, relevan, dan komunikatif, sehingga memudahkan *stakeholder* penegakan hukum dan pembuat kebijakan memahami tren *cybercrime* (Microsoft, 2024; Python Software Foundation, 2001; The pandas development team, 2020).

Selanjutnya, analisis *natural language processing* melalui *latent dirichlet allocation* (LDA) diterapkan pada dokumen hasil *web scraping* guna mengungkap topik tersembunyi, pola komunikasi, dan profil pelaku, yang mendukung investigasi serta pencegahan *cybercrime* secara lebih efektif. Peneliti menggunakan *software* Orange Data Mining untuk melakukan analisis *natural language processing* dengan pendekatan *latent dirichlet allocation* (Demšar et al., 2013).

1.8. Luaran Penelitan

Melalui publikasi di jurnal terakreditasi SINTA - Science and Technology Index, penelitian ini diharapkan memberikan kontribusi signifikan bagi berbagai kalangan. Bagi komunitas ilmiah, publikasi ini dapat berfungsi sebagai referensi penting dalam penelitian terkait kejahatan *carding*. Bagi analis digital forensik, hasil penelitian ini dapat diimplementasikan untuk memperkuat metode investigasi mereka. Selain itu, bagi peneliti dan institusi, publikasi ini dapat meningkatkan reputasi serta memperkokoh posisi mereka dalam bidang forensika digital.

1.9. Sistematika Penulisan

Penulisan laporan tesis terdiri dari beberapa bab yang masing-masing memiliki subbab dan pembahasan. Susunan dan sistematika penulisan tesis ini adalah sebagai berikut:

- BAB 1 Pendahuluan: Bab ini menguraikan tentang latar belakang, rumusan masalah, pertanyaan penelitian, batasan masalah, tujuan penelitian, manfaat penelitian, luaran penelitan, dan sistematika penulisan.
- BAB 2 Tinjauan Pustaka: Bab ini menguraikan mengenai beberapa teori yang digunakan dengan tambahan *review* dari penelitian terdahulu untuk mendukung penerapan penelitian.
- BAB 3 Metodologi Penelitian: Bab ini menguraikan tahapan-tahapan dalam pelaksanaan penelitian ini yang tahapan utamanya adalah tahapan pengumpulan data – *web scraping*; mengolah data; menginterpretasikan hasil pengolahan data; mengembangkan *framework* investigasi forensik *carding*; serta memberikan kesimpulan dan saran.
- BAB 4 Hasil dan Pembahasan: Bab ini membahas mengenai hasil analisis forensik serta interpretasinya. Bagian ini juga membahas mengenai *framework* investigasi forensik *carding* yang dikembangkan.
- BAB 5 Kesimpulan dan Saran: Bab ini membahas tentang kesimpulan yang merupakan temuan utama dari hasil penelitian yang dilakukan, dan juga berisi saran-saran terhadap penelitian selanjutnya.

BAB 2
Tinjauan Pustaka

2.1. Landasan Teori

2.1.1. *Web Scraping*

Web scraping, juga dikenal sebagai *web extraction* atau *harvesting*, adalah teknik untuk mengekstrak data dari *World Wide Web* (WWW) dan menyimpannya ke sistem file atau database untuk *retrieval* atau dianalisis setelahnya. Umumnya, data web di *scrapped* menggunakan *Hypertext Transfer Protocol* (HTTP) atau melalui *web browser*. Hal ini dilakukan baik secara manual oleh pengguna atau secara otomatis oleh bot atau *web crawler* (Zhao, 2017). Adapun faktanya bahwa sejumlah besar data heterogen secara kontinu di-*generate* di WWW, *web scraping* diakui secara luas sebagai teknik yang efektif dan efisien untuk mengumpulkan data besar (Bar-Ilan, 2001; Mooney et al., 2015). Teknik *web scraping* dapat digunakan untuk mengumpulkan data dari situs web secara otomatis (Guntara et al., 2024).

Pada penelitian ini teknik *web scraping* dilakukan dengan menggunakan *web scraping software*. Saat ini banyak *software* yang tersedia, yang dapat digunakan untuk solusi *customize web scraping*. *Software* ini mungkin mencoba mengenali struktur data suatu *web page* secara otomatis atau menyediakan *recording interface* yang menghilangkan kebutuhan untuk menulis kode *web scraping* secara manual, atau beberapa fungsi *scripting* yang dapat digunakan untuk mengekstrak dan mengubah konten, serta *database interfaces* yang dapat menyimpan data yang telah di *scraping* pada *local databases* (Saurkar et al., 2018).

Web Scraping Software adalah *tools* yang digunakan untuk mengotomatisasi pekerjaan *copy paste* manual untuk mengumpulkan data dalam jumlah besar dari situs web seperti situs direktori, situs *real estate*, situs web rahasia, dan situs lowongan pekerjaan. Misalnya, ketika ada kebutuhan untuk *scraping* rincian properti *real estate* di Inggris maka langkah berikutnya adalah dengan merekrut beberapa orang untuk pekerjaan *copy paste* manual atas rincian dari situs web ke excel dengan mengunjungi setiap *property page*. Dengan menggunakan cara ini, tentunya diperlukan waktu berhari-hari bahkan berbulan-bulan agar data properti tersebut siap digunakan. Jadi, *web scraping* dapat mengotomatisasi pekerjaan manual secara terprogram dengan mengunjungi setiap *page* dan mengekstrak data dari *page* dan *parsing html pages* (Sirisuriya, 2015).

Web scraping software berfungsi sebagai, bot atau *web crawler* yang mengakses data web secara langsung menggunakan *Hypertext Transfer Protocol*, atau melalui *web browser* dan mengekstrak data yang tepat dari *web page* tersebut. Data yang diekstraksi ini kemudian disimpan ke dalam *central local database* atau *spreadsheet* untuk digunakan atau dianalisis di kemudian waktu (Saurkar et al., 2018). Pada penelitian ini *web scraping* digunakan sebagai teknik untuk investigasi forensik tren aktivitas *cybercrime* di *website carding forum* dan *carding shop* (@cashout vendors, 2020b; Altenen, n.d.; Astradumps, 2023; Invision Community, n.d.). Adapun *web scraping software* yang digunakan dalam penelitian ini adalah WebHarvy Version 7.3.0.222 (SysNucleus, 2024).

2.1.2. Kejahatan *Carding*

Dalam arti sempit, istilah *"carding"* mengacu pada penggunaan informasi rekening kartu kredit dan debit secara tidak sah untuk membeli barang dan jasa secara curang. Dalam arti yang lebih luas *carding* dapat didefinisikan sebagai proses penipuan berupa pencurian, penjualan kembali, dan pada akhirnya menggunakan informasi pembayaran dalam jumlah besar untuk melakukan penipuan (Li et al., 2016; Peretti, 2009).

Istilah *carding* ini telah berkembang dalam beberapa tahun terakhir, namun mencakup berbagai kegiatan seputar pencurian dan penipuan nomor rekening kartu kredit dan debit termasuk peretasan komputer, *phishing*, pencairan nomor rekening yang dicuri, skema pengiriman ulang, dan penipuan lelang melalui internet. Individu yang terlibat dalam kegiatan kriminal *carding* disebut sebagai *"carders"* (Peretti, 2009). Para *carder* tentunya memiliki media yang dapat diandalkan untuk melancarkan kegiatan *carding* diantaranya adalah *carding forum* dan *carding shop*. Penelitian ini menggunakan teknik *web scraping* untuk menginvestigasi tren aktivitas *cybercrime* pada keseluruhan *website carding forum* dan *carding shop*.

Carding forum dapat didefinisikan sebagai suatu forum di Internet yang digunakan oleh para *carder* untuk *sharing* informasi, *tutorial*, dan transaksi *illegitimate* mengenai kartu kredit. Pada forum ini, para *carder sharing* informasi tentang cara: mendapatkan kartu kredit ilegal, menggunakan *tools carding*, serta mengenai transaksi jual-beli kartu kredit ilegal. Forum ini juga menjadi tempat untuk berbagi tips dan tautan yang berkaitan dengan kegiatan *carding*. Selain menyediakan forum perdagangan *online* atas informasi akun yang dicuri, *carding forum* juga menyediakan forum perdagangan berbagai dokumen identitas palsu. Disamping itu, ketika para *carder* bertemu melalui forum ini, sering kali para *carder* tersebut

bergabung bersama dalam melakukan penipuan keuangan atau aktivitas kriminal tertentu (Peretti, 2009).

Adapun *carding shop* dapat didefinisikan sebagai *platform* daring menjual data kartu kredit dan/atau akses ilegal. Menurut Benjamin et al. (2015) *carding shop* adalah bagian penting lainnya dari komunitas *hacker* dan *cybercriminal* global. *Carding shop* membantu memfasilitasi kejahatan *cyber carding* karena *carding shop* menyediakan rantai pasokan bagi *carder* yang ingin menjual kartu curian. Pada penelitian ini, situs web *carding forum* dan *carding shop* yang dijadikan objek penelitian adalah: (1) Altenen Forums-Images & Videos & Porn Accounts 18 Section; (2) carding.store-Cracking Tutorials Section; (3) Astradumps Shop; dan (4) Money-Heist.org Shop (@cashout vendors, 2020b; Altenen, n.d.; Astradumps, 2023; Invision Community, n.d.).

2.1.3. *Natural Language Processing*

Menurut Liddy (2001) *natural language processing* adalah serangkaian teknik komputasi yang dimotivasi secara teoritis untuk menganalisis dan merepresentasikan teks yang terjadi secara alami pada satu atau lebih tingkat analisis linguistik untuk tujuan mencapai pemrosesan bahasa mirip manusia untuk berbagai tugas atau aplikasi. Tujuan NLP sebagaimana dinyatakan di atas adalah "untuk mencapai pemrosesan bahasa yang mirip manusia". Pemilihan kata 'memproses sangat disengaja, dan tidak boleh diganti dengan 'pemahaman'. Meskipun bidang NLP awalnya disebut sebagai *natural language understanding* (NLU) pada masa awal AI, namun hari ini dapat disepakati bahwa meskipun tujuan NLP adalah NLU yang sebenarnya, tujuan tersebut belum tercapai. Sistem NLU yang lengkap akan mampu (Liddy, 2001):

- Parafrase teks masukan
- Menerjemahkan teks ke dalam bahasa lain
- Menjawab pertanyaan tentang isi teks
- Menarik kesimpulan dari teks tersebut

Meskipun NLP telah membuat terobosan serius dalam mencapai tujuan 1 hingga 3, fakta bahwa sistem NLP tidak dapat menarik kesimpulan dari teks dengan sendirinya, NLU masih tetap menjadi tujuan NLP (Liddy, 2001). *Natural language processing* (NLP) adalah kumpulan teknik komputasi untuk analisis otomatis dan representasi bahasa manusia, yang dimotivasi oleh teori. Namun, analisis teks secara otomatis, setara dengan manusia, memerlukan pemahaman yang lebih mendalam tentang bahasa alami oleh mesin, dan hal ini

masih jauh dari kenyataan. Ada banyak contoh NLP, seperti pengambilan informasi *online*, agregasi, dan tanya jawab, yang sebagian besar didasarkan pada algoritma yang mengandalkan representasi tekstual halaman web, serta NLP sampai batas tertentu. Algoritma seperti ini sangat baik dalam mengambil teks (IR), membaginya menjadi beberapa bagian, memeriksa ejaan, dan analisis tingkat kata, namun tidak berhasil untuk analisis pada tingkat kalimat dan paragraf. Oleh karena itu, ketika menyangkut pertanyaan tentang menafsirkan kalimat dan mengekstraksi informasi yang bermakna, kemampuan algoritma ini masih sangat terbatas (Chowdhary, 2020). NLP secara umum memerlukan kemampuan simbolik tingkat tinggi, yang meliputi hal-hal berikut (Chowdhary, 2020):

- Akses dan perolehan karakteristik leksikal, semantik, dan episodik,
- Penciptaan dan penyebaran ikatan dinamis,
- Manipulasi struktur rekursif konstituen,
- Koordinasi banyak modul pemrosesan dan pembelajaran,
- Identifikasi konstruksi bahasa dasar (misalnya objek dan tindakan) dan,
- Representasi konsep abstrak

Natural language processing (NLP) berada di tengah-tengah antara linguistik komputasi ilmu komputer, dan didedikasikan untuk konversi bahasa alami manusia secara tertulis dan lisan menjadi *structured mineable data*. Melalui kombinasi metode linguistik, statistik, dan AI, NLP dapat digunakan untuk menentukan makna sebuah teks atau bahkan untuk menghasilkan respons mirip manusia. NLP sudah menjadi bagian dari kehidupan kita sehari-hari karena diterapkan secara luas di *software* komputer atau di ponsel kita (Fanni et al., 2023). Pendekatan NLP yang digunakan dalam penelitian ini adalah deteksi bahasa yang berbasis algoritma *latent dirichlet allocation* (LDA).

2.1.4. *Latent Dirichlet Allocation*

Latent dirichlet allocation (LDA) adalah model probabilistik generatif dari sebuah korpus. Ide dasarnya adalah bahwa dokumen direpresentasikan sebagai campuran acak atas topik-topik laten, di mana setiap topik dicirikan oleh distribusi kata-kata. LDA mengasumsikan proses generatif berikut untuk setiap dokumen **w** di korpus *D* (Blei et al., 2003):

1. Memilih $N \sim$ Poisson (ξ).
2. Memilih $\theta \sim$ Dir (α).
3. Untuk masing-masing dari N kata w_n:
   (a) Pilih sebuah topik $z_n \sim$ Multinomial (θ).

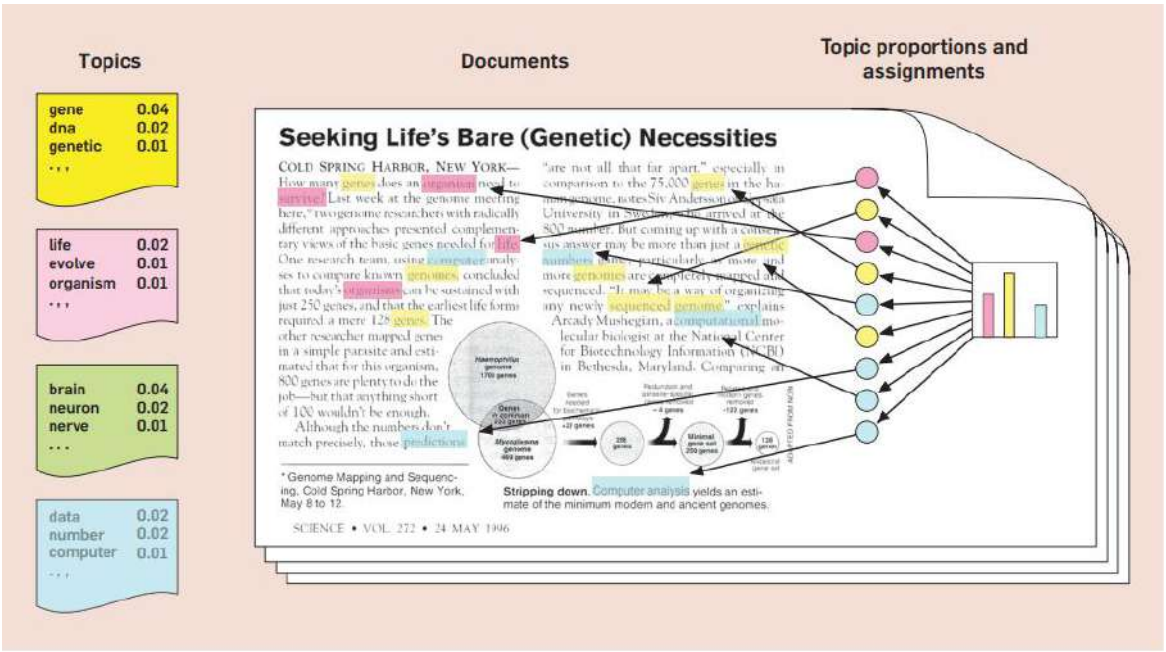
(b) Pilih sebuah kata w_n dari $p(w_n|z_n,\beta)$, probabilitas multinomial yang dikondisikan pada topik z_n (Blei et al., 2003).

LDA adalah model statistik kumpulan dokumen yang mencoba menangkap intuisi ini. Hal ini paling mudah dijelaskan melalui proses generatifnya, proses acak imajiner yang digunakan model untuk mengasumsikan munculnya dokumen. Secara formal dapat definisikan suatu topik sebagai distribusi atas kosakata tetap. Misalnya, topik genetika memuat kata-kata tentang genetika dengan probabilitas tinggi dan topik biologi evolusioner memuat kata-kata tentang biologi evolusioner dengan probabilitas tinggi. Oleh karenanya dapat diasumsikan bahwa topik ini ditentukan sebelum data apa pun dihasilkan. Lalu untuk setiap dokumen dalam koleksi, dapat dihasilkan kata-kata dalam proses dua tahap sebagai berikut (Blei, 2012):

- Pilih distribusi berdasarkan topik secara acak.
- Untuk setiap kata dalam dokumen

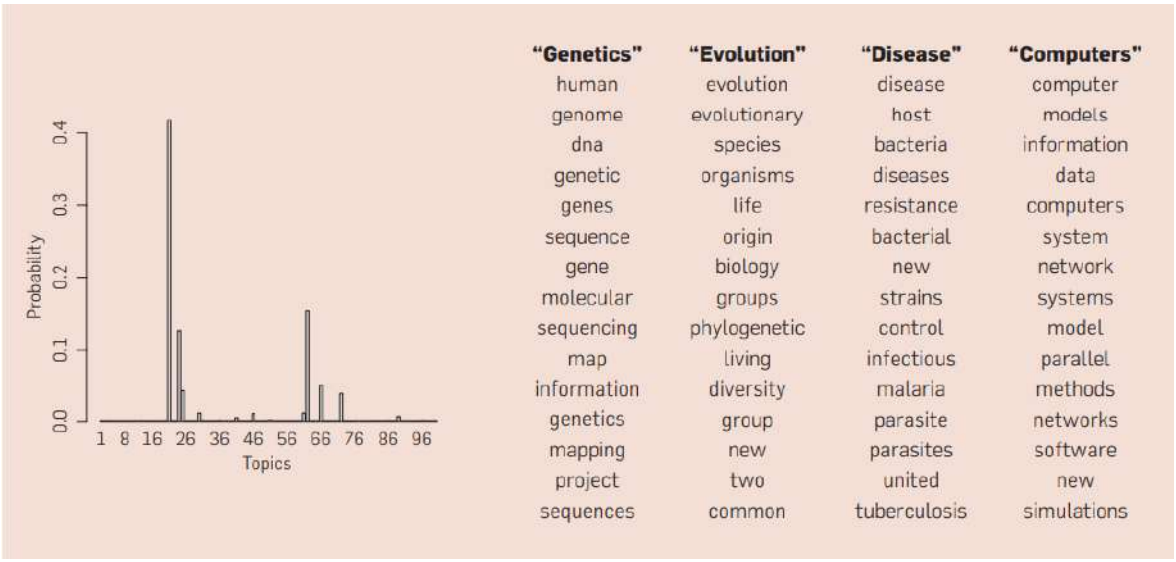
A. Pilih topik secara acak dari distribusi topik pada langkah #1.
B. Pilih secara acak sebuah kata dari distribusi kosakata yang sesuai.

Model statistik ini mencerminkan intuisi bahwa dokumen menunjukkan banyak topik. Setiap dokumen menampilkan topik dalam proporsi berbeda (langkah #1); setiap kata dalam setiap dokumen diambil dari salah satu topik (langkah #2b), di mana topik yang dipilih dipilih dari distribusi topik per dokumen (langkah #2a) (Blei, 2012). Berikut dibawah ini disajikan gambar 1.2 dan 1.3 terkait gambaran konsep dan model LDA:



Gambar 2.1 Intuisi dibalik *Latent Dirichlet Allocation* (LDA)

Sumber: Blei (2012)



Gambar 2.2 *Real Inference* dengan *Latent Dirichlet Allocation* (LDA)

Sumber: Blei (2012)

2.1.5. *Alexiou Principle*

Alexiou Principle (dinamai menurut penciptanya Mike Alexiou dari *IT infrastructure services provider* Terremark Worldwide, Inc.) (Greiner, 2009) memberikan pertanyaan mendasar, yang mana dapat digunakan oleh *digital forensic investigator* sebagai panduan dan arahan dalam melakukan *searching* dan manajemen investigasi. Pertanyaan tersebut adalah:

- *What question are you trying to answer?*
- *What data do you need to answer that question?*
- *How do you extract that data?*
- *What does the data tell you?* (Pogue, 2010)

Setiap elemen dari empat pertanyaan tersebut memiliki peran penting dalam penyusunan rencana investigasi, yang mencakup penjabaran tujuan investigasi. Tujuan ini sangat penting, karena tanpa mengetahui tujuan yang jelas maka, tidak mungkin investigator dapat menentukan apa yang harus dicari dalam proses investigasi. Rencana ini juga mendeskripsikan definisi keberhasilan investigasi dan memastikan adanya kesepahaman antara investigator digital forensik dan klien (organisasi yang menjadi korban pelanggaran keamanan) (Greiner, 2009).

2.2.*Review* Penelitian Terdahulu

Berikut dibawah ini disajikan tabel *review* peneltian terdahulu yang berkaitan dengan *web scraping forensic, web crawling forensic, profiling forensic, machine learning,* dan *latent dirichlet allocation* (LDA):

Tabel 2.1 *Literature Review* Kategori Investigasi Forensik Berbasis *Web Scraping*

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
1	Muehlethaler & Albert (2021)	Domain penelitian ini adalah pengumpulan data tekstil menggunakan *tools web crawling* dan *web scraping* untuk penelitian populasi serat dalam bidang forensik.	1) Penelitian ini menggunakan *web crawler* dan *web scraper* untuk mengumpulkan data dari situs web retailer pakaian secara otomatis. 2) *Web crawler* dikembangkan sendiri berdasarkan *open source tools* Anaconda Navigator 1.9.7 *(Python data science platform)*, Scrapy 1.5.2 *(web spider for crawling websites)* dan Elasticsearch 6.7.0/Kibana 6.7.0 *(monitoring of collected data)*.	1) Melalui *web crawling* dan *web scraping*, peneliti telah mengekstraksi 68 bidang berbasis teks yang menggambarkan 24.701 pakaian dari situs web *online retaliler* besar. Data ini diekstraksi dengan kecepatan sekitar 1000 *piece* pakaian per jam dan dikelola melalui *interface* Kibana/Elasticsearch. 2) Hasil penelitian menunjukkan bahwa kapas, poliester, *viscose*, dan *elastane* adalah 4 jenis serat utama yang digunakan dalam industri tekstil, dengan dominasi kapas dan poliester. *Elastane*, meskipun sangat populer dalam garmen (muncul di 32% garmen), jarang menyumbang lebih dari 10% massa sedangkan kapas sering kali melebihi 80%. *Weighted frequencies of apparition* benda tersebut masing-masing adalah 1,61% untuk *elastane* dan 49,24% untuk kapas. Statistik warna

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
				juga menunjukkan dominasi warna hitam, biru, dan putih, dengan ketergantungan frekuensi pada jenis serat. 3) Melalui penelitian ini, telah ditunjukkan kelayakan, kegunaan, dan kemudahan yang digunakan robot *web scraping* dalam menyediakan statistik waktu nyata mengenai populasi serat untuk membantu praktisi forensik dalam pekerjaan rutinnya.
2	Maybir & Chapman (2021)	Domain penelitian ini adalah pemanfaatan teknik *open-source intelligence* (OSINT) *web scraping* untuk memantau tren narkoba secara *real-time* dengan menganalisis data yang dilaporkan pengguna mengenai pil ekstasi dari situs web Pill	1) Penelitian ini memanfaatkan data yang diambil dari situs Pill Reports. Pill Reports merupakan basis data global dimana pengguna ekstasi melaporkan pil ekstasi berdasarkan analisis ilmiah (yaitu pengujian di tempat) serta pengalaman/gejala subjektif pengguna saat berada di bawah pengaruh narkoba	1) Data ekstasi yang bersifat *online*, *open-source*, dan dilaporkan oleh pengguna terbukti berguna untuk mengembangkan informasi tentang perdagangan narkoba di Australia dan kemungkinan besar akan terbukti bermanfaat di wilayah lain. Ringkasan data umum ini memberikan *insight* tentang jenis pil ekstasi yang beredar di dalam negeri selama 15 tahun terakhir dan sesuai dengan data yang ada dari *wastewater analysis* dan *population surveys*. Informasi ringkasan tambahan mengidentifikasi wilayah dan pengguna yang bertanggung jawab

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
		Reports untuk memerangi pasar narkoba.	tersebut. Semua data diekstraksi melalui pengumpulan pasif menggunakan *visual web scraping tool* yang didukung AI, ScrapeStorm, Versi 3.5.0. Data diambil pada 13 Mei 2020; laporan apa pun yang tercantum setelah tanggal ini tidak dimasukkan dalam analisis selanjutnya. 2) Kumpulan data laporan pengguna pil yang diekstraksi dianalisis menggunakan Microsoft Excel® (Versi 16.38). Pill Reports adalah *open-fill reporting system* dimana pengguna dapat melaporkan informasi dalam bentuk bebas tanpa pedoman	atas jumlah *listing* terbanyak dan tren temporal umum yang terkait dengan Pill Reports. 2) Tidak ada alur arus antar negara bagian yang signifikan (yaitu gerakan peredaran narkoba) yang dapat diidentifikasi menggunakan data yang dilaporkan pengguna, namun, terdapat bukti yang diidentifikasi yang menunjukkan bahwa bubuk MDMA atau pil ekstasi dipres diimpor ke Australia melalui NSW atau VIC. Hasil yang paling signifikan adalah penentuan periode jeda rata-rata dari timur ke barat yang menunjukkan bahwa kumpulan pil ekstasi yang dilaporkan di pantai timur membutuhkan waktu sekitar satu tahun untuk tiba di pantai barat. Secara keseluruhan, OSINT memberikan pendekatan berbasis lebih strategis dan dapat digunakan untuk memantau tren narkoba secara *real-time*.

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
			pelaporan ketat atau kategori yang telah ditentukan. Akibatnya, data dilaporkan dalam berbagai cara yang unik (yaitu istilah sinonim, singkatan, tambahan deskriptor awalan, kesalahan ejaan, dll.) meskipun semuanya mengacu pada fitur yang sama. 3) Empat dari enam variabel (lokasi, logo, warna dan bentuk) disaring secara manual dan diberi kode ulang untuk memastikan bahwa karakteristik yang sesuai dapat diidentifikasi menggunakan satu nilai alfanumerik unik sebelum analisis (yaitu istilah seperti Victoria, VIC, Melb, Melbourne semuanya mewakili	

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
			wilayah di negara bagian Victoria, sehingga semuanya diberi kode "VIC"). Fungsi excel '= COUNTIF' digunakan untuk menentukan jumlah laporan pada setiap variasi unik variabel tertentu (yaitu berapa kali pil dilaporkan berwarna hijau versus merah).	
3	Ghugare et al. (2024)	Domain penelitian ini terdapat pada tantangan etika dan praktis yang terkait dengan pelacakan aktivitas terlarang dalam *blockchain* Bitcoin. Penelitian ini menyoroti kesulitan yang dihadapi penegak hukum dalam meneliti transaksi	1) Penelitian ini menggunakan *web scraping* untuk mengumpulkan data dari berbagai *platform online*, forum, dan pasar untuk mengidentifikasi alamat Bitcoin yang terkait dengan aktivitas terlarang. 2) Pertimbangan etis mengenai pengumpulan data dari *blockchain* juga dieksplorasi,	1) Temuan penelitian ini tidak mengungkapkan aspek transaksi mata uang kripto namun juga mengusulkan metodologi yang kuat. Metodologi ini dapat dimanfaatkan oleh para pemangku kepentingan, termasuk badan-badan, lembaga penegak hukum, dan komunitas mata uang kripto yang lebih luas. Hal ini akan membantu dalam memantau, mengatur dan membina ekosistem yang sah, untuk *cryptocurrency*. 2) Penelitian ini menekankan perlunya upaya untuk menetapkan serangkaian peraturan yang tidak

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
		Bitcoin dan mengusulkan penggunaan *web scraping* untuk mengidentifikasi alamat yang terlibat dalam aktivitas penipuan. Selain itu, penelitian ini mengeksplorasi pertimbangan etis dalam pengumpulan data dan menekankan perlunya peraturan untuk mendorong ekosistem mata uang kripto yang sah.	dengan menekankan perlunya metodologi yang menghormati privasi.	hanya mencegah aktivitas ilegal tetapi juga mendorong integrasi mata uang kripto ke dalam sektor keuangan arus utama.
4	Narasimhan et al. (2023)	Domain penelitian ini adalah investigasi OSINT di Facebook dengan pengumpulan	1) Peneltian ini menerapkan metode *web scraping* yang mana objek *scraping* utamanya yaitu: (1) *Content scraping,* (2)	1) Penelitian ini menyimpulkan dan menjelaskan semua alat yang tersedia untuk mengumpulkan semua data berharga dan informasi penting dari *platform* Facebook.

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
		datanya dilakukan menggunakan *web scraping* dan *optical character recognition* (OCR).	*price scraping,* dan (3) *contact scraping*. 2) *Tools* yang digunakan untuk *scraping* adalah Selenium, APIFY, OSINT Combine, Sowsearch, LookupID, Whopostedwhat, Nairaland, dan Facepager. 3) Penelitian ini juga menggunakan *optical character recognition* (OCR) dengan *small open-source algorithm* Tesseract untuk membaca teks karakter pada foto hasil *scraping* dari Facebook.	2) Contoh praktis kecil telah didemonstrasikan oleh peneliti, seperti mengambil foto iklan dari *career opportunities group* melalui Selenium dengan Python dan mendownloadnya ke folder *local*. 3) Kemudian peneliti menerapkan, *small open-source algorithm* Tesseract digunakan pada gambar yang di-*scraped* dan di-*download* untuk mendeteksi kalimat yang tertulis pada gambar dan menambahkannya ke file teks. 4) Penelitian ini adalah beberapa investigasi *open-source intelligence* seperti berapa banyak iklan yang memiliki semua informasi berharga dan tepat yang dibutuhkan untuk Rekrutmen Pekerjaan yang tertulis di poster.

Tabel 2.2 *Literature Review* Kategori Investigasi Forensik Berbasis *Machine Learning* & *Latent Dirichlet Allocation* (LDA)

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
1	Bergman & Popov (2022)	Domain penelitian ini berfokus pada pemanfaatan teknik *machine learning* untuk menganalisis kumpulan data besar terkait kejahatan dunia maya dan forensik digital. Hal ini melibatkan pengembangan alat untuk membuat anotasi dan menyimpan konten *dark web* tertentu termasuk diantaranya *web scraping*, sehingga memungkinkan kolaborasi antar lembaga penegak hukum.	1) Metode yang diterapkan pada penelitian terdiri dari lima kegiatan (atau tahapan) yaitu: (1) Penjelasan Masalah, (2) Pemunculan Persyaratan, (3) Pengembangan Artefak, (4) Demonstrasi Artefak, dan (5) Evaluasi Artefak. 2) Guna mewujudkan kesesuaian tujuan perangkat, peneliti menggunakan kumpulan datanya sebagai data pelatihan untuk model klasifikasi berbasis *machine learning*. Teknik *five cross-fold validation* digunakan untuk mengevaluasi *classifiers*.	1) Penelitian ini menyajikan alat anotasi konten web yang dikembangkan sebagai *plugin* Tor Browser (dan Firefox), yang mendukung investigator lembaga penegak hukum yang menangani *cybercrime*. *Plugin* ini berfungsi sebagai anotasi konten web manual, kategorisasi, dan alat pengumpulan yang berfungsi sebagai *input* untuk *supervised machine learning algorithms*. 2) *Algorithms learn* mengklasifikasikan konten web yang relevan dengan investigasi *cybercrime* dari kumpulan konten web yang dibuat *user*. Alat anotasi secara otomatis menyinkronkan anotasi dan konten *web page* dengan server pusat, memungkinkan kolega kerja untuk berbagi materi satu sama lain dan memperluas *database*.

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
2	U. Agarwal et al. (2024)	Domain penelitian ini berfokus pada deteksi dan pencegahan penipuan mata uang kripto, mengatasi meningkatnya insiden penipuan, *phishing*, dan aktivitas kriminal dalam lanskap mata uang kripto yang terdesentralisasi. Penelitian ini menggunakan kombinasi AI, *blockchain*, dan *machine learning* untuk mengembangkan sistem deteksi penipuan yang efektif, menekankan integrasi teknologi	1) Penelitian ini menggunakan pendekatan komprehensif yang menggabungkan pengembangan taksonomi dan desain arsitektur. 2) Peneliti melakukan evaluasi empiris menggunakan *machine learning* (ML) *classification algorithms*, khususnya *random forest* (RF), untuk menilai efektivitas. 3) Data deteksi penipuan mata uang kripto disimpan dengan aman menggunakan teknologi blockchain dan IPFS.	1) Penelitian ini mencapai hasil yang signifikan. Arsitektur yang diusulkan secara efektif mendeteksi penipuan mata uang kripto, dengan *random forest classifier* menunjukkan akurasi tertinggi sebesar 97,5%. 2) *Detected fraud instances* disimpan dengan aman menggunakan teknologi *blockchain* dan IPFS, sehingga memungkinkan akses penegakan hukum. 3) Namun, tantangan seperti masalah privasi dan kesulitan membedakan penipuan di antara berbagai mata uang kripto telah teridentifikasi, sehingga menekankan perlunya penelitian lebih lanjut dalam *blockchain* dan forensik kripto.

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
		untuk penyimpanan data yang aman dan akses penegakan hukum.		
3	Sonmez & Codal (2024)	Domain penelitian ini melibatkan eksplorasi aktivitas kriminal di *dark web*, khususnya berfokus pada diskusi terkait terorisme. Dengan menggunakan pendekatan pemodelan topik berbasis LDA, bertujuan untuk mengidentifikasi topik yang dibahas, terutama berfokus pada komunikasi jihadis untuk mendeteksi perilaku yang *abnormal* dan terkait terorisme.	1) Artikel ini menggunakan pendekatan pemodelan topik berbasis LDA untuk mengidentifikasi topik yang dibahas dalam diskusi di *dark web*. 2) Tujuan utamanya adalah untuk menyajikan gambaran komunikasi para jihadis di *cyberspace* untuk mendeteksi perilaku yang tidak biasa atau tujuan terkait terorisme.	1) Penelitian ini mengidentifikasi topik-topik utama dalam komunikasi jihadis di *dark web*, terutama berfokus pada perekrutan dan propaganda. 2) Berdasarkan temuan, percakapan dalam konteks rekrutmen dan propaganda mendominasi forum tersebut. 3) Meskipun tidak ditemukan bukti langsung adanya kolaborasi teroris, kehadiran alat propaganda dan rekrutmen menyoroti risiko aktivitas teroris yang sedang berlangsung di *platform dark web* tersebut.

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
4	Gong et al. (2025)	Domain penelitian ini berada pada persimpangan antara keamanan siber, *victimology digital,* dan pemetaan spasial, dengan fokus khusus pada penipuan lowongan kerja *(employment scams)* dalam ruang *hybrid* (gabungan fisik dan virtual), menggunakan pendekatan *artificial intelligence* untuk mendeteksi, mengklasifikasi, dan mengurangi risiko korban.	Penelitian ini menggunakan pendekatan berbasis *artificial intelligence* (AI) untuk: 1) Menganalisis konsistensi spasial antara informasi lokasi fisik dan virtual dalam lowongan kerja. 2) Mengembangkan mekanisme deteksi dan mitigasi penipuan kerja dengan mempertimbangkan *hybrid space* (kombinasi lokasi fisik dan deskripsi *online*). 3) Melibatkan analisis spasial dan klasifikasi berbasis AI untuk membedakan antara lowongan palsu dan asli.	1) Hasil penelitian menunjukkan bahwa konsistensi informasi geografis dalam *hybrid space* lowongan kerja palsu lebih rendah daripada lowongan kerja yang sah, dan terdapat heterogenitas spasial dalam distribusi lokasi fisik lowongan kerja palsu. 2) Konsistensi ini, serta lokasi fisik yang terperinci, berkontribusi signifikan terhadap identifikasi dan klasifikasi lowongan kerja asli dan palsu. 3) Dengan mengintegrasikan berbagai disiplin ilmu, penelitian ini meningkatkan pemahaman tentang prevalensi, dampak, faktor-faktor yang berkontribusi, dan strategi mitigasi yang terkait dengan viktimisasi siber selama masa kerja. 4) Penelitian ini juga berkontribusi pada pengembangan metodologi dan pendekatan baru untuk mendeteksi, memitigasi, dan mencegah kejahatan siber.

Tabel 2.3 *Literature Review* Kategori Investigasi Forensik Berbasis *Web Crawling & Machine Learning*

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
1	Jin et al. (2024)	Domain penelitian ini berfokus pada tantangan pelacakan pelaku kriminal dan aktivitas ilegal di *dark web*. Laporan ini menyoroti isu-isu seperti sifat *dark web* yang selalu berubah, dampak kesehatan mental pada investigator karena konten yang mengganggu, dan anonimitas yang menghalangi identifikasi kriminal. penelitian ini memperkenalkan model *advanced crawler* dan	1) Penelitian ini mengembangkan *advanced crawler* untuk mengumpulkan data dari *dark web*, mengatasi tantangan seperti perubahan domain yang sering terjadi dan konten yang mengganggu. 2) Model *machine learning* diterapkan untuk mendeteksi konten semacam itu, sehingga melindungi kesehatan mental investigator. 3) Tiga studi kasus memvalidasi metodologi penelitian yang diusulkan.	1) Penelitian ini berhasil dalam penggunaan *advanced crawler* untuk mengumpulkan data dari *dark web*, mengatasi tantangan seperti seringnya perubahan domain dan konten yang mengganggu. 2) Model *machine learning* secara efektif mendeteksi konten *dark web* yang menggangu, sehingga menjaga kesehatan mental investigator. 3) Dalam artikel ini, keadaan *dark web* saat diperkenalkan dengan menganalisis 14.993 *crawled dark websites*. Dengan menyajikan tiga studi kasus, terbukti bahwa metodologi investigasi yang peneliti usulkan dapat mengidentifikasi operator *illegal dark websites* dengan menghubungkan *dark web* dengan *surface websites* yang sesuai.

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
		machine learning yang untuk mengatasi tantangan ini, dan menunjukkan efektivitasnya melalui tiga studi kasus.		
2	Bergman & Popov (2023)	Domain penelitian ini mengkaji tantangan yang dihadapi penegakan hukum akibat enkripsi dan anonimitas dalam kejahatan dunia maya. *Systematic literature review*-nya menganalisis penelitian yang ada tentang *dark web crawlers*, prevalensi,	1) Metode penelitian ini melibatkan melakukan *systematic literature review* untuk mengkaji penelitian yang ada tentang *dark web crawling*. Dari 58 *peer-reviewed articles* yang menyebutkan *crawling* dan *dark web*, 34 artikel tetap diputuskan dikaji setelah *excluding* artikel yang tidak relevan. 2) Langkah pertama, (1) perencanaan, mencakup persiapan *systematic literature*	1) Pengetahuan yang dikumpulkan dari *systematic literature review* digunakan untuk mengembangkan *Tor-based web crawling model* menjadi *software* yang sudah ada dan *customised* untuk investigasi berbasis ACN. 2) Dari pengetahuan ini, *dark web crawler* dikembangkan agar sesuai dengan *dark web cybercrime toolset* yang sudah ada sebelumnya yang disebut D3. 3) Dikombinasikan dengan *annotation-based machine learning classifier* di perangkat D3, *crawler* yang dikembangkan dan disajikan dalam penelitian ini akan memberdayakan perangkat untuk secara otomatis mengumpulkan dan

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
		karakteristik, dan metodologinya.	*review*: membuat sketsa latar belakang penelitian, pertanyaan penelitian, pemilihan penelitian dan kriteria penilaian kualitas penelitian, serta ekstraksi data dan strategi diseminasi. 3) Tahap kedua (2), melakukan *literature review*. Tahap kedua meliputi kegiatan: (1) seleksi penelitian, (2) penilaian kualitas penelitian, (3) ekstraksi data, dan (4) sintesis data. Setiap kegiatan disajikan dalam empat bagian mendatang. 4) Tahap ketiga, (3) pelaporan, mencakup penetapan mekanisme diseminasi serta format dan evaluasi laporan.	mengklasifikasikan konten web berdasarkan *web page* yang diberi anotasi sebelumnya. 4) Pada akhirnya, hal ini akan menghemat tenaga kerja manual bagi *cybercrime investigators* dalam memeriksa konten web dalam jumlah besar tanpa kehilangan kendali atas proses perayapan. 5) Terakhir, performa model diperiksa melalui serangkaian eksperimen. Hasilnya menunjukkan bahwa *crawler* yang dikembangkan berhasil mengambil konten web dari *clear web page* dan *dark web page*, serta *scraping dark marketplaces* di jaringan Tor.

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
			Hal ini tersirat karena sifat laporan ini, yang pada dasarnya merupakan artikel penelitian *peer-reviewed* yang menjalani evaluasi dan disebarluaskan ke publik.	
3	Rao et al. (2021)	Domain penelitian ini berfokus pada pembuatan *framework* terpadu untuk *web scraping* dan *text analysis*, memungkinkan analisis *skalabel* terhadap kumpulan *dataset* besar, *tracking* perubahan dari waktu ke waktu, dan memfasilitasi interpretasi teks berbasis web dengan	Pada penelitian ini *framework* yang diusulkan terdiri dari langkah-langkah sebagai berikut: 1) *Pre-Processing: Cleaning* HTML dengan menghapus tag. Ekstrak teks yang relevan. Pre-process: *remove stop words, lowercase, lemmatize,* dan *stem for analysis*. 2) *Working of the Web-Scraper: Library* scraper web pada penelitian ini dibuat menggunakan bahasa pemrogramman C, soket Linux	1) *Web Scraping* dilakukan dengan menghubungkan ke *server* situs web, membuat koneksi SSL, dan men-*download web page*. Algoritma pencarian sub string digunakan untuk mengisolasi elemen yang dituju. 2) *Preliminary analysis* diperlukan sebelum *scraping* situs web. Bergantung pada aplikasinya, model harus di-*training* pada kumpulan data yang diperlukan untuk melakukan analisis spesifik seperti analisis sentimen dan *cosine similarity*. 3) Pada analisis sentimen, peneliti menggunakan *linear classifiers* seperti regresi linier. Dengan menggunakan *tools* tersebut, peneliti telah menulis

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
		mudah, khususnya *profiling* untuk ulasan produk, hiburan, dan berita.	dan OpenSSL. Hal itu memerlukan request IP situs web via DNS, membuat koneksi TCP, men-*downlad* HTML via SSL, menutup koneksi. 3) *Text Summarization:* Prosesnya melibatkan cleaning dan *pre-processing* teks HTML. Kemudian dibagi menjadi kalimat dan kata, dan frekuensi *term* juga dihitung. Kalimat diberi skor yang sesuai, dengan mempertimbangkan frekuensi kata dan panjang yang dinormalisasi. Terakhir, *summary sentences* dipilih berdasarkan kriteria tertentu.	dan telah mengkonstruksi Google news scraper untuk menganalisis laporan berita. 4) Peneliti telah menjalankan *tool* pada pencarian 'Dell XPS i3 Laptop Reviews'. Berdasarkan sentimen dari hasil tersebut, peneliti memperoleh persentase artikel positif sebesar 80.

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
			4) *Sentiment Analysis:* Dokumen melalui proses *cleaning* dan *pre-processing*, penghapusan tag HTML dan tanda baca. *Stop words* dihilangkan, dan kata-kata diberi *lemmatized*. Melalui pengkodean *one-hot*, dokumen dikonversi menjadi vektor. Peneliti telah melatih model regresi linier pada kumpulan data setelah pengkodean one-hot kumpulan data tersebut. 5) *TF-IDF Scoring:* Bergantung pada frekuensi *term* setiap kata dalam dokumen dan jumlah dokumen yang diambil oleh *scraper*, lalu dihitung untuk skor TF-IDF untuk dokumen tersebut.	

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
			6) *Topic Modeling:* Pada bagian ini peneliti melakukan *topic modeling* menggunakan *latent semantic analysis* (LSA). LSA cocok dengan semua kata dalam dokumen setelah *pre-processing* menjadi Model *Bag of Words*, hingga pada akhirnya peneliti kemudian memilih kata dengan frekuensi *term* tertinggi dalam topik tersebut sebagai nama topik. Pada tahap *experimentation* Teknologi Python digunakan. Teknologi Python: Penerjemah adalah jenis program pemrograman yang mengeksekusi proyek pilihan. Saat peneliti membuat program Python, *source code* yang disusun dengan panduan insinyur diubah	

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
			menjadi *middle language* yang cukup diubah menjadi local *language/contraption language* dengan harapan dapat dieksekusi. Eksperimen dilakukan menggunakan prosesor corei3 3 Ghz menggunakan Anaconda (Python 3.8), Linux Shell/Windows Command-Prompt, dan python-requests, pip versi 20.0.3.	

Tabel 2.4 *Literature Review* Kategori Investigasi Forensik Berbasis *Web Scraping*, *Profiling* Forensik & *Machine Learning*

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
1	Nurseno et al. (2024)	Domain Penelitian ini mengkaji website dengan domain go.id yang telah disusupi dengan URL tersembunyi yang berafiliasi dengan situs judi *online*. Metode yang digunakan dalam penelitian ini adalah eksperimen dengan algoritma *web scraping* yang dikembangkan dalam bahasa pemrogramman Python serta menggunakan *dataset* aFOFA.info yang berisi daftar	1) Penelitian ini menggunakan pendekatan kuantitatif berdasarkan metode eksperimental dimana algoritma *web scraping* dikembangkan dengan bahasa pemrogramman Python untuk mendeteksi URL perjudian *online* tersembunyi dari domain .go.id. 2) Algoritma *web scraping* yang dikembangkan dengan Python digunakan untuk mengidentifikasi situs web yang berpotensi disusupi dari daftar target dengan menganalisis kata kunci terkait perjudian dalam bahasa lokal,	1) Dengan model DESLOT, peneliti dapat mengidentifikasi situs berdomain .go.id yang telah disusupi URL tersembunyi untuk promosi judi *online*. Namun, masih ada beberapa *false positives* dalam pelaksanaannya. Peningkatan strategis algoritma peneliti, dicapai melalui upaya kolaboratif, termasuk integrasi lapisan tambahan validasi otomatis. 2) Analisis cermat terhadap struktur HTML yang berisi URL yang terkait dengan perjudian *online* memainkan peran penting dalam menyempurnakan keakuratan algoritma peneliti. 3) Dengan menggunakan pendekatan *two-fold approach*, algoritman tersebut berhasil meneliti seluruh konten HTML dan mengintensifkan analisisnya pada segmen URL. Hasilnya menunjukkan bahwa 958 situs web .go.id telah disusupi dari 1.482 situs web yang dicurigai dan memberikan akurasi 99,1%.

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
		lengkap 450.000 domain .go.id.	seperti 'slot', 'judi', 'gacor', dan 'togel'. 3) Dalam penelitian ini, dikonseptualisasikan bahwa *attackers* terlibat dalam *website defacement* untuk memasukkan konten ilegal, seperti situs perjudian *online* ilegal yang tersembunyi, untuk meningkatkan peringkat atau visibilitas aktivitas perjudian *online*.	4) Setelah itu, peneliti menggali lebih dalam ke lanskap *online* Indonesia, yang mengungkap meluasnya penggunaan *advanced HTML coding techniques* dan menyoroti praktik *black hat SEO* dalam domain ".go.id". 5) Domain-domain yang saling terhubung yang diamati dalam penelitian ini menyoroti upaya terkoordinasi untuk mengeksploitasi situs web ".go.id" yang telah disusupi, sehingga menimbulkan kekhawatiran serius tentang keamanan dan integritas domain-domain ini. 6) Penyisipan tautan tersembunyi secara strategis ke dalam situs perjudian *online* muncul sebagai penemuan penting, yang menggarisbawahi perlunya kewaspadaan dalam melawan taktik semacam itu.
2	Christian et al. (2022)	Domain penelitian ini berfokus pada pemberantasan lonjakan *cybercrimes* di tengah	1) Dalam sistem yang diusulkan, peneliti mengkategorikan/*profiling cybercrimes* untuk	1) Penelitian ini berfokus pada klasifikasi/*profiling cybercrimes* dari portal berita yang tersedia. Dengan pengetahuan ini, peneliti telah

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
		meningkatnya aktivitas *online*, khususnya selama pandemi. Penelitian ni melibatkan *web scraping* untuk mengumpulkan berita terkait *cybercrime* untuk klasifikasi yang membantu penegakan hukum dan kesadaran masyarakat.	menciptakan kesadaran dan membantu penegakan hukum. 2) Informasi mengenai *cybercrimes* tersebut akan dikumpulkan dari portal berita berbahasa Inggris terkemuka di India. 3) Kemudian data yang telah di-*scraped* harus di-*filter* untuk menghapus semua data yang tidak perlu dan tidak konsisten. Dan pada akhirnya, *user* akan dapat melihat jenis *cybercrimes* yang terjadi menurut jangka waktu dan lokasinya. 4) Libraries Python yaitu scrapy dan BeautifulSoup4 akan digunakan untuk *scraping* data dari portal berita. Penggunaan *frameworks* ini akan	mengusulkan sistem *web scraping* dimana peneliti mengumpulkan artikel berita terkait *cybercrimes*. 2) Dari data hasil *scraping* tersebut, peneliti dapat mengklasifikasikan/*profiling* kejahatan-kejahatan tersebut ke dalam kategorinya masing-masing sesuai dengan wilayah dan jangka waktunya. Hal ini dapat membantu penegakan hukum dan juga menciptakan kesadaran di kalangan masyarakat umum.

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
			memastikan bahwa peneliti tidak mengumpulkan data sensitif apa pun saat data tersebut mengikuti *robots* .txt *security configuration*. 5) URL artikel berita terkait *cybercrimes* akan dikumpulkan terlebih dahulu. Kemudian judul dan teks artikel akan di-*scraped* dan disimpan. Data yang disimpan ini kemudian dapat diakses melalui aplikasi web GUI. Melalui aplikasi web ini pengguna akan dapat melihat secara visual jenis-jenis kejahatan menurut kota, negara bagian, bahkan tanggal tertentu.	

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
3	Yang diusulkan peneliti	Pada penelitian ini, peneliti melakukan proses *web scraping* terhadap situs *carding forum* dan *carding shop* untuk memperoleh dokumen digital yang mencerminkan aktivitas, diskusi, serta pola transaksi pelaku kejahatan *carding*. Dokumen hasil *scraping* tersebut kemudian dianalisis menggunakan pendekatan *profiling* forensik dan *natural language processing* (NLP), dengan fokus pada teknik *topic modelling* berbasis algoritma *latent dirichlet allocation* (LDA). Penerapan NLP dan LDA dalam konteks forensik digital *carding* telah digunakan secara luas dalam berbagai penelitian sebelumnya untuk mengekstraksi tema dominan dari data teks tak terstruktur di *dark web* atau forum *cybercrime* (Bergman & Popov, 2022; Sonmez & Codal, 2024). Penelitian ini mengadopsi pendekatan tersebut dan mengadaptasinya secara kontekstual dalam investigasi digital terhadap *carding shop*, dengan mempertimbangkan keterkaitan antara topik diskusi dan profil pelaku. Adapun kontribusi utama penelitian ini dibagi ke dalam tiga aspek sebagai berikut: 1) Aspek Analisis Profiling Forensik: Melalui analisis deskriptif berbasis *profiling* forensik atas dokumen hasil *web scraping*, penelitian ini mampu menggambarkan karakteristik umum pengguna *carding forum,* pola aktivitas mereka, serta modus operandi yang sering digunakan. Temuan ini sejalan dengan hasil penelitian (Christian et al., 2022; Nurseno et al., 2024), yang menunjukkan bahwa *profiling* berbasis teks dapat membantu klasifikasi tipe kejahatan serta mengidentifikasi pola geografis dan temporal dalam aktivitas siber. 2) Aspek Pendekatan NLP dan LDA: Penggunaan algoritma LDA dalam penelitian ini mempermudah ekstraksi topik-topik utama yang sering muncul dalam *carding forum*, seperti metode pembayaran ilegal, *tool carding* terbaru, hingga strategi menghindari deteksi. Teknik ini terbukti efektif dalam mengurai kumpulan data besar dan tidak terstruktur, sebagaimana ditunjukkan oleh penelitian (Rao et al., 2021; Sonmez & Codal, 2024). Dengan demikian,		

No	Peneliti	Domain Penelitian	Pendekatan/Metode Penelitian	Hasil Penelitian
		kontribusi LDA dalam penelitian ini bukan merupakan klaim sepihak, melainkan diperkuat oleh referensi ilmiah dan bukti empiris. 3) Aspek Penyajian Hasil dalam Kerangka Investigasi Forensik: Pendekatan yang digunakan dalam penelitian ini dirancang dalam kerangka *forensic investigation framework* yang mempertimbangkan tidak hanya aspek teknis tetapi juga prosedur legal penegakan hukum. Kontribusi ini memperkaya literatur yang selama ini mayoritas lebih berfokus pada pengembangan *framework scraping* dan analisis teks (Rao et al., 2021; Shahbazi & Byun, 2022), dengan menunjukkan bagaimana *web scraping*, NLP, dan *profiling* forensik dapat diintegrasikan secara legal dan aplikatif dalam konteks investigasi *carding*. Dengan memperhatikan hasil penelitian terdahulu serta pembahasan pada Bab 4, penelitian ini menyajikan kombinasi pendekatan teknis dan prosedural yang aplikatif, tidak hanya terbatas pada eksplorasi data, tetapi juga dapat dimanfaatkan langsung oleh penyidik dan penegak hukum dalam menangani kasus kejahatan siber secara lebih sistematis dan terarah.		

2.3.Kesimpulan Tinjauan Pustaka

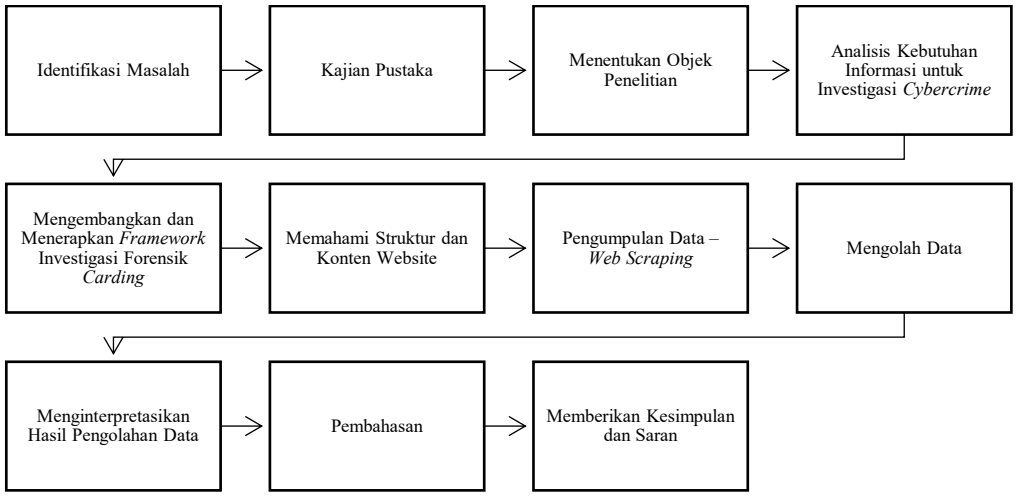
Berdasarkan pemaparan tinjauan pustaka yang telah dilakukan, maka kesimpulannya adalah sebagai berikut:

1. Penelitian ini menerapkan teknik investigasi *web scraping* pada *carding forum* dan *carding shop*, lalu menganalisisnya dengan pendekatan analisis deskriptif *profiling* forensik dan *latent dirichlet allocation*. Tiga landasan teori yang relevan telah diuraikan, yaitu *web scraping*, kejahatan *carding*, *natural language processing*, dan *latent dirichlet allocation*.
2. Berbagai teknik investigasi *web scraping forensic* dan metode analisis datanya telah diuraikan dalam penelitian terdahulu yang telah direview. Terdapat beberapa irisan antara *framework* investigasi forensik pada penelitian terdahulu dengan penelitian ini. Namun, fokus utama penelitian ini adalah analisis dokumen hasil web scraping pada *carding forum* dan *carding shop* menggunakan pendekatan analisis deskriptif *profiling* forensik dan *latent dirichlet allocation*.

BAB 3
Metodologi

3.1.Tahapan Penelitian

Agar dapat melaksanakan penelitian yang berkualitas, diperlukan adanya langkah-langkah yang terencana agar penelitian tersebut lebih terstruktur dan tepat sasaran. Pada Bab 3 ini, peneliti menguraikan tahapan-tahapan dalam melaksanakan penelitian ini. Secara garis besar langkah-langkah dalam penelitian ini dijelaskan pada gambar 3.1. dibawah berikut:



Gambar 3.1 Tahapan Penelitian

3.2. Identifikasi Masalah

Pada tahapan ini peneliti melakukan identifikasi masalah yang berasal dari referensi ilmiah, media massa, media ilmiah, dan laporan statistik resmi terkait kejahatan *cyber carding*, *carding forum, carding shop,* teknik investigasinya menggunakan *web scraping;* serta analisisnya menggunakan pendekatan analsis deskriptif *profiling* forensik dan *natural language processing*.

3.3. Kajian Pustaka

Kajian pustaka adalah suatu kegiatan penelitian yang bertujuan melakukan kajian secara spesifik dan mendalam tentang teori-teori dan konsep-konsep yang berkaitan dengan topik yang akan diteliti sebagai dasar dalam melangkah pada tahap penelitian selanjutnya. Sebuah kajian pustaka dianggap penting karena digunakan sebagai landasan dalam penyusunan laporan penelitian dan merupakan langkah pencegahan terhadap adanya duplikasi dari sebuah penelitian (Ridwan et al., 2021; Sugiyono, 2019). Kajian pustaka yang akan

dilaksanakan akan berfokus pada tema kejahatan *cyber carding*, *carding forum, carding shop,* teknik investigasinya menggunakan *web scraping;* analisisnya menggunakan pendekatan analsis deskriptif *profiling* forensik dan *natural language processing;* serta literatur yang masih berkaitan dengan penelitian ini.

3.4. Menentukan Objek Penelitian

Pada tahapan ini peneliti melakukan penentuan objek penelitian. Objek penelitian ini adalah *carding forum*, *carding shop,* serta tren aktivitas *cybercrime* yang ada didalamnya. Pada penelitian ini, situs web *carding forum* dan *carding shop* yang dijadikan objek penelitian adalah:

1. Altenen Forums-Images & Videos & Porn Accounts 🔞 Section (https://altenens.is/forums/images-videos-porn-accounts 🔞.469197/) (Altenen, n.d.)
2. carding.store-Cracking Tutorials Section (https://carding.store/forum/20-cracking-tutorials/) (Invision Community, n.d.)
3. Astradumps Shop (https://astradumps.com/shop/) (Astradumps, 2023)
4. Money-Heist.org Shop (https://money-heist.org/shop/) (@cashout vendors, 2020b)

Alasan peneliti memilih Altenen Forums-Images & Videos & Porn Accounts 🔞 Section dan carding.store-Cracking Tutorials Section sebagai objek penelitian karena kedua *carding forum* memiliki jumlah *total visits*, *bounce rate*, *page per visit*, dan *average visit duration* yang tinggi, menunjukkan potensi untuk mengumpulkan data yang kaya dan relevan (Altenen, n.d.; Invision Community, n.d.; Similarweb LTD, 2024). Berikut dibawah ini disajikan tabel terkait indikator kedua *carding forum* tersebut:

Tabel 3.1 Indikator *Carding Forum*

No	***Carding Forum***	***Total Visits***	***Bounce Rate***	***Pages per Visit***	***Avg Visit Duration***
1	Altenen Forums	355,1K	35,66%	17,46	00:11:38
2	carding.store	9K	18%	1,47	00:01:35

Sumber: Similarweb LTD (2024)

Tingginya aktivitas pengguna menandakan signifikansi kedua forum tersebut dalam komunitas *carding*, sehingga *web scraping* pada kedua forum tersebut dapat memberikan *insight* yang berharga tentang tren dan aktivitas *cybercrime* yang berkembang. Adapun alasan peneliti memilih Astradumps Shop dan Money-Heist.org Shop sebagai objek penelitian karena kedua *carding shop* tersebut memiliki banyak variasi barang dan jasa ilegal yang dijual, struktur dan konten web yang bisa diidentifikasi dengan baik (@cashout

vendors, 2020b; Astradumps, 2023). Banyaknya variasi barang dan jasa ilegal yang dijual; serta struktur dan konten web yang jelas membantu menghasilkan data *web scraping* yang kaya dan terstruktur. Hal ini memungkinkan analisis *profiling* forensik yang mendalam untuk mengidentifikasi tren, pola, serta strategi penanganan *cybercrime*. Keempat objek penelitian tersebut akan diinvestigasi menggunakan teknik *web scraping*.

3.5. Analisis Kebutuhan Informasi untuk Investigasi *Cybercrime*

Pada tahapan ini peneliti melakukan analisis kebutuhan informasi untuk investigasi *cybercrime* pada *carding forum* dan *carding shop* berbasis 5W1H dan *Alexiou Principle* (Han et al., 2020; Pogue, 2010) yang disajikan pada tabel 3.2 dan tabel 3.3 dibawah ini:

Tabel 3.2 Analisis Kebutuhan Informasi untuk Investigasi *Cybercrime* pada *Carding Forum* dan *Carding Shop* Berbasis 5W1H

Unsur 5W1H	Pertanyaan	Kebutuhan Informasi
What	Apa jenis aktivitas *cybercrime* yang terjadi?	Informasi tentang berbagai aktivitas ilegal seperti *carding, hacking & cracking*, penjualan konten ilegal, dan lainnya.
Who	Siapa pelaku yang terlibat dalam aktivitas *cybercrime* tersebut?	*Profiling* pengguna, pelaku aktif, dan jaringan kriminal yang terlibat dalam *carding* forum, *hacking*, atau konten ilegal.
When	Kapan aktivitas *cybercrime* tersebut dilakukan?	Waktu dan frekuensi aktivitas ilegal berdasarkan *log* dan jejak digital yang diambil dari *forum* atau *shop*.
Where	Dimana aktivitas *cybercrime* tersebut dilakukan?	Lokasi *server* atau domain *carding forum*, *carding shop*, dan tempat penyimpanan konten serta transaksi ilegal.
Why	Mengapa aktivitas *cybercrime* tersebut dilakukan?	Motivasi di balik aktivitas, seperti keuntungan finansial, pembelajaran *hacking*, atau distribusi konten terlarang.
How	Bagaimana aktivitas *cybercrime* tersebut dilakukan?	Teknik *carding, hacking,* distribusi konten ilegal, metode pembayaran, dan cara penyamaran identitas pelaku.

Tabel 3.3 Analisis Kebutuhan Informasi untuk Investigasi *Cybercrime* pada *Carding Forum* dan *Carding Shop* Berbasis *Alexiou Principle*

Alexiou Principle	Kebutuhan Informasi
What question are you trying to answer?	Identifikasi aktivitas ilegal utama di *carding forum* dan *carding shop* (*carding, hacking,* penjualan konten ilegal, dll.).
What data do you need to answer that question?	*Log* aktivitas pengguna, transaksi, pesan dalam forum, data identitas, serta tautan atau jejak transaksi digital.

Alexiou Principle	Kebutuhan Informasi
How do you extract that data?	Penggunaan teknik *web scraping* untuk mengumpulkan data dari *carding forum* dan *carding shop*, serta analisis jejak digital.
What does that data tell you?	Mengungkap pola, tren, dan jaringan *cybercrime*, serta memberikan *insight* untuk tindakan pencegahan dan penegakan hukum.

3.6. Mengembangkan dan Menerapkan Framework Investigasi Forensik *Carding*

Pengembangan *framework* investigasi forensik *carding* melibatkan perumusan struktur tahapan investigasi, definisi prosedur pengumpulan bukti, dan penetapan indikator penting. Langkah ini memastikan *framework* dapat digunakan secara sistematis untuk menangani kejahatan *carding* secara efektif. Keseluruhan tahapan tersebut dikembangkan menjadi sebuah *framework* investigasi forensik *carding* yang dirancang untuk memberikan panduan sistematis dalam memperkuat penindakan hukum terhadap aktivitas *cybercrime* di *carding forum* dan *carding shop*.

3.7. Memahami Struktur dan Konten *Website*

Pada tahapan ini peneliti memahami struktur dan konten *website* sebelum memulai *web scraping.* Tahapan ini melibatkan analisis elemen *web page*, seperti *layout*, *pagination, element click, tag* HTML, pola URL, serta identifikasi data yang akan diekstraksi untuk memastikan efisiensi dan keberhasilan pengumpulan data.

3.8. Pengumpulan Data – *Web Scraping*

Pada tahapan ini peneliti melakukan pengumpulan data menggunakan *web scraping software* untuk mengambil data dari *web page carding forum* dan *carding shop*, memilih elemen yang relevan, mengekstraksi data, dan menyimpannya dalam format Microsoft Excel (.xlsx) yang dapat dianalisis secara lebih lanjut. Pengumpulan data dengan *web scraping* dilakukan dua hari secara terpisah yaitu pada tanggal 5 Desember 2024 dan 27 Agustus 2025 sehingga, perubahan konten website pada tanggal lainnya tidak dimasukkan dalam analisis selanjutnya.

Penelitian ini menggunakan WebHarvy Versi 7.3.0.222 untuk *web scraping*, sebuah *GUI-based tool* yang memungkinkan ekstraksi data yang efisien tanpa memerlukan keahlian pemrograman. Performanya yang berkecepatan tinggi memungkinkan pengambilan *dataset* besar, seperti lebih dari 10.000 entri dari *carding forums* dan *carding shops*, sementara

pattern recognition-nya meningkatkan presisi dalam mengekstraksi teks, gambar, dan elemen lain dari struktur web yang kompleks (SysNucleus, 2024).

Selain itu, WebHarvy Versi 7.3.0.222 mendukung *multiple output formats*, termasuk CSV, Excel, XML, dan SQL, serta mampu menangani *JavaScript-based dynamic pages* dengan *scraping* otomatis terjadwal. Alat ini juga menawarkan mekanisme untuk *bypass anti-scraping* seperti CAPTCHA dan *IP blocking*, sehingga sangat efisien untuk pengumpulan data skala besar. Namun, keterbatasannya mencakup berkurangnya efektivitas terhadap pertahanan *advanced anti-scraping defenses*, seperti *AI-driven restrictions*, dan lebih sedikit fleksibilitas dibandingkan dengan *scripting-based tools* untuk ekstraksi data yang *highly customized* atau *unstructured data extraction* (SysNucleus, 2024).

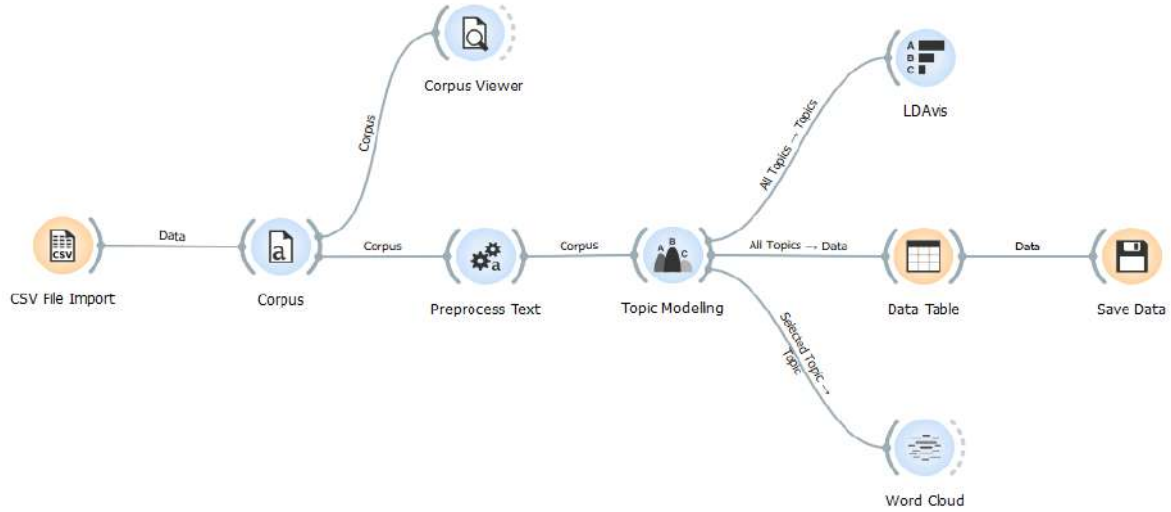
3.9. Pengolahan Data

Setelah data diperoleh dari tahapan *web scraping,* maka selanjutnya dilakukan proses pengolahan data. Pengolahan data dalam penelitian ini berfokus pada analisis deskriptif *profiling* forensik dan NLP *topic modelling* dengan pendekatan deteksi bahasa yang berbasis algoritma *latent dirichlet allocation* (LDA).

1. Analisis Deskriptif *Profiling* Forensik: Analisis deskriptif *profiling* forensik adalah teknik untuk mengidentifikasi pola dan karakteristik penting dari data forensik melalui penggunaan basis kata kunci dan visualisasi data secara detail dalam investigasi kriminal atau keamanan digital (Maybir & Chapman, 2021; Muehlethaler & Albert, 2021). Alat bantu yang digunakan untuk mengolah data dalam analisis deskriptif *profiling* forensik penelitian ini adalah Microsoft Excel serta *packages* Python yaitu Pandas (Microsoft, 2024; Python Software Foundation, 2001; The pandas development team, 2020). *Profiling* forensik yang dimaksud adalah pemetaan top 10 tematik *thread* populer berbasis analisis pandas, lalu dianalisis kontribusi penulis, tema dominan, dan interaksi *(views/replies)*, sehingga terbentuk gambaran pola aktivitas komunitas untuk kepentingan forensik. Dalam penelitian ini, proses *profiling forensics* dilakukan dengan menggunakan analisis deskriptif *univariate*. Analisis ini dipilih karena fokus utama terletak pada penggambaran karakteristik masing-masing elemen data hasil *web scraping* secara mandiri, tanpa melihat hubungan antar variabel. Pada analisis deskripitif profiling forensik, data yang digunakan hanya mencakup data pada tanggal waktu tunggal yaitu 5 Desember 2024. Data yang dianalisis mencakup atribut-atribut diantaranya *replies, views, time published, thread writer member,* serta jenis barang dalam *carding shop*. Melalui pendekatan *univariate*, setiap atribut dianalisis secara

statistik untuk mengidentifikasi pola distribusi, frekuensi kemunculan, dan kecenderungan aktivitas pengguna yang dapat mencerminkan peran atau tingkat keterlibatan dalam ekosistem *carding*. Pendekatan ini sesuai dengan tujuan *forensic profiling* yang bertumpu pada pengenalan struktur dasar perilaku pelaku melalui ciri-ciri individu yang terekam dalam data digital. Dengan demikian, analisis deskriptif *univariate* memberikan landasan awal yang kuat dalam menyusun profil digital pelaku tanpa harus melakukan inferensi terhadap hubungan antar variabel.

2. *Topic Modelling* — Algoritma *Latent Dirichlet Allocation*: Tahapan pengolahan data diuraikan berdasarkan Orange Data Mining *Workflow* dibawah ini:



Gambar 3.2. Orange Data Mining *Workflow*

Penelitian ini menggunakan *software* Orange Data Mining untuk menganalisis data teks hasil investigasi *web scraping* forensik pada *carding forum* dan *carding shop* (Demšar et al., 2013). Data *web scraping* yang digunakan dalam analisis *latent dirichlet allocation* dilakukan secara *temporal* dua hari secara terpisah untuk menguji:

— Efek Sebelum *(Pre-Effect)*: 5 Desember 2024

— Efek Sesudah *(Post-Effect)*: 27 Agustus 2025

Proses dimulai dengan mengimpor data dari file CSV yang berisi hasil *web scraping*, di mana data tersebut diatur dengan *cell delimiter* berupa *semicolon*. Setiap respons dalam file CSV diorganisasikan dalam baris-baris yang merepresentasikan jawaban individu dari partisipan. Langkah berikutnya adalah menggunakan modul *"Corpus"* di Orange Data Mining untuk mengumpulkan teks-teks dari file CSV menjadi sebuah korpus. Korpus ini kemudian dilihat melalui *"Corpus Viewer"*, yang memungkinkan peneliti untuk meninjau dan memastikan bahwa teks telah diimpor dengan benar serta memahami struktur data yang ada. Setelah korpus terbentuk, dilakukan praproses teks dengan menggunakan modul

"Preprocess Text". Pada tahap ini, teks diolah dengan berbagai teknik untuk membersihkan dan menstandarisasi data. Proses ini meliputi konversi semua teks menjadi huruf kecil *(lowercase)*, penghapusan aksen, *parsing HTML*, penghapusan URL, penghapusan tanda baca pada kata, penggunaan Porter Stemmer untuk mengurangi kata-kata ke bentuk dasarnya, serta penghapusan kata-kata umum yang tidak penting *(stopwords)*. Selain itu, digunakan juga kamus *(lexicon)* untuk mengidentifikasi kata-kata khusus, penghapusan angka, dan penggunaan ekspresi reguler *(regexp)* untuk membersihkan pola-pola tertentu dalam teks.

Setelah teks diproses, langkah berikutnya adalah pemodelan topik dengan menggunakan teknik *Latent Dirichlet Allocation* (LDA). *Latent Dirichlet Allocation* (LDA) adalah model probabilistik generatif dari sekumpulan *corpus*, Ide dasarnya adalah bahwa dokumen dapat direpresentasikan sebagai model campuran dari berbagai topik yang disebut juga laten, di mana setiap topik dikarakateristikan oleh kata (Blei et al., 2003; Zulhanif, 2016). LDA adalah metode statistik yang digunakan untuk menemukan pola-pola dalam kumpulan teks dengan mengidentifikasi topik-topik tersembunyi yang mungkin ada dalam korpus. Dengan LDA, setiap dokumen dalam korpus diwakili sebagai campuran dari berbagai topik, dan setiap topik sebagai campuran dari berbagai kata (Blei et al., 2003).

Hasil dari pemodelan LDA ini kemudian divisualisasikan menggunakan *"Latent Dirichlet Allocation Visualizer"*. Visualisasi ini membantu peneliti untuk melihat distribusi topik dalam korpus dan memahami bagaimana setiap topik diwakili oleh kata-kata tertentu. Visualisasi ini sangat berguna untuk menginterpretasikan hasil analisis topik secara lebih intuitif dalam kaitan tren dan aktivitas *cybercrime* yang berkembang pada *carding forum* dan *carding shop*. Sebagai tambahan, dibuat juga visualisasi dalam bentuk *"Wordcloud"*. *Wordcloud* adalah representasi visual dari kata-kata yang paling sering muncul dalam korpus, dengan ukuran kata yang lebih besar menunjukkan frekuensi kemunculan yang lebih tinggi. Wordcloud ini memberikan gambaran cepat tentang tema utama dan kata-kata penting yang ada dalam teks, yang dapat memberikan *insight* yang berharga tentang tren dan aktivitas *cybercrime* yang berkembang. Setelah semua analisis dilakukan, hasilnya ditampilkan dalam *"Data Table"*. Tabel ini memuat ringkasan dari data yang telah dianalisis, termasuk topik-topik yang diidentifikasi dan distribusi kata-kata dalam setiap topik. Ini memungkinkan peneliti untuk melihat data dalam bentuk yang lebih terstruktur dan mudah dipahami.

Langkah terakhir adalah menyimpan data yang telah dianalisis menggunakan modul *"Save Data"*. Data yang telah diolah dan dianalisis ini disimpan dalam format yang sesuai

untuk dokumentasi lebih lanjut atau untuk dianalisis lebih lanjut di masa depan. Secara keseluruhan, *workflow* ini memungkinkan peneliti untuk mengimpor, memproses, menganalisis, dan memvisualisasikan data teks dari kuesioner terbuka dengan cara yang sistematis dan efisien, sehingga dapat memperoleh *insight* yang bermakna dari data tersebut.

3.10. Menginterpretasikan Hasil Pengolahan Data

Menginterpretasikan hasil pengolahan data adalah proses menganalisis dan memahami informasi yang dihasilkan dari data yang telah diolah. Proses interpretasi ini melibatkan penafsiran hasil statistik, identifikasi pola, dan menyampaikan *insight* berdasarkan hasil yang telah diperoleh.

3.11. Pembahasan

Aspek yang paling penting dalam sebuah penelitian adalah pembahasan, karena pada tahap ini hasil penelitian tidak hanya disajikan, tetapi juga diinterpretasikan berdasarkan landasan teori dan penelitian terdahulu. Dalam konteks forensik *carding*, pembahasan berfungsi untuk menjembatani hasil analisis dengan *insight* yang dapat memperkaya pemahaman mengenai modus operandi, pola kejahatan, dan strategi pencegahannya. Pada bagian pembahasan ini juga dilakukan validasi kualitatif konfirmasi data dan evaluasi kinerja *framework*. Validasi kualitatif konfirmasi data dilakukan dengan teknik triangulasi data. Evaluasi kinerja *framework* dilakukan untuk menilai sejauh mana rancangan sistem ini efektif dan efisien dalam mendukung proses investigasi *forensik carding* berbasis data dari *carding forum* dan *carding shop*. Parameter evaluasi diturunkan dari hasil pembahasan dan dilakukan berdasarkan indikator berikut:

1. Efektivitas Investigasi
    - Kesesuaian hasil dengan tujuan investigatif: *Framework* berhasil mengidentifikasi pola perilaku pelaku, jenis data yang dicari, serta modus operandi umum seperti alat carding yang digunakan dan taktik penghindaran deteksi.
    - Kedalaman *insight* yang dihasilkan: Analisis menggunakan *latent dirichlet allocation* mampu mengungkap topik-topik utama dalam forum, seperti diskusi tentang strategi, alat bantu *carding*, hingga isu terkait akun pornografi alternatif, yang menunjukkan kemampuan *framework* dalam mengeksplorasi tema tersembunyi secara tematik.
    - Keterhubungan hasil analisis dengan tindakan hukum: Temuan dari analisis digunakan untuk menyusun laporan forensik yang kemudian diserahkan kepada

aparat penegak hukum, yang menjadi dasar bagi tindak lanjut hukum secara sah dan prosedural.

2. Efisiensi Proses Investigasi
   - Minimnya keterlibatan langsung dalam ekosistem ilegal: Penggunaan *web scraping* dan *profiling* digital memungkinkan pengambilan data dilakukan tanpa interaksi langsung dengan pelaku, sehingga mengefisienkan proses secara etis dan legal.
   - Reduksi beban manual dalam pengolahan data: *Framework* ini menggantikan proses identifikasi topik, pola, dan metadata secara manual, yang biasanya memerlukan waktu lama. Meskipun waktu secara kuantitatif tidak diukur langsung dalam penelitian ini, efisiensi dicapai melalui otomasi analisis berbasis *natural language processing* yang terbukti mempercepat proses dibandingkan pendekatan manual sebagaimana diperlihatkan dalam penelitian (Gazeau et al., 2024).
   - Penyusunan laporan yang sistematis dan berbasis bukti digital: Laporan temuan yang dihasilkan telah mengkompilasi pola-pola kunci dalam format visual dan tekstual yang komunikatif, sehingga memudahkan proses koordinasi dan penindakan dari aparat hukum.
3. Kepatuhan terhadap prosedur hukum
   - *Framework* diawali dengan surat permohonan resmi dari kepolisian sebagai dasar legal, dan seluruh proses investigasi, mulai dari ekstraksi data hingga pelaporan, mengikuti prosedur hukum yang berlaku, memperkuat integritas dan validitas hasil investigasi.

Dengan menggunakan parameter-parameter tersebut, evaluasi tidak hanya dilakukan terhadap apa yang ditemukan, tetapi juga terhadap bagaimana temuan diperoleh, serta bagaimana hasil tersebut dimanfaatkan dalam konteks penegakan hukum. Penelitian ini tidak mengklaim efisiensi dalam bentuk persentase waktu secara kuantitatif, namun menilai efisiensi dan efektivitas berdasarkan ketercapaian tujuan investigasi, kecepatan relatif proses otomatisasi, serta relevansi hasil terhadap kebutuhan praktis penegakan hukum, hal ini sejalan dengan pandangan (Dunsin et al., 2024; Gazeau et al., 2024) mengenai peran *artificial intelligence* dan *natural language processing* dalam mempercepat dan memperkuat analisis forensik digital modern.

3.12. Memberikan Kesimpulan dan Saran

Pada tahapan ini peneliti akan memberikan kesimpulan atas hasil penelitian. Pada bagian ini peneliti juga akan memaparkan keterbatasan penelitan dan saran guna pengembangan penelitian selanjutnya khususnya di bidang investigasi pada *carding forum* dan *carding shop* di masa yang akan datang.

BAB 4
Hasil dan Pembahasan

4.1.Pengembangan dan Penerapan *Framework* Investigasi Forensik *Carding*

Pada bagian ini, peneliti telah merumuskan *framework* investigasi forensik *carding* yang dirancang dan diterapkan dalam penelitian ini sebagai berikut:

Tabel 4.1 *Framework* Investigasi Forensik *Carding*

Tahapan	Sub-Tahapan	Definisi
Surat Permohonan Pemeriksaan Digital Forensik	Penerimaan Surat Permohonan	Menerima surat permohonan dari pihak kepolisian untuk melakukan pemeriksaan digital forensik terkait *carding forum* dan *carding shop*.
Pengumpulan Data	*Web Scraping*	Mengumpulkan data dari *carding forum* dan *carding shop* sesuai dengan izin yang diberikan dalam surat permohonan, untuk memperoleh dokumen dan transaksi yang relevan.
Pengolahan Data	*Profiling* Forensik	Menganalisis data yang dikumpulkan untuk mengidentifikasi pola dan karakteristik pelaku *carding*.
	Natural Language Processing (NLP)	Menggunakan teknik *latent dirichlet allocation* (LDA) untuk mengeksplorasi topik-topik utama dalam diskusi di *carding forum* dan *carding shop*.
Penegakan Hukum	Penyusunan Laporan	Menyusun laporan temuan berdasarkan hasil pengolahan data untuk mendukung proses hukum.
	Koordinasi dengan Penegak Hukum	Berkoordinasi dengan pihak kepolisian dan otoritas hukum untuk melaporkan temuan dan mendukung langkah-langkah berikutnya.
Keselarasan dengan *Alexiou Principle*	Pertanyaan yang Ingin Dijawab	Apa pola dan modus operandi pelaku *carding*?
	Data yang Diperlukan	Data dari *carding forum, carding shop,* dan transaksi terkait.

Tahapan	Sub-Tahapan	Definisi
	Metode Ekstraksi Data	Bagaimana cara mengumpulkan data melalui *web scraping* sesuai izin dari surat permohonan, serta menganalisis dengan NLP?
	Makna Data	Apa *insight* yang dapat diperoleh mengenai perilaku pelaku *carding* dan topik-topik utama dalam diskusi mereka?
Pengambilan Keputusan	Rekomendasi Tindakan	Memberikan rekomendasi untuk langkah-langkah investigasi berikutnya berdasarkan temuan.
	Validasi Kualitatif Konfirmasi Data	Pada tahapan ini dilakukan validasi kualitatif konfirmasi data triangulasi data, yaitu membandingkan hasil analisis LDA dengan data asli forum *carding*, literatur terkait, dan interpretasi manual, sehingga diperoleh konsistensi temuan, mengurangi bias peneliti, serta memperkuat validitas interpretasi kualitatif.
	Evaluasi Kinerja *Framework*	Menilai efektivitas *framework* dalam mendukung investigasi dan penegakan hukum. Dengan pengukuran yaitu: 1) Efektivitas investigasi, 2) efisiensi proses investigasi, dan 3) kepatuhan terhadap prosedur hukum seperti yang sudah dijelaskan pada Bab 3.11.

Pembuatan *framework* investigasi forensik *carding* ini merupakan implementasi dari Bab 3. Metode Penelitian – 3.6. Mengembangkan dan Menerapkan *Framework* Investigasi Forensik *Carding*, yang berfokus pada pengembangan dan implementasi *framework* terstruktur untuk melakukan investigasi forensik dalam kasus-kasus terkait *carding*. *Framework* ini menggabungkan prinsip dan metodologi forensik yang telah kuat dan mapan untuk memastikan pengumpulan data, analisis, dan koordinasi yang menyeluruh dengan penegak hukum.

Asal mula tahapan dalam *framework* investigasi forensik *carding* berakar pada *best practices* dalam investigasi forensik digital yang mencakup pengumpulan bukti yang sah *(web scraping)*, analisis data melalui teknik NLP (untuk pemahaman konten), dan penerapan prinsip-prinsip investigasi sistematis Alexiou untuk memastikan keberlanjutan hukum dan efektivitas proses investigasi. Proses ini disusun dengan mengacu pada literatur dan metodologi yang digunakan dalam investigasi kejahatan siber dan forensik digital seperti (Amato et al., 2019; Jin et al., 2024; Pogue, 2010; Rodrigues et al., 2024; Sonmez & Codal, 2024).

4.2.Penerimaan Surat Permohonan

Pada tahapan ini peneliti menerima surat permohonan dari pihak kepolisian untuk melakukan pemeriksaan digital forensik terkait *carding forum* dan *carding shop*.

Tabel 4.2 Surat Permohonan Pemeriksaan Digital Forensik

Kepolisian Negara X Alamat: [Alamat Lengkap] Telepon: [Nomor Telepon] Email: [Alamat Email]
Surat Permohonan Pemeriksaan Digital Forensik Nomor: 001/DPF/XYZ/2024 Kepada Yth. Fikri Irfan Adristi Ahli Digital Forensik Di Tempat Dengan hormat, Sehubungan dengan penyelidikan yang sedang kami lakukan terkait berbagai tindak pidana *cybercrime*, kami dari Kepolisian Negara X memohon bantuan Anda sebagai ahli digital forensik untuk melakukan pemeriksaan dan analisis forensik terhadap beberapa website *carding forum* dan *carding shop* yang kami duga terlibat dalam aktivitas *cybercrime*. Adapun website-website yang dimaksud adalah sebagai berikut: 1. Altenen Forums-Images & Videos & Porn Accounts 🔞 Section (https://altenens.is/forums/images-videos-porn-accounts 🔞.469197/)

Kepolisian Negara X

Alamat: [Alamat Lengkap]

Telepon: [Nomor Telepon]

Email: [Alamat Email]

2. carding.store-Cracking Tutorials Section (https://carding.store/forum/20-cracking-tutorials/)
3. Astradumps Shop (https://astradumps.com/shop/)
4. Money-Heist.org Shop (https://money-heist.org/shop/)

Kami berharap Anda dapat membantu dengan melakukan analisis forensik terhadap data dan dokumen yang telah kami kumpulkan dari situs-situs tersebut. Pemeriksaan ini bertujuan untuk mendalami lebih lanjut potensi keterlibatan situs-situs ini dalam berbagai aktivitas ilegal yang kami selidiki. Hasil pemeriksaan dan temuan yang Anda berikan akan sangat membantu dalam pengumpulan bukti-bukti yang diperlukan dalam proses penyidikan lebih lanjut.

Demikian permohonan ini kami ajukan. Atas perhatian dan kerja sama yang diberikan, kami ucapkan terima kasih.

Hormat kami,

Penyidik Kepolisian Negara X

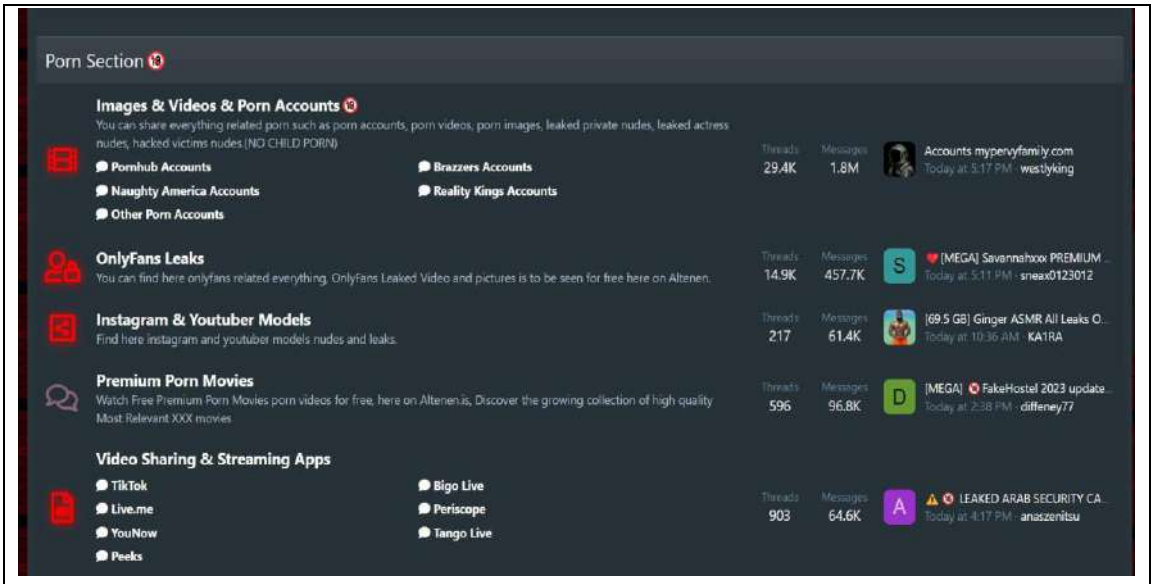
(Nama dan Jabatan)

(Tanggal)

4.3.*Web Scraping*

Sebelum melakukan proses *web scraping*, peneliti perlu memahami struktur dan konten dari *website carding forum* serta *carding shop* yang menjadi objek penelitian. Berikut disajikan hasil *capture* dari *website carding forum* dan *carding shop* tersebut beserta nilai *hash* masing-masing gambar, yang diperoleh menggunakan aplikasi HashMyFiles:

Tabel 4.3 *Capture Website Carding Forum* dan *Carding Shop*

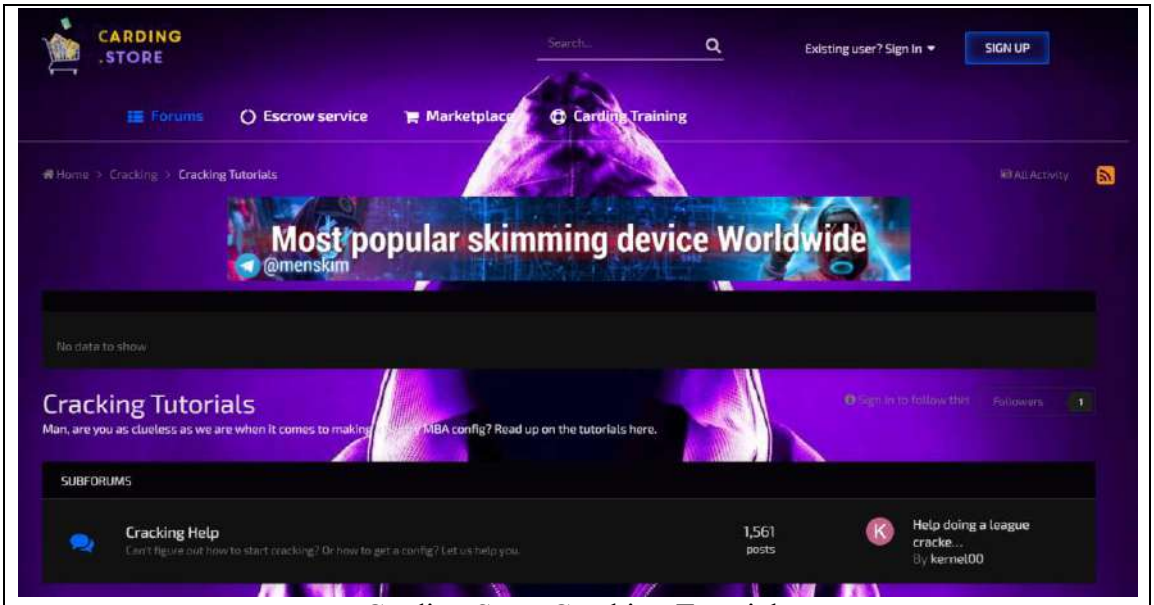


Altenens.is

==========

Filename : Altenens Porn Section Capture.jpg
MD5 : c634b28820ddd39443e509f48d761a32
SHA1 : 4bc14ccaa44fe9513be0a0cffdb49bc9820fca1d
CRC32 : 98375107
SHA-256 : 422856dbe0af3ad4d710863af6ecf869eafa630046d96782b27473ca428154b6
SHA-512 : 59193426d4651bfa6eea6f35f02dae9fc7db1745c836b77998a2168fc8cb92b91f3f3ec4fb04db864bb6695262c05f2a80b4b812a7a0aa2144111d6d6636061e
SHA-384 : 1051cdd60c255ea5fd498b8526bedac7fcdd98e3b573c07ba95bf5128192647cbb742cd2e3fbcdf86407571fb54ccbc5
Full Path : C:\Users\LENOVO\Downloads\Tesis Fikri\Gambar Capture Website\Altenens Porn Section Capture.jpg
Modified Time : 05/12/2024 01:11:00
Created Time : 05/12/2024 01:11:00
Entry Modified Time: 05/12/2024 07:07:46
File Size : 273.864
File Version :
Product Version :
Identical :
Extension : jpg
File Attributes : A

==========

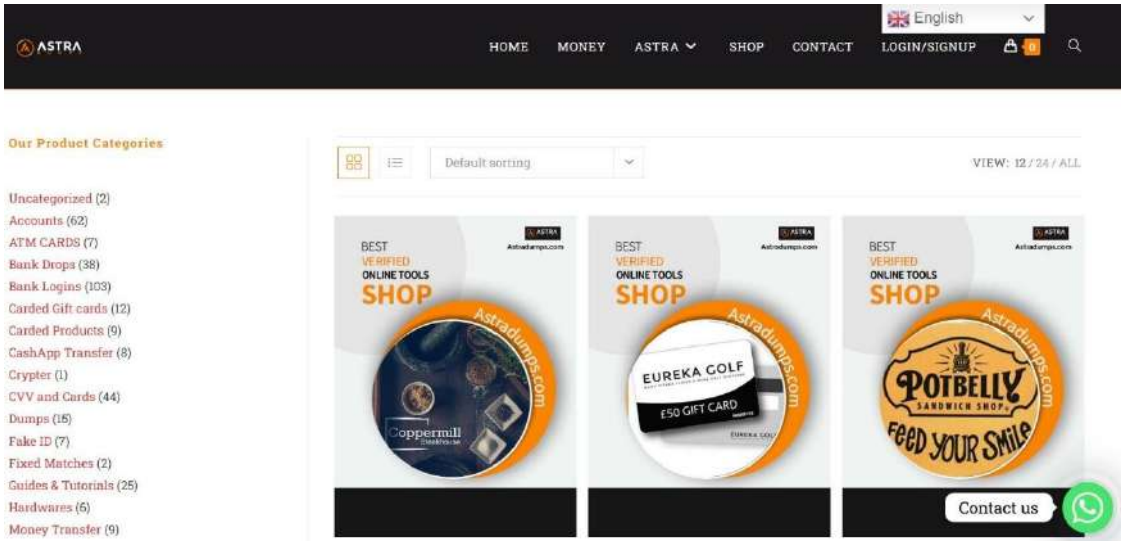


Carding.Store Cracking Tutorials

==========

Filename : Carding.Store Cracking Tutorials Capture.jpg
MD5 : 78ad14dc419a7df985393c18ec771c23
SHA1 : bf165da8806d5b4c0494d75314c104030e408eaa
CRC32 : b815a4e2
SHA-256 : a12821657a803282917b2c35ba2b0196d7a25c7418862a0cb905f75295a16f68
SHA-512 : ae01028ef8dd41f2f6f820683f4fc89087a28dee91b6aea694a7af0951c9d2b569eb616fcfce8343f683b06a9ed62d476c6d8c9ccc7ca27e8e53a10638401eb9
SHA-384 : 36576a769afc852f42bd501f26782fa6b052fe6d254e16dec48ab20cf85e122c78ac9b97574579059cdfac00ec2c0581
Full Path : C:\Users\LENOVO\Downloads\Tesis Fikri\Gambar Capture Website\Carding.Store Cracking Tutorials Capture.jpg
Modified Time : 05/12/2024 01:11:00
Created Time : 05/12/2024 01:11:00
Entry Modified Time: 05/12/2024 07:07:46
File Size : 199.835
File Version :
Product Version :
Identical :
Extension : jpg
File Attributes : A

==========



Astradumps Shop

==========

Filename : Astradumps Shop Capture.jpg
MD5 : f2be7eca8e74dd2eba70307a72bb1fc1
SHA1 : 2e2eff71f50c4fa5a752417bbabde8ec3b8c11f7
CRC32 : 1cc3a590
SHA-256 : 0f9b94db982a5f00dfb429ac8bd3daffbd511a7b8907dc419434848054be14c5
SHA-512 : 1b4d39a5a12ffba3773a4b08d27af012c954256d380097175977b4e294a301ab2fc57d3699e5f595d5ff67999e5d7ce68beddf25cc7ea6f86307cf57059a67a3
SHA-384 : 0533f3d6aefde588c61dc85a218c2aa5b7a3cff4e81d37e39978ab8bbcb32dd4ef91d8ec617cb54997e784168fe6f55d
Full Path : C:\Users\LENOVO\Downloads\Tesis Fikri\Gambar Capture Website\Astradumps Shop Capture.jpg
Modified Time : 05/12/2024 01:11:00
Created Time : 05/12/2024 01:11:00
Entry Modified Time: 05/12/2024 07:07:46
File Size : 217.450
File Version :
Product Version :
Identical :
Extension : jpg
File Attributes : A

==========

Money-Heist Shop

==========

Filename : Money-Heist Capture.jpg
MD5 : fb4e7f2b611d329273145560de270c9f
SHA1 : 46f17b6632b660fedf9c6ad21a7d5116f6606d52
CRC32 : 6cbc1413
SHA-256 : c5f2117a310f0874a150045389bfe5288eb03a312d38f40d2baf4b94925397e3
SHA-512 : aba9e6ed06c1886a5a211d560283f377b294a5a561597e585d4419f10d643f79f75ab8b797cfcb91b5292aa15ed53e31a154a67d371e3d3b0c9ceaec084e3642
SHA-384 : b6fdb5705b42050cad737da0df5f3bd76fec0add7aa3876e52ff680ee82e2f2af10d1891a4a656316e85a9ac2f61d1aa
Full Path : C:\Users\LENOVO\Downloads\Tesis Fikri\Gambar Capture Website\Money-Heist Capture.jpg
Modified Time : 05/12/2024 01:11:00
Created Time : 05/12/2024 01:11:00
Entry Modified Time: 05/12/2024 07:07:46
File Size : 251.432
File Version :
Product Version :
Identical :
Extension : jpg
File Attributes : A

==========

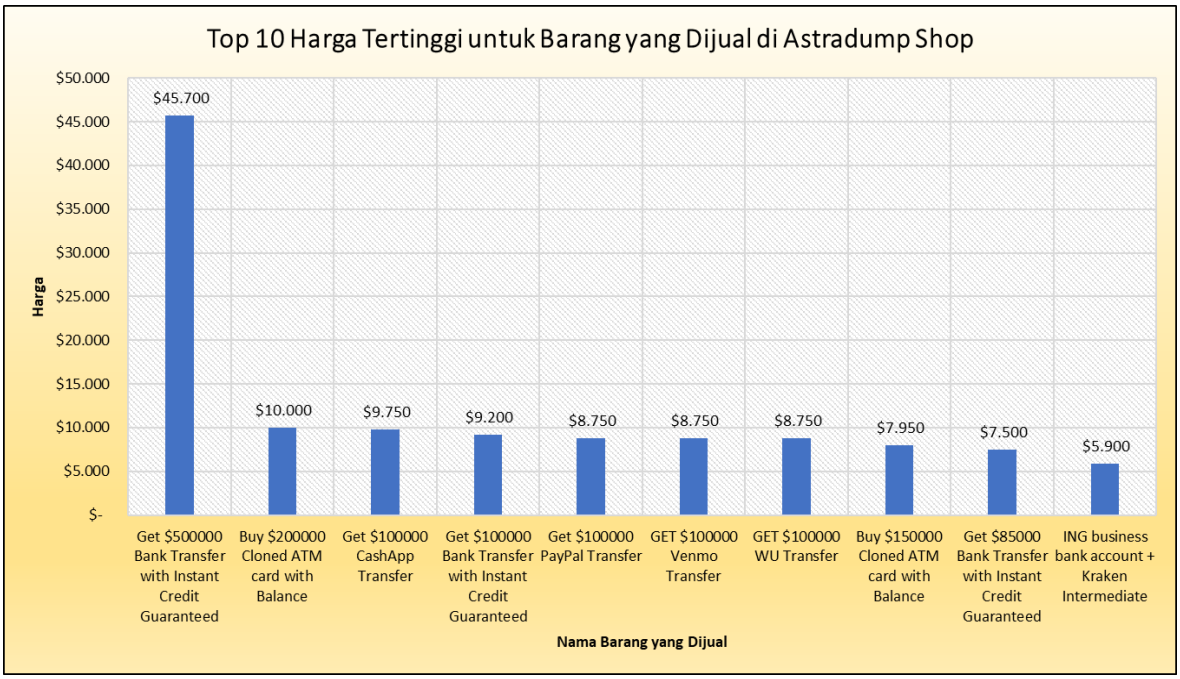
Sumber: Data Diolah - Adristi (2024)

Setelah peneliti memahami struktur dan konten dari website *carding forum* serta *carding shop* yang menjadi objek penelitian maka, peneliti mengumpulkan data dari *carding forum* dan *carding shop* sesuai dengan izin yang diberikan dalam surat permohonan investigasi, guna memperoleh dokumen dan transaksi yang relevan. Data hasil *web scarping* tersebut dapat diakses melalui akun Github peneliti (https://github.com/451Fikrie/Tesis-Magister-Informatika-Fikri) (Adristi, 2024).

4.4.Hasil Pengolahan Data

4.4.1. Analisis *Profiling* Forensik

Pada bagian ini, penulis akan menguraikan hasil analisis *profiling* forensik yang diperoleh dari analisis infografis berupa bagan di Microsoft Excel serta *output* dari Pandas. Berikut adalah penjelasan lebih lanjut mengenai hasil tersebut:

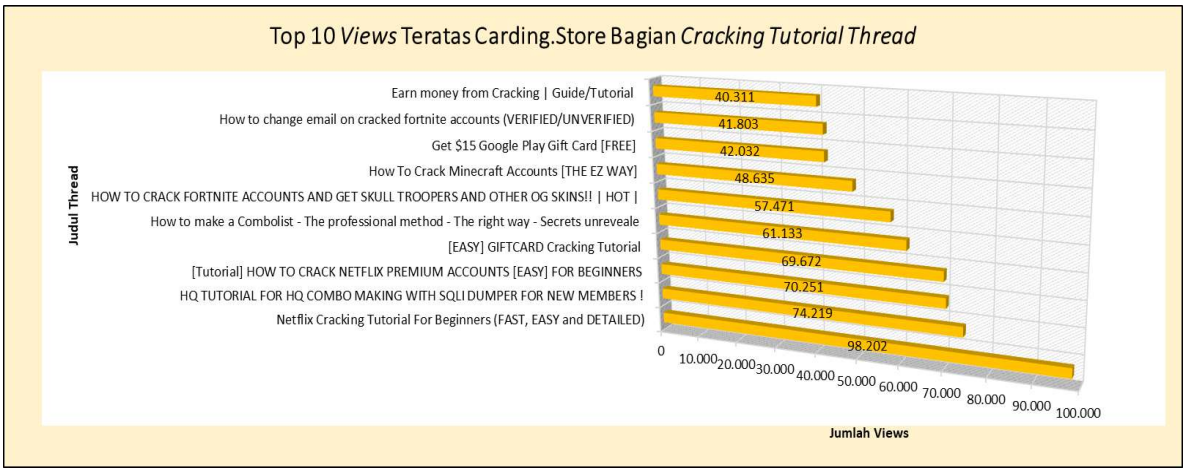


Gambar 4.1 Top 10 Harga Tertinggi untuk Barang yang Dijual di Astradump Shop

Sumber: Data Diolah - Adristi (2024)

Top 10 barang yang dijual di Astradump Shop (Astradumps, 2023) pada gambar 4.1 di atas menunjukkan transaksi ilegal dengan harga bervariasi. Barang yang termahal adalah transfer bank sebesar $500.000 dengan harga $45.700. Barang lainnya, seperti kartu ATM yang dikloning dan transfer melalui *platform* seperti CashApp, PayPal, dan Venmo, dijual dengan harga antara $5.900 hingga $10.000.

Berdasarkan 10 item teratas yang dijual di Astradump Shop (Astradumps, 2023) pada gambar 4, implikasi dari strategi investigasi forensik digital mencakup analisis mendalam terhadap transaksi ilegal pada *platform* seperti PayPal, Venmo, dan CashApp, dengan fokus pada pola transfer yang mencurigakan dan jumlah yang besar, seperti $45.700 untuk transfer $500.000. Artefak digital seperti *server logs, browser history*, dan *hardware* pelaku dapat dianalisis lebih lanjut untuk mengungkap jaringan kriminal. Menelusuri jejak transaksi ini memerlukan kerja sama antara perusahaan korban dan investigator seperti dalam penelitian Xiaoyu (2024) untuk membantu memahami modus operandi pelaku dan mengidentifikasi titik masuk untuk investigasi lebih lanjut.

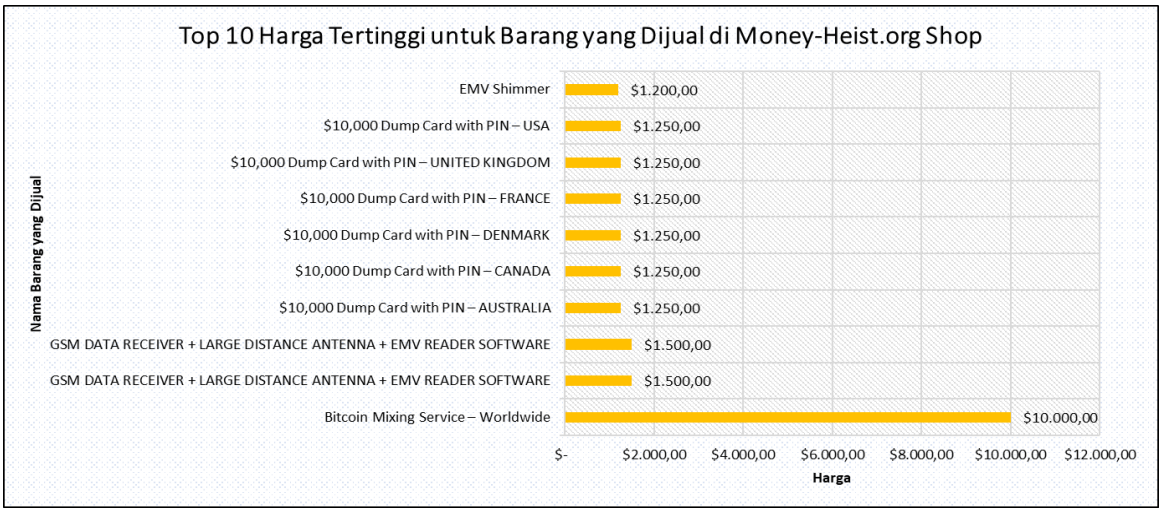


Gambar 4.2 Top 10 *Views* Teratas Carding.Store Bagian *Cracking Tutorial Thread*

Sumber: Data Diolah - Adristi (2024)

Top 10 *views* pada Carding.Store Bagian *Cracking Tutorial Thread* (Invision Community, n.d.) pada gambar 4.2 di atas menunjukkan *tutorial cracking* dengan tema populer seperti Netflix Cracking Tutorial For Beginners (98.202 views) dan HQ Combo Making with SQLI Dumper (74.219 views). Fokus utama adalah *cracking* akun premium, *gift card*, dan panduan membuat *combolist*, dengan *views* berkisar 40.311–98.202.

Strategi investigasi forensik digital mencakup analisis *metadata* pada *popular threads* seperti tutorial *cracking* Netflix dan *combolist creation*. Investigasi lebih lanjut berfokus *thread metadata, user activity logs, IP access data, upload history,* dan perangkat pelaku untuk membantu mengidentifikasi pelaku seperti dalam penelitian (Hidayat, 2020; You et al., 2016). Modus operandinya melibatkan *premium account cracking guides* dan *gift cards* melalui forum-forum dengan jumlah *views* tinggi (40.311–98.202), yang menargetkan *user* baru untuk memperluas jaringan aktivitas ilegal.



Gambar 4.3 Top 10 Harga Tertinggi untuk Barang yang Dijual di Money-Heist.org Shop

Sumber: Data Diolah - Adristi (2024)

Top 10 harga tertinggi di Money-Heist.org Shop (@cashout vendors, 2020b) didominasi layanan Bitcoin Mixing seharga $10.000. Produk lain berupa GSM receiver ($1.500) dan $10.000 Dump Card with PIN untuk berbagai negara ($1.250). Item tambahan seperti EMV Shimmer ($1,200) melengkapi daftar produk *cybercrime* berteknologi tinggi ini.

Strategi investigasi forensik digital dapat berfokus pada pelacakan transaksi terkait produk kejahatan siber berteknologi tinggi di Money-Heist.org Shop (@cashout vendors, 2020b), seperti Bitcoin Mixing ($10.000) dan Dump Card with PIN for various countries ($1.250). Investigasi dan analisis metadata transaksi pembayaran dapat mengungkap pola jaringan pelaku seperti dalam penelitian (Darmadi & Dananjaya, 2024; Jamil et al., 2024; Koo et al., 2024). Modus operandinya melibatkan penjualan barang ilegal, seperti penerima GSM receivers dan EMV Shimmer, untuk mendukung aktivitas kejahatan siber, dengan menggunakan instrumen pembayaran kripto untuk mengaburkan jejak kejahatan siber.

Tabel 4.4 *Altenen Porn Section Pandas Forensic Profiling*

```
=== Altenen Porn Section ===
Columns: Index(['Thread', 'Rating & Votes', 'Replies', 'Views',
'Time Published',
       'Thread Writer Member'],
      dtype='object')
Summary:
                                                      Thread Rating &
Votes  \
count                                                 14387
14391
unique                                                13390
481
top     ★[MEGA]★ \♡REAL GIRLS PACK LEAK GIRLS PRIVATE ...
freq                                                     18
10907

           Replies      Views Time Published Thread Writer Member
count        14391      14391          14391                14391
unique         717        818           1826                 5693
top     Replies\n0  Views\n1K   Dec 15, 2022         GhostlyGamer
freq          2266       1010             94                  332
Top Thread Writers: Thread Writer Member
GhostlyGamer     332
Ninjavu          264
huju66           241
QuickPhantasm    230
RushWay          197
Name: count, dtype: int64
Most Active Threads:
Thread       Replies  \
6535                      Busty Blonde Milf Gets Railed
Replies\n997
```

```
=== Altenen Porn Section ===
Columns: Index(['Thread', 'Rating & Votes', 'Replies', 'Views',
'Time Published',
       'Thread Writer Member'],
      dtype='object')
```

```
2308                    ⛔ REAL MOM LEAKED [SUPER PACK] ⛔
Replies\n995
5713      ⚡💟 LEAKED TEEN SNAPCHAT VIDEOS 💟⚡ @1njector
Replies\n99
12940  TEEN GIRL WITH THE BIGGEST TITS NUDE VIDEOS LE...
Replies\n99
12528          MULTIPLE TEEN SLUTS SNAPCHAT NUDES LEAKED
Replies\n99

            Views
6535     Views\n3K
2308     Views\n9K
5713    Views\n989
12940   Views\n980
12528   Views\n971
```

Sumber: Data Diolah - Adristi (2024)

Hasil analisis menunjukkan bahwa kolom *Thread* memiliki tingkat keberagaman yang tinggi dengan 13.390 entri unik dari 14.387 total *thread*. Thread dengan judul menarik seperti "⭐[MEGA]⭐ \❤️REAL GIRLS PACK LEAK GIRLS PRIVATE..." adalah yang paling sering muncul. Kolom *Rating & Votes* sebagian besar kosong atau tidak digunakan secara informatif, karena lebih dari 75% entri berasal dari kategori "kosong".

Kolom Replies dan Views didominasi oleh beberapa entri populer seperti "Replies 997" dan "Views 3K". Meski ada variasi, pola aktivitas menunjukkan preferensi terhadap konten yang eksplisit dan bersifat "sensasi". Penulis *thread* terbanyak adalah GhostlyGamer dengan kontribusi 332 *thread*, diikuti oleh Ninjavu (264) dan lainnya. Adanya *thread* yang sangat aktif seperti "Busty Blonde Milf Gets Railed" menunjukkan korelasi tinggi antara jumlah *thread* populer dan topik tertentu.

Thread populer memiliki jumlah balasan dan jumlah tampilan yang signifikan, menjadikannya pusat perhatian pengguna. Beberapa judul *thread* bahkan menunjukkan upaya eksplisit dalam menarik audiens dengan gaya *clickbait*. Secara keseluruhan, hasil ini menyoroti pola kontribusi komunitas dalam kategori ini yang bisa dimanfaatkan untuk analisis lebih lanjut dengan *Natural Language Processing* (NLP) guna mengungkap pola, tren, dan potensi aktivitas ilegal di forum ini.

Tabel 4.5 *Astradumps Pandas Forensic Profiling*

```
=== Astradumps ===
 Columns: Index(['Product Categories', 'Name of Item Sold', 'Price',
                  'Description'], dtype='object')
Summary:
        Product Categories                                    Name of
Item Sold  \
count                  400
400
unique                  24
392
top           Bank Logins  USA NON-VBV CARDS (CREDIT/DEBIT CARD) +
FREE C...
freq                   103
2
mean                   NaN
NaN
std                    NaN
NaN
min                    NaN
NaN
25%                    NaN
NaN
50%                    NaN
NaN
75%                    NaN
NaN
max                    NaN
NaN
               Price
Description
count     400.000000
400
unique           NaN
398
top              NaN  Description\nNEWLY ITALY CC + CVV +
INFO  $50K...
freq             NaN
2
mean      863.832500
NaN
std      2667.089616
NaN
min         0.000000
NaN
25%       217.500000
NaN
50%       300.000000
NaN
75%       550.000000
NaN
max     45700.000000
NaN
Top Categories: Product Categories
Bank Logins           103
Accounts               62
CVV and Cards          44
```

```
=== Astradumps ===
 Columns: Index(['Product Categories', 'Name of Item Sold', 'Price',
                  'Description'], dtype='object')
Bank Drops             38
Guides & Tutorials     25
Name: count, dtype: int64
Most Expensive Items:
          Product Categories  \
346          Money Transfer
65                ATM Cards
234        CashApp Transfer
340          Money Transfer
395  Western Union Transfer

                                    Name of Item Sold  Price  \
346  Get $500000 Bank Transfer with Instant Credit ...  45700
65            Buy $200000 Cloned ATM card with Balance  10000
234                      Get $100000 CashApp Transfer   9750
340  Get $100000 Bank Transfer with Instant Credit ...   9200
395                           GET $100000 WU TRANSFER   8750

                                           Description
346  Description\nGet $500000 Bank Transfer with In...
65   Description\nBuy $200000 Cloned ATM card with ...
234  Description\nGet $100000 CashApp Transfer\n\nO...
340  Description\nGet $100000 Bank Transfer with In...
395  Description\nGET $100,000 WU TRANSFER\n\nWe do...
```

Sumber: Data Diolah - Adristi (2024)

Dataset Astradumps terdiri dari 400 baris dan 4 kolom yang mencakup kategori produk *(Product Categories)*, nama item yang dijual *(Name of Item Sold)*, harga *(Price)*, dan deskripsi produk *(Description)*. Kolom *Price* memiliki nilai numerik, dengan harga rata-rata sebesar 863,83; yang menunjukkan bahwa harga produk bervariasi cukup luas, dengan harga terendah sebesar 0 dan harga tertinggi mencapai 45.700. Kolom *Product Categories* berisi 24 kategori unik, dengan kategori yang paling banyak muncul adalah *Bank Logins* yang ditemukan 103 kali, diikuti oleh *Accounts* sebanyak 62 kali dan *CVV and Cards* sebanyak 44 kali. Kolom *Name of Item Sold* mencatat 392 item unik, dengan item yang paling sering muncul adalah "USA NON-VBV CARDS (CREDIT/DEBIT CARD) + FREE C...", yang ditemukan 2 kali. Kolom *Description* juga mencatat 398 deskripsi unik, dengan deskripsi yang paling sering muncul berkaitan dengan "NEWLY ITALY CC + CVV + INFO $50K". Dataset ini tidak memiliki nilai yang hilang/*missing value*, sehingga dapat langsung digunakan untuk analisis lebih lanjut.

Terkait dengan harga produk, beberapa produk dengan harga tertinggi ditemukan, seperti "Get $500000 Bank Transfer with Instant Credit" yang dihargai 45.700, "Buy $200000 Cloned ATM card with Balance" seharga 10.000, dan produk lain yang mencakup layanan transfer uang dan kartu ATM dengan harga tinggi. Dengan data harga ini, kita bisa

mengidentifikasi produk-produk premium yang mungkin menarik untuk analisis lebih mendalam, termasuk pola harga dalam kategori tertentu. Secara keseluruhan, dataset ini memberikan gambaran yang jelas tentang produk yang dijual di *website* Astradumps, baik dari segi kategori, harga, maupun deskripsi produk, yang bisa digunakan untuk analisis lebih lanjut mengenai tren penjualan dan harga.

Tabel 4.6 *Carding.Store Cracking Tutorial Section Pandas Forensic Profiling*

```
=== Cracking Tutorial Section ===
Columns: Index(['Thread Title', 'First Thread Writer', 'Replies',
'Views',
       'Second Thread Writer'],
      dtype='object')
```

```
Summary:
        Thread Title First Thread Writer      Replies         Views
\
count           1030                 1030  1030.000000   1030.000000
unique           999                  472          NaN           NaN
top       [supreme]              JuNaiD™          NaN           NaN
freq              16                   43          NaN           NaN
mean             NaN                  NaN    73.105825   3802.318447
std              NaN                  NaN   185.255135   7421.333137
min              NaN                  NaN     0.000000    522.000000
25%              NaN                  NaN     3.000000    832.000000
50%              NaN                  NaN    19.000000   1409.000000
75%              NaN                  NaN    70.000000   3747.000000
max              NaN                  NaN  2630.000000  98202.000000

       Second Thread Writer
count                  1030
unique                  883
top                 JuNaiD™
freq                     19
mean                    NaN
std                     NaN
min                     NaN
25%                     NaN
50%                     NaN
75%                     NaN
max                     NaN
Top Thread Writers: First Thread Writer
JuNaiD™    43
Like it    41
Lincoln    33
upgruad    28
Merged     27
Name: count, dtype: int64
Most Viewed Threads:
                                          Thread Title  Views
34   Netflix Cracking Tutorial For Beginners (FAST ...  98202
213  HQ TUTORIAL FOR HQ COMBO MAKING WITH SQLI DUMP...  74219
93   [Tutorial] HOW TO CRACK NETFLIX PREMIUM ACCOUN...  70251
185                [EASY] GIFTCARD Cracking Tutorial   69672
11   How to make a Combolist - The professional met...  61133
```

Sumber: Data Diolah - Adristi (2024)

Dataset Cracking Tutorial Section mencakup 1.030 entri yang terdiri dari lima kolom: judul thread *(Thread Title)*, penulis *thread* pertama *(First Thread Writer)*, jumlah balasan *(Replies)*, jumlah tampilan *(Views)*, dan penulis *thread* kedua *(Second Thread Writer)*. Kolom *Replies* memiliki rata-rata 73,11 balasan per *thread*, dengan variasi yang cukup besar, mulai dari 0 hingga 2.630 balasan. Kolom *Views* menunjukkan rata-rata 3.802 tampilan per *thread*, dengan beberapa *thread* yang mendapatkan tampilan yang sangat tinggi, bahkan mencapai 98.202 tampilan, menunjukkan bahwa topik tertentu memiliki daya tarik yang jauh lebih besar dibandingkan yang lainnya.

Kolom *First Thread Writer* mencatat bahwa penulis yang paling aktif adalah "JuNaiD™", yang menulis 43 *thread*, diikuti oleh "Like it" dengan 41 *thread* dan penulis lainnya dengan jumlah *thread* yang lebih sedikit. Hal ini menunjukkan bahwa beberapa penulis memiliki kontribusi yang lebih besar dalam menciptakan konten dalam forum ini. Kolom *Second Thread Writer* mencatat 883 penulis unik, dengan "JuNaiD™" juga mencatatkan frekuensi tertinggi di kolom ini, yang menunjukkan kontribusi signifikan dari penulis tersebut pada *thread-thread* yang ditulis oleh orang lain.

Dalam analisis ini, kita dapat melihat bahwa ada sejumlah thread dengan jumlah tampilan yang sangat tinggi, seperti "Netflix Cracking Tutorial For Beginners (FAST...)" yang mendapatkan 98.202 tampilan, yang dapat menarik perhatian lebih dalam analisis popularitas dan dampak topik yang dibahas. Dengan demikian, informasi ini memberikan gambaran yang jelas tentang dinamika penulisan *thread* dan interaksi dalam *tutorial cracking forum*, yang dapat digunakan untuk mengidentifikasi konten yang paling populer dan penulis yang paling produktif.

Tabel 4.7 *Money-Heist Pandas Forensic Profiling*

```
=== Money-Heist ===
Columns: Index(['Categories', 'Name of Item Sold', 'Price',
'Description'], dtype='object')
Summary:
                 Categories                                      Name
of Item Sold  \
count                   229
229
unique                   27
214
top     Credit/Debit Cards  ACR38 R4 RFID Smart Contact Chip EMV
Card Read...
freq                     46
2
mean                    NaN
NaN
std                     NaN
NaN
```

```
=== Money-Heist ===
Columns: Index(['Categories', 'Name of Item Sold', 'Price',
'Description'], dtype='object')
min                    NaN
NaN
25%                    NaN
NaN
50%                    NaN
NaN
75%                    NaN
NaN
max                    NaN
NaN

               Price
Description
count     229.000000
229
unique           NaN
192
top              NaN  PLEASE READ THIS CAREFULLY \n\nDue to
receivin...
freq             NaN
15
mean      316.379913
NaN
std       723.793739
NaN
min        20.000000
NaN
25%        40.000000
NaN
50%       150.000000
NaN
75%       350.000000
NaN
max     10000.000000
NaN
Top Categories: Categories
Credit/Debit Cards     46
Dumps                  24
Burners                23
Carded Products        19
Carded E-Gift Cards    17
Name: count, dtype: int64
Most Expensive Items:
               Categories                                      Name of
Item Sold  \
14   BTC Mixing Services                 Bitcoin Mixing Service -
Worldwide
175             Hardware  GSM DATA RECEIVER + LARGE DISTANCE
ANTENNA + E...
6           Atm Skimmers  GSM DATA RECEIVER + LARGE DISTANCE
ANTENNA + E...
152                Dumps              $10,000 Dump Card with PIN -
DENMARK
154                Dumps        $10,000 Dump Card with PIN - UNITED
KINGDOM
```

```
=== Money-Heist ===
Columns: Index(['Categories', 'Name of Item Sold', 'Price',
'Description'], dtype='object')
```

```
     Price                                          Description
14   10000  DESCRIPTION PLEASE READ THIS CAREFULLY We are ...
175   1500  The description GSM DATA RECEIVER + LARGE DIST...
6     1500  DESCRIPTION\nThe description GSM DATA RECEIVER...
152   1250  PLEASE READ THIS CAREFULLY \n\nThis listing is...
154   1250  PLEASE READ THIS CAREFULLY \n\nThis listing is...
```

Sumber: Data Diolah - Adristi (2024)

Dataset Money-Heist berisi 229 entri dengan empat kolom: kategori produk *(Categories)*, nama item yang dijual *(Name of Item Sold)*, harga *(Price)*, dan deskripsi produk *(Description)*. Kolom *Categories* mencatatkan 27 kategori unik, dengan kategori yang paling banyak muncul adalah *"Credit/Debit Cards"* yang tercatat 46 kali, diikuti oleh kategori *"Dumps"* (24 kali) dan *"Burners"* (23 kali). Hal ini menunjukkan bahwa kategori produk tertentu lebih dominan di *platform* ini.

Kolom *Name of Item Sold* mencatatkan 214 nama produk unik, dengan produk yang paling sering muncul adalah "ACR38 R4 RFID Smart Contact Chip EMV Card Reader Writer", yang terjual dua kali. Rata-rata harga produk dalam dataset ini adalah sekitar 316,38; dengan harga yang bervariasi secara signifikan, mulai dari 20 hingga 10.000, menunjukkan adanya produk dengan harga yang sangat tinggi. Kolom Description berisi 192 deskripsi unik, dengan satu deskripsi muncul 15 kali, menunjukkan adanya produk dengan penawaran yang hampir serupa namun berbeda dalam konteks atau detail penawaran.

Beberapa item yang paling mahal, seperti "Bitcoin Mixing Service – Worldwide" dengan harga 10.000, dan/atau "GSM DATA RECEIVER + LARGE DISTANCE ANTENNA + EMV" dengan harga 1.500, menyoroti adanya transaksi dengan nilai tinggi yang terkait dengan produk spesifik, terutama yang terkait dengan layanan dan peralatan untuk kegiatan teknis. Analisis ini menggambarkan variasi produk, harga, dan kategori yang dijual dalam *platform* tersebut, serta potensi minat dan fokus dari para pembeli dan penjual.

4.4.2. *Natural Language Processing - Latent Dirichlet Allocation*

Pada bagian ini disajikan hasil analisis Orange Data Mining - *natural language processing* dengan pendekatan *latent dirichlet allocation* (LDA) untuk mengidentifikasi dan mendistribusikan topik secara sistematis berdasarkan data yang tersedia:

4.4.2.1.Analisis Efek Sebelum *(Pre-Effect)* pada Tanggal 5 Desember 2024

Berikut dibawah ini disajikan hasil analisis *latent dirichlet allocation* untuk *pre-effect:*

Tabel 4.8 *Topic Modelling: Latent Dirichlet Allocation - Altenen Porn Section*

Topic Modelling: Latent Dirichlet Allocation Topic Evaluation Log perplexity: 50,24009 Topic Coherence: 0,42245 Number of topics: 5
1: leak, teen, mega, nude, girl, snapchat, hot, ⭐, pack, video 2: mega, premium, collect, ❤, nz, porn, leak, pack, gb, exclus 3: video, nude, porn, girl, 🔞, 🔥, big, n, photo, sex 4: sexi, mega, account, hot, super, onli, ⚠, ❤, 【, ❄, 】 ✨ 5: leak, onlyfan, girl, new, 🔞, pack, collect, amateur, free, video

Sumber: Data Diolah - Adristi (2024)

Analisis *text mining* menggunakan metode *Topic Modelling: Latent Dirichlet Allocation* (LDA) pada tabel 4.8 menghasilkan lima topik utama dengan kata kunci tertentu. Topik pertama berfokus pada kebocoran konten pribadi, khususnya yang melibatkan remaja dan penggunaan *platform* media sosial seperti Snapchat. Kata kunci seperti *leak, teen, nude,* dan snapchat menunjukkan bahwa topik ini berkaitan dengan penyebaran konten eksplisit yang bersifat sensitif. Selain itu, istilah seperti *pack* dan video mengindikasikan bahwa distribusi konten ini dilakukan dalam bentuk koleksi atau paket melalui *platform* berbagi file seperti Mega.nz.

Topik kedua mengangkat tema distribusi konten eksklusif atau premium secara ilegal. Kata kunci seperti *premium, collect, exclus,* dan *mega* menunjukkan bahwa konten-konten ini berasal dari layanan berbayar dan kemudian disebarluaskan tanpa izin. *Platform* seperti Mega.nz terlihat menjadi media utama untuk menyimpan dan mendistribusikan koleksi konten ini. Simbol seperti ❤ juga mencerminkan adanya strategi pemasaran atau daya tarik emosional yang digunakan untuk mempromosikan konten tersebut.

Topik ketiga menyoroti distribusi konten pornografi dalam berbagai format, seperti video, foto, atau gambar eksplisit. Kata kunci seperti *nude, porn, girl,* dan *sex* menunjukkan bahwa fokus utama topik ini adalah materi dewasa yang ditujukan untuk konsumsi publik. Simbol seperti 🔞 dan 🔥 menguatkan kesan eksplisit dan mengisyaratkan bahwa konten ini dikategorikan sebagai materi yang membutuhkan perhatian khusus terkait usia.

Topik keempat berfokus pada promosi akun atau layanan yang menawarkan konten eksklusif. Kata kunci seperti *account, sexi, hot,* dan *super* menunjukkan bahwa konten ini dipasarkan dengan menggunakan istilah yang menarik perhatian. Elemen pemasaran

diperkuat dengan penggunaan simbol seperti ⚠ dan ❤, yang menonjolkan konten tersebut sebagai sesuatu yang istimewa atau membutuhkan akses tertentu untuk dinikmati.

Topik kelima mengangkat isu kebocoran konten dari *platform* OnlyFans. Kata kunci seperti *leak, onlyfan, new,* dan *amateur* menunjukkan bahwa konten ini melibatkan kreator amatir yang memanfaatkan *platform* tersebut untuk distribusi konten berbayar. Kebocoran ini sering kali disertai dengan koleksi konten *(pack)* yang diunggah secara gratis di internet, mengindikasikan adanya pelanggaran privasi dan hak cipta.

Berrdasarkan lima topik ini, analisis menunjukkan pola distribusi konten eksplisit, penyalahgunaan *platform* berbagi file, serta isu privasi dan pelanggaran hak digital dalam *website carding forum Altenen Porn Section.*

Tabel 4.9 *Marginal Topic Probability - Altenen Porn Section*

5 instances (no missing data) 6712 features No target variable. 2 meta attributes	
Topics	Marginal Topic Probability
Topic 1	0,323916
Topic 2	0,182307
Topic 3	0,153554
Topic 4	0,186736
Topic 5	0,153399

Sumber: Data Diolah - Adristi (2024)

Berdasarkan tabel 4.9, dataset terdiri dari 5 dokumen *(instances)* tanpa adanya nilai yang hilang *(no missing data)*, dengan jumlah fitur sebanyak 6712. Tidak ada variabel target dalam analisis ini, sehingga interpretasi berfokus pada distribusi topik berdasarkan probabilitas marginal yang dihasilkan. Selain itu, terdapat 2 atribut meta yang mungkin memberikan informasi tambahan tentang data, seperti identitas dokumen atau waktu pembuatan.

Topik pertama memiliki probabilitas tertinggi yaitu 0,323916 atau sekitar 32,39%. Hal ini menunjukkan bahwa topik ini adalah tema yang paling dominan dalam dataset. Mengingat interpretasi sebelumnya, topik pertama berfokus pada kebocoran konten pribadi yang melibatkan remaja dan *platform* seperti Snapchat. Dominasi probabilitas ini mencerminkan bahwa isu kebocoran konten pribadi adalah masalah utama dalam data.

Topik kedua memiliki probabilitas sebesar 0,182307 atau sekitar 18,23%. Topik ini menyoroti distribusi konten premium secara ilegal melalui *platform* berbagi file seperti Mega.nz. Probabilitas ini menunjukkan bahwa meskipun bukan yang paling dominan,

distribusi konten premium adalah isu yang cukup signifikan dalam dataset. Topik ketiga memiliki probabilitas sebesar 0,153554 atau 15,36%. Topik ini berkaitan dengan distribusi konten pornografi dan visual eksplisit. Walaupun probabilitasnya lebih kecil dibandingkan topik pertama dan kedua, hal ini menunjukkan bahwa konten eksplisit tetap merupakan bagian penting dari pola distribusi data.

Topik keempat memiliki probabilitas sebesar 0,186736 atau 18,67%. Topik ini fokus pada promosi akun atau layanan yang menawarkan konten eksklusif. Probabilitasnya yang sedikit lebih besar dari topik kedua menunjukkan bahwa aktivitas promosi ini memiliki relevansi yang cukup tinggi dalam konteks data. Topik kelima memiliki probabilitas terendah yaitu 0,153399 atau 15,34%. Topik ini menyoroti kebocoran konten dari *platform* OnlyFans, termasuk distribusi konten amatir. Walaupun probabilitasnya paling rendah, topik ini tetap relevan sebagai salah satu tema yang muncul dalam dataset.

Berdasarkan hasil analisis *latent dirichlet allocation* (LDA) pada tabel 4.8 dan tabel 4.9, topik kebocoran konten pribadi remaja di media sosial mengharuskan investigator forensik digital untuk memprioritaskan pengumpulan bukti dari *platform* seperti Snapchat dan Mega.nz. Pada distribusi konten premium ilegal, fokus investigator dapat mengarah ke *tracking file metadata* dan transaksi digital yang diperlukan.

Sementara itu, distribusi konten eksplisit memerlukan pendekatan untuk melindungi privasi dan identifikasi individu. Investigasi terhadap OnlyFans menekankan pentingnya menganalisis pelanggaran hak cipta dan privasi. Strategi ini mendukung mitigasi risiko privasi dan pelanggaran hukum. Strategi investigasi forensik digital dalam penelitian ini, sejalan dengan yang dilakukan dalam penelitian (Alyahya & Kausar, 2017; Griné & Teixeira Lopes, 2023; Huie et al., 2024; Noval et al., 2023).

Gambar 4.4 *Altenen Porn Section Word Cloud*

Sumber: Data Diolah - Adristi (2024)

Word cloud pada gambar 4.4. Altenen Porn Section di atas, menunjukkan fokus pada distribusi konten eksplisit, dengan *"video"* (5%) sebagai kata yang paling dominan, mencerminkan format utama konten yang dibagikan. Kata *"nude"* (3%), *"porn"* (2%), dan *"girl"* (2%) menyoroti tema eksplisit utama.

Simbol seperti " 🔞 " dan " 🔥 " menggambarkan penekanan pada konten dewasa atau yang dianggap menarik perhatian; menunjukkan upaya strategi pemasaran pelaku kejahatan siber untuk menarik *user* baru dengan cara yang melanggar norma hukum, khususnya dalam (Undang-undang (UU) No. 44 Tahun 2008 Tentang Pornografi, 2008; Undang-Undang Republik Indonesia Nomor 1 Tahun 2024 Tentang Perubahan Kedua Atas Undang-Undang Nomor 11 Tahun 2008 Tentang Informasi dan Transaksi Elektronik, 2024; Undang-undang Republik Indonesia Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik, 2008; Undang-Undang Republik Indonesia Nomor 19 Tahun 2016 Tentang Perubahan Atas Undang-Undang Nomor 11 Tahun 2008 Tentang Informasi dan Transaksi Elektronik, 2016). Kata seperti *"big", "n", "photo",* dan *"se"* memperlihatkan variasi dalam jenis konten visual yang didistribusikan. Strategi ini menunjukkan pelanggaran privasi, hak cipta, dan eksploitasi digital sebagai kegiatan inti pelaku kejahatan siber.

Tabel 4.10 *Topic Modelling: Latent Dirichlet Allocation - Carding Shop & Cracking Tutorial Threads*

Topic Modelling: Latent Dirichlet Allocation Topic Evaluation Log perplexity: 79,52553 Topic Coherence: 0,33847 Number of topics: 5
1: crack, account, hq, tutori, fortnit, spotifi, guid, make, premium, checker 2: free, get, crack, account, method, proxi, rdp, work, tutori, hq 3: crack, method, netflix, work, get, free, tutori, make, card, new 4: dork, sqli, use, get, tutori, dumper, databas, hq, crack, best 5: hq, make, dork, privat, keyword, combo, use, sqli, get, method

Sumber: Data Diolah - Adristi (2024)

Analisis *text mining* menggunakan metode *Topic Modelling: Latent Dirichlet Allocation* (LDA) pada tabel 4.10 menghasilkan lima topik utama dengan kata kunci tertentu. Topik pertama berfokus pada panduan dan tutorial untuk membobol akun-akun tertentu seperti Fortnite, Spotify, dan layanan premium lainnya. Kata kunci seperti *crack, account, hq, tutori,* dan *premium* menunjukkan bahwa topik ini berkaitan dengan metode atau alat untuk mendapatkan akses tidak sah ke akun berbayar. Kata *checker* dan *guid* mengindikasikan adanya alat bantu untuk memverifikasi keberhasilan pembobolan tersebut.

Topik kedua menyoroti penggunaan metode untuk mendapatkan akun secara gratis. Kata kunci seperti *free, get, crack,* dan *method* mengindikasikan bahwa topik ini fokus pada cara-cara mendapatkan akses ilegal ke akun dengan bantuan teknologi seperti *proxies* dan RDP. Keberadaan istilah seperti *tutori* dan *hq* menunjukkan bahwa terdapat panduan berkualitas tinggi untuk mempermudah proses ini.

Topik ketiga berfokus pada pembobolan layanan streaming populer seperti Netflix. Kata kunci seperti netflix, *crack, method,* dan *free* menunjukkan bahwa topik ini berkaitan dengan cara mendapatkan akses tidak sah ke layanan streaming melalui metode atau alat tertentu. Kata *card* dan *new* menunjukkan bahwa sering kali ada penggunaan kartu tertentu atau teknik baru dalam proses ini.

Topik keempat menyoroti penggunaan *dork* dan *SQLi* (SQL injection) sebagai metode untuk membobol *database* atau akun. Kata kunci seperti *dork, sqli, use,* dan *dumper* menunjukkan bahwa topik ini berfokus pada eksploitasi kelemahan sistem untuk mendapatkan data atau akses ilegal. Istilah seperti *tutori* dan *hq* menunjukkan adanya tutorial berkualitas tinggi yang mendukung eksploitasi ini.

Topik kelima mengangkat tema tentang pembuatan kombinasi kata kunci dan alat untuk eksploitasi. Kata kunci seperti *hq, make, dork,* dan *keyword* menunjukkan bahwa topik

ini berkaitan dengan penggunaan kata kunci khusus dan kombinasi SQL injection untuk mendapatkan akses atau data. Istilah seperti *combo* dan *method* menunjukkan fokus pada alat dan teknik yang digunakan dalam aktivitas ini.

Kelima topik ini menunjukkan pola yang konsisten tentang aktivitas pembobolan akun, layanan *streaming*, dan eksploitasi *database* menggunakan teknik seperti *dork, SQLi,* dan alat bantu lainnya. Fokus utama tampaknya pada penyebaran panduan dan alat berkualitas tinggi untuk mendukung tindakan ilegal, dengan beberapa topik menonjolkan layanan tertentu seperti Netflix dan Spotify.

Tabel 4.11 *Marginal Topic Probability - Carding Shop & Cracking Tutorial Threads*

5 instances (no missing data) 6712 features No target variable. 2 meta attributes	
Topics	Marginal Topic Probability
Topic 1	0,185420
Topic 2	0,341975
Topic 3	0,173992
Topic 4	0,122227
Topic 5	0,176387

Sumber: Data Diolah - Adristi (2024)

Berasarkan tabel 4.11, dataset terdiri dari 5 dokumen tanpa nilai yang hilang *(no missing data)* dengan 6712 fitur. Tidak ada variabel target dalam dataset ini, sehingga fokus utama adalah distribusi probabilitas topik. Selain itu, terdapat 2 atribut meta yang mungkin memberikan informasi tambahan tentang data. Topik pertama memiliki probabilitas sebesar 0,185420 atau sekitar 18,54%. Topik ini berfokus pada tutorial pembobolan akun seperti Fortnite, Spotify, dan layanan premium lainnya. Probabilitas ini menunjukkan bahwa topik pertama memiliki peran yang cukup signifikan, namun tidak dominan dibandingkan dengan topik kedua.

Topik kedua memiliki probabilitas tertinggi sebesar 0,341975 atau sekitar 34,20%. Topik ini menyoroti metode untuk mendapatkan akun gratis melalui teknologi seperti proxies dan RDP. Dengan probabilitas yang paling dominan, topik ini mencerminkan bahwa aktivitas untuk mendapatkan akses tidak sah ke akun gratis adalah perhatian utama dalam dataset ini. Topik ketiga memiliki probabilitas sebesar 0,173992 atau sekitar 17,40%. Topik ini berfokus pada pembobolan layanan *streaming* populer seperti Netflix. Probabilitas ini menunjukkan bahwa meskipun bukan tema utama, pembobolan layanan streaming tetap menjadi isu yang cukup penting dalam dataset.

Topik keempat memiliki probabilitas terendah, yaitu 0,122227 atau sekitar 12,22%. Topik ini terkait dengan eksploitasi database menggunakan *dork* dan *SQLi*. Walaupun memiliki probabilitas terendah, topik ini masih relevan sebagai salah satu aspek yang muncul dalam dataset, yang mencerminkan bahwa metode ini mungkin lebih spesifik atau kurang umum dibandingkan topik lainnya. Topik kelima memiliki probabilitas sebesar 0,176387 atau sekitar 17,64%. Topik ini berfokus pada penggunaan kata kunci khusus dan kombinasi alat untuk eksploitasi. Probabilitasnya yang sedikit lebih tinggi dari topik ketiga dan keempat menunjukkan bahwa aktivitas ini memiliki relevansi yang cukup tinggi dalam konteks data.

Berdasarkan hasil analisis *latent dirichlet allocation* (LDA) pada tabel 4.10 dan tabel 4.11, strategi investigasi forensik digital dapat berfokus pada identifikasi dan penelusuran alat serta teknik yang digunakan dalam peretasan, seperti proxies, RDP, dorks, dan SQLi. Dalam kasus topik dominan, terkait peretasan akun premium dan layanan *streaming* seperti Netflix, investigasi sebaiknya melibatkan penelusuran distribusi tutorial dan alat yang digunakan dalam aktivitas ilegal tersebut. Prioritas juga sebaiknya diberikan pada pemantauan aktivitas ilegal terkait akses tidak sah ke akun atau basis data menggunakan metode teknis tertentu. Contoh penerapannya terdapat pada penelitian (Lo et al., 2025; Mechri et al., 2025; Takyi et al., 2025).

Gambar 4.5 *Carding Shop & Cracking Tutorial Threads Word Cloud*

Sumber: Data Diolah - Adristi (2024)

Gambar 4.5 *word cloud* di atas, menunjukkan fokus utama pada aktivitas *cracking* dengan kata *"crack"* (8%) sebagai yang paling dominan, diikuti oleh *"account"* (6%) yang mengindikasikan target pembobolan akun. Kata seperti *"hq", "tutori",* dan *"guid"* menunjukkan adanya panduan berkualitas tinggi untuk memfasilitasi peretasan, sementara *"fortnit", "spotifi",* dan *"premium"* mencerminkan layanan spesifik yang ditargetkan, yang mencerminkan strategi pelaku dalam mengeksploitasi celah keamanan di *platform* populer. Temuan distribusi *word cloud* ini juga sejalan dengan penelitian (J. Chen et al., 2023; Flowers, 2008).

4.4.2.2.Analisis Efek Sesudah *(Post-Effect)* pada Tanggal 27 Agustus 2025

Berikut dibawah ini disajikan hasil analisis *latent dirichlet allocation* untuk *post-effect*:

Tabel 4.12 *Topic Modelling: Latent Dirichlet Allocation - Altenen Porn Section*

Topic Modelling: Latent Dirichlet Allocation Topic Evaluation Log perplexity: 55.37507 Topic Coherence: 0.24367 Number of topics: 5
1: view, repli, mega, collect, jun, sep, jul, onlyfan, premium, leak 2: view, repli, mega, sexi, n, hot, ghostlygam, k, c, © 3: repli, view, may, star, vote, apr, nov, account, sep, mega 4: view, repli, star, vote, jun, mega, girl, oct, video, leak 5: repli, view, leak, teen, mega, nude, girl, feb, snapchat, cybercaliph

Sumber: Data Diolah - Adristi (2024)

Analisis *topic modelling: latent dirichlet allocation* (LDA) pada tabel 4.12 diatas menghasilkan lima topik utama dengan kata kunci spesifik. Hasil evaluasi model menunjukkan nilai *log perplexity:* 55,37507 dan topic coherence: 0,24367; yang menandakan model cukup mampu mengelompokkan tema besar meski kohesinya masih moderat. Topik pertama berfokus pada distribusi konten premium dan onlyfans. Kata kunci seperti *mega, collect, onlyfan, premium,* dan leak menunjukkan bahwa topik ini berfokus pada penyebaran ilegal konten eksklusif atau berbayar, khususnya yang berasal dari *platform* OnlyFans. Pola waktu *(jun, sep, jul)* mengindikasikan distribusi konten dilakukan secara periodik atau berdasarkan bulan rilis. Platform Mega.nz muncul sebagai media utama untuk penyimpanan dan penyebaran.

Topik kedua mengangkat tema Promosi konten eksplisit dengan *branding & pseudonim*. Kata kunci *sexi, hot, ghostlygam, k, c,* © mengindikasikan bahwa topik ini berhubungan dengan pemasaran akun atau koleksi konten eksplisit. Nama pengguna atau

alias seperti ghostlygam berfungsi sebagai identitas pemasaran. Simbol © menunjukkan adanya *branding* atau klaim kepemilikan atas konten yang disebarkan. Mega kembali menjadi sarana distribusi. Topik ketiga menyoroti aktivitas forum dan *voting* terhadap akun. Kata kunci seperti *account, star, vote, may, apr, nov, sep* menunjukkan topik ini terkait dinamika forum, di mana pengguna saling menilai, memberi bintang *(star)*, atau melakukan *voting* terhadap akun/konten yang dibagikan. Elemen mega tetap ada, menandakan bahwa distribusi file tetap menjadi bagian penting meski fokus utama adalah interaksi komunitas.

Topik keempat berfokus ke konten eksplisit berbasis media (video & foto). Kata kunci *girl, video, leak, star, vote, oct, jun* mengindikasikan distribusi konten eksplisit berupa video atau foto dengan fokus pada figur perempuan. Aktivitas forum *(star, vote)* memperkuat kesan bahwa konten ini dipromosikan melalui sistem reputasi atau penilaian anggota. Mega kembali berfungsi sebagai repositori utama. Topik kelima mengangkat isu kebocoran konten remaja & platform sosial. Kata kunci *teen, nude, girl, snapchat, cybercaliph* menunjukkan fokus pada konten sensitif yang melibatkan remaja. Kehadiran snapchat mempertegas bahwa *platform* media sosial menjadi salah satu sumber kebocoran. Istilah *cybercaliph* bisa merepresentasikan nama pengguna atau grup yang terlibat dalam distribusi konten ini. Isu privasi, pelecehan digital, dan eksploitasi anak menjadi sorotan besar dalam topik ini.

Berdasarkan lima topik yang dihasilkan, analisis menunjukkan adanya pola distribusi konten premium dan OnlyFans secara ilegal, promosi konten eksplisit dengan *branding* tertentu, aktivitas komunitas forum melalui sistem *voting* dan reputasi, penyebaran media visual berupa foto maupun video, serta kebocoran konten remaja dari *platform* media sosial. Secara keseluruhan, temuan ini mengindikasikan penyalahgunaan *platform file sharing*, praktik pemasaran konten terlarang, serta isu serius terkait privasi dan pelanggaran hak digital dalam *website carding forum* Altenen Porn Section.

Tabel 4.13 *Marginal Topic Probability - Altenen Porn Section*

5 instances (no missing data) 11411 features No target variable 2 meta attributes	
Topics	Marginal Topic Probability
Topic 1	0,16081
Topic 2	0,172937
Topic 3	0,132899
Topic 4	0,202062
Topic 5	0,328928

Sumber: Data Diolah - Adristi (2024)

Berdasarkan hasil analisis pada tabel 4.13, dataset terdiri dari 5 dokumen *(instances)* tanpa nilai yang hilang *(no missing data)*, dengan jumlah fitur sebanyak 11.411. Tidak terdapat variabel target, sehingga interpretasi berfokus pada distribusi probabilitas marginal dari setiap topik. Selain itu, terdapat 2 atribut *meta* yang dapat memberikan informasi tambahan, seperti identitas dokumen atau waktu publikasi.

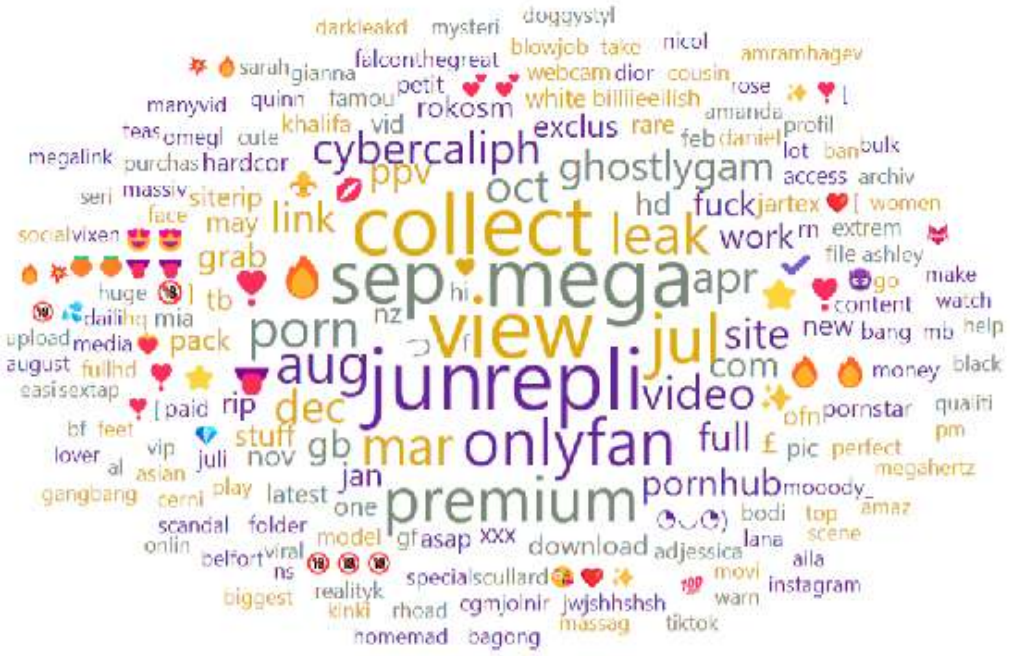
Topik pertama memiliki probabilitas sebesar 0,16081 atau sekitar 16,08%. Tema ini berhubungan dengan distribusi konten premium dan OnlyFans yang disebarkan secara ilegal melalui *platform file sharing* seperti Mega.nz. Walaupun proporsinya tidak dominan, distribusi konten berbayar tetap signifikan dalam membentuk pola data. Topik kedua memiliki probabilitas sebesar 0,172937 atau sekitar 17,29% dan menyoroti strategi promosi konten eksplisit dengan *branding* serta penggunaan *pseudonim*. Nilai ini lebih tinggi dibanding topik pertama sehingga menunjukkan bahwa promosi akun melalui citra eksklusif menjadi salah satu aktivitas penting di forum.

Topik ketiga memiliki probabilitas 0,132899 atau sekitar 13,29% yang merupakan nilai terendah di antara semua topik. Hal ini mengindikasikan bahwa interaksi komunitas seperti *voting* dan pemberian reputasi terhadap akun memang ada, tetapi kontribusinya relatif kecil dibandingkan penyebaran konten eksplisit. Sementara itu, topik keempat memiliki probabilitas 0,202062 atau sekitar 20,21% dan menegaskan bahwa distribusi konten eksplisit berupa foto maupun video perempuan menjadi isu signifikan dalam dataset, apalagi ditunjang dengan mekanisme reputasi forum yang digunakan untuk memperkuat promosi konten.

Topik kelima memiliki probabilitas paling tinggi yaitu 0,328928 atau sekitar 32,89%. Dengan hampir sepertiga distribusi, tema ini menjadi isu dominan dalam dataset. Topik tersebut berhubungan dengan kebocoran konten pribadi remaja di media sosial, khususnya Snapchat, yang menegaskan adanya masalah serius terkait privasi, eksploitasi anak, dan penyebaran data sensitif.

Secara keseluruhan, distribusi probabilitas marginal memperlihatkan bahwa isu kebocoran konten remaja di media sosial menjadi dominan, diikuti dengan penyebaran konten visual eksplisit, promosi dengan *branding* tertentu, distribusi konten premium, dan interaksi komunitas forum. Jika dikaitkan dengan interpretasi topik sebelumnya, hal ini mengindikasikan bahwa investigator forensik digital perlu memprioritaskan pengumpulan bukti dari *platform* media sosial seperti Snapchat, repositori *file sharing* seperti Mega.nz, serta menelusuri pola distribusi, promosi, dan interaksi komunitas yang mendukung penyebaran konten. Pendekatan ini penting untuk mendukung upaya perlindungan privasi,

pencegahan eksploitasi anak, serta penegakan hukum terhadap pelanggaran hak cipta dalam konteks kejahatan siber. Strategi investigasi forensik digital dalam penelitian ini, sejalan dengan yang dilakukan dalam penelitian (Daryabar et al., 2017; Gupta & Zhang, 2025; A. W. Malik et al., 2024; S. Mishra & Bajahzar, 2024; Serttaş & Al-Turjman, 2025).



Gambar 4.6 *Altenen Porn Section Word Cloud*

Sumber: Data Diolah - Adristi (2024)

Word cloud pada Altenen Porn Section memperlihatkan bahwa *"view"* (8%) dan *"repli"* (8%) menjadi kata dominan, menunjukkan bagaimana popularitas thread di forum diukur dari jumlah tayangan dan balasan. Kata *"mega"* (3%) dan *"collect"* (3%) menegaskan peran layanan Mega.nz sebagai repositori utama untuk pengumpulan serta distribusi file ilegal, sementara kata penanda waktu seperti *"jun", "jul",* dan *"sep"* memperlihatkan pola distribusi yang berlangsung secara periodik. Kehadiran kata "*onlyfan", "premium",* dan *"leak"* semakin menekankan praktik kebocoran konten eksklusif yang seharusnya berbayar, sehingga forum tidak hanya menjadi sarana berbagi tetapi juga pusat promosi dan pengarsipan konten eksplisit.

Aktivitas ini jelas melanggar hak cipta, privasi individu, serta regulasi hukum di Indonesia, khususnya Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan

Transaksi Elektronik (UU ITE) yang mengatur distribusi ilegal konten digital, serta Undang-Undang Nomor 44 Tahun 2008 tentang Pornografi yang melarang penyebaran konten cabul (Undang-undang (UU) No. 44 Tahun 2008 Tentang Pornografi, 2008; Undang-Undang Republik Indonesia Nomor 1 Tahun 2024 Tentang Perubahan Kedua Atas Undang-Undang Nomor 11 Tahun 2008 Tentang Informasi dan Transaksi Elektronik, 2024; Undang-undang Republik Indonesia Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik, 2008; Undang-Undang Republik Indonesia Nomor 19 Tahun 2016 Tentang Perubahan Atas Undang-Undang Nomor 11 Tahun 2008 Tentang Informasi dan Transaksi Elektronik, 2016). Dengan demikian, *word cloud* ini menggambarkan pola kejahatan digital yang sistematis, di mana *platform file sharing* dan forum daring dimanfaatkan untuk melanggengkan praktik distribusi konten eksplisit yang ilegal dan merugikan banyak pihak.

Tabel 4.14 *Topic Modelling: Latent Dirichlet Allocation - Carding Shop & Cracking Tutorial Threads*

Topic Modelling: Latent Dirichlet Allocation Topic Evaluation Log perplexity: 83.68025 Topic Coherence: 0.46944 Number of topics: 5
1: card, number, cvv, account, credit, use, code, get, dump, thi 2: thi, drop, bank, pleas, order, account, use, us, work, place 3: card, thi, reader, atm, skimmer, emv, smart, omnikey, use, data 4: bank, account, avail, order, pleas, log, us, thi, usd, state 5: thi, card, account, us, pleas, order, transfer, use, place, know

Sumber: Data Diolah - Adristi (2024)

Analisis *topic modelling: latent dirichlet allocation* (LDA) pada tabel 4.14 diatas menghasilkan lima topik utama dengan kata kunci yang berkaitan dengan aktivitas *carding*. Nilai *log perplexity:* 83,68025 dan *topic coherence:* 0,46944 menunjukkan bahwa model cukup baik dalam mengelompokkan tema dengan kohesi yang relatif lebih tinggi dibandingkan hasil sebelumnya. Topik pertama berfokus pada penjualan data kartu kredit & cvv. Kata kunci *card, number, cvv, account, credit, dump, code* menunjukkan topik ini berfokus pada penyediaan informasi kartu kredit yang dicuri. Istilah *dump* menandakan penjualan data mentah hasil *skimming,* sedangkan *cvv* dan *code* mengindikasikan data digunakan untuk transaksi *online*.

Topik kedua menyoroti jasa drop & penggunaan rekening bank. Kata kunci *drop, bank, order, account, us, work, place* menyoroti layanan *drop* (rekening perantara) untuk pencairan hasil kejahatan. Pengguna forum menawarkan atau mencari jasa ini untuk memfasilitasi transfer dana ilegal dari kartu curian. Topik ketiga berfokus pada temuan

skimmer, ATM, dan teknologi EMV. Kata kunci *reader, atm, skimmer, emv, smart, omnikey,* data menunjukkan adanya diskusi maupun perdagangan perangkat keras untuk *skimming* kartu. Istilah seperti EMV *(chip-based card)* dan *omnikey* mengindikasikan fokus pada teknologi *cloning* kartu canggih yang digunakan untuk mengakses data dari *chip* kartu kredit/debit.

Topik keempat menyoroti akses log akun bank & penawaran. Kata kunci *bank, account, avail, order, log, usd, state* menandakan topik ini terkait dengan penjualan akun bank *(bank logs)* yang sudah diretas. Adanya istilah *usd* dan *state* mengindikasikan bahwa transaksi dan harga dikaitkan dengan nilai tukar dolar AS dan target bank di wilayah tertentu. Topik kelima mengangkat tema tentang transaksi & transfer ilegal. Kata kunci *transfer, place, know, us, order, pleas* menyoroti layanan transfer uang hasil *carding*. Forum digunakan untuk mengoordinasikan transaksi, negosiasi, serta berbagi informasi terkait jalur pencairan dana ilegal.

Dengan lima topik yang dihasilkan, analisis menunjukkan pola perdagangan data kartu kredit dan CVV, penyediaan layanan drop untuk pencairan dana, penggunaan perangkat skimmer ATM dan teknologi EMV, penjualan akun bank hasil peretasan, serta transaksi dan transfer ilegal. Secara keseluruhan, temuan ini mengindikasikan eksistensi ekosistem *carding* yang terorganisir, mencakup penyalahgunaan data keuangan, peredaran perangkat kriminal, dan praktik pencucian uang digital dalam website *carding forum* dan *carding shop*.

Tabel 4.15 *Marginal Topic Probability - Carding Shop & Cracking Tutorial Threads*

5 instances (no missing data) 4802 features No target variable 2 meta attributes	
Topics	Marginal Topic Probability
Topic 1	0,175677
Topic 2	0,122264
Topic 3	0,143273
Topic 4	0,133996
Topic 5	0,421363

Sumber: Data Diolah - Adristi (2024)

Berdasarkan hasil analisis pada tabel 4.15, dataset terdiri dari 5 dokumen *(instances)* tanpa nilai yang hilang *(no missing data)*, dengan jumlah fitur sebanyak 4.802. Tidak terdapat variabel target, sehingga interpretasi berfokus pada distribusi probabilitas marginal dari setiap topik. Selain itu, terdapat 2 atribut *meta* yang dapat memberikan informasi tambahan terkait identitas dokumen atau waktu publikasi.

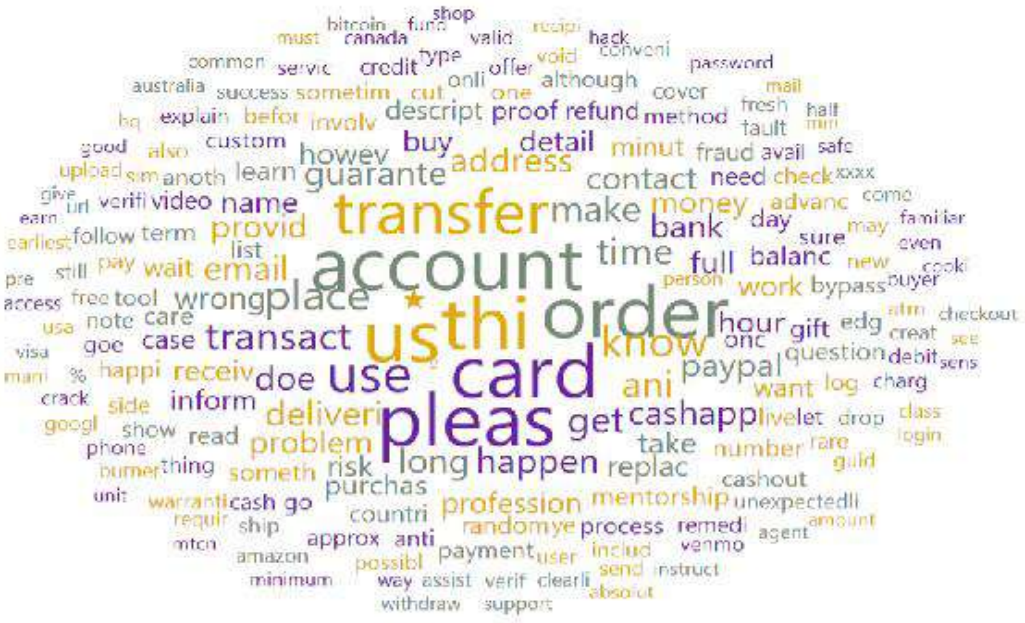
Topik pertama memiliki probabilitas sebesar 0,175677 atau 17,57% dan berfokus pada penjualan data kartu kredit dan CVV. Kata kunci seperti *card, number, cvv, account, credit, dump,* dan *code* menunjukkan bahwa data finansial hasil curian menjadi komoditas utama dalam forum. Kehadiran istilah dump mengindikasikan perdagangan data mentah hasil *skimming*, sedangkan *cvv* menegaskan peran data tersebut dalam memfasilitasi transaksi ilegal berbasis *online*.

Topik kedua menempati probabilitas 0,122264 atau 12,23% dan menyoroti jasa *drop* serta penggunaan rekening bank perantara. Kata kunci *drop, bank, order, account, us, work,* dan *place* menandakan adanya layanan untuk memfasilitasi pencairan hasil *carding*. Hal ini memperlihatkan bagaimana forum menyediakan infrastruktur yang mendukung pelaku kejahatan dalam menyamarkan identitas serta mengurangi risiko deteksi.

Topik ketiga memiliki probabilitas sebesar 0,143273 atau 14,33% dan berkaitan dengan perangkat *skimmer* ATM serta teknologi EMV. Kata kunci *reader, atm, skimmer, emv, smart, omnikey,* dan data memperlihatkan adanya diskusi atau perdagangan perangkat keras yang digunakan untuk pencurian data. Fokus pada istilah EMV dan omnikey menunjukkan tingkat teknis yang cukup tinggi, menandakan keterlibatan pelaku dengan kemampuan *cloning* kartu canggih.

Topik keempat menyumbang probabilitas sebesar 0,133996 atau 13,40% dan berfokus pada penjualan akses log akun bank. Kata kunci bank, *account, avail, order, log, usd,* dan *state* mengindikasikan bahwa *bank logs* hasil peretasan ditawarkan di forum dengan harga yang dikaitkan pada dolar AS. Istilah *state* menunjukkan bahwa target peretasan mencakup bank di wilayah tertentu, memperlihatkan dimensi internasional dari aktivitas *carding*.

Topik kelima merupakan yang paling dominan dengan probabilitas 0,421363 atau 42,13% dan berfokus pada transaksi serta transfer ilegal. Kata kunci *transfer, order, pleas, place,* dan *know* memperlihatkan koordinasi sistematis untuk mengolah dan mencairkan hasil *carding*. Dominasi topik ini menegaskan bahwa tujuan utama aktivitas *carding* adalah monetisasi, dengan forum berfungsi sebagai ruang negosiasi, pengaturan transaksi, dan pencucian uang digital. Secara keseluruhan, kelima topik ini mencerminkan ekosistem *carding* yang terorganisir, mulai dari pencurian data kartu kredit, penyediaan perangkat kriminal, hingga mekanisme pencairan dana dan transfer ilegal. Temuan ini menegaskan bahwa *forum carding* berfungsi sebagai pasar gelap digital dengan rantai pasok kejahatan finansial yang lengkap. Temuan atas penelitian ini selaras dengan penelitian (Alamsyah et al., 2025; Parida et al., 2025; Wang et al., 2025).



Gambar 4.7 *Carding Shop & Cracking Tutorial Threads Word Cloud*

Sumber: Data Diolah - Adristi (2024)

Word cloud pada gambar 4.7 *Carding Shop & Cracking Tutorial Threads* di atas memperlihatkan fokus pada aktivitas kejahatan siber di bidang keuangan, dengan kata *"thi"* (3%) muncul dominan sebagai pola linguistik umum dalam komunikasi forum. Kata *"card"* (3%) dan "account" (3%) menegaskan bahwa inti pembahasan forum adalah perdagangan data kartu kredit dan akun keuangan hasil kompromi. Kehadiran kata *"us"* (2%), *"order"* (2%), dan *"pleas"* (2%) memperlihatkan pola transaksi yang berlangsung dalam forum, di mana pelaku menawarkan jasa maupun produk *carding* untuk pengguna lain, sering kali dengan target bank dan sistem keuangan di wilayah Amerika Serikat.

Selain itu, kata *"transfer"* (2%), *"place"* (1%), dan *"know"* (1%) menandakan adanya praktik transfer dana ilegal dan koordinasi lokasi untuk mencairkan hasil *carding*. Kata *"use"* (1%) memperkuat indikasi bahwa data curian tidak hanya diperjualbelikan, tetapi juga digunakan secara langsung dalam berbagai modus transaksi ilegal. Pola ini menunjukkan bahwa aktivitas dalam forum melibatkan penyalahgunaan data finansial, praktik pencucian uang, dan peredaran akses akun secara sistematis, yang jelas melanggar ketentuan hukum siber dan keuangan di Indonesia, termasuk UU ITE serta UU No. 8 Tahun 2010 tentang Tindak Pidana Pencucian Uang (TPPU) (Undang-undang (UU) No. 8 Tahun

2010 Tentang Pencegahan dan Pemberantasan Tindak Pidana Pencucian Uang, 2010; Undang-Undang Republik Indonesia Nomor 1 Tahun 2024 Tentang Perubahan Kedua Atas Undang-Undang Nomor 11 Tahun 2008 Tentang Informasi dan Transaksi Elektronik, 2024; Undang-undang Republik Indonesia Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik, 2008; Undang-Undang Republik Indonesia Nomor 19 Tahun 2016 Tentang Perubahan Atas Undang-Undang Nomor 11 Tahun 2008 Tentang Informasi dan Transaksi Elektronik, 2016).

4.4.2.3.Analisis Konsistensi & Pergeseran *(Pre vs Post)*

Pada bagian Altenen Porn Section, hasil *pre-analysis* (Desember 2024) menampilkan lima topik utama yang menekankan kebocoran konten pribadi remaja, distribusi konten premium secara ilegal, promosi konten eksplisit, hingga penyebaran konten dari OnlyFans. Model ini cukup jelas menyoroti pola content leakage yang dipaketkan dalam bentuk koleksi *(pack)* melalui Mega.nz dan diekspos dengan strategi promosi simbolis seperti 🔞, ❤️, dan 🔥. Isu dominan terletak pada eksploitasi remaja, penyalahgunaan konten premium, dan mekanisme pemasaran forum melalui simbol atau label.

Sementara itu, hasil *post-analysis* (Agustus 2025) masih memperlihatkan konsistensi tema besar, yakni distribusi konten premium, OnlyFans, kebocoran konten remaja, serta penggunaan Mega.nz sebagai repositori utama. Namun, terjadi pergeseran pola: selain distribusi, forum mulai menunjukkan adanya aktivitas komunitas berupa peningkatan pada sistem *vote, star rating,* dan penggunaan *branding/pseudonim* (misalnya ghostlygam, cybercaliph). Dengan kata lain, praktik tidak hanya berfokus pada distribusi konten, tetapi juga menguat pada aspek interaksi sosial antar pengguna yang memperkuat reputasi dan pemasaran dalam komunitas forum. Dari sisi risiko, hal ini menandakan pelembagaan struktur komunitas yang lebih matang—tidak sekadar berbagi, tapi juga membangun *trust* dan *engagement* dalam ekosistem ilegal.

Pada bagian Carding Shop & Cracking Tutorial Threads, hasil *pre-analysis* (Desember 2024) menampilkan fokus utama pada distribusi *tutorial cracking* akun premium (Spotify, Netflix, dll.), metode dork/SQLi, serta pembuatan kombinasi kata kunci eksploitasi. Hal ini menunjukkan tahap awal berupa *knowledge sharing* (*tutorial, tools, metode*) dengan orientasi pada layanan digital populer. Hasil *post-analysis* (Agustus 2025) memperlihatkan pola yang lebih transaksional dan profesional. Tema besar bergeser dari sekadar tutorial menjadi ekosistem perdagangan data finansial: penjualan kartu kredit & CVV, jasa *drop account,* perangkat *skimmer* EMV/ATM, penjualan akun bank *(bank logs)*,

hingga layanan transfer dana ilegal. Dengan kata lain, forum ini bertransformasi dari ruang edukasi eksploitasi menjadi pasar kriminal terorganisir. Hal ini juga tercermin dari nilai *coherence* yang lebih tinggi (0,46944), menandakan struktur tema yang lebih konsisten dan mengerucut pada aktivitas *carding* murni.

Tabel 4.16. Ringkasan Analisis Konsistensi & Pergeseran *(Pre vs Post)*

Analisis Konsistensi & Pergeseran *(Pre vs Post)*	
***Pre* (5 Des 2024)**	***Post* (27 Ags 2025)**
Altenen Porn Section – Distribusi konten premium & pribadi ilegal – Kebocoran konten remaja via Snapchat – Fokus utama: *file-sharing* & eksploitasi konten	Altenen Porn Section – Distribusi konten premium & OnlyFans – Dimensi komunitas (*voting*, reputasi) – Strategi *branding* & *pseudonim* – Fokus utama: kombinasi konten + interaksi sosial
Carding Shop & Cracking Threads – Tutorial *cracking* (Fortnite, Spotify, Netflix) – Metode teknis (SQLi, dork, combo) – Fokus utama: edukasi & *sharing tools*	Carding Shop & Cracking Threads – Penjualan data kartu kredit & CVV – Layanan *drop* & transaksi ilegal – Perangkat *skimmer* & *cloning* EMV – Fokus utama: komersialisasi & pasar kriminal

Sumber: Data Diolah - Adristi (2024)

4.5.Pembahasan

4.5.1. Penegakan Hukum: Penyusunan Laporan & Koordinasi dengan Penegak Hukum

Tahapan selanjutnya dalam penelitian ini adalah menyusun laporan hasil analisis berdasarkan standar laporan digital forensik (DF) dan melakukan koordinasi dengan penegak hukum sesuai permintaan resmi yang diterima. Penyusunan laporan dilakukan dengan mempertimbangkan tiga jenis utama laporan digital forensik (DF), yaitu laporan teknis, laporan investigasi, dan laporan evaluatif. Setiap jenis laporan memiliki fungsi khusus untuk mendukung proses investigasi secara komprehensif dan sistematis. Pada laporan teknis, fokus utama adalah penyajian hasil pengolahan data yang rinci, mencakup metode yang digunakan, seperti *Topic Modelling* dengan *Latent Dirichlet Allocation* (LDA), serta visualisasi data berupa *word cloud* dan distribusi probabilitas topik. Laporan ini bertujuan untuk mendokumentasikan proses teknis pengolahan data, sehingga dapat menjadi dasar bagi proses verifikasi lebih lanjut.

Laporan investigasi disusun berdasarkan salah satu struktur laporan yang relevan, baik itu Report Structure Version 1, Report Structure Version 2 (ISO 27042), maupun Report

Structure Version 3 (PUSFID) (Luthfi, 2024). Laporan ini mencakup latar belakang kasus, analisis bukti digital yang diperoleh, interpretasi hasil, serta potensi pelanggaran hukum yang teridentifikasi, seperti aktivitas *carding, cracking,* dan distribusi konten ilegal. Pemilihan struktur laporan akan disesuaikan dengan kebutuhan pihak penyidik dan standar yang berlaku dalam yurisdiksi setempat. Laporan evaluatif digunakan untuk memberikan analisis terhadap dampak yang dihasilkan dari temuan ini, termasuk implikasi hukum dan rekomendasi kebijakan. Laporan ini mencakup penilaian terhadap efektivitas metode analisis yang digunakan, serta saran untuk langkah mitigasi di masa mendatang.

Koordinasi dengan penegak hukum dilakukan dalam bentuk pengajuan laporan tersebut dan penyusunan Berita Acara Pemeriksaan Saksi Ahli (BAPSA). Pada umumnya berita acara pemeriksaan adalah suatu keterangan yang diberikan oleh saksi, tersangka, maupun saksi ahli yang oleh undang-undang diberi nilai sebagai bukti yang dapat menjadi landasan hakim untuk menentukan hasil akhir dari proses persidangan tindak pidana yang terjadi (Soesilo, 1985). BAP (Berita Acara Pemeriksaan) merupakan rahasia negara yang tidak boleh dipublikasikan untuk masyarakat umum (Arifisnti et al., 2024). Pada tahapan ini, peneliti bertindak sebagai saksi ahli yang memberikan keterangan teknis dan mendukung proses investigasi melalui penjelasan hasil analisis secara formal. BAPSA berfungsi untuk memastikan bahwa hasil analisis dapat digunakan sebagai bukti yang sah di pengadilan.

Seluruh proses penyusunan laporan dan koordinasi dilakukan dengan tetap mematuhi standar etika dan hukum yang berlaku, termasuk menjaga kerahasiaan data yang tidak relevan dengan kasus. Dengan pendekatan ini, penelitian tidak hanya memberikan kontribusi teknis tetapi juga mendukung proses investigasi secara profesional dan terintegrasi, sesuai permintaan resmi dari pihak kepolisian.

4.5.2. Keselarasan dengan *Alexiou Principle*

A. Pola dan Modus Operandi Pelaku *Carding, Cracking,* dan Pornografi Alternatif

Pola dan modus operandi pelaku *carding, cracking,* serta konten pornografi alternatif dapat dianalisis dengan melihat pola interaksi dalam *carding forum* dan *carding shop* yang menjadi objek penelitian. Pelaku *carding* mengandalkan teknik untuk mencuri dan memanfaatkan data kartu kredit, sementara pelaku *cracking* sering mendiskusikan cara-cara untuk meretas *software* dan sistem keamanan. Di sisi lain, forum terkait pornografi alternatif sering kali berbicara tentang distribusi konten terlarang atau eksploitasi. Semua ini memerlukan pendekatan analisis yang lebih luas yang mencakup pemahaman mengenai teknik yang

digunakan dan cara mereka beroperasi dalam ruang ilegal. Modus operandi mereka juga menonjolkan layanan tertentu seperti Ebay, Amazon, Netflix dan Spotify (Adristi, 2024).

B. Data yang Diperlukan

Data yang diperlukan untuk analisis ini termasuk data yang diambil dari *carding shop* dan *carding forum*. Pada *carding shop*, data mencakup informasi terkait pencurian data dan transaksi kartu. Sementara itu, pada *carding forum*, data yang diperlukan berfokus pada pembahasan teknik meretas perangkat lunak dan keamanan serta, sub-forum pornografi alternatif memuat data terkait eksploitasi konten yang ilegal atau berbentuk eksploitasi visual lainnya. Pada penelitian ini, situs web *carding forum* dan *carding shop* yang dijadikan objek penelitian untuk diambil datanya adalah: (1) Altenen Forums-Images & Videos & Porn Accounts 🔞 Section; (2) carding.store-Cracking Tutorials Section; (3) Astradumps Shop; dan (4) Money-Heist.org Shop (@cashout vendors, 2020b; Altenen, n.d.; Astradumps, 2023; Invision Community, n.d.).

C. Metode Ekstraksi Data

Proses pengumpulan data menggunakan *web scraping*, sesuai dengan izin yang diberikan melalui surat permohonan pemeriksaan digital forensik, merupakan metode yang digunakan untuk mendapatkan data dari *carding forum* dan *carding shop* yang menjadi objek penelitian. *Web scraping* memungkinkan pengumpulan informasi dari *carding forum* dan *carding shop* untuk menganalisis modus operandi dari berbagai pelaku tersebut. Agar memastikan data yang dikumpulkan sah, teknik *web scraping* dilakukan dengan mematuhi regulasi hukum yang berlaku, menjaga keamanan dan integritas data yang diperoleh. *Web scraping software* yang digunakan dalam penelitian ini adalah WebHarvy Version 7.3.0.222 (SysNucleus, 2024).

D. Makna Data

Berbasis penggunaan analisis NLP - *topic modeling* LDA maka, data yang dianalisis dapat memberikan *insight* tentang perilaku pelaku dalam tiga kategori aktivitas. Pada aktivitas *carding*, ini termasuk teknik transaksi ilegal dan metode penyembunyian identitas. Pada aktifitas *cracking*, *insight* yang diperoleh berfokus pada teknik peretasan yang paling populer, serta eksploitasi kelemahan perangkat lunak. Pada *insight* pornografi alternatif, analisis akan menggali pola distribusi konten ilegal dan eksploitatif. Hasil *topic modeling* dapat mengidentifikasi topik utama dalam forum-forum ini, yang memberikan pemahaman yang lebih dalam mengenai tren perilaku pelaku, dan dapat membantu mengembangkan strategi untuk deteksi dini dan pencegahan kejahatan siber dalam berbagai bentuk.

Analisis NLP berbasis algoritma *latent dirichlet allocation* (LDA) pada teks bahasa informal, seperti terlihat pada tabel 4.8 sampai dengan tabel 4.11, rentan terhadap bias. Kata-kata *slang*, singkatan, atau simbol emosional (emoji) yang sering digunakan dalam percakapan di *carding forums* dan *carding shops* dapat menyebabkan model LDA menghasilkan topik yang kurang representatif atau bahkan sulit ditafsirkan. Model *latent dirichlet allocation* (LDA) ini mengandalkan frekuensi kata untuk mengidentifikasi topik, yang berarti kata-kata yang sering muncul, bahkan dalam konteks yang tidak signifikan, dapat mendominasi interpretasi. Selain itu, pemilihan kata kunci dalam dataset juga dapat menciptakan bias interpretasi, karena model mungkin tidak sepenuhnya menangkap nuansa dan konteks sosial atau emosional yang terkandung dalam bahasa informal. Akibatnya, meskipun topik dengan probabilitas tinggi muncul, topik tersebut tidak selalu mencerminkan isu yang paling relevan atau penting dalam konteks tersebut. Interpretasi ini mengandalkan model LDA yang diterapkan pada dataset informatif. Potensi bias dalam analisis ini dapat muncul karena keterbatasan representasi konteks dan penggunaan bahasa informal dalam teks yang dianalisis.

4.5.3. Pengambilan Keputusan: Rekomendasi Tindakan

Berdasarkan hasil analisis menggunakan metode *topic modelling latent dirichlet allocation* (LDA) dan visualisasi melalui *word cloud*, beberapa rekomendasi tindakan dapat disusun untuk menangani aktivitas ilegal yang teridentifikasi pada *carding forum, carding shop*, serta aktivitas terkait lainnya. Analisis ini telah mengungkapkan pola utama dalam aktivitas ilegal, termasuk pembobolan akun *(cracking)*, penyebaran konten eksplisit *(pornography)*, penyalahgunaan data pribadi, dan eksploitasi sistem melalui teknik seperti *SQL injection* dan penggunaan *dork*.

A. Penindakan Teknis oleh Penegak Hukum

Hasil analisis menunjukkan keterkaitan yang jelas antara aktivitas ilegal dan penggunaan *platform daring* seperti *carding forum* atau *cracking shop*. Oleh sebab itu, diperlukan langkah konkret berupa pelacakan IP, pengumpulan bukti digital tambahan seperti *log* aktivitas, serta identifikasi pengguna utama pada forum-forum ini. Penegak hukum dapat bekerja sama dengan penyedia layanan internet (ISP) dan *platform* terkait untuk menghapus konten ilegal, menutup akses ke situs atau forum berbahaya, serta mengidentifikasi pihak yang terlibat (Benhamou, 2017; Hidayat, 2020). Implementasi rill-nya seperti saat proses penutupan Alphabay, '*dark market' online* terbesar yang melibatkan *multi-stakeholder* (Baraniuk, 2017; U.S. Attorney's Office, 2017).

Jika alamat web URL dari keempat *carding forum* dan *carding shop* tersebut berubah, maka teknis analisis investigasi forensiknya dilakukan dengan pendekatan *URL resolution* dan artefak jejak digital (Kailas & Roopalakshmi, 2025; Nasraoui & Krishnapuram, 2002; Yapici, 2025). Pertama, peneliti melakukan pelacakan melalui *historical domain records* (misalnya *WHOIS history, DNSdumpster, SecurityTrails*) untuk mengidentifikasi perubahan nama *domain* atau *hosting* (Bracciale et al., 2025; Fernandez et al., 2024; P et al., 2021; Rashid et al., 2019). Kedua, digunakan *search engine footprinting* dan *web archive (Wayback Machine)* untuk menemukan *snapshot* lama maupun redireksi URL baru (Bowyer, 2021; Hegarty, 2023). Ketiga, dilakukan analisis struktur HTML dan *fingerprint server* agar meski domain berubah, pola *file path, layout,* atau *hash* dari *resource* tetap bisa dikenali. Keempat, jika *carding forum* atau *carding shop* tersebut berpindah ke domain baru, investigator dapat memanfaatkan *crawling* berbasis *keyword* unik (misalnya nama *thread, vendor,* produk) untuk menautkan kembali konten lama ke lokasi baru (Rajiv & Navaneethan, 2021). Dengan teknik ini, meskipun URL berubah, kesinambungan analisis forensik tetap terjaga dan validasi data tetap dapat dilakukan.

B. Penguatan Keamanan Digital oleh Pemangku Kepentingan Teknologi

Aktivitas ilegal seperti *cracking* akun dan pembobolan sistem dapat diatasi dengan meningkatkan keamanan *platform daring*. Rekomendasi teknis mencakup penerapan *multifactor authentication* (MFA), pembaruan rutin perangkat lunak untuk menutup kerentanan *(patch management)*, serta peningkatan algoritma deteksi aktivitas mencurigakan, seperti penggunaan *proxy* atau *dumper tools* yang terdeteksi dalam analisis (Cheng et al., 2024; Coscia et al., 2024; Ruiz et al., 2023). Pemangku kepentingan teknologi, termasuk pengelola *platform* seperti Netflix, Spotify, dan lainnya, harus memperkuat kerja sama dengan tim keamanan siber untuk melindungi data pengguna mereka.

C. Edukasi Publik dan Pencegahan

Kata-kata kunci seperti *"tutori", "guid",* dan *"method"* yang sering muncul menunjukkan adanya penyebaran panduan untuk aktivitas ilegal di forum-forum ini. Oleh sebab itu, diperlukan kampanye edukasi yang menargetkan masyarakat umum, terutama generasi muda, mengenai bahaya dan konsekuensi hukum dari keterlibatan dalam aktivitas *cracking*, *carding*, dan eksploitasi lainnya. Edukasi ini dapat disampaikan melalui program literasi digital di sekolah atau media daring.

Pentingnya pendidikan juga ditekankan dalam penelitian Ayanwale et al. (2023) yang menunjukkan bahwa meningkatkan kesadaran keamanan siber sejak dini dapat memberikan dampak signifikan pada kehidupan pribadi dan masyarakat. Penelitian ini

menekankan perlunya dukungan kepada guru pra-jabatan di Lesotho untuk memahami implikasi keamanan siber dan mendorong penerapan literasi digital yang lebih luas. Temuan ini relevan dengan upaya peningkatan kesadaran di kalangan calon pemimpin dan karyawan perusahaan.

Selaras dengan hal tersebut, penelitian Bhagat & Pravin (2023) menegaskan bahwa manusia akan tetap menjadi pertahanan utama (atau *firewall* terakhir) aset siber, namun merekalah yang paling rentan terhadap berbagai serangan siber. Penelitian ini menawarkan model yang memperluas kerangka *knowledge-attitude-behavior*, menyoroti pentingnya pelatihan keamanan siber, pemahaman atas insiden, dan sikap proaktif dalam membentuk respons adaptif terhadap risiko siber. Hal ini menunjukkan bahwa literasi keamanan siber tidak hanya relevan bagi generasi muda tetapi juga bagi berbagai kalangan profesional. Oleh karena itu, literasi keamanan siber yang masyarakat umum terutama pada generasi muda melalui kampanye edukasi tidak hanya penting untuk mencegah keterlibatan dalam aktivitas ilegal seperti *cracking, carding,* dan eksploitasi, tetapi juga untuk mengamankan diri dari ancaman tersebut dan menciptakan masyarakat yang lebih sadar akan risiko serta tanggung jawab hukum di era digital.

D. Regulasi dan Kebijakan Penegakan Hukum yang Lebih Kuat

Temuan ini juga mengindikasikan perlunya penguatan regulasi terkait keamanan digital, perlindungan data pribadi, dan penyalahgunaan sistem teknologi informasi. Pemerintah perlu mengimplementasikan kebijakan yang mendukung penegakan hukum, termasuk penyediaan sumber daya yang memadai untuk unit *cybercrime*, serta kolaborasi lintas negara untuk menangani forum-forum yang berbasis internasional.

Seperti halnya penanganan kasus *carding forum* dan *carding shop* ini secara yurisdiksi antarnegara melibatkan kerja sama antara negara-negara yang terlibat baik secara bilateral maupun multilateral. Hal ini dapat dilakukan melalui pertukaran informasi dan data, serta koordinasi antara lembaga penegak hukum dari masing-masing negara. Disamping itu, beberapa negara telah menandatangani perjanjian internasional untuk memerangi kejahatan siber, seperti (Convention on Cybercrime, 2001). Konvensi ini menetapkan kerangka hukum internasional untuk memerangi kejahatan siber dan mendorong kerja sama internasional dalam penanganannya. Adapun, penelitian Aminu (2024) merekomendasikan agar negara-negara anggota Interpol, mitra, dan pemangku kepentingan perlu memenuhi kewajiban keuangan mereka sesuai dengan Konstitusi Interpol agar dapat melaksanakan program dan kegiatan yang efektif dalam memerangi ancaman *cybercrime* seperti di Nigeria. Penelitian Buçaj & Idrizaj (2025) menyoroti pentingnya menyelaraskan definisi dan praktik hukum di

berbagai yurisdiksi, baik di negara-negara yang menganut sistem hukum *common law* maupun *civil law*, dalam menangani kejahatan siber yang sifatnya tanpa batas. Implikasi peneltiannya menjelaskan perlunya peningkatan kolaborasi antara sektor publik dan swasta dalam investigasi kejahatan siber juga ditekankan, di samping pentingnya menetapkan standar pengumpulan dan penyebaran data yang etis.

E. Pengawasan terhadap Distribusi Konten Eksplisit

Forum yang menyebarkan konten eksplisit dan ilegal harus ditangani secara tegas. Penyedia layanan *cloud storage* seperti Mega.nz dapat diminta untuk meningkatkan sistem pengawasan otomatis terhadap file yang melanggar hukum seperti dengan cara moderasi konten. Moderasi konten adalah proses memastikan konten yang dipublikasikan tidak melanggar aturan mana pun (ARTICLE 19, 2023). Selain itu, perlu ada sistem pelaporan publik yang memudahkan masyarakat melaporkan temuan konten ilegal kepada pihak berwenang.

Dengan langkah-langkah di atas, hasil analisis dapat digunakan secara efektif untuk memberantas aktivitas ilegal, melindungi masyarakat, dan menciptakan lingkungan digital yang lebih aman. Rekomendasi ini diharapkan dapat menjadi panduan bagi penegak hukum, pemangku kebijakan, dan pihak terkait lainnya dalam mengambil tindakan yang terarah dan komprehensif.

4.5.4. Validasi Kualitatif Konfirmasi Data

Pada bagian ini disajikan hasil validasi kualitatif konfirmasi berbasis triangulasi data. Triangulasi data adalah metode validasi kualitatif dengan membandingkan hasil analisis LDA terhadap berbagai sumber, seperti data asli *forum carding*, literatur penelitian terdahulu, dan hasil analisis manual, guna memastikan konsistensi, keandalan, serta mengurangi potensi bias interpretasi. Berikut pada tabel 4.17 dibawah ini dijelaskan secara lebih komprehensif:

Tabel 4.17 Hasil Validasi Kualitatif Konfirmasi

Carding Forum / Carding Shop	**Data Asli Forum (Kutipan)**	**Literatur / Penelitian Terdahulu**	**Temuan Analisis**
Altenen (Mega.nz leaks) (FORUM SxTN, 2025)	*"🔥 [PAID][DAILY][MEGA] 🔥 ❤️ ULTIMATE Snapchat Leaks , Over 2000X GIRLS ❤️ Part 120 –*	1. Daryabar et al. (2017): *MEGA app forensics*, artefak login, *upload, download, share, timestamp.*	Pada tahap ini terlihat pola penggunaan *cloud storage* Mega.nz sebagai sarana distribusi konten

Carding Forum / Carding Shop	**Data Asli Forum (Kutipan)**	**Literatur / Penelitian Terdahulu**	**Temuan Analisis**
	2025. Promosi akses gratis VPN (Express, NordVPN), Xbox GamePass, dan BINs 2025" (FORUMSxTN, 2025).	2. H. Mishra et al. (2022): *forensic cloud storage* (MEGA, Android & Windows 10). 3. Kang et al. (2024): *forensic methodology* pada E2EE *cloud storage* (MEGA), *metadata & selective acquisition*.	ilegal yang diperkuat dengan iklan layanan tambahan (VPN, BINs). Analisis forensik mengindikasikan bahwa aktivitas ini dapat ditelusuri melalui artefak digital berupa metadata *file*, catatan *login*, serta pola *sharing* tautan. Temuan ini menunjukkan keterhubungan erat antara forum kejahatan siber dengan pemanfaatan layanan penyimpanan berbasis awan untuk memfasilitasi anonimitas, monetisasi, dan distribusi konten terlarang.
Carding.store (Yagami, n.d.)	*"Amazon Giftcard Carding Tutorial… Buy fresh CC, use RDP/VPN matching CC owner's location, clear cookies, sign up Outlook, create Amazon account with CC, cart <$40, wait 48h, then purchase e-giftcards < $400 and send to own email."* (Yagami, n.d.)	1. Nguyen et al. (2025): Serangan modern mengeksploitasi infrastruktur AD, termasuk simulasi *real-world attack* untuk *fraud planning*. 2. Logie & Das (2025): *Darknet forum* adalah sarana *criminogenic learning, fraud planning,* dan *knowledge sharing* (termasuk pembayaran & *giftcards*). 3. Beju & Făt (2023): Evolusi *fraud* kartu	Tutorial menguraikan langkah teknis *carding* Amazon: mulai pembelian CC curian, penggunaan VPN, rekayasa akun, hingga konversi ke *e-giftcard* yang mudah dicairkan ke Bitcoin. Temuan: forum berfungsi sebagai *school of crime* dengan pembelajaran kolektif. Validasi

Carding Forum / Carding Shop	**Data Asli Forum (Kutipan)**	**Literatur / Penelitian Terdahulu**	**Temuan Analisis**
		meningkat pasca-pandemi; *internet payment* memperluas metode *carding*. 4. Siu et al. (2021): Forum bawah tanah menghubungkan *carding* dengan *currency exchange* (PayPal, Bitcoin). 5. S. Agarwal & Vasek (2025): *Smishing* dan *CNP fraud* memanfaatkan detail kartu curian untuk pembelian *daring*, pola serupa dengan *giftcard fraud.*	literatur mendukung bahwa carding *giftcard* adalah bentuk CNP *fraud* terstruktur, memanfaatkan *knowledge sharing forum* untuk memperluas skema kriminal ke arah konversi aset digital.
Astradumps.com (Astradumps, 2024)	*"EMV X2 Smart Card Reader Writer – Full Setting. Package includes drivers, SDK, software for card duplication."* (Astradumps, 2024)	1. Sheehan (2025): SDK EMV memberi keleluasaan *developer*, tetapi di *dark market* dijadikan paket kriminal. 2. Basin et al. (2021): EMV *flaws* dapat dimanfaatkan untuk *merchant fraud* & *cardholder fraud*. 3. Radu et al. (2022): *Relay attack & offline bypass* dapat dijalankan lewat kombinasi *reader* dan *smartphone.* 4. Nezhad et al. (2025): Analisis perbandingan Visa–Mastercard menyoroti celah keamanan sistem *open-loop contactless*.	Situs menekankan *driver*, SDK, dan *software* sebagai bagian dari paket. Temuan: *dark marketplace* ini meniru model e-*commerce* legal, tetapi mengedarkan *toolkit* EMV untuk kejahatan finansial. Validasi literatur menunjukkan penggunaan SDK & *software* EMV dalam konteks ilegal membuka peluang serangan *relay, cloning,* hingga *bypass otorisasi.*
Money-Heist.org (@cashout vendors, 2020a)	*"EMV X2 Smart Card Chip Reader Writer + Full Setting Package… compatible with JCOP J2A040/72K, J3A081,*	1. Sheehan (2025): *EMV software* mengelola transaksi *chip card* dengan *cryptogram* unik, namun SDK & *hosted solution* bisa	Situs menjual paket lengkap *cloning* EMV (*hardware* + *software* + kompatibilitas JCOP). Temuan:

Carding Forum / Carding Shop	Data Asli Forum (Kutipan)	Literatur / Penelitian Terdahulu	Temuan Analisis
	etc." (@cashout vendors, 2020a)	disalahgunakan untuk *cloning*. 2. Basin et al. (2021): *Flaw* EMV memungkinkan transaksi tanpa PIN & *offline fraud* via Android PoC. 3. Radu et al. (2022): *Relay attack* memungkinkan *pickpocket* nirkabel, bahkan *bypass* Apple Pay *lock screen* dengan *smartphone rooted*. 4. Nezhad et al. (2025): Identifikasi 7 vektor serangan pada EMV *contactless*, meliputi aplikasi, autentikasi, dan otorisasi transaksi.	*carding shop* mengkomodifikasi EMV kit sebagai "produk siap pakai", sehingga aktor kriminal dapat langsung melakukan *card cloning & bypass* otorisasi. Hal ini sejalan dengan temuan literatur mengenai kerentanan otorisasi, *relay*, dan *fraud contactless* pada EMV.

4.5.5. Pengambilan Keputusan: Evaluasi Kinerja *Framework*

Framework investigasi forensik *carding* yang digunakan dalam penelitian ini telah efektif dalam mendukung proses investigasi kejahatan *carding*. Langkah pertama dalam *framework* ini adalah penerimaan surat permohonan pemeriksaan digital forensik yang diberikan oleh pihak kepolisian, yang memberikan landasan hukum bagi proses investigasi. Proses pengumpulan data melalui *web scraping* dari *carding forum* dan *carding shop* sesuai dengan izin yang diberikan dalam surat permohonan, memberikan akses ke informasi yang relevan mengenai transaksi dan pola perilaku pelaku *carding*.

Proses pengumpulan data ini mengarah pada analisis lebih mendalam menggunakan teknik *profiling* forensik, yang memanfaatkan data yang diperoleh untuk mengidentifikasi pola perilaku dan karakteristik pelaku *carding*. Metode ini memungkinkan investigator untuk menggali lebih dalam mengenai modus operandi yang sering digunakan oleh pelaku carding, seperti cara mereka bertransaksi, jenis data yang mereka cari, serta alat dan teknik yang digunakan untuk menyembunyikan jejak digital mereka. Selain itu, penggunaan *natural language processing* (NLP) melalui teknik *latent dirichlet allocation* (LDA) dalam menganalisis data diskusi dan deskripsi pada *carding forum* dan *carding shop* turut

membantu untuk mengidentifikasi topik-topik utama yang sering dibahas, seperti strategi pengelolaan data curian, alat *carding* terbaru, taktik untuk menghindari deteksi serta akun pornografi alternatif.

Setelah data dianalisis, langkah berikutnya adalah penyusunan laporan temuan berdasarkan hasil analisis forensik, yang mendokumentasikan pola, karakteristik, dan modus operandi pelaku *carding* yang ditemukan selama proses investigasi. Laporan ini kemudian digunakan untuk koordinasi dengan penegak hukum dan pihak kepolisian, yang memungkinkan tindakan hukum selanjutnya untuk diambil berdasarkan bukti yang ditemukan. Koordinasi ini juga memastikan bahwa langkah-langkah investigasi dan penegakan hukum dilakukan secara tepat sesuai dengan regulasi yang berlaku.

Langkah terakhir dalam *framework* ini adalah evaluasi kinerja yang dilakukan untuk menilai efektivitas setiap tahap dalam mendukung investigasi dan penegakan hukum. Evaluasi ini mencakup penilaian terhadap metode ekstraksi data, pengolahan data, serta penyusunan laporan, yang berperan penting dalam mempercepat proses identifikasi dan penangkapan pelaku *carding*. Berdasarkan hasil evaluasi, dapat disimpulkan bahwa *framework* ini telah berhasil memberikan wawasan yang berharga dalam memahami perilaku pelaku carding dan mengidentifikasi tren terbaru dalam dunia kejahatan siber ini.

Secara keseluruhan, *framework* investigasi ini tidak hanya efektif dalam menjawab pertanyaan yang ingin dijawab, seperti pola dan modus operandi pelaku carding, tetapi juga mampu memberikan *insight* yang mendalam tentang cara pelaku *carding* beroperasi, serta topik-topik utama yang mereka diskusikan dalam forum dan situs *web carding*. Dengan demikian, *framework* ini sangat berperan dalam mendukung investigasi forensik yang lebih efisien dan terarah, sekaligus mempercepat langkah-langkah hukum dalam memberantas kejahatan *carding* termasuk koordinasi dengan lembaga penegak hukum untuk penangkapan dan pencegahan kerugian lebih lanjut. *Framework* ini berhasil mengidentifikasi pola diskusi kejahatan dunia maya, membantu regulator dan penegak hukum dalam memahami perilaku kriminal. *Framework* ini memiliki potensi aplikasi di dunia nyata, seperti memetakan aktivitas pencurian kartu di Kepulauan Canary meskipun investigasi lebih lanjut diperlukan untuk implementasi khusus kasus (Civil, n.d.; Lanzarote, 2024). Dengan memanfaatkan pengumpulan data dan teknik NLP, termasuk LDA, pendekatan ini mendukung investigasi komprehensif dan meningkatkan koordinasi untuk penangkapan dan pencegahan kejahatan, memperkuat kegunaannya dalam investigasi forensik digital.

Penerapan *framework* dalam penelitian ini juga sejalan dengan temuan pada penelitian Dunsin et al. (2024) yang menggarisbawahi pentingnya integrasi AI dan ML

dalam digital forensik, yang menawarkan *insight* tentang manfaat, kekurangan, dan implikasi yang lebih luas untuk mengatasi ancaman siber modern. Hasil penelitian ini juga sejalan dengan hasil penelitian Gazeau et al. (2024) yang menunjukkan kemampuan *parser* untuk menyederhanakan proses pengambilan informasi dengan memberikan penghematan waktu sebesar 92,12% dibandingkan dengan pendekatan manual.

Berbeda dengan penelitian yang dilakukan oleh Shahbazi & Byun (2022), yang lebih menekankan pada pengembangan *framework* teknis berbasis *machine learning* dan *blockchain*, penelitian ini mengintegrasikan dimensi penegakan hukum dalam pendekatan forensik digital. Hal ini menjadi pembeda utama dibandingkan dengan karya (Rao et al., 2021; Sonmez & Codal, 2024), yang cenderung fokus pada aspek teknis semata tanpa mempertimbangkan konteks legal dan operasional penegakan hukum. *Framework* ini menawarkan pendekatan yang lebih holistik dalam penanggulangan kejahatan *carding* karena mengintegrasikan aspek hukum ke dalam metodologi forensik, sehingga memperkuat efektivitas upaya penegakan hukum secara menyeluruh.

BAB 5
Kesimpulan dan Saran

5.1.Kesimpulan

Kesimpulan dari penelitian ini menunjukkan bahwa penerapan teknik investigasi *web scraping* pada *carding forum* dan *carding shop* berbasis *framework* investigasi *carding* telah efektif dalam mengumpulkan data yang relevan dan menganalisis aktivitas pelaku *cybercrime* dengan tepat. Melalui pengumpulan data menggunakan *web scraping*, peneliti dapat mengakses informasi dalam diskusi-diskusi *online* yang sering kali menjadi sumber utama bagi pelaku *cybercrime*. Selanjutnya, analisis deskriptif *profiling forensik* yang diterapkan pada dokumen hasil web scraping telah memberikan *insight* yang lebih dalam mengenai pola dan perilaku pelaku *carding*, yang pada gilirannya dapat membantu dalam merancang strategi pencegahan dan penanggulangan kejahatan yang lebih efektif.

Penerapan analisis *natural language processing* (NLP) dengan algoritma *latent dirichlet allocation* (LDA) juga memberikan kontribusi signifikan dalam menyaring topik-topik utama yang sering dibahas dalam *carding forum* dan *carding shop*. Hal ini memungkinkan para peneliti dan profesional digital forensik untuk lebih fokus pada isu-isu yang paling relevan, serta memahami dinamika percakapan yang terjadi di dalam komunitas tersebut. Berbasis analisis *natural language processing* (NLP) dengan algoritma *latent dirichlet allocation* (LDA), peneliti dapat mengidentifikasi pola-pola diskusi yang dapat menjadi indikasi adanya aktivitas kejahatan.

Berdasarkan hasil analisis konsistensi dan pergeseran *(pre vs post)*, penelitian ini juga menemukan dinamika penting dalam ekosistem forum ilegal. Pada Altenen Porn Section, pola utama sejak Desember 2024 hingga Agustus 2025 tetap konsisten pada kebocoran konten premium dan eksploitasi remaja, tetapi terdapat pergeseran menuju pelembagaan komunitas dengan sistem *vote, rating,* dan *branding* pengguna. Sementara itu, pada Carding Shop & Cracking Tutorial Threads, pola awal berupa *knowledge sharing* dalam bentuk *tutorial cracking bergeser* menjadi ekosistem pasar kriminal yang lebih profesional, dengan fokus pada perdagangan data finansial, layanan *drop account,* dan instrumen *carding* lainnya. Temuan ini menandakan adanya transformasi signifikan dari ruang berbagi menuju struktur komunitas dan pasar kejahatan siber yang lebih matang dan terorganisir.

Lebih lanjut, pengembangan dan penerapan *framework* investigasi forensik *carding* yang berbasis pada analisis dokumen hasil *web scraping* telah berhasil memberikan arahan yang sistematis dalam proses investigasi. *Framework* ini menyatukan seluruh tahapan investigasi mulai dari penerimaan surat permohonan investigasi, pengumpulan data, analisis, hingga evaluasi efektivitas *framework*, yang semuanya dilakukan secara terstruktur dan sesuai dengan standar prosedur yang berlaku.

Framework ini membantu meningkatkan efektivitas dan efisiensi proses investigasi *cybercrime*, serta memfasilitasi koordinasi yang lebih baik antara pihak-pihak terkait, seperti penegak hukum dan investigator digital forensik. Secara ilmiah, penelitian ini memberikan kontribusi penting dalam pengembangan pendekatan baru berbasis *data science* untuk investigasi forensik digital. Kebaruan utama terletak pada (1) integrasi teknik *web scraping* dengan *natural language processing* dan *descriptive forensic profiling* secara terpadu; (2) pemanfaatan algoritma *latent dirichlet allocation* (LDA) untuk mengungkap topik-topik tersembunyi dalam komunikasi *carding*; dan (3) penyajian hasil analisis dalam bentuk visualisasi data forensik yang komunikatif. Selain itu, penggunaan data dari *platform* internasional yang jarang diteliti menambah nilai literatur dan praktik forensik digital global, sekaligus memberikan kontribusi sosial dengan menyediakan metode investigasi efisien yang mengurangi risiko hukum dan etika bagi penyidik.

Secara keseluruhan, penelitian ini berhasil mengembangkan sebuah pendekatan yang lebih terorganisir dan berbasis data dalam menangani kejahatan di *carding forum* dan *carding shop*, yang dapat menjadi acuan bagi penelitian dan penerapan selanjutnya di bidang investigasi forensik digital.

5.2.Saran

Saran untuk penelitian selanjutnya adalah untuk memperluas cakupan penelitian dengan melibatkan analisis yang lebih mendalam terhadap jenis-jenis situs web ilegal lainnya, seperti forum atau toko daring yang terkait dengan aktivitas ilegal, termasuk perdagangan data pribadi, penipuan finansial, atau kegiatan *cybercrime* lainnya yang semakin berkembang, termasuk pada *platform* yang lebih sulit diakses atau tersembunyi, seperti *dark web* dan *deep web*. Penelitian selanjutnya juga dapat menggali lebih jauh mengenai perbedaan dalam pola dan modus operandi pelaku kejahatan di berbagai situs web ilegal tersebut, serta bagaimana perilaku mereka mungkin berbeda dari apa yang ditemukan di website ilegal yang lebih umum. Selain itu, penelitian selanjutnya juga dapat mencakup pengembangan algoritma yang lebih kompleks dalam analisis *natural language processing*

(NLP), seperti *dark web crawlers*, analisis metadata, dan *deep learning-based models*, untuk meningkatkan akurasi ekstraksi informasi dan identifikasi pola perilaku *cybercriminals*.

Penerapan metode yang lebih lanjut dalam analisis deskriptif *profiling* forensik juga dapat diperluas dengan memasukkan elemen-elemen baru, seperti analisis jejaring sosial pelaku yang terhubung melalui berbagai situs web ilegal, yang dapat memberikan *insight* lebih luas tentang kolaborasi antara pelaku dan cara mereka berkomunikasi untuk merencanakan kejahatan. Selain itu, penelitian mendatang juga dapat berfokus pada pengujian dan evaluasi *framework* investigasi forensik *cybercrime* dalam konteks yang lebih nyata, dengan melibatkan kolaborasi dengan instansi penegak hukum untuk menguji keefektifan implementasi *framework* dalam kasus-kasus dunia nyata.

Daftar Pustaka

@cashout vendors. (2020a). *EMV X2 Smart Card Chip Reader/Writer + Full Setting Package*. @cashout vendors. https://money-heist.org/product/emv-x2-smart-card-chip-reader-writer-full-setting-package/

@cashout vendors. (2020b). *Money-Heist.org Shop*. Money-Heist.org. https://money-heist.org/shop/

Adristi, F. I. (2024). *Tesis Magister Informatika Fikri*. Github. https://github.com/451Fikrie/Tesis-Magister-Informatika-Fikri

Agarwal, A., Iqbal, M., & Mitra, B. (2020). Survey of Various Techniques used for Credit Card Fraud Detection. *International Journal for Research in Applied Science & Engineering Technology*, *8*(7), 1642–1646. https://www.ijraset.com/fileserve.php?FID=30614

Agarwal, S., & Vasek, M. (2025). Card-Not-Present Fraud resulting from Smishing Attacks: An Experimental Study. *Proceedings of the New Security Paradigms Workshop (NSPW) 2025*, In press. https://discovery.ucl.ac.uk/id/eprint/10210719/

Agarwal, U., Rishiwal, V., Tanwar, S., & Yadav, M. (2024). Blockchain and crypto forensics: Investigating crypto frauds. *International Journal of Network Management*, *34*(2), e2255. https://doi.org/10.1002/nem.2255

Agrawal, H., Singh, S. P., Dixit, S., Nagdev, P., P., V., & Thaseen, S. (2024). A Digital Forensic Analysis of Profiling and Avoidance of Websites Disseminating Disinformation. *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*, 1–9. https://doi.org/10.1109/ic-ETITE58242.2024.10493661

Ahmad, W. (2008). Is Credit Card Fraud a Real Crime? Does it Really Cripple the E-Commerce Sector of E-Business? *2008 International Conference on Management of e-Commerce and e-Government*, 364–370. https://doi.org/10.1109/ICMECG.2008.99

Alam, T., & Gupta, R. (2024). Reviewing the Framework of Blockchain in Fake News Detection. *JOIN (Jurnal Online Informatika)*, *9*(2), 286–296. https://doi.org/10.15575/join.v9i2.1349

Alamsyah, A., Santoso, E., & Pranadita, N. (2025). Kajian Terhadap Kejahatan Carding Sebagai Bentuk Cybercrime Di Indonesia. *Iustitia Omnibus: Jurnal Ilmu Hukum*, *6*(2), 60–68. https://jurnal-pasca.unla.ac.id/iustitiaomnibus/article/view/189

Alsaraireh, N. (2025). Digital Forensics in the Age of Cybercrime: Challenges and Future

Directions. In R. K. Hamdan (Ed.), *Tech Fusion in Business and Society : Harnessing Big Data, IoT, and Sustainability in Business: Volume 2* (hal. 555–564). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-84636-6_48

Altenen. (n.d.). *Altenen Forums - Images & Videos & Porn Accounts?* Altenen. Diambil 21 Maret 2024, dari https://altenens.is/forums/images-videos-porn-accounts 🔞 .469197/

Alyahya, T., & Kausar, F. (2017). Snapchat Analysis to Discover Digital Forensic Artifacts on Android Smartphone. *Procedia Computer Science*, *109*, 1035–1040. https://doi.org/10.1016/j.procs.2017.05.421

Amato, F., Cozzolino, G., Moscato, V., & Moscato, F. (2019). Analyse digital forensic evidences through a semantic-based methodology and NLP techniques. *Future Generation Computer Systems*, *98*, 297–307. https://doi.org/https://doi.org/10.1016/j.future.2019.02.040

Aminu, A. M. (2024). International Criminal Police Organisation and the Challenges in the Fight against Cybercrime in Nigeria. *Kashere Journal of Politics and International Relations*, *2*(1), 48–56. https://journals.fukashere.edu.ng/index.php/kjpir/article/view/178

Arifisnti, I., Kustriyono, E., & Pramitasari, A. (2024). Pola Interogasi Penyidik terhadap Tersangka pada Berita Acara Pemeriksaan Kasus Delik Aduan Tinjauan Linguistik Forensik. *Parafrasa: Jurnal Bahasa, Sastra, dan Pengajaran*, *6*(1), 1–10. https://jurnal.unikal.ac.id/index.php/parafrasa/article/view/4668

ARTICLE 19. (2023). *Buku Panduan Moderasi Konten dan Kebebasan Berekspresi*. Uni Eropa & UNESCO. https://www.article19.org/wp-content/uploads/2024/03/BAHASA-Final-SM4P-Content-moderation-handbook-7-Aug-ID-translated-revised-022924_YHM.pdf

Ashraf, S. J., & Tilawat, M. (Ed.). (2024). *Credit Card Fraud Statistics: Losses to Explode $41 Billion by 2025!* VPNRanks.com. https://www.vpnranks.com/resources/credit-card-fraud-statistics/

Astradumps. (2023). *Astra Dumps Shop*. Astra Dumps. https://astradumps.com/shop/

Astradumps. (2024). *EMV X2 Smart Card Chip Reader/Writer + Full Setting*. Astradumps. https://astradumps.com/product/emv-x2-smart-card-chip-reader-writer-full-setting/

Ayanwale, M. A., Sanusi, I. T., Molefi, R. R., & Otunla, A. O. (2023). A Structural Equation Approach and Modelling of Pre-service Teachers' Perspectives of Cybersecurity Education. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-11973-5

Azhan, M., & Meraj, S. (2020). Credit Card Fraud Detection using Machine Learning and Deep Learning Techniques. *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, 514–518. https://doi.org/10.1109/ICISS49785.2020.9316002

Bar-Ilan, J. (2001). Data collection methods on the Web for infometric purposes — A review and analysis. *Scientometrics*, *50*(1), 7–32. https://doi.org/10.1023/A:1005682102768

Baraniuk, C. (2017). *AlphaBay and Hansa dark web markets shut down*. BBC. https://www.bbc.com/news/technology-40670010

Barker, K. J., D'Amato, J., & Sheridon, P. (2008). Credit card fraud: awareness and prevention. *Journal of Financial Crime*, *15*(4), 398–410. https://doi.org/10.1108/13590790810907236

Basheer, R., & Alkhatib, B. (2024). Conceptualizing Discussions on the Dark Web: An Empirical Topic Modeling Approach. *Complexity*, *2024*(1), 2775236. https://doi.org/https://doi.org/10.1155/2024/2775236

Basin, D., Sasse, R., & Toro-Pozo, J. (2021). The EMV Standard: Break, Fix, Verify. *2021 IEEE Symposium on Security and Privacy (SP)*, 1766–1781. https://doi.org/10.1109/SP40001.2021.00037

Beju, D.-G., & Făt, C.-M. (2023). Frauds in Banking System: Frauds with Cards and Their Associated Services. In M. V. Achim (Ed.), *Economic and Financial Crime, Sustainability and Good Governance. Contributions to Finance and Accounting* (hal. 31–52). Springer International Publishing. https://doi.org/10.1007/978-3-031-34082-6_2

Benhamou, Y. (2017). Website Blocking Injunctions Under Swiss Law: From Civil and Administrative Injunctions to Criminal Seizure or Forfeiture. *Expert Focus*, *11*, 885–893. http://archive-ouverte.unige.ch/unige:98862

Benjamin, V., Li, W., Holt, T., & Chen, H. (2015). Exploring threats and vulnerabilities in hacker web: Forums, IRC and carding shops. *2015 IEEE International Conference on Intelligence and Security Informatics (ISI)*, 85–90. https://doi.org/10.1109/ISI.2015.7165944

Bergman, J., & Popov, O. B. (2022). The Digital Detective's Discourse - A toolset for forensically sound collaborative dark web content annotation and collection. *Journal of Digital Forensics, Security and Law*, *17*, 1–25. https://doi.org/10.15394/jdfsl.2022.1740

Bergman, J., & Popov, O. B. (2023). Exploring Dark Web Crawlers: A Systematic Literature

Review of Dark Web Crawlers and Their Implementation. *IEEE Access*, *11*, 35914–35933. https://doi.org/10.1109/ACCESS.2023.3255165

Bhagat, S., & Pravin, D. P. (2023). Cybersecurity Awareness and Adaptive Behavior: Does Prior Exposure Lead to Adaptive Behavior? *AMCIS 2023 Proceedings*, 23.

Blei, D. M. (2012, April). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022. https://jmlr.csail.mit.edu/papers/v3/blei03a.html

Böhm, F., Englbrecht, L., & Pernul, G. (2020). Designing a Decision-Support Visualization for Live Digital Forensic Investigations. In A. Singhal & J. Vaidya (Ed.), *Data and Applications Security and Privacy XXXIV* (hal. 223–240). Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-030-49669-2_13

Bollikonda, V. B., & Kiran, K. (2024). Reconnaissance on Dark Web Trades and Traders Activities for Investigation. *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)*, 1649–1653. https://doi.org/10.23919/INDIACom61295.2024.10498813

Bowyer, S. (2021). The Wayback Machine: notes on a re-enchantment. *Archival Science*, *21*(1), 43–57. https://doi.org/10.1007/s10502-020-09345-w

Bracciale, L., Coni, M., Loreti, P., Raso, E., & Bianchi, G. (2025). Forgotten & Reclaimed: Detecting and Preventing Subdomain Takeover in the Italian Medical Landscape. *Joint National Conference on Cybersecurity (ITASEC & SERICS 2025)*. https://ceur-ws.org/Vol-3962/paper37.pdf

Broucke, S. vanden, & Baesens, B. (2018). From Web Scraping to Web Crawling. In S. vanden Broucke & B. Baesens (Ed.), *Practical Web Scraping for Data Science: Best Practices and Examples with Python* (hal. 155–172). Apress. https://doi.org/10.1007/978-1-4842-3582-9_6

Buçaj, E., & Idrizaj, K. (2025). The need for cybercrime regulation on a global scale by the international law and cyber convention. *Multidisciplinary Reviews*, *8*(1), 2025024. https://doi.org/10.31893/multirev.2025024

Buil-Gil, D., Lord, N., & Barrett, E. (2021). The Dynamics of Business, Cybersecurity and Cyber-Victimization: Foregrounding the Internal Guardian in Prevention. *Victims & Offenders*, *16*(3), 286–315. https://doi.org/10.1080/15564886.2020.1814468

Chen, H., He, M., & Peng, L. (2025). Understanding online shopping fraud among Chinese

elderly: Extending routine activity theory in the online context. *Telematics and Informatics*, *96*, 102208. https://doi.org/10.1016/j.tele.2024.102208

Chen, J., He, S., & Yang, X. (2023). Platform Loophole Exploitation, Recovery Measures, and User Engagement: A Quasi-Natural Experiment in Online Gaming. *Information Systems Research*, *35*(4), 1609–1633. https://doi.org/10.1287/isre.2020.0416

Cheng, S.-T., Horng, G.-J., Hsu, C.-W., & Su, Z.-Y. (2024). Per-user network access control kernel module with secure multifactor authentication. *The Journal of Supercomputing*, *80*(1), 970–1008. https://doi.org/10.1007/s11227-023-05480-0

Chowdhary, K. R. (2020). Natural Language Processing. In K. R. Chowdhary (Ed.), *Fundamentals of Artificial Intelligence* (hal. 603–649). Springer Nature India Private Limited. https://doi.org/10.1007/978-81-322-3972-7_19

Christian, J., Valiveti, S., & Jain, S. (2022). Profiling Cyber Crimes from News Portals Using Web Scraping. In P. K. Singh, S. T. Wierzchoń, J. K. Chhabra, & S. Tanwar (Ed.), *Futuristic Trends in Networks and Computing Technologies* (hal. 1007–1016). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-5037-7_72

Civil, G. (n.d.). *Police Warn of 'Carding' scam in the Canary Islands costing victims thousands*. Cana. https://www.canarianweekly.com/posts/Police-warn-of-carding-scam-in-the-Canary-Islands-costing-victims-thousands

Convention on Cybercrime, Pub. L. No. Budapest, 23.XI.2001 (2001). https://rm.coe.int/1680081561

Coscia, A., Dentamaro, V., Galantucci, S., Maci, A., & Pirlo, G. (2024). PROGESI: A PROxy Grammar to Enhance Web Application Firewall for SQL Injection Prevention. *IEEE Access*, *12*, 107689–107703. https://doi.org/10.1109/ACCESS.2024.3438092

Cullen, R., Heitkemper, E., Backonja, U., Bekemeier, B., & Kong, H.-K. (2024). Designing an infographic webtool for public health. *Journal of the American Medical Informatics Association*, *31*(2), 342–353. https://doi.org/10.1093/jamia/ocad105

Darmadi, A. A. N. O. Y., & Dananjaya, N. S. (2024). Authority of the Financial Transaction Analysis Reporting Center in Tracing Hidden Trading Crimes. *Sociological Jurisprudence Journal*, *7*(1), 8–14. https://doi.org/10.22225/scj.7.1.2024.8-14

Daryabar, F., Dehghantanha, A., & Choo, K.-K. R. (2017). Cloud storage forensics: MEGA as a case study. *Australian Journal of Forensic Sciences*, *49*(3), 344–357. https://doi.org/10.1080/00450618.2016.1153714

Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., Štajdohar, M., Umek, L., Žagar, L., Žbontar, J.,

Žitnik, M., & Zupan, B. (2013). Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research*, *14*(71), 2349–2353. https://www.jmlr.org/papers/v14/demsar13a.html

Dunsin, D., Ghanem, M. C., Ouazzane, K., & Vassilev, V. (2024). A comprehensive analysis of the role of artificial intelligence and machine learning in modern digital forensics and incident response. *Forensic Science International: Digital Investigation*, *48*, 301675. https://doi.org/10.1016/j.fsidi.2023.301675

Fanni, S. C., Febi, M., Aghakhanyan, G., & Neri, E. (2023). Natural Language Processing. In M. E. Klontzas, S. C. Fanni, & E. Neri (Ed.), *Introduction to Artificial Intelligence* (hal. 87–99). Springer International Publishing. https://doi.org/10.1007/978-3-031-25928-9_5

Federal Trade Commission. (2025). *Identity Theft Reports*. Public Tableau. https://public.tableau.com/app/profile/federal.trade.commission/viz/IdentityTheftReports/TheftTypesOverTime

Fernandez, S., Hureau, O., Duda, A., & Korczynski, M. (2024). WHOIS Right? An Analysis of WHOIS and RDAP Consistency. In P. Richter, V. Bajpai, & E. Carisimo (Ed.), *Passive and Active Measurement. PAM 2024. Lecture Notes in Computer Science, vol 14537* (hal. 206–231). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-56249-5_9

Flowers, S. (2008). Harnessing the hackers: The emergence and exploitation of Outlaw Innovation. *Research Policy*, *37*(2), 177–193. https://doi.org/10.1016/j.respol.2007.10.006

FORUMSxTN. (2025). ?*[PAID][DAILY][MEGA]? 💔? ULTIMATE Snapchat Leaks , Over 2000X GIRLS 💔? Part 120 - 2025*. Altenen. https://altenens.is/threads/fire-paid-daily-mega-fireheart-on-fire-ultimate-snapchat-leaks-over-2000x-girlsheart-on-fire-part-120-2025.2738395/

Frosio, G., & Bulayenko, O. (2021). Website blocking injunctions in flux: static, dynamic and live. *Journal of Intellectual Property Law & Practice*, *16*(10), 1127–1143. https://doi.org/10.1093/jiplp/jpab107

Gazeau, V., Gupta, K., & An, M. K. (2024). Enhancing Social Media Data Collection for Digital Forensic Investigations: A Web Parser Approach. *2024 International Conference on Computer, Information and Telecommunication Systems (CITS)*, 1–7. https://doi.org/10.1109/CITS61189.2024.10607983

Ghugare, R., Thakare, A., Deshmukh, P., Bhatia, Y., & Dhengre, A. R. (2024). Web Scraping Illicit Bitcoin Addresses. *2024 2nd International Conference on Computer, Communication and Control (IC4)*, 1–4. https://doi.org/10.1109/IC457434.2024.10486746

Gong, W., Lee, C. S., Li, S., Adkison, D., Li, N., Wu, L., & Ye, X. (2025). Cyber victimization in hybrid space: an analysis of employment scams using natural language processing and machine learning models. *Journal of Crime and Justice*, 1–22. https://doi.org/10.1080/0735648X.2024.2448804

Grabosky, P. (2014). *The evolution of cybercrime, 2004-2014* (2014/58; RegNet Research Paper No. 2014/58). https://doi.org/10.2139/ssrn.2535605

Greiner, L. (2009). Sniper Forensics. *NetWorker*, *13*(4), 8–10. https://doi.org/10.1145/1655737.1655740

Griné, T., & Teixeira Lopes, C. (2023). A Social Media Tool for Domain-Specific Information Retrieval - A Case Study in Human Trafficking. In I. Koprinska, P. Mignone, R. Guidotti, S. Jaroszewicz, H. Fröning, F. Gullo, P. M. Ferreira, D. Roqueiro, G. Ceddia, S. Nowaczyk, J. Gama, R. Ribeiro, R. Gavaldà, E. Masciari, Z. Ras, E. Ritacco, F. Naretto, A. Theissler, P. Biecek, … S. Pashami (Ed.), *Machine Learning and Principles and Practice of Knowledge Discovery in Databases. ECML PKDD 2022. Communications in Computer and Information Science, vol 1752* (hal. 23–38). Springer Nature Switzerland.

Guntara, R. G., Kashira, F. B., Amri, T. K., Restu, L. B., & Susanto, F. R. (2024). Analisis Penjualan Handphone di Tokopedia dengan Teknik Web Scraping Menggunakan Python pada Google Colab. *ULIL ALBAB : Jurnal Ilmiah Multidisiplin*, *3*(4), 69–75. https://journal-nusantara.com/index.php/JIM/article/view/3200

Gupta, V., & Zhang, Y. (2025). From Snap to Evidence: Snapchat Footprint on iOS. *2025 13th International Symposium on Digital Forensics and Security (ISDFS)*, 1–6. https://doi.org/10.1109/ISDFS65363.2025.11012030

Han, J., Kim, J., & Lee, S. (2020). *5W1H-based Expression for the Effective Sharing of Information in Digital Forensic Investigations*. https://doi.org/10.48550/arXiv.2010.15711

Hegarty, Kieran. (2023). Imagining permanence on the web: Tracing the meanings of long-term preservation among the subjects of web archives. *New Media & Society*, *27*(2), 898–913. https://doi.org/10.1177/14614448231187031

Hidayat, E. A. (2020). Kewenangan Penyadapan Badan Narkotika Nasional dalam

Perspektif Undang-Undang Narkotika dan Undang-Undang Informasi dan Transaksi Elektronik. *Tadulako Master Law Journal*, *4*(2), 129–145. http://103.245.72.41/index.php/TMLJ/article/view/197

Huie, K., Butler, M., & Percy, A. (2024). Identifying trends and patterns in offending and victimization on Snapchat: a rapid review. *Security Journal*, *37*(3), 903–920. https://doi.org/10.1057/s41284-023-00400-6

Invision Community. (n.d.). *carding.store - Cracking Tutorials*. carding.store. Diambil 8 Juli 2024, dari https://carding.store/forum/20-cracking-tutorials/

Jamil, A. bin, Johari, R. J., Zarefar, A., & Yudi, M. M. (2024). An analysis of suspicious transaction reporting decisions in Malaysia's money services business. *Edelweiss Applied Science and Technology*, *8*(1), 24–32. https://doi.org/10.55214/25768484.v8i1.413

Jin, P., Kim, N., Lee, S., & Jeong, D. (2024). Forensic investigation of the dark web on the Tor network: pathway toward the surface web. *International Journal of Information Security*, *23*(1), 331–346. https://doi.org/10.1007/s10207-023-00745-4

Jirovský, V., Pastorek, A., Mühlhäuser, M., & Tundis, A. (2018). Cybercrime and Organized Crime. *Proceedings of the 13th International Conference on Availability, Reliability and Security*, 1–5. https://doi.org/10.1145/3230833.3233288

Kailas, S., & Roopalakshmi, R. (2025). 'Think Before You Click' - Malicious URL Detection in Cybersecurity: A Systematic Review and Research Roadmap. *IEEE Access*, 1. https://doi.org/10.1109/ACCESS.2025.3601387

Kalacska, M., & Bouchard, M. (2011). Using police seizure data and hyperspectral imagery to estimate the size of an outdoor cannabis industry. *Police Practice and Research*, *12*(5), 424–434. https://doi.org/10.1080/15614263.2010.536722

Kang, J., Kim, J., Lee, S., & Park, J. (2024). Forensic Approaches for End-to-End Encryption Cloud Storage Services: MEGA as a Case Study. *Arab Journal of Forensic Sciences and Forensic Medicine*, *6*(Special Issue (ASFSFM 2023)), 171–190. https://repository.nauss.edu.sa//handle/123456789/67232

Khder, M. A. (2021). Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application. *International Journal of Advances in Soft Computing and its Applications*, *13*(3), 144–168. https://doi.org/10.15849/IJASCA.211128.11

Koo, K., Park, M., & Yoon, B. (2024). A Suspicious Financial Transaction Detection Model Using Autoencoder and Risk-Based Approach. *IEEE Access*, *12*, 68926–68939. https://doi.org/10.1109/ACCESS.2024.3399824

Koops, B.-J. (2011). *The Internet and its Opportunities for Cybercrime* (09/2011; Tilburg Law School Legal Studies Research Paper Series No. 09/2011). https://doi.org/10.2139/ssrn.1738223

Kopel, K. (2013). Operation Seizing Our Sites: How the Federal Government is Taking Domain Names Without Prior Notice. *Berkeley Technology Law Journal*, *28*(4), 859–900. https://doi.org/10.15779/Z384Q3M

Kshetri, N. (2006). The simple economics of cybercrimes. *IEEE Security & Privacy*, *4*(1), 33–39. https://doi.org/10.1109/MSP.2006.27

Lanzarote. (2024). *Police Warn of 'Carding' Scam in The Canary Islands Costing Victims Thousands*. Canarian Weekly. https://www.canarianweekly.com/posts/Police-warn-of-carding-scam-in-the-Canary-Islands-costing-victims-thousands

Levy, Y., & Gafni, R. (2021). Introducing the concept of cybersecurity footprint. *Information & Computer Security*, *29*(5), 724–736. https://doi.org/10.1108/ICS-04-2020-0054

Li, W., Chen, H., & Nunamaker Jr., J. F. (2016). Identifying and Profiling Key Sellers in Cyber Carding Community: AZSecure Text Mining System. *Journal of Management Information Systems*, *33*(4), 1059–1086. https://doi.org/10.1080/07421222.2016.1267528

Liddy, E. D. (2001). Natural Language Processing. In *Encyclopedia of Library and Information Science* (2 ed.). Marcel Decker, Inc. https://surface.syr.edu/istpub/63/

Lo, R.-T., Hwang, W.-J., & Tai, T.-M. (2025). SQL Injection Detection Based on Lightweight Multi-Head Self-Attention. In *Applied Sciences* (Vol. 15, Nomor 2, hal. 571). https://doi.org/10.3390/app15020571

Logie, K., & Das, S. (2025). Lessons learned from Dread darknet communities: How and why are fraudsters targeting the elderly to be victims or accomplices? *Criminology & Public Policy*, *24*(2), 237–271. https://doi.org/10.1111/1745-9133.12684

Luthfi, A. (2024). *Documentation and Reporting ISO 27042:2015* (hal. 29). Universitas Islam Indonesia.

Malik, A. W., Bhatti, D. S., Park, T. J., Ishtiaq, H. U., Ryou, J. C., & Kim, K.-I. (2024). Cloud Digital Forensics: Beyond Tools, Techniques, and Challenges. *Sensors*, *24*(2), 433. https://doi.org/10.3390/s24020433

Malik, M. S., & Islam, U. (2019). Cybercrime: an emerging threat to the banking sector of Pakistan. *Journal of Financial Crime*, *26*(1), 50–60. https://doi.org/10.1108/JFC-11-2017-0118

Maybir, J., & Chapman, B. (2021). Web scraping of ecstasy user reports as a novel tool for detecting drug market trends. *Forensic Science International: Digital Investigation*, *37*, 301172. https://doi.org/10.1016/j.fsidi.2021.301172

Mazurczyk, W., & Caviglione, L. (2021, Februari). Cyber reconnaissance techniques. *Communications of the ACM*, *64*(3), 86–95. https://doi.org/10.1145/3418293

Mechri, A., Ferrag, M. A., & Debbah, M. (2025). SecureQwen: Leveraging LLMs for vulnerability detection in python codebases. *Computers & Security*, *148*, 104151. https://doi.org/10.1016/j.cose.2024.104151

Microsoft. (2024). *Microsoft Excel* (365). Microsoft. https://www.microsoft.com/en-in/microsoft-365/excel

Mishra, H., Sihag, V., Choudhary, G., Dragoni, N., & You, I. (2022). Cloud Storage Client Forensic: Analysis of MEGA Cloud. In P. K. Singh, S. T. Wierzchoń, J. K. Chhabra, & S. Tanwar (Ed.), *Futuristic Trends in Networks and Computing Technologies* (hal. 1099–1110). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-5037-7_79

Mishra, S., & Bajahzar, M. A. (2024). Cloud Forensic Artefacts: Digital Forensics Registry Artefacts discovered from Cloud Storage Application. *International Journal of Computing and Digital Systems*, *16*(1), 13–27. https://iiict.uob.edu.bh/IJCDS/papers/IJCDS160102_1570906411.pdf

Mooney, S. J., Westreich, D. J., & El-Sayed, A. M. (2015). Commentary: Epidemiology in the era of big data. *Epidemiology (Cambridge, Mass.)*, *26*(3), 390–394. https://doi.org/10.1097/EDE.0000000000000274

Muehlethaler, C., & Albert, R. (2021). Collecting data on textiles from the internet using web crawling and web scraping tools. *Forensic Science International*, *322*, 110753. https://doi.org/10.1016/j.forsciint.2021.110753

Narasimhan, P. K., Bhosale, C., Pervez, M. H., Naqvi, N. Z., Ecevit, M. I., Schwarz, K., & Creutzburg, R. (2023). Open-Source Intelligence (OSINT) Investigation in Facebook. *Electronic Imaging*, *35*(3), 357-1-357–12. https://doi.org/10.2352/EI.2023.35.3.MOBMU-357

Nasraoui, O., & Krishnapuram, R. (2002). AN EVOLUTIONARY APPROACH TO MINING ROBUST MULTI-RESOLUTION WEB PROFILES AND CONTEXT SENSITIVE URL ASSOCIATIONS. *International Journal of Computational Intelligence and Applications*, *2*(3), 339–348. https://doi.org/10.1142/S1469026802000646

Ndubuisi, O. J., Adene, G., Sunday, B. T., Mbonu, C. E., & Gift-Adene, A. U. (2024). Digitally improving UK police surveillance and incidence response using real-time crowd reporting app: Digipolice. *Global Journal of Engineering and Technology Advances*, *18*(3), 124–138. https://doi.org/10.30574/gjeta.2024.18.3.0048

Nezhad, M. M., Hao, F., Epiphaniou, G., Maple, C., & Yunusov, T. (2025). *SoK: Security of EMV Contactless Payment Systems*. https://doi.org/10.48550/arXiv.2504.12812

Nguyen, G.-H., Nguyen, K., Luu, M.-T., Nguyen, A.-N., & Ngo, T.-S. (2025). Automated Framework for Active Directory Security and Compliance through Continuous Monitoring and Attack Path Validation. *2025 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)*, 287–292. https://doi.org/10.1109/ECTIDAMTNCON64748.2025.10962057

Nigam, H., & Biswas, P. (2021). From Web Scraping to Web Crawling. In A. Choudhary, A. P. Agrawal, R. Logeswaran, & B. Unhelkar (Ed.), *Applications of Artificial Intelligence and Machine Learning* (hal. 97–112). Springer Singapore.

Noval, S. M. R., Soecipto, Jamaludin, A., Saputra, D. D., Munifah, N., Nurhasanah, Raswanti, P. S., & Lestia, S. N. (2023). The Fusion of Blockchain, Pornography and Human Trafficking in A Global Digital Dragnet That Forms The Online Child Sex Trafficking. *Russian Law Journal*, *11*(5s), 1–19. https://cyberleninka.ru/article/n/the-fusion-of-blockchain-pornography-and-human-trafficking-in-a-global-digital-dragnet-that-forms-the-online-child-sex-trafficking

Nurseno, M., Aditiawarman, U., Maarif, H. A. Q., & Mantoro, T. (2024). Detecting Hidden Illegal Online Gambling on .go.id Domains Using Web Scraping Algorithms. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *23*(2), 365–378. https://doi.org/10.30812/matrik.v23i2.3824

P, B. S., V, S., U, V., & S, Y. (2021). Identification of URL Fuzzing and Subdomain Enumeration Using Raccoon Tool. *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, 1620–1625. https://doi.org/10.1109/ICOEI51242.2021.9453002

Parida, S., Neeraj, Rawat, H., & Agrawal, R. (2025). Decoding Cyber Fraud Schemes and the Rise of 'Fraud-as-a-Service.' *2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN)*, 538–542. https://doi.org/10.1109/CICTN64563.2025.10932616

Peretti, K. (2009). Data Breaches: What the Underground World of Carding Reveals. *Santa*

Clara High Tech Law Journal, *25*(2), 375–413. https://digitalcommons.law.scu.edu/chtlj/vol25/iss2/4/

Pogue, C. (2010). Sniper Forensics “One Shot, One Kill.” In *DEFCON 18* (hal. 1–33). DEF CON Communications, Inc. https://archives.sector.ca/presentations09/Sector_SniperForensics92909_final%282%29.pdf

Pringle, J. K., Ruffell, A., Styles, P., Stringfellow, M., Stimpson, I. G., Banham, S. G., Wisniewski, K. D., Owen, S., Hobson, L., & Thompson, J. (2024). Forensic geoscience non-invasive detection and characterisation of underground clandestine complexes, bunkers, tunnels and firing ranges. *Forensic Science International*, *359*, 112033. https://doi.org/10.1016/j.forsciint.2024.112033

Python Software Foundation. (2001). *Python* (3.12.2). Python Software Foundation. https://www.python.org/

Radu, A.-I., Chothia, T., Newton, C. J. P., Boureanu, I., & Chen, L. (2022). Practical EMV Relay Protection. *2022 IEEE Symposium on Security and Privacy (SP)*, 1737–1756. https://doi.org/10.1109/SP46214.2022.9833642

Rajiv, S., & Navaneethan, C. (2021). Keyword weight optimization using gradient strategies in event focused web crawling. *Pattern Recognition Letters*, *142*, 3–10. https://doi.org/10.1016/j.patrec.2020.12.003

Rao, G. M., Reddy, B. R., & Vishnu, P. (2021). Smart Web Investigation Framework. In J. Singh, S. Kumar, & U. Choudhury (Ed.), *Innovations in Cyber Physical Systems* (hal. 305–314). Springer Singapore. https://doi.org/10.1007/978-981-16-4149-7_27

Rashid, S. M. Z. U., Kamrul, M. I., & Islam, A. (2019). Understanding the Security Threats of Esoteric Subdomain Takeover and Prevention Scheme. *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 1–4. https://doi.org/10.1109/ECACE.2019.8679122

Ridwan, M., AM, S., Ulum, B., & Muhammad, F. (2021). Pentingnya Penerapan Literature Review pada Penelitian Ilmiah. *Jurnal Masohi*, *2*(1), 42–51. https://doi.org/10.36339/jmas.v2i1.427

Rodrigues, F. B., Giozza, W. F., Albuquerque, R. de O., & Villalba, L. J. G. (2024). Natural Language Processing Applied to Forensics Information Extraction With Transformers and Graph Visualization. *IEEE Transactions on Computational Social Systems*, *11*(4), 4727–4743. https://doi.org/10.1109/TCSS.2022.3159677

Ruiz, R., Winter, R., Rosa, F. de F., Shukla, P., & Kazemian, H. (2023). Brazil Method of

Anti-Malware Evaluation and Cyber Defense Impacts. *Journal of Applied Security Research*, *18*(4), 925–941. https://doi.org/10.1080/19361610.2022.2104104

Saurkar, A. V, Pathare, K. G., & Gode, S. A. (2018). An Overview On Web Scraping Techniques And Tools. *International Journal on Future Revolution in Computer Science & Communication Engineering*, *4*(4), 363–367. http://www.ijfrcsce.org/index.php/ijfrcsce/article/view/1529

ScrapeOps. (2024). *Differences of Web Scraping Vs Web Crawling Explained*. ScrapeOps. https://scrapeops.io/web-scraping-playbook/web-scraping-vs-web-crawling/

Serttaş, Z., & Al-Turjman, F. (2025). Usability of Cloud-Based Applications in Digital Forensics: An Experimental Study on Image Acquisition and Digital Evidence Preservation Processes. *Near East University Journal for Artificial Intelligence and Internet of Things*, *4*(2). https://dergi.neu.edu.tr/index.php/aiit/article/view/1092

Shahbazi, Z., & Byun, Y.-C. (2022). NLP-Based Digital Forensic Analysis for Online Social Network Based on System Security. *International Journal of Environmental Research and Public Health*, *19*(12), 7027. https://doi.org/10.3390/ijerph19127027

Sharma, D., Mittal, R., Sekhar, R., Shah, P., & Renz, M. (2023). A bibliometric analysis of cyber security and cyber forensics research. *Results in Control and Optimization*, *10*, 100204. https://doi.org/10.1016/j.rico.2023.100204

Sharmila, S., & Aparna, C. (2024). Tracing footprints of anti-forensics and assuring secured data transmission in the cloud using an effective ECCDH and Kalman Filter. *Journal of Network and Computer Applications*, *221*, 103762. https://doi.org/10.1016/j.jnca.2023.103762

Sheehan, A. (2025). *What Is EMV Software and Why Do You Need It?* Shopify. https://www.shopify.com/retail/emv-software

Similarweb LTD. (2024). *Top 10 altenens.is Competitors*. Similarweb LTD. https://www.similarweb.com/website/altenens.is/competitors/

Siricharoen, W. V., & Siricharoen, N. (2018). Infographic Utility in Accelerating Better Health Communication. *Mobile Networks and Applications*, *23*(1), 57–67. https://doi.org/10.1007/s11036-017-0900-3

Sirisuriya, S. de S. (2015). A Comparative Study on Web Scraping. *Proceedings of 8th International Research Conference*, 135–140. http://192.248.104.6/bitstream/handle/345/1051/com-059.pdf?sequence=1&isAllowed=y

Siu, G. A., Collier, B., & Hutchings, A. (2021). Follow the money: The relationship between

currency exchange and illicit behaviour in an underground forum. *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 191–201. https://doi.org/10.1109/EuroSPW54576.2021.00027

Smadi, B. Al, & Min, M. (2020). A Critical review of Credit Card Fraud Detection Techniques. *2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 732–736. https://doi.org/10.1109/UEMCON51285.2020.9298075

Soesilo, R. (1985). *Membuat Berita Acara dan Laporan Polisi Menurut KUHAP*. Politeia.

Sonmez, E., & Codal, K. S. (2024). Analyzing a Dark Web forum page in the context of terrorism: a topic modeling approach. *Security Journal*. https://doi.org/10.1057/s41284-024-00421-9

Steinmetz, K. F., Schaefer, B. P., Brewer, C. G., & Kurtz, D. L. (2023). The Role of Computer Technologies in Structuring Evidence Gathering in Cybercrime Investigations: A Qualitative Analysis. *Criminal Justice Review*, 07340168231161091. https://doi.org/10.1177/07340168231161091

Sugiyono. (2019). *Metode Penelitian Pendidikan (Kuantitatif, Kualitatif, Kombinasi, R&D dan Penelitian Tindakan)* (A. Nuryanto (Ed.); 3 ed.). Alfabeta.

SysNucleus. (2024). *WebHarvy* (7.3.0.222). SysNucleus. https://www.webharvy.com/download.html

Szigeti, Á., Frank, R., & Kiss, T. (2024). Contribution to the harm assessment of darknet markets: topic modelling drug reviews on Dark0de Reborn. *Crime Science*, *13*(1), 13. https://doi.org/10.1186/s40163-024-00211-z

Takyi, K., Gyening, R.-M. O. M., Kobinnah, M., Boateng, M. A., & Boadu-Acheampong, S. (2025). Enhancing SQL Injection Detection with Long Short-Term Memory Networks in Deep Learning. *International Journal of Open Information Technologies*, *13*(1), 7–13. http://www.injoit.org/index.php/j1/article/view/1978

Teodoro, N., Gonçalves, L., & Serrão, C. (2015). NIST CyberSecurity Framework Compliance: A Generic Model for Dynamic Assessment and Predictive Requirements. *2015 IEEE Trustcom/BigDataSE/ISPA*, *1*, 418–425. https://doi.org/10.1109/Trustcom.2015.402

The pandas development team. (2020). *pandas-dev/pandas: Pandas* (2.2.1). Zenodo. https://doi.org/10.5281/zenodo.3509134

U.S. Attorney's Office, N. D. of G. (2017). *DOJ Announces Today the Takedown of Alphabay, the Largest Online 'Dark Market.'* U.S. Attorney's Office, Northern District

of Georgia. https://www.justice.gov/usao-ndga/pr/doj-announces-today-takedown-alphabay-largest-online-dark-market

Undang-undang (UU) No. 44 Tahun 2008 Tentang Pornografi, Pub. L. No. UU No. 44 Tahun 2008, 23 (2008). https://peraturan.bpk.go.id/Details/39740

Undang-undang (UU) No. 8 Tahun 2010 Tentang Pencegahan dan Pemberantasan Tindak Pidana Pencucian Uang, Pub. L. No. UU Nomor 8 Tahun 2010 (2010). https://peraturan.bpk.go.id/Details/38547/uu-no-8-tahun-2010

Undang-Undang Republik Indonesia Nomor 1 Tahun 2024 Tentang Perubahan Kedua Atas Undang-Undang Nomor 11 Tahun 2008 Tentang Informasi dan Transaksi Elektronik, Pub. L. No. UU Nomor 1 Tahun 2024 (2024). https://peraturan.bpk.go.id/Details/274494/uu-no-1-tahun-2024

Undang-undang Republik Indonesia Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik, Pub. L. No. UU Nomor 11 Tahun 2008 tentang ITE (2008). https://peraturan.bpk.go.id/Details/37589/uu-no-11-tahun-2008

Undang-Undang Republik Indonesia Nomor 19 Tahun 2016 Tentang Perubahan Atas Undang-Undang Nomor 11 Tahun 2008 Tentang Informasi dan Transaksi Elektronik, Pub. L. No. UU Nomor 19 Tahun 2016 (2016). https://peraturan.bpk.go.id/Details/37582/uu-no-19-tahun-2016

Valiño Ces, A. (2024). The Importance of the Computer Undercover Agent as an Investigative Measure Against Cybercrime: A Special Reference to Child Pornography Crimes. In F. A. C. P. de Andrade, P. M. F. Freitas, & J. R. de S. C. de Abreu (Ed.), *Legal Developments on Cybersecurity and Related Fields* (1 ed., hal. 145–165). Springer International Publishing. https://doi.org/10.1007/978-3-031-41820-4_9

Vasoya, Y., Modi, T., Patel, O., Kotak, D., Shah, K., & Sabale, K. (2024). A Comprehensive Exploration to Cybercrimes Investigation Techniques. *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)*, 1046–1053. https://doi.org/10.23919/INDIACom61295.2024.10498752

Wang, Fangzhou, Dickinson, Timothy, & Ghazi-Tehrani, Adam. (2025). Not All Money Is the Same: The Meanings of Money in Online Fraud. *Crime & Delinquency*, 00111287251321442. https://doi.org/10.1177/00111287251321442

Whittaker, Jack M, McGuire, Michael R, & Lazarus, Suleman. (2025). Conversations with deviant website developers: A case study of online shopping fraud enablers. *Journal of Criminology*, 26338076251321844. https://doi.org/10.1177/26338076251321844

Wiratmoko, G., Thamrin, H., & Pamungkas, E. W. (2025). Performance of Machine

Learning Algorithms on Automatic Summarization of Indonesian Language Texts. *JOIN (Jurnal Online Informatika)*, *10*(1), 196–204. https://doi.org/10.15575/join.v10i1.1506

Xiaoyu, J. (2024). Legal and Regulatory Research on the Involvement of Third Parties in Criminal Electronic Data Forensics. *Science of Law Journal*, *3*(1), 20–24. https://doi.org/10.23977/law.2024.030104

Yagami, L. (n.d.). *(AMAZON) GIFTCARD CARDING TUTORIAL*. carding.store. Diambil 27 Agustus 2025, dari https://carding.store/topic/88554-amazon-giftcard-carding-tutorial/

Yapici, M. M. (2025). URL-Based Phishing Detection and Comparison of Encoding Approaches. *2025 17th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 1–7. https://doi.org/10.1109/ECAI65401.2025.11095572

Yip, M., Webber, C., & Shadbolt, N. (2013). Trust among cybercriminals? Carding forums, uncertainty and implications for policing. *Policing and Society*, *23*(4), 516–539. https://doi.org/10.1080/10439463.2013.780227

You, Q., Wu, O., Luo, G., & Hu, W. (2016). Metadata-Based Clustered Multi-task Learning for Thread Mining in Web Communities. In P. Perner (Ed.), *Machine Learning and Data Mining in Pattern Recognition. MLDM 2016. Lecture Notes in Computer Science(), vol 9729* (hal. 421–434). Springer International Publishing. https://doi.org/10.1007/978-3-319-41920-6_33

Zahra, H., & Urumsah, D. (2025). Financial cybercrime avoidance behavior among employees of financial sector companies in Indonesia. *The Indonesian Accounting Review*, *14*(2), 223–238. https://register-jobfair.perbanas.ac.id/index.php/tiar/article/view/4596

Zhao, B. (2017). Web Scraping. In L. A. Schintler & C. L. McNeely (Ed.), *Encyclopedia of Big Data* (hal. 1–3). Springer International Publishing. https://doi.org/10.1007/978-3-319-32001-4_483-1

Zulhanif. (2016). Pemodelan Topik dengan Latent Dirichlet Allocation. *Seminar Nasional Pendidikan Matematika 2016*, 1–8. https://publikasiilmiah.ums.ac.id/handle/11617/7633

Lampiran

Repository Github Peneliti

Dataset hasil *web scraping* forensik serta file hasil pengolahan data yang digunakan dalam penelitian tesis ini dapat diakses melalui repositori GitHub peneliti pada berikut:

Komponen	Deskripsi	Detail
Nama Akun GitHub	Akun GitHub Peneliti	451Fikrie
Nama Repository	Repositori untuk dataset dan file pengolahan	Tesis-Magister-Informatika-Fikri
URL Repository	Tautan untuk mengakses repositori	https://github.com/451Fikrie/Tesis-Magister-Informatika-Fikri