



Klasifikasi dan Analisis Semantik Cyberbullying Sosial Media X: Integrasi Web Scraping dan Natural Language Processing (NLP)

Syifa Aulia Azzahra*, Nuur Wachid Abdul Majid

Pendidikan Sistem Teknologi Informasi, Universitas Pendidikan Indonesia, Bandung, Indonesia

***Corresponding Author:**

syifaazzahra@upi.edu

Article History:

Received 2025-01-11

Revised 2025-04-04

Accepted 2025-04-15

Keywords:

Cyberbullying;

Analyze Semantic;

Web Scraping;

NLP;

Media Social X;

Abstract

Cyberbullying on social media, particularly on platform X, has become a critical issue with significant psychological impacts. This study analyzes the extent to which cyberbullying persists on platform X using a semantic approach. Data was collected through web scraping with Selenium, employing specific categories and keywords such as "gendut" (overweight) and "bodoh" (stupid) during December 2024. A total of 700 data points were obtained after deduplication, meeting Slovin's formula criteria with a 3.77% margin of error. The analysis involved Natural Language Processing (NLP) techniques, including text cleaning, lowercasing, normalization, tokenization, stopword removal, and classification using a fine-tuned BERT model to identify cyberbullying comments. Keywords were mapped to 8 predefined categories, such as "racism" and "ethnicity, religion, race, and intergroup relations (SARA)." Results revealed that 55.4% of the data contained cyberbullying indicators, with the "sexual" category being the most dominant at 26.6%. The keyword "anjing" (dog) appeared 99 times, and certain negative terms exhibited fluctuating temporal patterns, peaking in Week 4 with the highest cyberbullying intensity (79.1%). These findings confirm that cyberbullying remains a significant phenomenon on platform X. Consequently, stricter content moderation policies and the development of automated machine learning-based detection systems are recommended to mitigate cyberbullying more effectively.

Abstrak

Cyberbullying di media sosial, khususnya X, telah menjadi isu kritis dengan dampak psikologis yang signifikan. Studi ini menganalisis sejauh mana cyberbullying masih terjadi di platform X dengan pendekatan semantik. Data dikumpulkan melalui proses web scraping menggunakan Selenium dengan menggunakan kategori dan kata kunci spesifik seperti "gendut" dan "bodoh" selama periode Desember 2024. Sebanyak 700 data berhasil dikumpulkan setelah melalui proses deduplikasi, yang mana memenuhi kriteria Slovin (*margin of error* 3.77%). Proses analisis melibatkan Natural Language Processing (NLP), termasuk *text-cleaning*, *lowercasing*, *normalization*, *tokenization*, *stopword removal*, klasifikasi model menggunakan model BERT yang telah di-fine-tune untuk memastikan program mengenali sebuah komentar termasuk cyberbullying atau tidak, serta pemetaan kata kunci ke 8 kategori, seperti "rasisme" dan "sara". Hasil menunjukkan bahwa sebanyak 55,4% mengandung indikasi cyberbullying, dengan kategori seksual sebagai yang paling dominan dengan 26,6%, serta kata kunci anjing yang disebut 99 kali. Kata-kata negatif tertentu menunjukkan pola temporal yang fluktuatif, di mana intensitas cyberbullying mencapai puncak pada Minggu 4 dengan persentase tertinggi (79,1%). Temuan ini mengonfirmasi bahwa cyberbullying masih menjadi fenomena signifikan di platform X, oleh karena itu diperlukan kebijakan moderasi konten yang lebih ketat serta pengembangan sistem deteksi otomatis berbasis *machine learning* untuk mitigasi cyberbullying secara lebih efektif.

PENDAHULUAN

Perkembangan media sosial telah mencapai skala global yang signifikan. Analisis terbaru oleh Kepios (2024) menunjukkan bahwa terdapat 5,22 miliar pengguna media sosial di seluruh dunia pada Oktober 2024, setara dengan 63,8% populasi global. Di Indonesia, platform X menjadi salah satu media sosial yang dominan, dengan pengguna aktif mencapai 618,9 juta secara global (We Are Social, 2024). Namun, peningkatan penggunaan ini diiringi oleh risiko cyberbullying yang semakin mengkhawatirkan. Cyberbullying merujuk pada bentuk kekerasan verbal di dunia maya yang dilakukan melalui platform digital, seperti penghinaan, ancaman, atau pelecehan, dengan dampak psikologis serius, termasuk gangguan kecemasan hingga ideasi bunuh diri (Balet et al., 2023).



Seiring dengan meningkatnya penggunaan media sosial, fenomena cyberbullying semakin marak terjadi dan biasanya menyerang korban secara verbal di platform seperti X, Facebook, dan Instagram (Winarno et al., 2023). Cyberbullying umumnya terjadi di kalangan dewasa muda usia 18-24 tahun, dengan pelecehan menjadi bentuk cyberbullying yang paling sering dialami. Laporan Office for National Statistics (ONS, 2023) menunjukkan bahwa 19.1% anak-anak usia 10 hingga 15 tahun mengalami perilaku perundungan daring (online bullying) dalam satu tahun terakhir. Selain itu, 9.5% anak usia 13 hingga 15 tahun menerima pesan bernuansa seksual secara daring, dan lebih dari tiga perempat (76.7%) dari mereka menerima pesan tersebut lebih dari sekali. Fenomena ini memiliki dampak yang signifikan terhadap kesejahteraan mental korban, yang dalam beberapa kasus dapat menyebabkan gangguan psikologis serius hingga bunuh diri (Balet et al., 2023). Penelitian lain juga menyebutkan dampak cyberbullying menyebabkan gangguan kesehatan mental seperti depresi dan kecemasan yang berlarut-larut, serta memperburuk hubungan sosial antar individu. Korban cyberbullying memiliki peluang yang lebih tinggi untuk mengalami gangguan mental, dan dalam kasus ekstrem, dapat mengarah pada tindak bunuh diri (Hinduja, & Patchin. 2018). Oleh karena itu, diperlukan strategi yang efektif untuk mengidentifikasi dan mengatasi cyberbullying guna menciptakan lingkungan digital yang lebih aman bagi para pengguna.

Berbagai penelitian telah dilakukan untuk memahami dan mengatasi masalah ini. Penelitian sebelumnya telah mencoba berbagai pendekatan dalam mendeteksi cyberbullying. Beberapa penelitian menggunakan metode berbasis pembelajaran mesin seperti Support Vector Machine (SVM), Naïve Bayes, dan Random Forest untuk mengidentifikasi ujaran kebencian dalam pesan-pesan media sosial (Ula & Fachrurrazi, 2023). Teknik-teknik ini telah menunjukkan keberhasilan dalam mendeteksi pola-pola ujaran kebencian, meskipun masih menghadapi kendala dalam memahami konteks sosial dan budaya dari suatu pernyataan. Penelitian yang dilakukan oleh Winarno et al. (2023) mengusulkan pendekatan berbasis web service yang mengimplementasikan algoritma SVM dan Random Forest dalam deteksi cyberbullying, yang menunjukkan tingkat akurasi yang cukup tinggi.

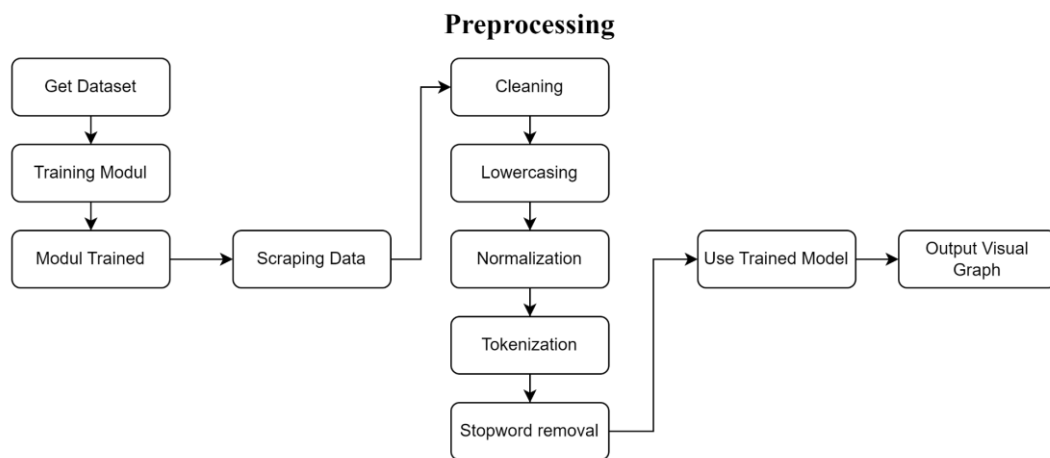
Penelitian ini berfokus pada analisis semantik menggunakan Natural Language Processing (NLP) untuk mendeteksi cyberbullying. Semantik, sebagai cabang linguistik yang mempelajari makna dan interpretasi kalimat dalam konteks tertentu (Kholis et al., 2024), diimplementasikan melalui model BERT. Data yang digunakan berupa teks komentar publik di platform X yang dikumpulkan secara sistematis menggunakan teknik web scraping berbasis Selenium, dengan filter kata kunci seperti "rasis", "seksual", dan "gendut". Proses scraping ini memungkinkan pengambilan 700 data tweet dalam periode spesifik, yang kemudian diproses melalui tahapan NLP (cleaning, lowercasing, normalization, tokenization, stopword removal). Analisis semantik tidak hanya mendeteksi kata kunci eksplisit berkonotasi kasar, tetapi juga mengidentifikasi distribusi cyberbullying berdasarkan kategori seperti rasisme, seksual, dan body shaming. Lalu cakupan data yang digunakan ialah teks berupa tweet sehingga tidak memasukkan konten berupa gambar atau video, guna memfokuskan analisis pada aspek linguistik murni.

Tujuan utama yang ingin dicapai yaitu untuk memberikan kontribusi signifikan terhadap pemahaman tentang cyberbullying di Indonesia, dengan menyajikan solusi berbasis teknologi yang dapat membantu mendeteksi dan mengklasifikasikan ujaran yang merugikan secara lebih akurat dan efisien. Selain itu, penelitian ini juga diharapkan dapat memberikan gambaran yang lebih jelas tentang sejauh mana cyberbullying mempengaruhi pengguna media sosial di Indonesia, serta seberapa efektif algoritma pemrograman dalam menangani permasalahan ini di dunia maya. Dengan demikian, penelitian ini dapat menjadi langkah awal dalam menciptakan lingkungan digital yang lebih aman dan mendukung kesehatan mental penggunanya.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis teks berbasis Natural Language Processing (NLP) untuk mengidentifikasi dan mengklasifikasikan ujaran cyberbullying pada platform X secara objektif dan sistematis. Model BERT yang telah di-*fine-tune* untuk bahasa Indonesia digunakan dalam analisis semantik guna menangkap makna kontekstual dari komentar yang dikumpulkan. Sesuai dengan

pandangan Marta et al. (2024), pendekatan kuantitatif memungkinkan pengukuran fenomena sosial melalui data numerik yang dalam hal ini digunakan untuk mengukur prevalensi dan pola ujaran bermuatan cyberbullying sehingga hasil penelitian dapat digeneralisasi dalam konteks media sosial di Indonesia.



Gambar 1. Flowchart Analisis Semantik Sistem Deteksi Cyberbullying

Berdasarkan flowchart yang telah disusun, penelitian ini mencakup dua tahap utama, yaitu prosedur pengumpulan data dan teknik analisis data. Tahap pertama melibatkan proses akuisisi data melalui metode web scraping, sedangkan tahap kedua berfokus pada penerapan teknik pemrosesan dan analisis semantik dengan NLP untuk mengidentifikasi pola ujaran cyberbullying. Penjelasan lebih rinci mengenai kedua tahap tersebut akan dibahas berikut.

Prosedur Pengumpulan Data

Dataset untuk pelatihan model klasifikasi diperoleh dari repositori publik GitHub, yakni proyek *Dataset Sentimen Analisis Bahasa Indonesia*, yang berisi 650 komentar berlabel "Cyberbullying" dan "Non-Cyberbullying". Data ini dipilih karena telah melalui proses anotasi manual dan relevan dengan konteks ujaran negatif dalam bahasa Indonesia. Setiap entri terdiri dari teks komentar dan label klasifikasinya.

Sebagai contoh, komentar seperti "yang aku suka dari dia adalah selalu cukur jembut sebelum manggung" dikategorikan sebagai cyberbullying (label: 1), sedangkan "Kaka tidur yaa, udah pagi, gaboleh capek2" sebagai non-cyberbullying (label: 0).

Model analisis semantik dilatih menggunakan arsitektur IndoBERT-base-p1, yang dioptimalkan untuk pemrosesan bahasa Indonesia. BERT dipilih karena kemampuannya memahami konteks kalimat secara mendalam, termasuk makna implisit dan penggunaan frasa bermakna ganda. Setelah data diproses, komentar dikonversi menjadi representasi numerik untuk dianalisis lebih lanjut oleh model klasifikasi.

Tabel 1. Kategori dan Kata Kunci Bahasa Kasar

Kategori	Kata Kunci
Rasisme	hitam, bule, cina, pribumi, etnis, ras, melayu, yahudi, turunan, minoritas
Fisik	gendut, gemuk, kurus, cungring, pendek, botak, jelek, cacat, pincang, pesek, sipit, dower, jenggot, item, jerawat
Seksual	kontol, memek, pentil, ngentot, ngewe, jembut, fuck, penis, vagina, peler, puki, lonte, silit, pelacur
Hewan	babi, anjing, monyet, asu
Umum	goblok, tolol, bodoh, bangsat, brengsek, mampus
Agama	kafir, munafik, sesat, atheis, haram, fanatik
Keluarga	yatim, piatu, janda, duda, haram, durhaka, cerai

Selanjutnya pengumpulan data dilakukan dengan teknik web scraping dengan mengumpulkan komentar publik di platform X selama periode Desember 2024. Data dikumpulkan secara mingguan dalam empat tahap pengambilan, dengan batasan hanya mengambil tweet yang menggunakan bahasa Indonesia. Kriteria pemilihan data mencakup penggunaan kata kunci spesifik yang telah dikategorikan sebagai mana disajikan di tabel 1.

Sebanyak 700 komentar berhasil dikumpulkan melalui proses scraping dari platform X, setelah dilakukan deduplikasi dan seleksi berdasarkan relevansi terhadap kategori cyberbullying. Jumlah data ini ditentukan dengan mempertimbangkan validitas statistik menggunakan perhitungan margin of error sebesar 3,77%, sehingga jumlah sampel dinilai cukup representatif terhadap populasi komentar selama periode penelitian.

Proses pengumpulan data dalam penelitian ini dilakukan melalui *web scraping* dengan menggunakan Selenium, sebuah pustaka Python yang memungkinkan otomatisasi interaksi di browser. Program dirancang untuk mensimulasikan perilaku pengguna di platform X, dimulai dari proses login, pencarian berdasarkan kata kunci, hingga merekap halaman secara otomatis untuk memperoleh lebih banyak data. Selama proses ini, sistem mengumpulkan informasi penting dari setiap tweet, seperti nama pengguna, isi komentar, serta waktu unggahan.

Setelah data berhasil dikumpulkan, sistem menyimpannya dalam format CSV untuk memudahkan proses pembersihan data dan analisis lanjutan menggunakan model NLP berbasis BERT. Pendekatan ini memungkinkan peneliti memperoleh data terkini dan relevan dari media sosial secara langsung, sekaligus memastikan bahwa data yang digunakan dalam analisis merupakan representasi nyata dari percakapan publik yang terjadi di platform tersebut.

Teknik Analisis Data

Setelah data komentar dikumpulkan melalui proses web scraping, tahap awal yang dilakukan adalah *preprocessing* teks untuk mempersiapkan data sebelum dimasukkan ke dalam model klasifikasi. Proses ini mencakup lima tahap utama, yaitu: (1) *Text cleaning*, yang menghapus elemen tidak relevan seperti URL, mention, tanda baca berlebihan, dan emoji; (2) *Lowercasing*, yaitu mengubah seluruh huruf menjadi kecil untuk konsistensi pemrosesan; (3) *Normalization*, mengganti kata tidak baku atau slang menjadi bentuk standar, seperti "yg" menjadi "yang", dan "wkwk" menjadi "tertawa"; (4) *Tokenization*, yaitu memecah kalimat menjadi unit kata; serta (5) *Stopword removal*, yaitu menghapus kata-kata umum tanpa nilai analitis seperti "yang", "di", dan "ini".

Hasil preprocessing berupa teks bersih dan terstruktur kemudian dianalisis menggunakan model *IndoBERT-base-p1*, sebuah model BERT yang telah di-*fine-tune* khusus untuk pemrosesan bahasa Indonesia. Model ini memanfaatkan mekanisme *self-attention* untuk memahami konteks kalimat secara mendalam, termasuk mendeteksi ironi, pelecehan implisit, atau makna tersirat lainnya. Setiap komentar diproses dan diberi label 1 (Cyberbullying) atau 0 (Non-Cyberbullying), disertai skor probabilitas sebagai indikator tingkat kepercayaan model terhadap hasil klasifikasinya.

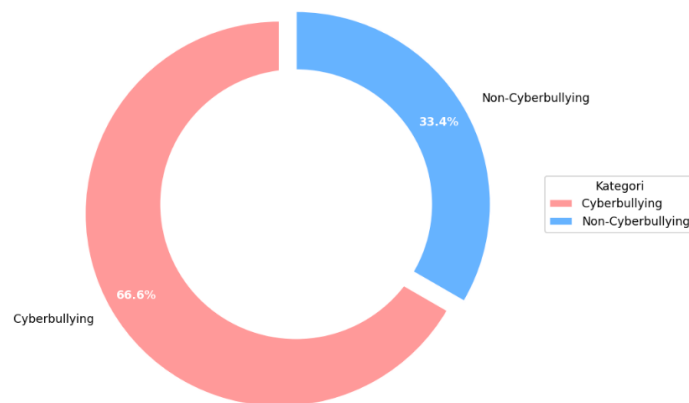
Sebagai ilustrasi, kalimat seperti "*lonte caper jir, biar rame aja itu akunya wkwk*" diklasifikasikan sebagai *Cyberbullying* dengan skor kepercayaan 95.02%, sementara komentar "*Transformasi Diet Pria yang Dulunya Dekil dan Gemuk Ini Bikin Takjub*" dikategorikan sebagai *Non-Cyberbullying* dengan skor 90.33%. Seluruh hasil klasifikasi direkapitulasi dan divisualisasikan menggunakan pustaka *matplotlib* dalam bentuk grafik dan tabel, guna memberikan gambaran proporsi cyberbullying di platform X selama periode penelitian.

HASIL DAN PEMBAHASAN

Berdasarkan Gambar 2. 66,6% komentar teridentifikasi sebagai cyberbullying, sementara 33,4% termasuk non-cyberbullying. Tingginya persentase cyberbullying ini mengonfirmasi bahwa platform X masih rentan terhadap konten negatif, meskipun telah ada upaya moderasi. Hasil ini sejalan dengan penelitian Adellia, et al. (2024), yang menemukan bahwa 64% siswa berada pada kategori sedang menerima perilaku cyberbullying.

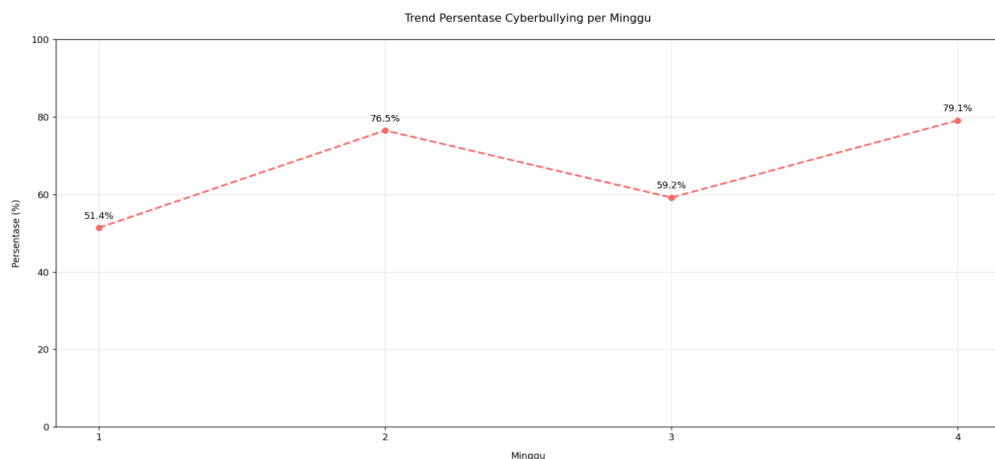
Temuan tersebut menunjukkan bahwa sebagian besar siswa mengalami atau terlibat dalam cyberbullying pada tingkat sedang, mencerminkan masih tingginya prevalensi fenomena ini di lingkungan digital. Model BERT yang *di-fine-tune* terbukti efektif mengidentifikasi konteks negatif secara semantik, terutama pada kata-kata implisit seperti sindiran atau metafora.

Hasil Analisis Semantik: Cyberbullying vs Non-Cyberbullying



Gambar 2. Hasil Analisis Semantik

Hasil ini menegaskan cyberbullying menjadi ancaman serius di ruang digital platform X. Kebebasan berpendapat melenceng terlalu jauh sehingga menjadi ruang tidak nyaman dalam bersosial media. Jika dirincikan lebih lanjut, berikut adalah tren per-minggu dalam bulan Desember 2024.



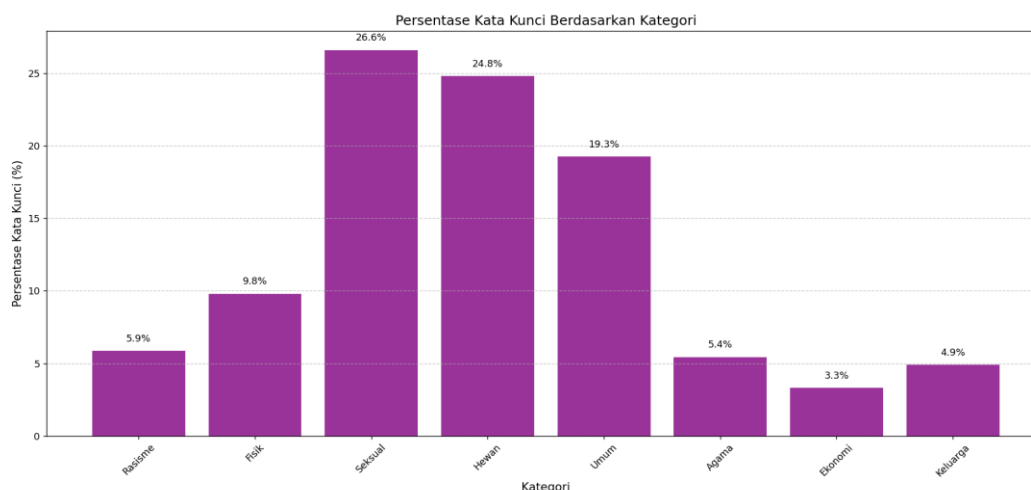
Gambar 3. Tren Per-minggu

Analisis tren persentase cyberbullying per minggu menunjukkan dinamika yang fluktuatif selama periode penelitian. Pada minggu pertama, persentase cyberbullying tercatat sebesar 51,4%, yang kemudian melonjak signifikan menjadi 76,5% di minggu kedua. Lonjakan ini diduga terkait dengan peningkatan aktivitas pengguna pada akhir pekan atau adanya isu viral yang memicu interaksi negatif. Namun, di minggu ketiga, persentase turun menjadi 59,2%. Penurunan ini tidak bertahan lama, karena di minggu keempat persentase kembali naik ke level tertinggi, yaitu 79,1%. Pola naik-turun-naik ini mengindikasikan bahwa cyberbullying tidak hanya dipengaruhi oleh faktor insidental, tetapi juga memiliki kecenderungan periodik yang perlu diwaspadai seperti tren yang berkelanjutan.

Fluktuasi tersebut memperlihatkan kompleksitas faktor pendorong cyberbullying. Lonjakan di minggu kedua dan keempat berkorelasi dengan momen emosional kolektif, seperti kontroversi publik atau tren yang sedang berlangsung, yang meningkatkan interaksi negatif. Sering kali pelaku cyberbullying melakukan dengan sengaja dengan terkonsentrasi pada periode engagement tinggi. Sementara itu, penurunan di minggu ketiga bisa mencerminkan efektivitas sementara upaya mitigasi, seperti penghapusan konten atau kesadaran

pengguna. Namun, ketidakmampuan mempertahankan tren penurunan hingga minggu keempat mengisyaratkan bahwa upaya tersebut belum holistik atau kurang berkelanjutan. Selain itu, ketiadaan data timestamp harian membatasi analisis untuk memastikan apakah lonjakan memang terjadi pada akhir pekan, meskipun pola tersebut secara teoritis sejalan dengan karakteristik aktivitas media sosial.

Selain itu, cyberbullying erat kaitannya dengan jenis konten dan pola linguistik yang digunakan. Analisis lebih mendalam terhadap kategori dan kata kunci dominan mengungkap bahwa cyberbullying di platform X cenderung terfokus pada serangan bernuansa seksual, dehumanisasi, dan penggunaan kata-kata vulgar. Hal ini memperkuat temuan bahwa pelaku tidak hanya memanfaatkan momentum interaksi tinggi, tetapi juga secara strategis memilih kata-kata dengan dampak psikologis maksimal untuk mempermalukan korban.



Gambar 4. Diagram Batang Persentase Kategori

Dominasi kategori seksual (26,6%) sebagai jenis cyberbullying terbanyak mencerminkan kecenderungan pelaku untuk menyerang identitas gender atau orientasi seksual korban, sering kali menggunakan kata-kata vulgar seperti "kontol" atau "lonte". Fenomena ini sejalan dengan studi Copp et al. (2021) yang menemukan bahwa dari 37% korban yang melapor, 15% terkena pengalaman pelecehan seksual, terutama terhadap perempuan. Kategori hewan (24,8%), seperti penggunaan kata "anjing", menunjukkan praktik dehumanisasi untuk merendahkan martabat korban, sebuah taktik yang umum dalam konflik identitas sosial. Sementara itu, kategori umum (19,3%) mencakup kata-kata seperti "tolol" atau "bodoh", yang meskipun terkesan ringan, berpotensi menormalisasi budaya penghinaan secara sistematis. Temuan ini mengonfirmasi bahwa cyberbullying tidak hanya bersifat insidental, tetapi juga terstruktur melalui pemilihan kategori yang spesifik untuk memperkuat dampak psikologis.



Gambar 5. Tag Cloud Persentase Kategori

Analisis frekuensi kata kunci mengungkap preferensi pelaku dalam menggunakan istilah bernuansa dehumanisasi dan vulgaritas. Kata "anjing" yang muncul 99 kali menjadi metafora dominan untuk merendahkan korban, menggambarkan pola othering (pengasingan) yang lazim dalam konteks kekerasan verbal. Sementara

itu, kata vulgar seperti "kontol" (66 kali) dan "lonte" (58 kali) menunjukkan upaya untuk menyerang aspek privat korban dengan tujuan untuk menimbulkan rasa malu dan isolasi sosial. Kata "tolol" (67 kali), meskipun termasuk dalam kategori umum, sering kali berfungsi sebagai pemicu awal yang memungkinkan eskalasi ke bentuk bullying yang lebih parah. Pola penggunaan kata kunci ini tidak hanya merefleksikan kebencian, tetapi juga mengekspos kegagalan sistem moderasi dalam menyaring istilah-istilah yang telah teridentifikasi secara luas dalam literatur akademis sebagai high-risk keywords.

Meskipun analisis semantik menggunakan model BERT yang di-*fine-tune* menunjukkan keandalan dalam mengidentifikasi konteks negatif dan pola linguistik cyberbullying, terdapat beberapa aspek yang perlu mendapat perhatian lebih lanjut. Misalnya, dalam penelitian ini, data dikumpulkan secara otomatis melalui web scraping dan analisis NLP terhadap konten X, sementara Adellia et al. (2024) menggunakan pendekatan survei dan wawancara langsung pada siswa SMA. Perbedaan ini bukan hanya soal teknik, tetapi mencerminkan dua paradigma pendekatan: kuantifikasi berbasis data digital anonim versus pengalaman langsung individu dengan latar sosial yang jelas. Metode Adellia memungkinkan pemahaman yang lebih dalam tentang konteks psikologis korban dan pelaku, terutama saat membahas dampak relasional dari bullying. Sementara itu, pendekatan berbasis NLP pada data X lebih cocok untuk mendeteksi pola umum dan frekuensi, namun kurang mampu menangkap aspek emosional dan dinamika hubungan.

Selain itu, dominasi kategori serangan seksual dan dehumanisasi yang juga diungkap dalam studi Copp et al. (2021) menegaskan bahwa pemilihan kata kunci high-risk perlu dikaji secara mendalam dengan mempertimbangkan konteks sosial dan kultural yang lebih luas. Hal ini menunjukkan bahwa meskipun algoritma NLP mampu mendeteksi pola berdasarkan frekuensi dan tren temporal, keterbatasan dalam menangkap nuansa emosional serta dinamika interaksi antar pengguna masih ada. Oleh karena itu, integrasi metode analisis kuantitatif dengan pendekatan kualitatif, seperti studi kasus mendalam terhadap konteks pernyataan atau analisis percakapan, sangat disarankan untuk memperoleh gambaran yang lebih holistik dan valid mengenai fenomena cyberbullying di platform X.

KESIMPULAN

Penelitian ini mengungkap bahwa cyberbullying masih menjadi fenomena yang signifikan di platform X, dengan lebih dari setengah data yang dianalisis (55,4%) mengandung indikasi cyberbullying. Dari hasil analisis semantik menggunakan pendekatan NLP dan klasifikasi berbasis model BERT, ditemukan bahwa kategori seksual menjadi yang paling dominan (26,6%), dengan kata kunci tertentu seperti anjing muncul dalam frekuensi tinggi (99 kali). Selain itu, pola temporal menunjukkan fluktuasi intensitas cyberbullying, dengan puncak terjadi pada minggu keempat (79,1%).

Temuan ini menegaskan perlunya kebijakan moderasi konten yang lebih ketat untuk menekan penyebaran ujaran kebencian dan pelecehan di media sosial. Di samping itu, pengembangan sistem deteksi otomatis berbasis machine learning dapat menjadi solusi yang lebih efektif dalam mengidentifikasi dan menangani konten bermuatan cyberbullying secara real-time. Sebagai prospek pengembangan lebih lanjut, penelitian ini dapat diperluas dengan analisis yang lebih mendalam terhadap korelasi antara pola temporal dan faktor eksternal seperti peristiwa sosial atau tren media. Selain itu, peningkatan akurasi model deteksi dengan fine-tuning lebih lanjut atau penggunaan dataset yang lebih besar dapat meningkatkan keandalan sistem dalam mengidentifikasi berbagai bentuk cyberbullying.

DAFTAR PUSTAKA

- Adellia, A. P., Sulistiyana, & Putro, H. Y. S. (2024). Studi komparatif: Bullying di dunia nyata dan dunia maya (cyberbullying). *Edukatif: Jurnal Ilmu Pendidikan*, 6(4), 4000-4007. <https://doi.org/10.31004/edukatif.v6i4.7240>
- Balet, T., Vo, Q., Salem, O., & Mehaoua, A. (2023). Cyberbullying detection on tweets from Twitter using machine learning algorithms. *2023 International Conference on Intelligent Computing, Communication, Networking and Services (ICCNS)*. <https://doi.org/10.1109/ICCNS58795.2023.10193450>

- Copp, J. E., Mumford, E. A., & Taylor, B. G. (2021). Online sexual harassment and cyberbullying in a nationally representative sample of teens: Prevalence, predictors, and consequences. *Journal of Adolescence*, 93, 202-211. <https://doi.org/10.1016/j.adolescence.2021.10.003>
- Fati, S. M., Muneer, A., Alwadain, A., & Balogun, A. O. (2023). Cyberbullying detection on Twitter using deep learning-based attention mechanisms and continuous bag of words feature extraction. *Mathematics*, 11(16), 1-21. <https://doi.org/10.3390/math11163567>
- Hinduja, S., & Patchin, J. W. (2018). *Cyberbullying: Identification, prevention, and response*. Cyberbullying Research Center. Retrieved from <https://cyberbullying.org/Cyberbullying-Identification-Prevention-Response-2018.pdf>
- Kepios. (2024). *Digital 2024 October Global Statshot Report*. We Are Social & Meltwater. <https://datareportal.com/reports/digital-2024-october-global-statshot>
- Kholis, K., Wiranata, T. D., Aisyah, S., & Asror, A. G. (2024). Semantik: Pengertian, teori, dan penerapannya dalam pembelajaran bahasa. *Prosiding Seminar Nasional Daring*, 4(1), 570-578. <https://prosiding.ikippgribojonegoro.ac.id/index.php/SPBSI/article/view/2727/pdf>
- Lalithia, N., Sumaya T. Sk., Tejaswini, N., Bhagya, Sri D., & Srivani, R. (2023). Enhancing cyberbullying detection on Twitter with psychological features and machine learning. *International Conference on Emerging Research in Computational Science (ICERCS)*. <https://doi.org/10.1109/ICERCS57948.2023.10434258>
- Laxmi, S. T., Rismala, R., & Nurrahmi, H. (2021). Cyberbullying detection on Indonesian Twitter using Doc2Vec and Convolutional Neural Network. *2021 International Conference on Information and Communication Technology (ICoICT)*. <https://doi.org/10.1109/ICoICT52021.2021.9527420>
- Marta, R. F., Kristina, Yulianto, A., & Febrianto, Y. (2024). *Metode penelitian: Memahami pendekatan kuantitatif, kualitatif, dan campuran*. PT Media Penerbit Indonesia. <http://repository.mediapenerbitindonesia.com/338/1/Naskah%20Fix%20K%202024%20-%20%28FINISH%20LAYOUT%29%20Metode%20Penelitian%20Memahami%20Pendekatan%20Kuantitatif%2C%20Kualitatif%2C%20dan%20Campuran.pdf>
- Nugraha, D., & Astuti, P. (2023). Analisis sentimen cyberbullying pada sosial media Instagram menggunakan metode Support Vector Machine. *Information System for Educators and Professionals*, 8(2), 152-164. <https://doi.org/10.51211/isbi.v8i2.2535>
- Office for National Statistics. (2023). *Children's online behaviour, England and Wales: Year ending March 2023*. Office for National Statistics. <https://www.ons.gov.uk/>
- Olasanmi, O. O., Agbaje, Y. T., & Adeyemi, M. O. (2020). Prevalence and prevention strategies of cyberbullying among Nigerian students. *Open Journal of Applied Sciences*, 10(6), 351-363. <https://doi.org/10.4236/ojapps.2020.106026>
- Rizalespe. (2019). *Dataset-Sentimen-Analisis-Bahasa-Indonesia* [Dataset]. GitHub. <https://github.com/rizalespe/Dataset-Sentimen-Analisis-Bahasa-Indonesia>
- Sandrila, W., & Wahyunengsih. (2023). Motives of cyberbullying behavior by teenage K-Pop fans on Twitter social media. *Jurnal Riset Rumpun Ilmu Sosial, Politik dan Humaniora (JURRISH)*, 2(2), 190-196. <https://doi.org/10.55606/jurrish.v2i2.1351>
- Ula, M., & Fachrurrazi, S. (2023). Analisis sentimen cyberbullying pada media sosial Twitter menggunakan metode Support Vector Machine dan Naïve Bayes Classifier. *TECHSI - Jurnal Teknik Informatika*, 14(2), 91. <https://doi.org/10.29103/techsi.v14i2.12103>
- We Are Social. (2024). *Digital 2024: Indonesia*. We Are Social & Meltwater. <https://datareportal.com/reports/digital-2024-indonesia>
- Winarno, Wiranto, & Harjito, B. (2023). Enhancing machine learning performance in cyberbullying detection through hyperparameter optimization. *2023 International Conference on Technology, Engineering, and Computing Applications (ICTECA)*. <https://doi.org/10.1109/ICTECA60133.2023.10490843>
- Yasirutomo. (2021). *text-normalization* [Normalization]. GitHub. <https://github.com/yasirutomo/text-normalization>