JURNAL INTI TALAFA ISSN: 2085-0573
Vol.17 No.2, Juli - Desember 2026, pp. 87 - 101 EISSN: 2829-1506

ANALISIS POLA PENYEBAB KARHUTLA KALIMANTAN TENGAH MENGGUNAKAN INDOBERT DAN CAUSAL PATTERN MINING

Fahrizal Maulana[1], Kusrini[2], Ika Safitiri Windiarti[3], Sutami[4]
Ilmu Komputer, Universitas Muhammadiyah Palangka Raya[1,3,4],
Informatika, Universitas Amikom Yogyakarta[2]

Abstrak

Kebakaran hutan dan lahan di Kalimantan Tengah merupakan permasalahan lingkungan yang dipengaruhi oleh faktor alam dan aktivitas manusia sehingga menghasilkan pola hubungan penyebab yang kompleks. Penelitian ini bertujuan menganalisis pola penyebab karhutla di Kalimantan Tengah menggunakan pendekatan NLP dan AI berbasis IndoBERT yang diintegrasikan dengan *causal pattern mining* pada artikel berita online. Data dikumpulkan melalui *web scraping* terhadap 6 portal berita lokal dan nasional menggunakan 9 *query* pencarian google terkait karhutla, menghasilkan 436 artikel relevan sebagai korpus utama. Tahapan metode meliputi *text preprocessing*, representasi semantik menggunakan model transformer IndoBERT berdimensi 768, *topic discovery* menggunakan BERTopic, serta analisis pola kausal melalui *co-occurrence analysis* dan *contextual relation mining*. Hasil *topic discovery* mengidentifikasi 10 topik utama, dengan topik "musim kemarau ekstrem" mendominasi sebanyak 104 data dan diikuti "cuaca panas dan kekeringan" (54 data). Faktor antropogenik seperti pembakaran lahan sengaja, aktivitas pembukaan lahan, dan kelalaian manusia juga teridentifikasi sebagai penyebab signifikan. Hasil *contextual relation mining* menunjukkan bahwa kata "lahan" (1.473 kemunculan), "asap" (938), "kemarau" (334), dan "gambut" (288) merupakan konteks penyebab paling dominan, sementara *co-occurrence analysis* menghasilkan 37.581 pasangan kata yang membentuk jaringan kausalitas. Penelitian selanjutnya disarankan melakukan integrasi data spasial, perluasan dataset serta penerapan *knowledge graph* untuk mendukung sistem *disaster intelligence* karhutla secara *real-time*.

Kata kunci: *karhutla, IndoBERT, BERTopic, causal pattern mining, Kalimantan Tengah*

Abstract

Forest and land fires in Central Kalimantan are environmental problems influenced by natural factors and human activities, resulting in complex causal relationship patterns. This study aims to analyze wildfire cause patterns in Central Kalimantan using an IndoBERT-based NLP and AI approach integrated with causal pattern mining on online news articles. Data were collected through web scraping from six local and national news portals using nine Google search queries related to wildfires, resulting in 436 relevant articles as the main corpus. The methodological stages included text preprocessing, semantic representation using a 768-dimensional IndoBERT transformer model, topic discovery using BERTopic, and causal pattern analysis through co-occurrence analysis and contextual relation mining. The topic discovery results identified 10 main topics, with the topic "extreme dry season" dominating with 104 data points, followed by "hot weather and drought" (54 data points). Anthropogenic factors such as intentional land burning, land clearing activities, and human negligence were also identified as significant causes. Contextual relation mining results showed that the words "land" (1,473 occurrences), "smoke" (938), "dry season" (334), and "peatland" (288) were the most dominant causal contexts, while co-occurrence analysis generated 37,581 word pairs forming a causal network. Future studies are recommended to integrate spatial data, expand the dataset, and implement knowledge graphs to support real-time wildfire disaster intelligence systems.

Keywords: *forest and land fire, IndoBERT, BERTopic, causal pattern mining, Central Kalimantan*

1. PENDAHULUAN

Kebakaran hutan dan lahan (karhutla) menjadi salah satu permasalahan lingkungan serius di berbagai wilayah dunia, termasuk Indonesia. Di Kalimantan, khususnya pada kawasan gambut, karhutla dipengaruhi oleh perubahan iklim dan aktivitas manusia. Penelitian [1] menunjukkan bahwa fenomena El Niño, peningkatan suhu udara, rendahnya curah hujan, serta konversi lahan gambut menjadi perkebunan dan lahan terbuka menyebabkan peningkatan frekuensi dan tingkat keparahan kebakaran. Fenomena serupa juga ditemukan pada penelitian [2] yang menunjukkan bahwa pola kebakaran dipengaruhi oleh musim, kekeringan, vegetasi, topografi, dan aktivitas manusia sehingga membentuk pola hotspot kebakaran secara spasio-temporal.

Karhutla merupakan fenomena kompleks yang dipengaruhi oleh faktor alam dan antropogenik secara bersamaan. Penelitian [3] menunjukkan bahwa faktor alami seperti suhu tinggi, vegetasi kering, curah hujan rendah, dan sambaran petir memiliki hubungan erat dengan faktor antropogenik seperti pembukaan lahan, pembangunan infrastruktur, dan kelalaian manusia. Kompleksitas hubungan antar faktor tersebut menyebabkan pola penyebab kebakaran sulit diidentifikasi secara langsung dan membutuhkan pendekatan analisis yang lebih sistematis.

Di sisi lain, perkembangan media digital menyebabkan informasi mengenai karhutla tersebar luas melalui artikel berita online. Berita daring tidak hanya memuat lokasi dan waktu kejadian kebakaran, tetapi juga menjelaskan kondisi cuaca, aktivitas pembukaan lahan, kebijakan pemerintah, dampak asap, dan respons masyarakat. Penelitian [4] menunjukkan bahwa jutaan artikel berita dapat dimanfaatkan untuk membangun sistem disaster intelligence berbasis AI yang mampu mengekstraksi informasi spasial dan temporal secara otomatis. Sementara itu, penelitian [5] menunjukkan bahwa berita online juga mengandung pola narasi dan hubungan antar isu yang dapat dianalisis menggunakan *topic modeling*. Meskipun informasi pada berita online sangat melimpah, proses analisis manual masih memiliki berbagai keterbatasan. Analisis secara manual membutuhkan waktu yang lama, rentan terhadap subjektivitas, dan sulit dilakukan secara konsisten pada ribuan dokumen berita yang terus bertambah setiap hari. Selain itu, teks berita bersifat tidak terstruktur sehingga menyulitkan proses identifikasi pola penyebab, pengelompokan isu dominan, dan hubungan antar faktor kebakaran secara konvensional. Kondisi tersebut menyebabkan pendekatan manual menjadi kurang efektif untuk mendukung analisis karhutla secara cepat dan komprehensif.

Perkembangan NLP (Natural Language Processing) dan AI (Artificial Intelligence) memberikan peluang baru dalam analisis data tekstual kebencanaan secara otomatis dan berskala besar. Penelitian [6] menunjukkan bahwa metode NLP mampu mengidentifikasi topik dan sentimen berita secara efektif menggunakan pendekatan topic modeling dan machine learning. Penelitian [7] juga menunjukkan bahwa kombinasi NLP, transformer models, sentiment analysis, dan geoparsing mampu digunakan untuk membangun sistem monitoring wildfire berbasis media sosial secara real-time. Dalam konteks bahasa Indonesia, penelitian [8] membuktikan bahwa IndoBERT memiliki performa lebih baik dibanding metode machine learning konvensional karena mampu memahami konteks bahasa Indonesia secara bidirectional dan kontekstual.

Berbagai penelitian sebelumnya telah banyak membahas prediksi kebakaran dan pola karhutla menggunakan pendekatan spasial, machine learning, dan deep learning. Penelitian [9], [10], [11], [12] menunjukkan bahwa metode AI mampu meningkatkan akurasi prediksi kebakaran berbasis data hotspot dan spasio-temporal. Pada bidang NLP, penelitian [5], [6], [7] berfokus pada topic modeling, sentiment analysis, dan monitoring bencana. Sementara itu, penelitian [13], [14], [15] menunjukkan bahwa causal mining mampu mengekstraksi hubungan sebab-akibat dari teks tidak terstruktur menggunakan semantic embeddings dan neural language models.

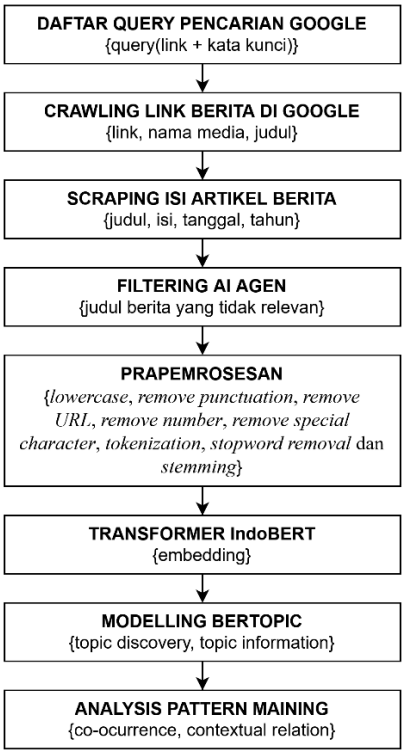
Meskipun demikian, masih terdapat *research gap* pada integrasi analisis teks berita karhutla berbasis bahasa Indonesia menggunakan semantic representation, topic discovery, dan causal pattern mining dalam satu framework analisis terpadu. Sebagian besar penelitian karhutla masih berfokus pada data spasial dan prediksi hotspot, sedangkan penelitian NLP kebencanaan lebih banyak membahas sentimen dan topic modeling tanpa analisis hubungan sebab-akibat secara mendalam. Selain itu, penerapan causal pattern mining pada berita karhutla berbahasa Indonesia masih relatif terbatas.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis pola penyebab karhutla di Kalimantan Tengah menggunakan pendekatan NLP berbasis IndoBERT yang mengintegrasikan *causal pattern mining* dalam satu alur analisis terpadu terhadap artikel berita online. Penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan disaster

intelligence berbasis AI untuk mendukung mitigasi dan pengambilan keputusan terkait kebakaran hutan dan lahan di Kalimantan Tengah.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan NLP dan AI untuk menganalisis pola penyebab kebakaran hutan dan lahan (karhutla) berdasarkan artikel berita online berbahasa Indonesia. Alur tahapan dalam penelitiin ini dimulai dari penentuan query pencarian google crawling link artikel berita, proses crawling, scraping isi artikel berita, *preprocessing*, *modelling* IndoBERT, *modelling* BERTTopic, lalu dilanjutkan dengan proses Causal Pattern Mining yang memiliki dua mekanisme yakni co-ocurrence analysis serta *contextual relation mining* dimana semua tahapan dalam penelitian digambarkan dalam alur diagram pada Gambar 1 di bawah ini.



Gambar 1. Alur Tahapan Penelitian

Sumber data penelitian diperoleh melalui proses *web scraping* terhadap portal berita lokal Kalimantan Tengah maupun media nasional yang aktif memberitakan kejadian karhutla. Media yang digunakan sebagai sumber data terdiri dari beberapa portal berita daring yang memiliki cakupan pemberitaan terkait kebakaran hutan dan lahan di Kalimantan Tengah. Daftar sumber media yang digunakan pada penelitian ini ditunjukkan pada Tabel 1.

Tabel 1. Sumber Portal Berita

No	Portal Berita
1	borneonews.co.id
2	tabengan.co.id
3	kaltengpost.jawapos.com
4	antaranews.com
5	kompas.com
6	cnnindonesia.com
7	cnnindonesia.com

Proses scraping dilakukan dalam dua tahapan utama. Tahap pertama adalah pengumpulan tautan berita (*link crawling*) menggunakan mesin pencari Google berdasarkan kata kunci tertentu yang berkaitan dengan karhutla di Kalimantan Tengah. Pada tahap ini, sistem melakukan pencarian artikel berita menggunakan kombinasi query yang telah ditentukan, kemudian mengambil data berupa judul berita dan tautan (*URL link*) artikel. Kata kunci dipilih untuk merepresentasikan berbagai konteks kebakaran hutan dan lahan seperti hotspot, kabut asap, kualitas udara, dan kebakaran lahan gambut. Daftar query yang digunakan pada proses pencarian ditunjukkan pada Tabel 2.

Tabel 2. Keyword yang Dipertahankan

No	Query
1	karhutla kalimantan tengah
2	karhutla kalteng
3	asap kalteng
4	hotspot kalteng
5	kebakaran lahan kalteng
6	kebakaran hutan kalteng
7	ispu kalteng
8	kabut asap
9	ahan gambut terbakar

Hasil proses *link crawling* berhasil memperoleh sebanyak 502 tautan artikel berita yang berkaitan dengan karhutla di Kalimantan Tengah. Selanjutnya dilakukan tahap kedua berupa *content scraping* terhadap seluruh tautan hasil tahap pertama. Pada tahapan ini, sistem melakukan ekstraksi isi lengkap artikel dari setiap tautan berita yang berhasil dikumpulkan. Data yang diambil pada tahap *content scraping* meliputi tanggal publikasi berita, tahun publikasi, dan isi lengkap artikel berita. Dari total 502 tautan berita yang berhasil dikumpulkan, sebanyak 460 artikel berhasil diekstraksi secara lengkap dan digunakan sebagai dataset awal penelitian. Dataset hasil scraping kemudian disusun ke dalam bentuk tabular sehingga setiap dokumen berita memiliki atribut utama berupa judul berita, tanggal, tahun, isi artikel, dan sumber media.

Sebelum dilakukan analisis utama, dataset terlebih dahulu melalui tahapan *preprocessing* untuk memastikan kualitas dan relevansi data. Tahap awal dilakukan penyaringan judul artikel berita yang dianggap keluar dari konteks pemeberitaan karhutla menggunakan bantuan agen AI ChatGPT dengan prompt "judul apa saja yg keluar konteks tidak membahas kebakaran hutan, asap tebal di kalteng (kalimantan tengah)". Hasil filtering dari AI berjumlah 33 judul yang dianggap keluar konteks pemberitaan, namun hal tersebut tidak langsung dihapus secara penuh daru corpus, tetapi kembali dilakukan proses validasi dan penyaringan manual untuk memastikan setiap artikel benar-benar relevan dengan konteks kebakaran hutan dan lahan di Kalimantan Tengah. Hasil akhir dari dokumen yang tetap dipertahankan setelah validasi manual adalah sebanyak 11 judul tetap dipertahanan dan 22 lainnya dihapus dari corpus. Setelah proses filtering selesai, dilakukan penghapusan data duplikat berdasarkan kesamaan judul artikel berita. Dari dua tahapan awal tersebut diperoleh sebanyak 436 artikel berita yang dinyatakan relevan dan digunakan sebagai korpus utama penelitian. Pada tahap analisis selanjutnya, atribut yang digunakan sebagai sumber teks utama adalah gabungan antara judul berita dan isi artikel sehingga setiap dokumen mampu merepresentasikan konteks berita secara lebih lengkap.

Tahapan berikutnya adalah *text preprocessing* yang bertujuan untuk membersihkan dan menormalisasi teks agar lebih mudah diproses pada tahap analisis NLP. Pada tahap ini, teks hasil gabungan judul dan isi artikel diproses melalui beberapa tahapan pembersihan data yang terdiri dari *lowercase*, *remove punctuation*, *remove URL*, *remove number*, *remove special character*, *tokenization*, *stopword removal* dan *stemming*.

Pada tahap *stopword removal* menerapkan dua mekanisme utama. Mekanisme pertama adalah mempertahankan kata-kata yang dianggap penting dalam pembentukan pola kausalitas karhutla sehingga kata tersebut tidak dihapus walaupun secara umum termasuk kata umum. Kata-kata tersebut dipertahankan karena memiliki peran penting dalam analisis hubungan sebab-akibat, faktor lingkungan, aktivitas manusia, dampak kesehatan, serta konteks spasial kebakaran. Khusus Spatial Context seluruh nama daerah di kalimantan tengah di tambahkan. Daftar keyword yang dipertahankan ditunjukkan pada Tabel 3.

Tabel 3. Keyword yang Dipertahankan

Kategori	Keyword
Core Causal Connectors	karena, akibat, menyebabkan, memicu, sehingga
Escalation/Event Dynamics	meningkat, meluas, menyebar, membesar
Core Wildfire Domain	karhutla, kebakaran, hotspot, asap
Environmental Factors	kemarau, panas, hujan, iklim
Peatland Factors	gambut, lahan, vegetasi, sawit
Human Activity Factors	pembakaran, aktivitas, perusahaan
Air Quality Impact	ispa, polusi, kesehatan
Disaster Response	water bombing, mitigasi, pemadaman
Temporal Sequence	sebelum, sesudah, selama
Impact Words	dampak, kerusakan, bencana
Spatial Context	kalteng, palangkaraya, kapuas, …

Mekanisme kedua adalah menetapkan kata-kata yang dianggap tidak memiliki nilai semantik penting terhadap analisis topik dan pola penyebab karhutla sehingga dapat dihapus dari korpus. Berikut daftar kata yang dihapus dirincikan pada Tabel 4.

Tabel 4. Daftar Stopword

Kategori	Keyword
Slang/Noise	yg, aja, dll, dong
Generic Connectors	dan, atau, untuk, dengan
Journalistic Words	ujar, ungkap, menurut
News Structure	berita, reporter, editor
Media Names	kompas, detik, cnn
Domain Noise	com, coid, news
Time References	senin, januari, oktober
Generic Fillers	masih, sudah, akan
Visual Navigation	foto, klik, video
Generic Location Words	wilayah, daerah, kawasan
Generic Government Words	pemerintah, dinas
Generic Verbs	melakukan, menggunakan
HTML Noise	nbsp, amp, http

Setelah seluruh korpus melewati tahap preprocessing, penelitian memasuki tahap *semantic text representation* menggunakan model transformer IndoBERT. Tahapan ini bertujuan untuk mengubah teks hasil preprocessing menjadi representasi numerik berbentuk *semantic embedding vector* yang mampu menangkap makna dan konteks semantik teks secara lebih mendalam. Setiap dokumen diproses menggunakan tokenizer IndoBERT dengan panjang maksimum 128 token. Selanjutnya model menghasilkan representasi embedding berdimensi 768 menggunakan token sebagai representasi utama dokumen. Tahap berikutnya adalah *topic discovery* menggunakan BERTopic untuk menemukan tema dominan penyebab karhutla berdasarkan representasi semantik IndoBERT. BERTopic melakukan clustering dokumen berdasarkan kemiripan makna semantik sehingga artikel dengan konteks serupa berada pada kelompok topik yang sama. Setelah proses clustering selesai, dilakukan *manual topic labeling* untuk memberikan interpretasi terhadap setiap cluster topik berdasarkan keyword dominan yang muncul. Label topik dirancang untuk merepresentasikan faktor penyebab karhutla, dampak, maupun respons penanganan bencana. Daftar label manual ditunjukkan pada Tabel 5.

Tabel 5. Manual Topic Labeling BERTopic

No	Label
1	musim kemarau ekstrem
2	cuaca panas dan kekeringan
3	angin dan percepatan penyebaran api
4	lahan gambut kering
5	vegetasi dan semak kering
6	minimnya sumber air
7	ispu kalteng
8	pembakaran lahan sengaja
9	aktivitas pembukaan lahan
10	kelalaian manusia
11	aktivitas perkebunan dan konsesi
12	penyebaran api dan asap
13	eskalasi kabut asap
14	gangguan kesehatan akibat asap
15	penurunan kualitas udara
16	upaya pemadaman karhutla
17	mitigasi dan penanganan bencana

Tahap terakhir adalah *causal pattern analysis* yang bertujuan untuk menemukan pola sebab-akibat terjadinya karhutla berdasarkan ekstraksi teks berita. Analisis dilakukan melalui tiga pendekatan utama, yaitu *co-occurrence analysis* dan *contextual relation mining*. Pada *co-occurrence analysis*, sistem menghitung frekuensi kemunculan pasangan kata dalam jendela token tertentu (*window size = 2*) untuk mengidentifikasi hubungan kata yang sering muncul secara bersamaan dalam korpus berita. Rumus untuk melakukan *co-occurrence analysis* di tuliskan pada Persamaan 1.

ISSN: 2085-0573
EISSN: 2829-1506

$$C(w_i, w_j) = \sum_{d \in D} \sum_{i=1}^{|T_d|} \sum_{j=i+1}^{\min(i+k, |T_d|)} I(w_i, w_j) \quad (1)$$

Keterangan:
$C(w_i, w_j)$ = frekuensi co-occurrence pasangan kata
D = kumpulan dokumen
T_d = token/kata pada dokumen
k = ukuran *window size*
$I(w_i, w_j)$ = pasangan kata yang muncul dalam window

Pada tahap *contextual relation mining*, sistem menghitung frekuensi kemunculan kata-kata penyebab dominan untuk menemukan pola kontekstual penyebab karhutla yang paling sering muncul dalam berita. Daftar keyword penyebab utama ditunjukkan pada Tabel 6 di bawah ini.

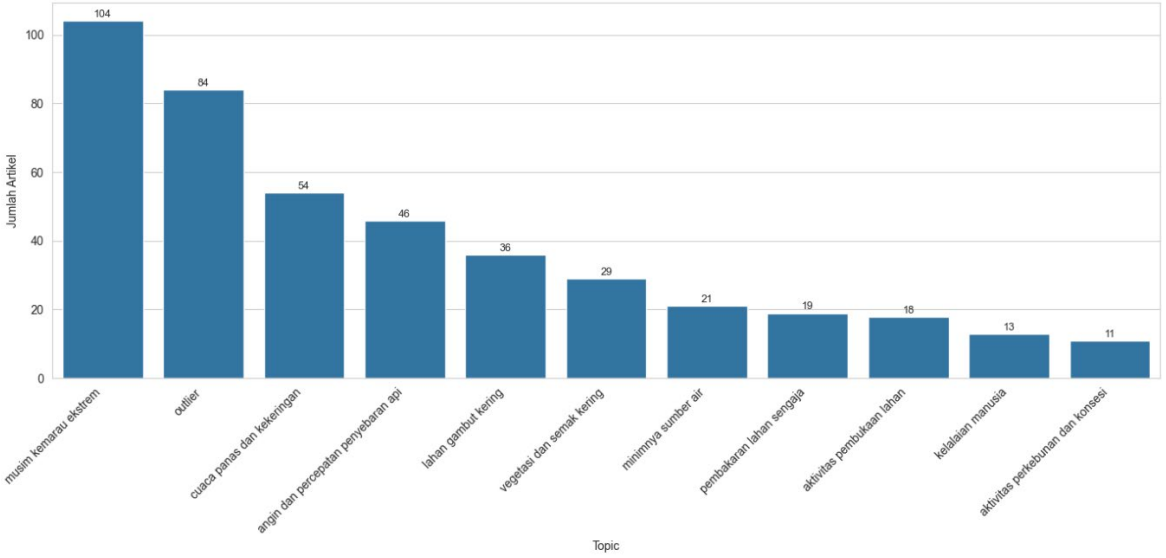
Tabel 6. Cause Keywords

No	Keywords
1	kemarau
2	gambut
3	lahan
4	angin
5	asap
6	manusia
7	pembakaran
8	perkebuan
9	kering

3. HASIL DAN PEMBAHASAN

3.1 Topic Discovery

Berdasarkan hasil proses *topic discovery* terhadap 435 artikel berita di Kalimantan Tengah, ditemukan beberapa topik utama yang menggambarkan faktor penyebab, kondisi lingkungan, serta dinamika penyebaran karhutla selama periode pengamatan. Proses identifikasi topik dilakukan menggunakan BERTopic berbasis semantic embedding IndoBERT sehingga setiap dokumen berita di cluster (dikelompokkan) berdasarkan kemiripan kata dan konteks pembahasannya. Dari hasil clustering tersebut diperoleh 10 topik utama yang ditampilkan pada Gambar 2.

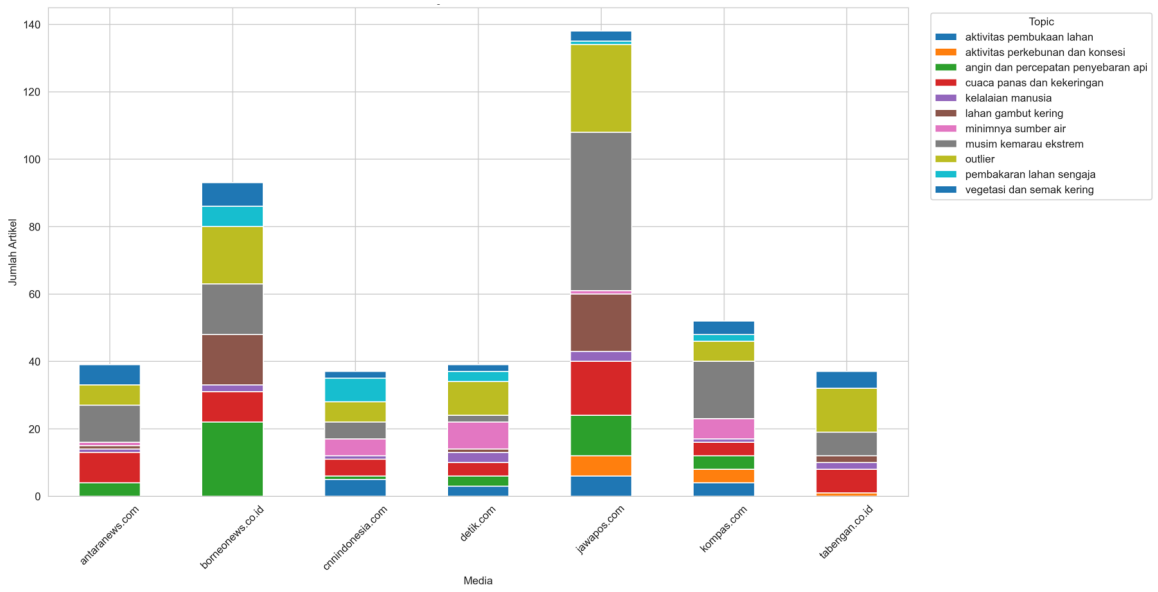


Gambar 2. Distribusi Topic

Topik dengan jumlah kemunculan tertinggi adalah topik “musim kemarau ekstrem” sebanyak 104 data. Dominasi topik ini menunjukkan bahwa kondisi cuaca menjadi faktor paling

ISSN: 2085-0573
EISSN: 2829-1506

sering dibahas dalam berita karhutla. Topik berikutnya yang juga cukup dominan adalah "cuaca panas dan kekeringan" sebanyak 54 data. Topik ini berkaitan dengan dampak langsung dari musim kemarau. Selain itu, ditemukan pula topik "angin dan percepatan penyebaran api" sebanyak 46 data yang menunjukkan bahwa arah dan kecepatan angin menjadi faktor penting dalam memperluas area kebakaran. Topik "lahan gambut kering" sebanyak 36 data menunjukkan bahwa karakteristik lahan gambut di Kalimantan Tengah memiliki kontribusi besar terhadap sulitnya proses pemadaman api. Selain itu, topik "vegetasi dan semak kering" sebanyak 29 data memperlihatkan bahwa semak belukar dan vegetasi yang mengalami kekeringan menjadi bahan bakar alami yang mempercepat munculnya titik api baru. Hasil lainnya juga menunjukkan adanya faktor manusia sebagai penyebab karhutla. Hal ini terlihat pada topik "pembakaran lahan sengaja", "aktivitas pembukaan lahan", "kelalaian manusia", serta "aktivitas perkebunan dan konsesi" dimana hal tersebut memperlihatkan bahwa aktivitas manusia tetap memiliki kontribusi signifikan terhadap munculnya kebakaran, terutama pada pembukaan lahan dengan cara membakar dan lemahnya pengawasan di area perkebunan. Selain itu, terdapat pula kelompok *outlier* yang merepresentasikan artikel berita dengan konteks campuran yang terlalu spesifik sehingga sulit dikelompokkan ke dalam cluster tertentu. Keberadaan *outlier* menunjukkan bahwa berita karhutla memiliki variasi narasi yang cukup tinggi dan tidak seluruhnya dapat direpresentasikan oleh topik mayoritas yang terdapat hampir di seluruh media berita yang ditunjukan pada Gambar 3.



Gambar 3. Distribusi Topik Berdasarkan Media

3.2 Topic Information

Untuk memaknai 10 temuan topik utama yang sudah dijelaskan pada sub bab sebelumnya yang terbentuk dari kumpulan kata kunci (keyword) yang memiliki keterkaitan konteks sehingga mampu merepresentasikan isu tertentu yang sering muncul dalam berita yang distribusi sebaran keyword ditampilkan pada Gambar 4 dibawah ini.

ISSN: 2085-0573
EISSN: 2829-1506

Gambar 4. Topic Word Score

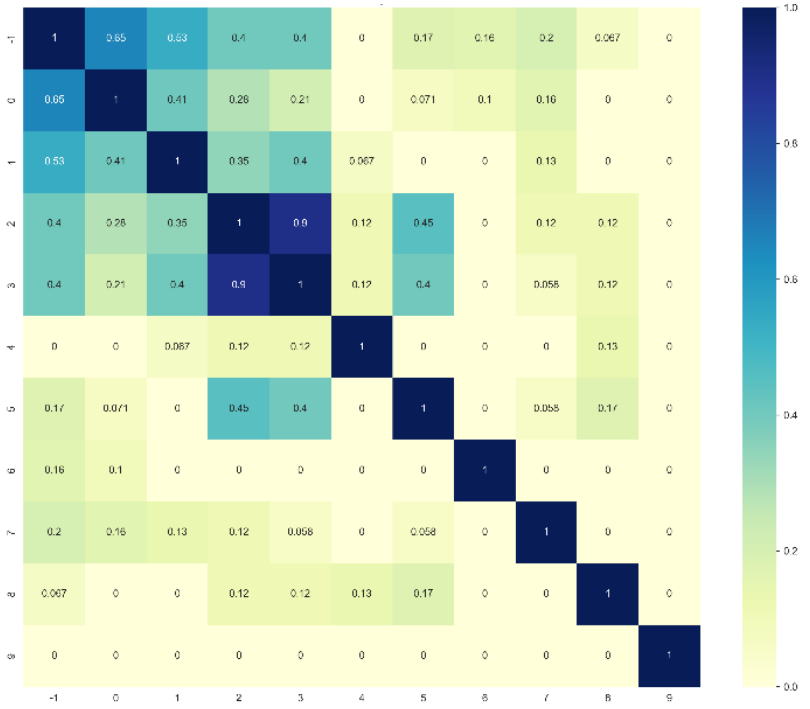
Dari visualisasi *Topic Word Scores* yang ditunjukan pada Gambar 3 diatas bahwasanya topik dengan jumlah data tertinggi adalah **Topik 0** dengan jumlah 108 data. Representasi kata pada topik ini didominasi oleh kata “karhutla”, “kalteng”, “bakar”, “bencana”, “lahan”, “kemarau”, “hutan”, “musim”, dan “darurat”. Dominasi kata-kata tersebut menunjukkan bahwa pemberitaan paling banyak berfokus pada kondisi darurat kebakaran hutan dan lahan di Kalimantan Tengah selama musim kemarau. Kata “darurat” mengindikasikan bahwa kebakaran telah berada pada kondisi serius yang memerlukan perhatian pemerintah dan masyarakat.

Selanjutnya **Topik 1** dengan jumlah 55 data berfokus pada proses kebakaran dan upaya pemadaman api. Kata dominan seperti “bakar”, “lahan”, “karhutla”, “hektare”, “madam”, “api”, dan “rawa” menunjukkan bahwa berita pada topik ini banyak membahas luas area kebakaran, lokasi lahan terbakar, serta proses pemadaman yang dilakukan oleh petugas.

Topik 2 dengan jumlah 46 data membahas dampak kabut asap. Representasi kata seperti “asap”, “palangka”, “kabut”, “kabut asap”, “raya”, dan “kalimantan” menunjukkan bahwa kabut asap menjadi isu utama yang mempengaruhi masyarakat di wilayah Palangka Raya dan sekitarnya. Topik ini menggambarkan bagaimana asap hasil kebakaran berdampak terhadap kualitas udara dan aktivitas masyarakat. **Topik 3** dengan jumlah 36 data berfokus pada masyarakat dan dampak sosial akibat karhutla. Kata seperti “masyarakat”, “karhutla”, “asap”, “kabut”, dan “lahan” menunjukkan bahwa masyarakat menjadi pihak yang paling terdampak oleh kebakaran. **Topik 4** dengan jumlah 29 data membahas kualitas udara dan kesehatan lingkungan. Kata dominan seperti “udara”, “kualitas”, “ISPU”, “alat”, “pm”, “pantau”, dan “cemar” menunjukkan bahwa berita banyak membahas pemantauan kualitas udara akibat asap kebakaran. Penggunaan istilah ISPU dan PM menunjukkan adanya perhatian terhadap tingkat pencemaran udara yang dapat membahayakan kesehatan masyarakat. **Topik 5** dengan jumlah 22 data berfokus pada kondisi kabut asap dan dampaknya terhadap jarak pandang masyarakat. Kata seperti “asap”, “kabut”, “warga”, “jarak pandang”, dan “sampit” menunjukkan bahwa kabut asap menyebabkan penurunan jarak pandang yang cukup signifikan sehingga mengganggu aktivitas masyarakat dan transportasi, khususnya di daerah Sampit. **Topik 6** dengan jumlah 22 data berkaitan dengan aspek hukum dan gugatan terhadap pemerintah. Kata dominan seperti “gugat”, “putusan”, “hukum”, “presiden”, “jokowi”, “menteri”, dan “majelis” menunjukkan adanya proses hukum serta tuntutan terhadap pemerintah terkait penanganan karhutla. Topik ini menggambarkan bahwa masalah karhutla telah menjadi isu nasional yang melibatkan kebijakan pemerintah dan tanggung jawab hukum. **Topik 7** dengan jumlah 18 data membahas hotspot atau titik panas. Kata seperti “hotspot”, “titik”, “panas”, “kalimantan”, dan “sumatera” menunjukkan adanya pemantauan titik api sebagai bagian dari mitigasi kebakaran. **Topik 8** dengan jumlah 13 data membahas dampak karhutla terhadap pendidikan. Kata “sekolah”, “libur”, “siswa”, “jam ajar”, dan “udara” menunjukkan bahwa kabut asap menyebabkan kegiatan belajar mengajar terganggu bahkan sekolah harus diliburkan demi menjaga kesehatan siswa. **Topik 9** dengan jumlah 11 data membahas gangguan akses internet di wilayah terdampak. Kata seperti “internet”, “desa”, “akses”, “blank spot”, dan “starlink” menunjukkan bahwa karhutla juga berdampak terhadap infrastruktur komunikasi, khususnya di daerah terpencil yang mengalami

ISSN: 2085-0573
EISSN: 2829-1506

keterbatasan akses jaringan. Meskipun Topik 9 ini sangat tidak relavan dengan pembahasan topik lainnya karna membahas diluar konteks karhutla yang dibuktikan pada tingkat similarity antar topik (heatmap similiraty) sangat rendah dengan nilai similirity nol terhadap hubungan konteks topik lainnya yang ditunjukan pada Gambar 5 di bawah ini.



Gambar 5. Heatmap Similarity Antar Topik

Dari hasil visualisasi heatmap similarity pada Gambar 4 diatas menunjukkan bahwa beberapa topik memiliki hubungan yang kuat, khususnya Topic 2 dan Topic 3 yang sama-sama membahas kabut asap dan dampak masyarakat. Hal ini terlihat dari nilai similarity yang tinggi serta posisi topik yang berdekatan pada intertopic distance map yang ditunjukan pada Gambar 6. Sebaliknya, Topic 9 memiliki hubungan yang sangat rendah dengan topik lain dan berada pada posisi terpisah dalam distance map, menunjukkan bahwa topik tersebut membahas isu berbeda.

 ISSN: 2085-0573
EISSN: 2829-1506

Gambar 6. Intertopic Distance Antar Topik

3.3 *Co-Occurrence Analysis*

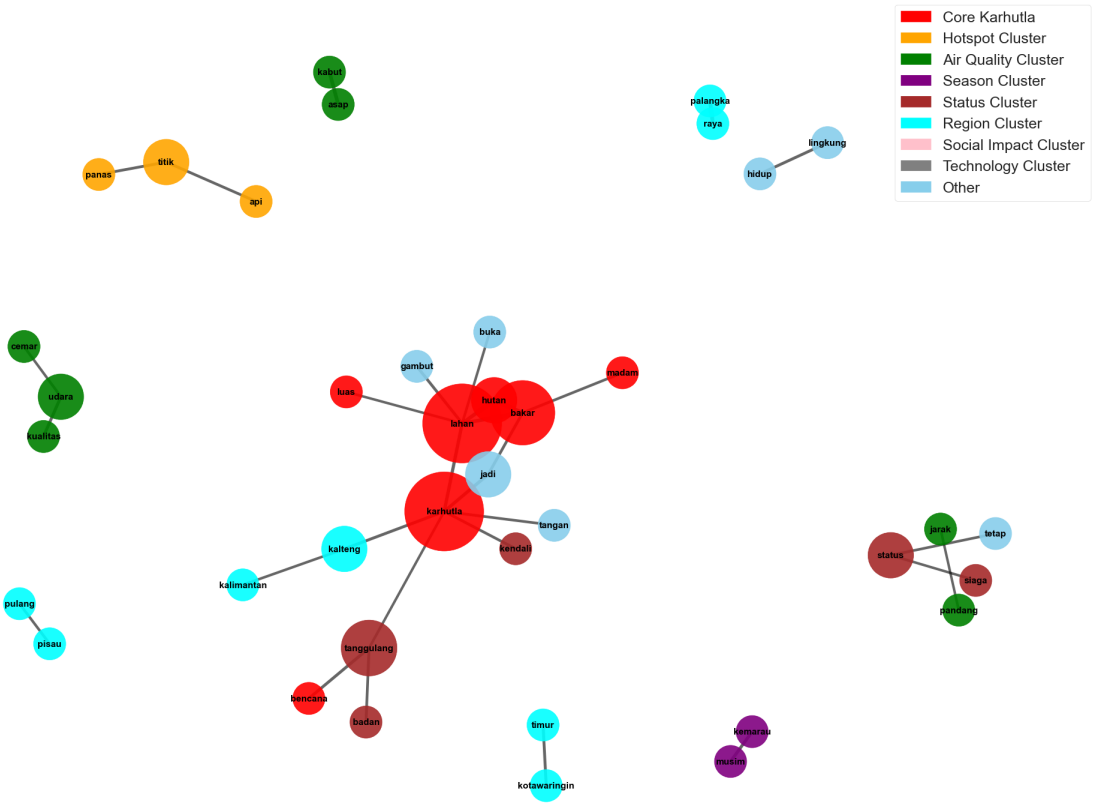
Hasil *co-occurrence analysis* dari lanjutan *topic discovery* dan *topic information*, menghasilkan 37.581 pasangan kata (word pair) yang muncul secara berdekatan antar dokumen dimana menunjukkan adanya hubungan kontekstual antar istilah yang berkaitan dengan kebakaran hutan, dampak lingkungan, wilayah terdampak, serta aktivitas masyarakat. Hasil analisis menunjukkan bahwa pasangan kata "palangka - raya" memiliki frekuensi tertinggi yaitu sebanyak 795 kemunculan. Tingginya frekuensi pasangan kata tersebut menunjukkan bahwa wilayah Palangka Raya menjadi lokasi yang paling sering dibahas dalam berita karhutla. Pasangan kata dengan frekuensi tertinggi berikutnya adalah "bakar - hutan" sebanyak 716 kemunculan, "hutan - lahan" sebanyak 613 dan "bakar - lahan" dengan frekuensi 534 kemunculan. Dominasi pasangan kata tersebut menunjukkan bahwa inti utama pemberitaan berkaitan dengan aktivitas pembakaran hutan dan lahan.Pada aspek lingkungan, pasangan kata "asap - kabut" muncul sebanyak 618 kali serta "kualitas - udara" dengan frekuensi 199 kemunculan. Tingginya kemunculan pasangan kata ini menunjukkan bahwa kabut asap dan kualitas udara merupakan dampak utama yang paling banyak dibahas dalam berita karhutla. Frekuansi pasangan kata yang muuncul ini di rincikan pada Tabel 8.

Tabel 8. Frekuensi Wordpair

No	Wordpair	Frekuensi
1	palangka – raya	795
2	bakar – hutan	716
3	asap – kabut	618
4	hutan – lahan	613
5	bakar – lahan	534
6	karthutla – lahan	313
7	jadi – karhutla	277
8	kemarau – musim	249
9	kualitas – udara	199
10	bencana – tanggulang	195
…	…	…

ISSN: 2085-0573
EISSN: 2829-1506

Pada *Co-Occurrence Word Network* yang divisualisasikan pada Gambar 7 memperlihatkan hubungan antar kata dalam bentuk jaringan (*network graph*). Pada grafik terlihat bahwa node "karhutla", "lahan", dan "bakar" menjadi pusat jaringan dengan ukuran node terbesar. Hal ini menunjukkan bahwa ketiga kata tersebut memiliki tingkat konektivitas paling tinggi terhadap kata lain dalam jaringan co-occurrence. Selain itu, grafik juga memperlihatkan beberapa cluster utama seperti cluster hotspot, kualitas udara, musim kemarau, wilayah geografis, dan status bencana. Cluster tersebut menunjukkan bahwa pemberitaan karhutla memiliki hubungan konteks yang kompleks dan multidimensional.

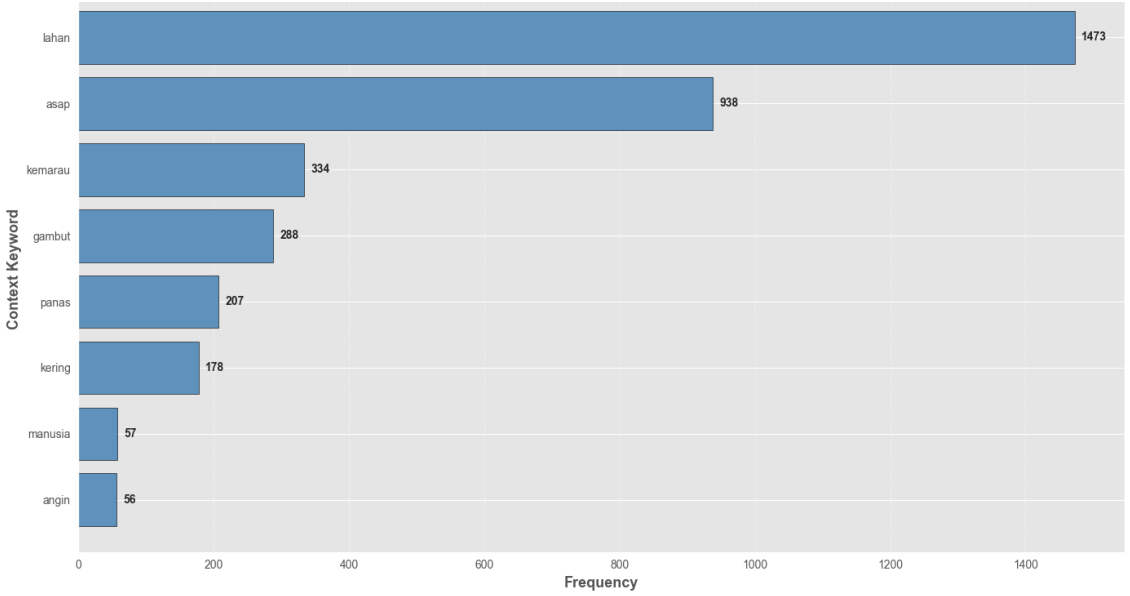


Gambar 7. Co-Ocorrence Word Network

3.4 Contextual Relation Mining

Analisis *contextual relation mining* dilakukan dengan menghitung frekuensi kemunculan kata-kata penyebab (*cause keywords*) pada seluruh dokumen hasil preprocessing dimana frekuensi kontekstual relasinya di visisualisasikan berupa grafik bar pada Gambar 8 di bawah ini.

ISSN: 2085-0573
EISSN: 2829-1506



Gambar 8. Contextual Relation Keyword Frequency

Pada Gambar 8 menunjukkan bahwa keyword “lahan” menjadi konteks penyebab paling dominan dengan total frekuensi sebanyak 1.473 kemunculan. Tingginya frekuensi keyword tersebut menunjukkan bahwa isu lahan menjadi fokus utama dalam pembahasan karhutla di Kalimantan Tengah. Keyword dominan berikutnya adalah “asap” dengan total 938 kemunculan. Tingginya kemunculan keyword ini menunjukkan bahwa asap menjadi salah satu dampak dan konteks utama yang paling sering dibahas dalam berita karhutla. Pada posisi berikutnya terdapat keyword “kemarau” dengan frekuensi 334 kemunculan. Hasil ini menunjukkan bahwa musim kemarau menjadi faktor penyebab yang sangat sering dikaitkan dengan meningkatnya kejadian kebakaran hutan dan lahan. Keyword “gambut” muncul sebanyak 288 kali dan menunjukkan bahwa lahan gambut memiliki keterkaitan yang kuat terhadap fenomena karhutla. Tingginya frekuensi keyword tersebut menunjukkan bahwa karakteristik lahan gambut yang mudah terbakar menjadi salah satu faktor penting dalam pemberitaan kebakaran hutan dan lahan di Kalimantan Tengah. Selain itu, keyword “panas” dengan 207 kemunculan dan “kering” sebanyak 178 kemunculan menunjukkan bahwa faktor suhu panas dan kondisi lingkungan yang kering menjadi konteks penyebab yang cukup dominan dalam berita. Sementara itu, keyword “manusia” sebanyak 57 kemunculan memperlihatkan bahwa aktivitas manusia juga menjadi salah satu faktor penyebab karhutla, meskipun frekuensinya lebih rendah dibandingkan faktor lingkungan dan cuaca. Keyword “angin” dengan 56 kemunculan menunjukkan bahwa faktor angin turut berkontribusi terhadap penyebaran api dan asap saat terjadi kebakaran hutan dan lahan.

4. KESIMPULAN

Pada penelitian ini telah menemukan 10 topik utama yang merepresentasikan berbagai aspek karhutla di Kalimantan Tengah. Topik "musim kemarau ekstrem" mendominasi pemberitaan dengan 104 data, diikuti oleh "cuaca panas dan kekeringan" serta "angin dan percepatan penyebaran api", yang mengindikasikan bahwa faktor meteorologis merupakan konteks paling sering dibahas dalam narasi berita karhutla. Sementara itu, faktor antropogenik seperti pembakaran lahan sengaja, aktivitas pembukaan lahan, dan kelalaian manusia juga ditemukan sebagai kontributor signifikan. Hasil *contextual relation mining* memperkuat temuan tersebut, di mana kata "lahan" (1.473 kemunculan), "asap" (938), "kemarau" (334), dan "gambut" (288) menjadi konteks penyebab paling dominan, menunjukkan bahwa karhutla di Kalimantan Tengah merupakan fenomena yang dipicu oleh interaksi kompleks antara faktor lingkungan, iklim, karakteristik lahan gambut, dan aktivitas manusia. Hasil *co-occurrence analysis* dengan 37.581 pasangan kata juga menunjukkan bahwa jaringan pemberitaan karhutla bersifat multidimensional dengan kluster-kluster konteks yang saling terhubung.

Kelebihan penelitian ini terletak pada pengembangan framework analisis terintegrasi yang menggabungkan representasi semantik IndoBERT, BERTopic, dan *causal pattern mining* untuk

memproses teks berita berbahasa Indonesia secara otomatis untuk mengatasi keterbatasan analisis manual yang selama ini digunakan. Pendekatan ini juga terbukti mampu menangkap hubungan antar faktor penyebab karhutla yang tidak dapat diperoleh dari analisis data spasial atau hotspot semata. Namun demikian, penelitian ini memiliki keterbatasan pada cakupan sumber data yang masih terbatas pada beberapa portal berita tertentu, sehingga berpotensi menimbulkan bias representasi. Selain itu, metode *causal pattern mining* yang diterapkan masih berbasis *co-occurrence* dan frekuensi kemunculan kata, sehingga belum sepenuhnya mampu menangkap struktur kausalitas yang kompleks dan berjenjang. Keberadaan topik *outlier*, khususnya Topik 9 yang membahas akses internet, juga menunjukkan bahwa proses *filtering* dan pembatasan konteks masih perlu ditingkatkan.

Penelitian selanjutnya dapat diarahkan pada peningkatan kapasitas ekstraks hubungan kausal menggunakan pendekatan *dependency parsing* atau *knowledge graph*, perluasan cakupan sumber data berita, serta integrasi dengan data spasial hotspot untuk membangun sistem *disaster intelligence* yang lebih komprehensif dan mendukung pengambilan keputusan mitigasi karhutla secara real-time.

DAFTAR PUSTAKA

[1] A. Schmidt *et al.*, “Fire frequency, intensity, and burn severity in Kalimantan’s threatened Peatland areas over two Decades,” *Frontiers in Forests and Global Change*, vol. 7, Feb. 2024, doi: 10.3389/ffgc.2024.1221797.

[2] Y. Bao *et al.*, “Satellite-Based Analysis of Spatiotemporal Wildfire Pattern in the Mongolian Plateau,” *Remote Sens. (Basel).*, vol. 15, no. 1, p. 190, Dec. 2022, doi: 10.3390/rs15010190.

[3] F. Sari, “Identifying anthropogenic and natural causes of wildfires by maximum entropy method-based ignition susceptibility distribution models,” *J. For. Res. (Harbin).*, vol. 34, no. 2, 2023, doi: 10.1007/s11676-022-01502-4.

[4] F. Sufi and M. Alsulami, “AI-Driven Global Disaster Intelligence from News Media,” *Mathematics*, vol. 13, no. 7, p. 1083, Mar. 2025, doi: 10.3390/math13071083.

[5] F. Sufi and M. Alsulami, “Disaster in the Headlines: Quantifying Narrative Variation in Global News Using Topic Modeling and Statistical Inference,” *Mathematics*, vol. 13, no. 13, 2025, doi: 10.3390/math13132049.

[6] Z. Gouliev and F. Perez-Tellez, “Topic Modelling and Sentiment Analysis via News Headlines, NLP Methods on Australian Broadcasting Commission,” *Computacion y Sistemas*, vol. 28, no. 4, 2024, doi: 10.13053/CyS-28-4-5305.

[7] C. Silva, I. Carvalho, B. Ferreira, J. C. Pinto, A. Cardoso, and H. Gonçalo Oliveira, “Twitter and Sentiment Analysis for Wildfire Heat Mapping,” *Expert Syst.*, vol. 42, no. 11, 2025, doi: 10.1111/exsy.70147.

[8] S. M. Anugerah, R. Wijaya, and M. A. Bijaksana, “Sentimen Analysis Social Media for Disaster using Naïve Bayes and IndoBERT,” *INTEK: Jurnal Penelitian*, vol. 11, no. 1, 2024, doi: 10.31963/intek.v11i1.4771.

[9] G. R. W. Syurifah and H. Fahmi, “Implementasi Metode ST-DBSCAN untuk Pengelompokan Pola Persebaran Titik Api pada Data Kebakaran Hutan di Indonesia,” *Jurnal Riset Mahasiswa Matematika*, vol. 3, no. 5, 2024, doi: 10.18860/jrmm.v3i5.27314.

[10] K. Milleville, S. Van Ackere, J. Verdoodt, S. Verstockt, P. De Maeyer, and N. Van de Weghe, “Exploring the potential of social media to study environmental topics and natural disasters,” *Journal of Location Based Services*, vol. 17, no. 4, 2023, doi: 10.1080/17489725.2023.2238663.

[11] F. Maulana and K. Kusrini, “HYPERPARAMETER MODEL LSTM-GRU UNTUK PREDIKSI PEMETAAN TINGKAT KEBAKARAN HUTAN,” *JIKA (Jurnal Informatika)*, vol. 9, no. 1, 2025, doi: 10.31000/jika.v9i1.12882.

[12] X. Lin, Z. Li, W. Chen, X. Sun, and D. Gao, “Forest Fire Prediction Based on Long- and Short-Term Time-Series Network,” *Forests*, vol. 14, no. 4, p. 778, Apr. 2023, doi: 10.3390/f14040778.

[13] T. Dasgupta, A. Naskar, L. Dey, and M. Shakir, “A Joint Model for Detecting Causal Sentences and Cause-Effect Relations from Text,” 2022. doi: 10.3233/ssw220021.

[14] O. Hassanzadeh *et al.*, "Causal knowledge extraction through large-scale text mining," in *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, 2020. doi: 10.1609/aaai.v34i09.7092.

[15] H. Q. Yu and S. Reiff-Marganiec, "Learning Disease Causality Knowledge From the Web of Health Data," *Int. J. Semant. Web Inf. Syst.*, vol. 18, no. 1, 2022, doi: 10.4018/IJSWIS.297145.