



Tren Riset Deteksi Ujaran Kebencian: Analisis Bibliometrik 2020–2025

Denina Nastiti Putri Amani¹, Syopiansyah Jaya Putra², Qurrotul Aini^{3*}

^{1,2,3} Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, UIN Syarif Hidayatullah Jakarta, Indonesia

Info Artikel

Riwayat Artikel:Diterima : **7 Desember 2025**Direvisi : **19 Desember 2025**Diterbitkan : **1 Juni 2026****Kata Kunci:**

Ujaran Kebencian,
Natural Language Processing,
Transformer,
Bibliometrik,
Multimodal dan Multibahasa

Keywords:

Hate Speech,
Natural Language Processing,
Transformer Model,
Bibliometric Analysis,
Multimodal and Multilingual Approaches

ABSTRAK

Perkembangan media sosial yang sangat cepat meningkatkan risiko munculnya ujaran kebencian (*hate speech*), sehingga mendorong penelitian dalam bidang *Natural Language Processing* (NLP) untuk mengembangkan sistem deteksi otomatis yang lebih akurat. Selama satu dekade terakhir, pendekatan deteksi ujaran kebencian mengalami evolusi signifikan, mulai dari metode *machine learning* klasik hingga arsitektur *deep learning* dan model *transformer* yang lebih mutakhir. Namun, kajian bibliometrik yang memetakan perkembangan metode dan domain implementasi masih terbatas. Penelitian ini menganalisis 1.335 publikasi dari basis data *Scopus* untuk mengidentifikasi tren teknik penelitian (misalnya SVM, *Naive Bayes*, LSTM, dan keluarga BERT) serta domain penerapannya (Twitter, Facebook, YouTube, dan konteks multibahasa). Analisis dilakukan menggunakan Python melalui ekstraksi kata kunci dan visualisasi tren temporal. Hasil menunjukkan dominasi model *transformer* sejak 2020 dan pergeseran fokus dari teks tunggal menuju pendekatan multimodal dan multilingual. Studi ini menegaskan arah pengembangan riset menuju integrasi *transformer*, multibahasa, dan *Explainable AI* (XAI) untuk meningkatkan transparansi deteksi ujaran kebencian

ABSTRACT

A The rapid growth of social media has increased the risk of hate speech proliferation, driving extensive research in Natural Language Processing (NLP) to develop more accurate automatic detection systems. Over the past decade, hate speech detection approaches have evolved significantly, shifting from traditional machine learning methods to deep learning architectures and advanced transformer-based models. However, comprehensive bibliometric studies that map methodological developments and implementation domains remain limited. This research analyzes 1,335 publications indexed in Scopus to identify trends in methodological approaches (e.g., SVM, Naive Bayes, LSTM, and BERT-family models) and application domains (Twitter, Facebook, YouTube, and multilingual contexts). Keyword extraction and temporal trend visualization were conducted using Python. The findings indicate that transformer models have dominated research since 2020, accompanied by a shift from single-text analysis toward multimodal and multilingual approaches. This study highlights future research directions involving transformer integration, multilingual processing, and Explainable AI (XAI) to enhance transparency in hate speech detection.

Penulis Korespondensi:

Qurrotul Aini,
Program Studi Sistem Informasi,
Universitas UIN Syarif Hidayatullah Jakarta
Email: qurrotul.aini@uinjkt.ac.id

This is an open access article under the [CC BY-SA](#) license



1. PENDAHULUAN

Perkembangan teknologi informasi dan media sosial telah membawa dampak signifikan terhadap pola komunikasi masyarakat modern [1]. Platform seperti X, Facebook, Instagram, dan YouTube telah menjadi ruang terbuka untuk mengekspresikan opini publik [2]. Namun, kebebasan berekspresi ini juga memunculkan tantangan baru, yaitu penyebaran ujaran kebencian (*hate speech*) yang dapat menimbulkan konflik sosial, diskriminasi, bahkan kekerasan berbasis identitas [3],[4]. Oleh karena itu, pengembangan sistem otomatis untuk deteksi ujaran kebencian menjadi fokus utama dalam bidang *Natural Language Processing* (NLP) dan *Artificial Intelligence* (AI).

Dalam satu dekade terakhir, penelitian di bidang deteksi ujaran kebencian menunjukkan peningkatan pesat baik dari segi jumlah publikasi maupun kompleksitas metode yang digunakan [5]. Pendekatan tradisional berbasis *machine learning* seperti Support Vector Machine (SVM), Naive Bayes, dan Logistic Regression secara bertahap bergeser menuju arsitektur *deep learning* seperti Convolutional Neural Network (CNN) dan Long Short-Term Memory (LSTM), serta model *transformer-based* yang lebih mutakhir. Model-model awal keluarga *transformer* seperti BERT, RoBERTa, dan XLNet banyak diadopsi dalam penelitian deteksi ujaran kebencian karena perannya sebagai arsitektur fundamental yang mendominasi literatur dan menjadi dasar pengembangan berbagai pendekatan lanjutan, khususnya dalam tugas klasifikasi teks dan pemahaman konteks semantik yang kompleks [6], [7].

Selain dari sisi metode, domain implementasi penelitian ini juga berkembang pesat. Awalnya, riset banyak berfokus pada deteksi ujaran kebencian di X, namun kini telah meluas ke berbagai platform dan bahasa, termasuk YouTube, Reddit, serta konteks multilingual dan multimodal [8], [9], [10], [11]. Perkembangan ini menunjukkan perlunya pemetaan yang komprehensif untuk memahami arah, intensitas, serta perubahan pola riset di bidang ini.

Dalam konteks tersebut, analisis bibliometrik menjadi pendekatan yang efektif untuk meninjau secara sistematis perkembangan suatu bidang penelitian [12]. Analisis bibliometrik bertujuan memberikan wawasan mendalam mengenai tren penelitian, perkembangan bidang, dan kolaborasi antar peneliti. Dengan memetakan publikasi yang ada, pendekatan ini dapat mengidentifikasi topik yang sedang berkembang serta kontribusi penulis maupun institusi tertentu [13]. Selain itu, analisis bibliometrik membantu para pemangku kepentingan untuk mengevaluasi kekuatan dan kelemahan literatur yang ada, serta mengidentifikasi celah pengetahuan (*research gaps*) yang dapat menjadi dasar penelitian berikutnya.

Beberapa penelitian sebelumnya juga telah menerapkan pendekatan serupa pada topik yang berkaitan dengan ujaran kebencian. Misalnya, [14] menyajikan tinjauan sistematis atas *benchmark* deteksi ujaran kebencian di media sosial menunjukkan pertumbuhan publikasi yang tajam sejak pertengahan 2010-an dan dominasi *supervised learning*. Sementara itu, [15] menelaah hubungan antara kebijakan moderasi platform dan dinamika ujaran kebencian, menyoroti tantangan bias data dan bahasa minoritas. Namun, sejauh pengetahuan penulis, belum ada studi yang secara khusus memetakan tren metode dan domain implementasi deteksi ujaran kebencian pada rentang waktu luas berbasis data bibliometrik besar.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk menganalisis secara bibliometrik perkembangan penelitian deteksi ujaran kebencian dengan meninjau dua dimensi utama yaitu evolusi metode penelitian, meliputi pendekatan *machine learning*, *deep learning*, dan *transformer-based models* serta domain implementasi, yang mencakup media sosial dan konteks multibahasa. Dengan menggunakan data dari basis Scopus yang terdiri atas 1.335 publikasi dan analisis berbasis Python, penelitian ini diharapkan dapat memberikan gambaran komprehensif mengenai tren riset, arah perkembangan, serta peluang penelitian masa depan dalam bidang deteksi ujaran kebencian.

2. METODE PENELITIAN

Metode penelitian yang diterapkan adalah analisis bibliometrik terhadap jurnal ilmiah yang membahas deteksi ujaran kebencian. Studi bibliometrik dilakukan melalui beberapa tahapan utama pencarian data, filtering, input pada perangkat analisis, dan interpretasi hasil sebagaimana panduan praktis terkini untuk penggabungan serta pembersihan data dan pemakaian multi alat bantu (*tools*). Alur penelitian ditampilkan pada Gambar 1.



Gambar 1. Alur Penelitian

2.1. Pengumpulan Data

Data penelitian diperoleh dari basis data Scopus dengan hasil pencarian awal sebanyak 1.731 entri BibTeX. Setelah melalui proses data cleaning untuk menghapus duplikasi dan entri tanpa metadata lengkap (misalnya tanpa tahun atau judul), diperoleh 1.335 publikasi valid yang dianalisis lebih lanjut menggunakan pendekatan bibliometrik berbasis Python [16].

Pengumpulan data memanfaatkan metadata dari Scopus. Pencarian menggunakan kata kunci utama "*hate speech detection*" dan "*offensive language detection*" untuk menjangkau lingkup yang luas, serta fitur *advanced search* Scopus guna memastikan relevansi terhadap tujuan penelitian. Setelah hasil pencarian diperoleh, dilakukan filtering untuk menyaring artikel sesuai kriteria penelitian. Batasan meliputi jenis publikasi *journal articles* dan *conference proceedings* serta rentang tahun 2020–2025 agar mencerminkan perkembangan penelitian terkini [17].

2.2. Memfilter dan Input Data

Data yang telah difilter kemudian diekspor dan dianalisis menggunakan kombinasi perangkat lunak analisis bibliometrik. Analisis dan visualisasi tren dilakukan dengan kombinasi R (*bibliometrix* atau *Biblioshiny*), VOSviewer, dan Python pada Google Colab untuk otomatisasi dan replikasi *workflow* [18], [19]. Perangkat-perangkat tersebut dimanfaatkan untuk menganalisis tren publikasi, distribusi metode penelitian, serta memvisualisasikan keterkaitan kata kunci dan jaringan penelitian. Penggunaan kombinasi multi-alat ini sejalan dengan praktik kontemporer dalam studi bibliometrik lintas domain [19], [20].

2.3. Interpretasi Data

Tahap interpretasi data dilakukan dengan menganalisis hasil *output* dari proses analisis bibliometrik dan visualisasi yang dihasilkan. Interpretasi difokuskan pada pemahaman tren publikasi, perkembangan metode penelitian, serta keterkaitan antar topik dalam bidang deteksi ujaran kebencian. Hasil visualisasi jaringan dan *density* digunakan untuk mengidentifikasi topik yang dominan, pola hubungan antar kata kunci, serta pergeseran fokus penelitian selama periode tertentu. Temuan pada tahap ini selanjutnya digunakan sebagai dasar dalam pembahasan hasil dan penarikan kesimpulan penelitian.

3. HASIL DAN ANALISIS

Bagian ini menyajikan hasil analisis bibliometrik terhadap publikasi ilmiah yang membahas deteksi ujaran kebencian (*hate speech detection*) pada periode 2020–2025. Analisis dilakukan untuk memetakan karakteristik data, tren publikasi, serta perkembangan metode dan domain penelitian yang dominan dalam literatur. Penyajian hasil diawali dengan pemaparan proses dan karakteristik data yang diperoleh dari basis data Scopus, kemudian dilanjutkan dengan analisis demografi literatur, distribusi bidang penelitian, frekuensi metode yang digunakan, serta keterkaitan topik penelitian melalui analisis jaringan kata kunci. Seluruh hasil dianalisis dan divisualisasikan untuk memberikan gambaran komprehensif mengenai arah dan dinamika riset deteksi ujaran kebencian.

3.1. Hasil Pengumpulan Data

Berdasarkan hasil pengumpulan data pada situs Scopus, diperoleh 1.335 publikasi terkait dengan deteksi ujaran kebencian, yang mencakup metadata penelitian seperti judul jurnal, abstrak, penulis, sitasi, DOI, serta berbagai elemen penting lainnya. Metadata ini diekstrak dalam format CSV, yang mencakup berbagai kolom yang relevan untuk analisis bibliometrik. Beberapa kolom utama dalam metadata tersebut antara lain adalah tahun publikasi, sumber publikasi, DOI, afiliasi penulis, kata kunci, dan referensi.

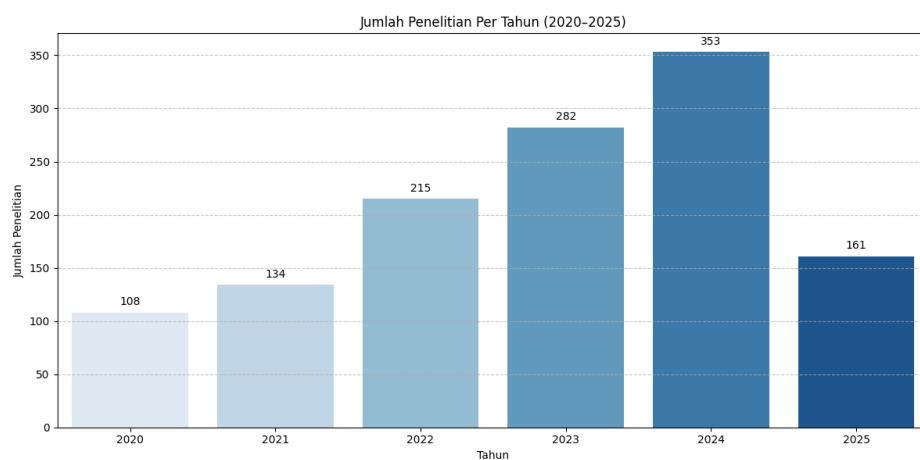
Berikut adalah contoh elemen-elemen metadata yang ada dalam *dataset*, yang diekstrak dan digunakan dalam analisis:

```
Index(['Authors', 'Author full names', 'Author(s) ID', 'Title', 'Year',
      'Source title', 'Volume', 'Issue', 'Art. No.', 'Page start', 'Page end',
      'Page count', 'Cited by', 'DOI', 'Link', 'Affiliations',
      'Authors with affiliations', 'Abstract', 'Author Keywords',
      'Index Keywords', 'References', 'Editors', 'Publisher', 'ISSN', 'ISBN',
      'CODEN', 'PubMed ID', 'Language of Original Document', 'Document Type',
      'Publication Stage', 'Open Access', 'Source', 'EID', 'Research Field'],
      dtype='object')
```

Elemen-elemen tersebut digunakan untuk memetakan tren publikasi, kolaborasi antar penulis, serta eksplorasi perkembangan bidang penelitian deteksi ujaran kebencian.

3.2. Demografi Literatur

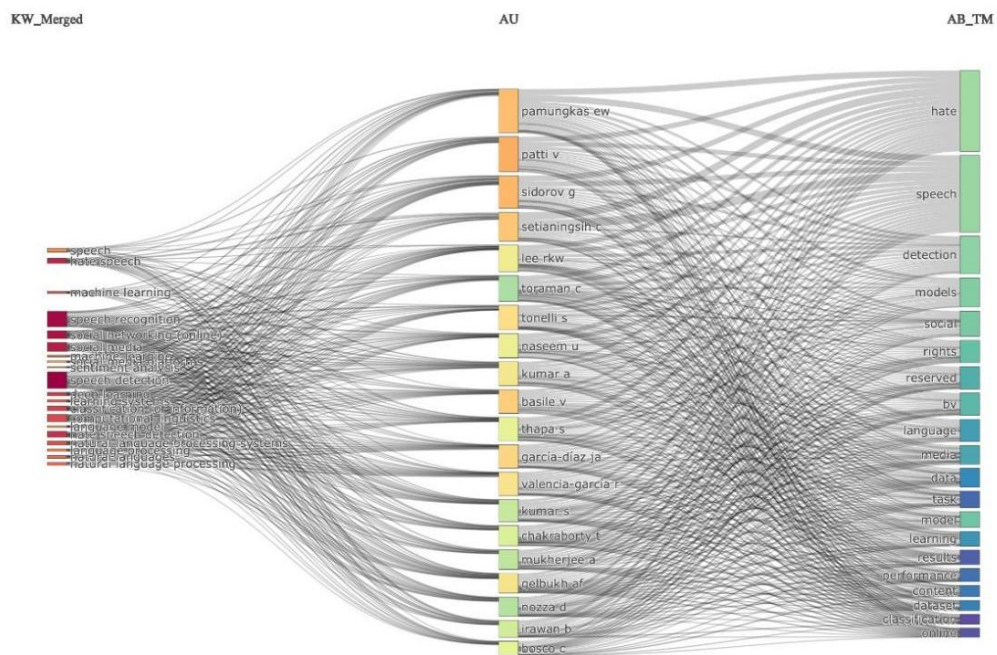
Analisis pertama merupakan analisis demografi literatur untuk mengetahui gambaran secara garis besar mengenai literatur yang membahas mengenai analisis *hate speech detection*. Analisis demografi pertama dilakukan untuk melihat sebaran penelitian tentang analisis ujaran kebencian pertahunnya pada Gambar 2.



Gambar 2. Jumlah Penelitian Per Tahun

Pada Gambar 2, terlihat bahwa jumlah penelitian terkait *hate speech detection* mengalami peningkatan yang signifikan dari tahun 2020 hingga 2024. Pada tahun 2020, jumlah publikasi tercatat sebanyak 108, kemudian meningkat menjadi 134 pada tahun 2021, 215 pada tahun 2022, dan 282 pada tahun 2023. Puncaknya terjadi pada tahun 2024, dengan total 353 publikasi, yang menunjukkan lonjakan signifikan dibandingkan tahun-tahun sebelumnya. Namun, pada tahun 2025, jumlah publikasi menurun menjadi 161, meskipun tren penelitian tetap menunjukkan perhatian yang tetap ada terhadap topik ini. Penurunan tersebut tidak mengurangi relevansi isu *hate speech detection*, yang tetap menjadi fokus penting dalam penelitian *Natural Language Processing* (NLP).

Analisis berikutnya merupakan analisis hubungan antara penerbit jurnal, kata kunci penelitian, dan negara peneliti dalam bidang analisis ujaran kebencian (*hate speech*). Analisis ini dilakukan untuk melihat keterkaitan antara topik penelitian dan kontribusi dari berbagai negara, yang ditunjukkan pada Gambar 3.



Gambar 3. Diagram Sankey

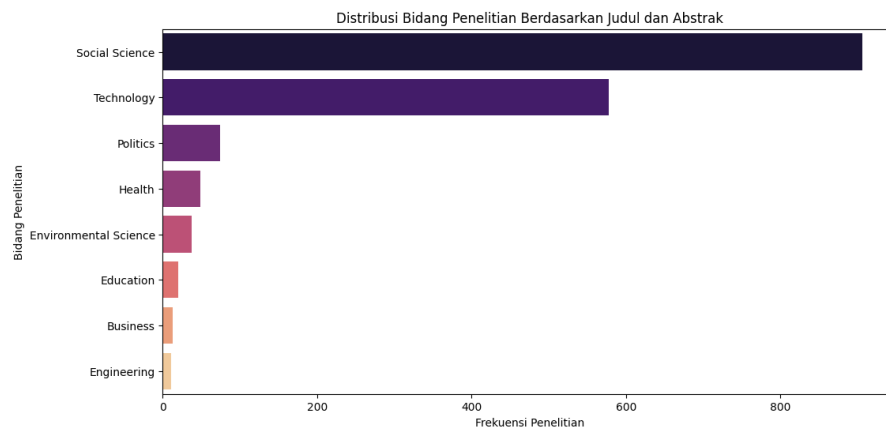
Pada Gambar 3, diagram Sankey menggambarkan hubungan antara penerbit jurnal, kata kunci penelitian, dan negara peneliti dalam analisis ujaran kebencian (*hate speech*). Interpretasi hasil diagram Sankey ini sebagai berikut:

- Distribusi topik penelitian, di mana kata kunci dominan yang muncul: *hate speech*, *speech recognition*, *machine learning*, *deep learning*, *social networking*, dan *sentiment analysis*. Hal ini menunjukkan fokus utama riset adalah pada deteksi ujaran kebencian berbasis NLP dan AI, dengan pendekatan teknis seperti klasifikasi teks dan analisis media sosial.
- Kontribusi negara, di mana negara yang paling banyak berkontribusi adalah Amerika Serikat, India, dan China. Hal ini menunjukkan bahwa negara-negara besar dengan ekosistem riset AI yang kuat menjadi kontributor utama dalam pengembangan metode deteksi *hate speech*.
- Keterkaitan antara penulis/peneliti dan kata Kunci, di mana nama-nama penulis (misalnya patti v, sidorov g, toraman c, basile v) terhubung dengan kata kunci seperti *hate speech*, *machine learning*, dan *social media*. Ini menunjukkan adanya jejaring riset internasional yang berfokus pada tema serupa, meskipun berasal dari institusi dan negara berbeda.
- Pola kolaborasi, di mana memperlihatkan aliran dari kata kunci ke negara dan penulis, sehingga dapat diinterpretasikan bahwa: topik populer seperti *hate speech detection* menarik kolaborasi lintas negara. Serta ada keterkaitan erat antara domain sosial (*social networking*) dan domain teknis (*machine learning/deep learning*).

3.3. Analisis Domain Penelitian

Selanjutnya, dilakukan analisis mengenai distribusi bidang penelitian berdasarkan judul dan abstrak. Gambar 4 menunjukkan kategori-kategori utama dalam domain penelitian, dengan bidang *social science* dan *technology* yang memiliki frekuensi penelitian yang dominan. Visualisasi ini memberikan wawasan mengenai fokus penelitian yang paling banyak dilakukan, serta membandingkan proporsi kontribusi berbagai bidang ilmu dalam topik-topik penelitian yang ada. Gambar 4 menunjukkan distribusi bidang penelitian berdasarkan judul dan abstrak, dengan *social science* dan *technology* sebagai dua bidang yang paling dominan, terlihat dengan frekuensi penelitian yang sangat tinggi. Bidang *social science* memiliki jumlah penelitian tertinggi, diikuti oleh *technology*. Bidang *social science* ini banyak berkaitan dengan pemahaman konteks sosial dalam

hate speech detection, seperti analisis sosial media, kajian tentang radikalisasi, dan dinamika sosial yang berperan dalam penyebaran ujaran kebencian.

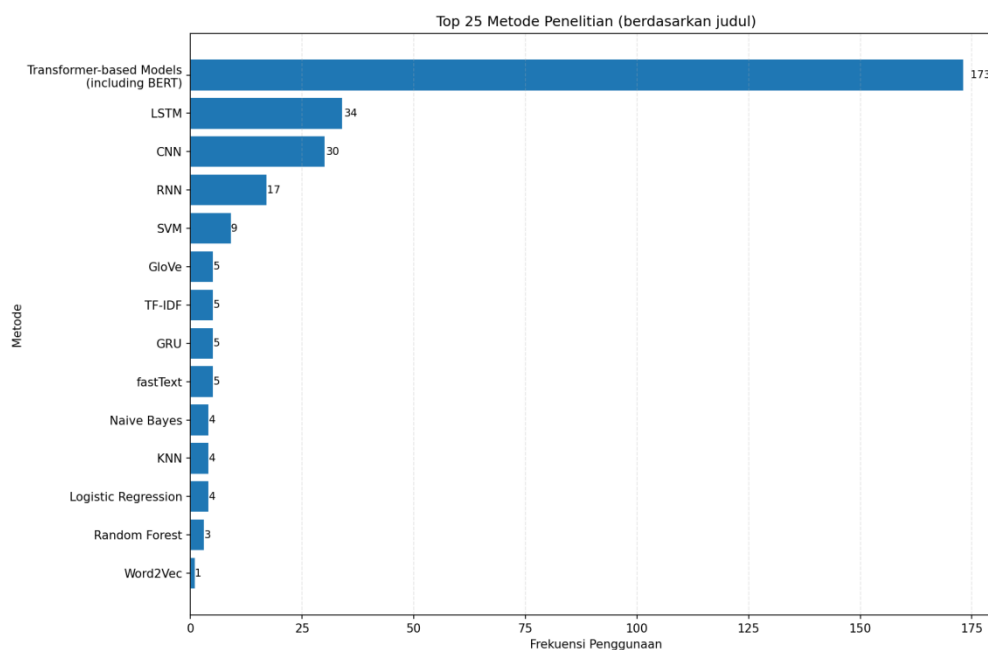


Gambar 4. Distribusi Bidang Penelitian Berdasarkan Judul dan Abstrak

Sementara itu, bidang *technology* berfokus pada pengembangan model teknis untuk deteksi otomatis, seperti *machine learning* dan *deep learning*. Bidang lainnya, seperti *politics*, *health*, dan *environmental science*, meskipun dengan frekuensi yang lebih rendah, turut memberikan kontribusi yang signifikan dalam penelitian terkait dampak sosial dari ujaran kebencian di berbagai sektor. Visualisasi ini memberikan gambaran mengenai fokus utama dari penelitian yang ada, dengan penekanan pada peran penting *social science* dalam mengidentifikasi dan memahami konteks sosial di balik fenomena *hate speech* serta pengembangan teknologi yang mendukung deteksinya.

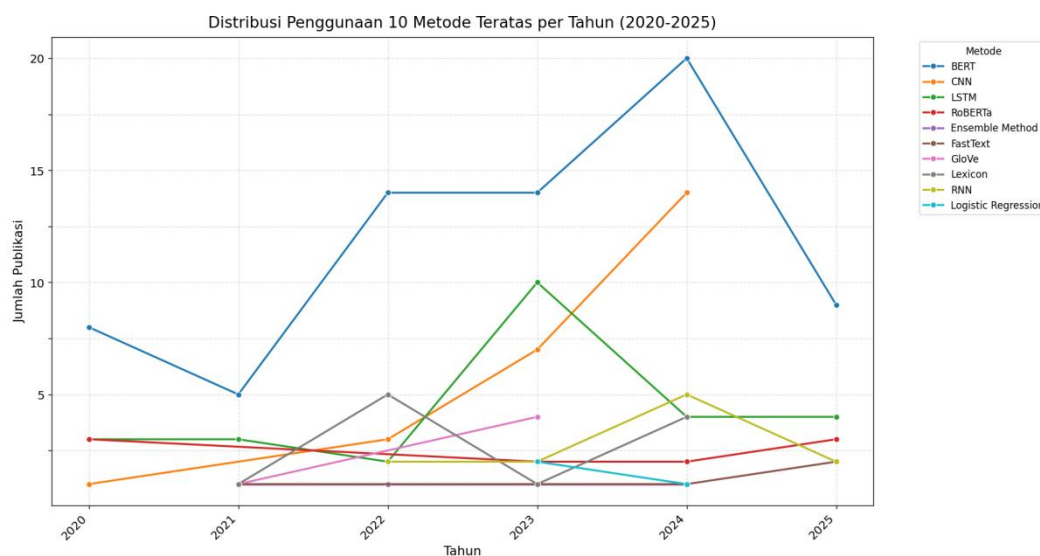
3.4. Analisis Frekuensi Metode Penelitian

Selanjutnya, dilakukan analisis mengenai distribusi frekuensi penggunaan metode penelitian yang paling banyak digunakan dalam topik *hate speech detection*. Gambar 5 menampilkan 'Top 25' metode penelitian berdasarkan judul.



Gambar 5. Top 25 Model Penelitian Berdasarkan Judul

Gambar 5 menunjukkan distribusi frekuensi penggunaan metode penelitian dalam deteksi *hate speech*, di mana *transformer-based models* (termasuk BERT) mendominasi dengan 173 frekuensi, jauh lebih tinggi dibandingkan metode lainnya. Metode lain seperti LSTM dan CNN juga memiliki kontribusi signifikan, masing-masing dengan 34 dan 30 frekuensi. Selain itu, metode RNN dan SVM mencatatkan frekuensi penggunaan masing-masing sebesar 17 dan 9, menggambarkan variasi pendekatan dalam penelitian ini. Selanjutnya, pada Gambar 6, dilakukan analisis distribusi penggunaan 10 metode teratas dalam penelitian *hate speech detection* selama periode 2020 hingga 2025. Penurunan penggunaan BERT di tahun 2025, karena bermunculan model baru seperti RoBERTa, XLNet, GPT, dan LLM generasi baru, sehingga banyak peneliti beralih ke model yang lebih efisien dan akurat, yang berdampak pada frekuensi publikasi BERT. Selain itu, fokus riset bergeser dari analisis teks tunggal ke multimodal (teks+gambar+video) dan multilingual. BERT kurang optimal untuk konteks ini, sehingga peneliti lebih memilih varian atau model lain yang mendukung *cross-lingual learning* dan *multimodal processing*.



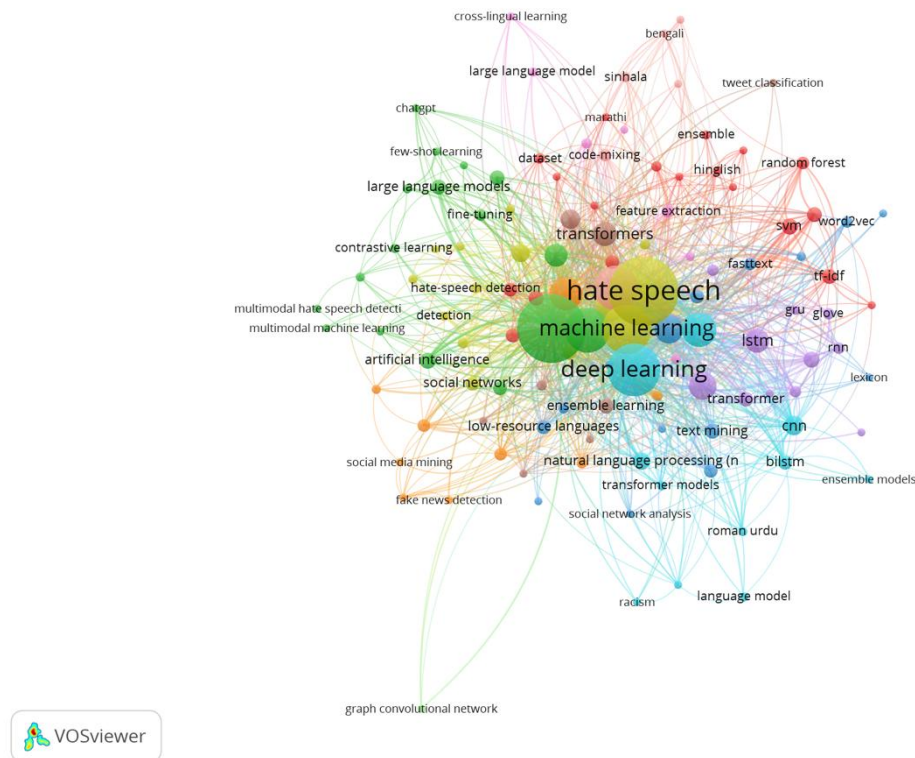
Gambar 6. Distribusi Penggunaan 10 Metode Teratas Tahun 2020-2025

Gambar 6 memperlihatkan tren penggunaan sepuluh metode utama dalam penelitian deteksi ujaran kebencian selama periode 2020–2025, di mana model berbasis *transformer* seperti BERT mendominasi sejak awal namun mulai menurun pada 2025 seiring munculnya model yang lebih mutakhir seperti RoBERTa dan pendekatan *ensemble*. Metode *deep learning* seperti CNN dan LSTM tetap konsisten digunakan meski dengan frekuensi lebih rendah, sementara metode klasik seperti Logistic Regression dan pendekatan berbasis leksikon semakin jarang dipakai hanya sebagai pembandingan. FastText dan GloVe muncul sebagai teknik representasi kata namun terbatas penggunaannya. Secara keseluruhan, tren ini menunjukkan pergeseran riset dari dominasi BERT menuju diversifikasi model *transformer*, multimodal, dan multilingual, dengan kecenderungan mengintegrasikan metode *ensemble* untuk meningkatkan akurasi dan transparansi sistem deteksi ujaran kebencian.

Mengacu pada distribusi penggunaan metode pada Gambar 6, terlihat bahwa penggunaan BERT pada tahun 2025 cenderung menurun dibandingkan tahun-tahun sebelumnya. Penurunan ini tidak menunjukkan berkurangnya relevansi BERT dalam penelitian deteksi ujaran kebencian. Seiring dengan kematangan bidang penelitian, BERT semakin diposisikan sebagai model *baseline*, sehingga banyak studi terbaru tidak lagi menyebutkannya secara eksplisit dalam judul atau kata kunci, melainkan menggunakan varian *transformer* yang lebih lanjut atau pendekatan berbasis integrasi model. Selain itu, meningkatnya perhatian terhadap pendekatan multimodal, multilingual, dan *Explainable AI* (XAI) turut mengalihkan fokus penelitian dari penggunaan satu model tertentu ke pengembangan arsitektur yang lebih kompleks. Faktor lain yang memengaruhi pola ini adalah keterbatasan data publikasi pada tahun 2025 yang masih bersifat parsial, sehingga jumlah publikasi yang tercatat relatif lebih rendah dibandingkan tahun-tahun sebelumnya.

3.5. Analisis Co-Occurance

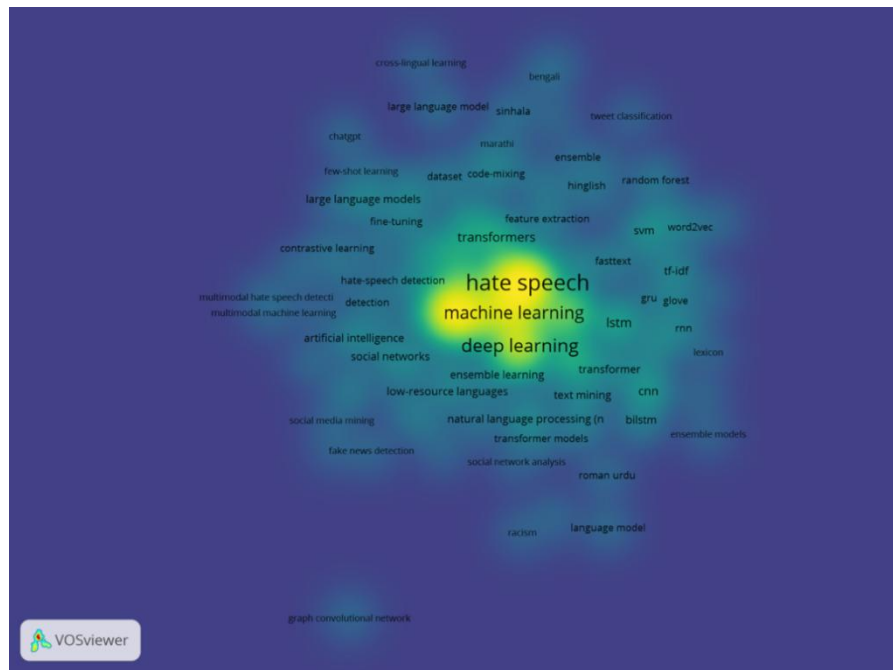
Analisis berikutnya merupakan visualisasi jaringan berdasarkan analisis *co-occurrence* yang menggambarkan keterkaitan antara berbagai topik penelitian dalam bidang *hate speech*. Gambar ini memberikan gambaran komprehensif tentang bagaimana topik-topik seperti *machine learning*, *deep learning*, *hate speech detection*, dan *social media mining* saling terkait satu sama lain, serta menunjukkan hubungan antara konsep-konsep utama yang sering muncul dalam penelitian mengenai ujaran kebencian.



Gambar 7. Analisis Jaringan Co-Occurance

Gambar 7 menampilkan hasil analisis jaringan *co-occurrence* yang memetakan keterkaitan antar kata kunci dalam penelitian deteksi ujaran kebencian. Visualisasi ini menunjukkan bahwa topik utama seperti *hate speech*, *machine learning*, *deep learning*, *social media mining*, dan *transformer models* saling terhubung erat, membentuk ekosistem penelitian yang kompleks. Selain itu, muncul pula kata kunci terkait seperti *cross-lingual learning*, *multimodal hate speech detection*, *ensemble learning*, dan *large language models*, yang menandakan arah perkembangan riset menuju pendekatan multibahasa, multimodal, serta integrasi model besar. Jaringan ini memperlihatkan bagaimana konsep-konsep teknis dan sosial berinteraksi, dengan topik dominan berada di pusat hubungan, sehingga memberikan gambaran komprehensif tentang fokus dan diversifikasi penelitian *hate speech detection* dalam beberapa tahun terakhir. Dalam analisis ini, kata kunci yang digunakan memiliki ambang frekuensi lebih dari lima kemunculan, sehingga menghasilkan 118 kata kunci yang relevan. Visualisasi jaringan menunjukkan bahwa topik *hate speech* berperan sebagai simpul utama yang memiliki keterhubungan erat dengan berbagai pendekatan metodologis seperti *machine learning*, *deep learning*, dan *transformer models*, serta domain aplikasi seperti *social media mining*. Klaster warna yang terbentuk mencerminkan pengelompokan topik berdasarkan pendekatan penelitian dan konteks penerapannya, yang menunjukkan bahwa metode komputasional dan domain aplikasi berkembang secara saling terintegrasi dalam riset deteksi ujaran kebencian.

Untuk melengkapi analisis keterkaitan antar topik tersebut, visualisasi *density* pada Gambar 8 digunakan untuk mengidentifikasi tingkat dominasi dan kepadatan kemunculan kata kunci dalam jaringan penelitian. Area dengan warna yang lebih terang menunjukkan topik-topik yang paling sering muncul dan memiliki hubungan paling kuat dengan kata kunci lainnya.



Gambar 8. Analisis Density

Terlihat pada Gambar 8, bahwa *hate speech*, *machine learning*, dan *deep learning* berada pada pusat kepadatan tertinggi, yang mengindikasikan bahwa topik-topik tersebut menjadi fokus utama dan landasan metodologis dalam penelitian deteksi ujaran kebencian selama periode analisis. Insight utama yang diperoleh dari analisis VOSviewer menunjukkan bahwa penelitian deteksi ujaran kebencian telah berkembang menjadi bidang yang relatif matang dan terintegrasi. Pengembangan metode komputasional tidak lagi berdiri sendiri, melainkan secara konsisten dikaitkan dengan konteks aplikasi dan dinamika media sosial. Keterhubungan antarkluster serta kepadatan topik inti mengindikasikan adanya konsolidasi pendekatan penelitian, dengan fokus yang semakin mengarah pada pemanfaatan metode pembelajaran mendalam dan model *transformer* sebagai fondasi utama dalam berbagai konteks implementasi deteksi ujaran kebencian.

4. KESIMPULAN

Penelitian ini menganalisis 1.335 publikasi tentang deteksi ujaran kebencian yang diambil dari Scopus dan menunjukkan bahwa sejak 2020, metode berbasis *transformer* seperti BERT dan RoBERTa mendominasi riset, menggantikan metode klasik seperti SVM dan Naive Bayes. Namun, tren penggunaan BERT menunjukkan kecenderungan menurun pada tahun 2025. Penurunan ini tidak serta-merta mencerminkan berkurangnya relevansi BERT, melainkan menunjukkan pergeseran fokus riset menuju pengembangan dan pemanfaatan varian *transformer* yang lebih lanjut, model multimodal dan multilingual, serta integrasi aspek *Explainable AI* (XAI). Selain itu, BERT semakin diposisikan sebagai model baseline, sehingga tidak selalu disebutkan secara eksplisit dalam judul atau kata kunci penelitian terbaru. Tren riset juga bergeser dari pendekatan teks tunggal menuju multimodal dan multilingual, dengan penekanan pada teknologi yang lebih adaptif terhadap konteks sosial. Bidang *social science* dan *technology* mendominasi penelitian ini, dengan fokus pada pemahaman dinamika sosial penyebaran ujaran kebencian serta pengembangan teknik deteksi otomatis. Meskipun penelitian ini terbatas pada data yang diambil dari Scopus dengan rentang waktu 2020–2025, hasil analisis memberikan gambaran komprehensif mengenai arah perkembangan dan tantangan riset deteksi ujaran kebencian di masa mendatang.

REFERENSI

- [1] M. S. Jahan and M. Oussalah, "A systematic review of hate speech automatic detection using natural language processing," *Neurocomputing*, vol. 546, Art. no. 126232, Aug. 2023, doi: 10.1016/j.neucom.2023.126232.

- [2] S. Mukherjee and S. Das, "Application of Transformer-based language models to detect hate speech in social media," *J. Comput. Cogn. Eng.*, vol. 2, no. 4, pp. 278–286, Dec. 2021, doi: 10.47852/bonviewJCCE2022010102.
- [3] P. Fortuna and S. Nunes, "A survey on automatic detection of hate speech in text," *ACM Comput Surv*, vol. 51, no. 4, pp. 1–30, Jul. 2018, doi: 10.1145/3232676.
- [4] J. S. Malik, G. Pang, and A. Van Den Hengel, "Deep learning for hate speech detection: A comparative study," *arXiv (Cornell University)*, Feb. 2022, doi: 10.48550/arxiv.2202.09517.
- [5] G. Ramos *et al.*, "A comprehensive review on automatic hate speech detection in the age of the transformer," *Soc. Netw. Anal. Min.*, vol. 14, no. 1, Art. no. 204, Oct. 2024, doi: 10.1007/s13278-024-01361-3.
- [6] G. K. Shahi, A. Dirkson, and T. A. Majchrzak, "An exploratory study of COVID-19 misinformation on Twitter," *Online Soc. Netw. Media*, vol. 22, Art. no. 100104, Mar. 2021, doi: 10.1016/j.osnem.2020.100104.
- [7] H. Saleh, A. Alhothali, and K. Moria, "Detection of hate speech using BERT and hate speech word embedding with deep model," *arXiv (Cornell University)*, Nov. 2021, doi: 10.48550/ARXIV.2111.01515.
- [8] V. Basile *et al.*, "SemEval-2019 Task 5: Multilingual detection of hate speech against immigrants and women in twitter," in *Proceedings of the 13th International Workshop on Semantic Evaluation*, Minneapolis, Minnesota, USA: Association for Computational Linguistics, 2019, pp. 54–63. doi: 10.18653/v1/S19-2007.
- [9] G. K. Shahi and T. A. Majchrzak, "Hate speech detection using cross-platform social media data in english and german language," *arXiv (Cornell University)*, 2024, doi: 10.48550/ARXIV.2410.05287.
- [10] H. Wang, T. R. Yang, U. Naseem, and R. K.-W. Lee, "MultiHateClip: A multilingual benchmark dataset for hateful video detection on YouTube and Bilibili," *arXiv (Cornell University)*, Jul. 2024, doi: 10.48550/ARXIV.2408.03468.
- [11] A. Chhabra and D. K. Vishwakarma, "A literature survey on multimodal and multilingual automatic hate speech identification," *Multimed. Syst.*, vol. 29, no. 3, pp. 1203–1230, Jun. 2023, doi: 10.1007/s00530-023-01051-8.
- [12] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *J. Bus. Res.*, vol. 133, pp. 285–296, Sep. 2021, doi: 10.1016/j.jbusres.2021.04.070.
- [13] S. Mariappanadar, "Do HRM systems impose restrictions on employee quality of life? Evidence from a sustainable HRM perspective," *J. Bus. Res.*, vol. 118, pp. 38–48, Sep. 2020, doi: 10.1016/j.jbusres.2020.06.039.
- [14] F. Poletto, V. Basile, M. Sanguinetti, C. Bosco, and V. Patti, "Resources and benchmark corpora for hate speech detection: a systematic review," *Lang. Resour. Eval.*, vol. 55, no. 2, pp. 477–523, Jun. 2021, doi: 10.1007/s10579-020-09502-8.
- [15] E. Chandrasekharan, S. Jhaver, A. Bruckman, and E. Gilbert, "Quarantined! Examining the effects of a community-wide moderation intervention on Reddit," *arXiv (Cornell University)*, Sep. 2020, doi: 10.48550/arXiv.2009.11483.
- [16] J. Baas, M. Schotten, A. Plume, G. Côté, and R. Karimi, "Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies," *Quant. Sci. Stud.*, vol. 1, no. 1, pp. 377–386, Feb. 2020, doi: 10.1162/qss_a_00019.
- [17] J. A. Moral-Muñoz, E. Herrera-Viedma, A. Santisteban-Espejo, and M. J. Cobo, "Software tools for conducting bibliometric analysis in science: An up-to-date review," *El Prof. Inf.*, vol. 29, no. 1, Art. no. e290103, Jan. 2020, doi: 10.3145/epi.2020.ene.03.
- [18] W. M. Lim, S. Kumar, and N. Donthu, "How to combine and clean bibliometric data and use bibliometric tools synergistically: Guidelines using metaverse research," *J. Bus. Res.*, vol. 182, Art. no. 114760, Sep. 2024, doi: 10.1016/j.jbusres.2024.114760.
- [19] J. D. Velasquez, "TechMiner: Analysis of bibliographic datasets using Python," *SoftwareX*, vol. 23, Art. no. 101457, Jul. 2023, doi: 10.1016/j.softx.2023.101457.
- [20] V. Pereira, M. P. Basilio, and C. H. T. Santos, "pyBibX -- A Python Library for Bibliometric and Scientometric Analysis Powered with Artificial Intelligence Tools," *Data Technol. Appl.*, vol. 59, no. 2, pp. 302–337, Apr. 2025, doi: 10.1108/DTA-08-2023-0461.