

Pemantauan Paparan Merek Terfokus Menggunakan Pemodelan Topik dan Formulasi Kueri

<http://dx.doi.org/10.28932/jutisi.v10i1.8395>

Riwayat Artikel

Received: 14 Februari 2024 | Final Revision: 17 Februari 2024 | Accepted: 18 Februari 2024

Creative Commons License 4.0 (CC BY – NC)



Iwan Santosa^{✉#1}, Bernard Renaldy Suteja^{#2}, Setia Budi^{#3}

[#] Program Magister Ilmu Komputer, Universitas Kristen Maranatha
Jl. Prof. drg. Surya Sumantri, M.P.H. No. 65, Bandung-40164, Indonesia

¹2279004@maranatha.ac.id

²bernard.rs@it.maranatha.edu

³setia.budi@it.maranatha.edu

[✉]Corresponding author: 2279004@maranatha.ac.id

Abstrak — *Brand exposure monitoring* adalah bentuk khusus dari *media monitoring* yang bertujuan memantau paparan merek dan *brand mention* pada media-media publisitas. Penelitian ini mengusulkan model kerangka kerja pemantauan *brand exposure* (paparan merek) melalui analisis pada media-media publisitas memanfaatkan formulasi kueri dan pemodelan topik, bertujuan mengekstraksi topik-topik utama yang terkandung dalam media eksternal atau media massa, dan membandingkannya dengan media internal institusi. Ekstraksi topik dilakukan menggunakan metode *Latent Dirichlet Allocation* (LDA) pada data teks berupa SERP (*Search Engine Results Page*) *snippets*. Pemrosesan berikutnya memanfaatkan penghitungan *Jaccard index* dan *cosine similarity*. Output dari kerangka kerja berupa visualisasi topik yang merepresentasikan tema dari artikel-artikel yang memaparkan *brand* institusi, dan metrik pengukuran yang berguna untuk analisis pengambilan keputusan terkait pengelolaan media dan komunikasi institusi. Pengujian rancangan kerangka kerja menggunakan sampel data menghasilkan output sesuai ekspektasi, yaitu kelompok topik yang terbentuk dapat mewakili topik-topik dominan dalam kedua kategori media pada periode pemantauan, dengan tingkat *coverage* tertinggi mencapai 82,69% dan similaritas 33,97%. Penggunaan metode LDA dalam penelitian ini memiliki keterbatasan, yaitu terbentuknya kelompok topik yang tidak murni berisi topik tunggal, melainkan terdiri dari beberapa subtopik. Namun demikian, hal ini tidak mengurangi kebermanfaatan kerangka kerja.

Kata kunci—formulasi kueri; *Latent Dirichlet Allocation* (LDA); *media monitoring*; paparan merek; pemodelan topik.

Focused Brand Exposure Monitoring Using Topic Modeling and Query Formulation

Abstract — *Brand exposure monitoring* is a specialized form of *media monitoring* that aims to monitor brand exposure and brand mentions in publicity media. This research proposes a framework model for monitoring brand exposure through analysis of publicity media utilizing query formulation and topic modeling, aimed at extracting important topics contained in external media or mass media, and comparing them with the institution's internal media. Topic extraction is performed using *Latent Dirichlet Allocation* (LDA) method on text data in the form of SERP (*Search Engine Results Page*) *snippets*. Subsequent processing utilizes

Jaccard index and cosine similarity calculations. The output of the framework includes visualizations of topics representing themes of articles exposing the institution's brand, and measurement metrics that are useful for decision-making analysis related to media management and institutional communication. Testing the proposed framework using data samples resulted in expected output, namely the topic groups formed can represent dominant topics in both media categories in a given monitoring period, with the highest coverage rate reaching 82.69% and similarity of 33.97%. The use of LDA method in this study does have its limitations, specifically the formation of topic groups that do not purely contain a singular topic, but rather consist of several subtopics. However, this does not diminish the usefulness of the framework.

Keywords— brand exposure; Latent Dirichlet Allocation (LDA); media monitoring; query formulation; topic modeling.

I. PENDAHULUAN

Lembaga pendidikan khususnya perguruan tinggi atau universitas swasta sebagai institusi penyelenggara pendidikan tinggi adalah institusi korporasi yang bergerak dalam bidang jasa. Sebagai entitas perusahaan, universitas harus melakukan pengelolaan *brand* untuk menjaga dan mengembangkan bisnisnya. Pengelolaan *brand* atau merek adalah bagian dari bauran komunikasi pemasaran, juga bagian dari pengelolaan reputasi institusi. Pengelolaan *brand communication* di universitas studi kasus, yaitu Universitas Kristen (UK) Maranatha, dilakukan terintegrasi sebagai bagian dari strategi komunikasi korporat terpadu. Salah satu faktor penting dalam strategi tersebut adalah pengelolaan *brand exposure* atau paparan merek.

Brand exposure pada universitas studi kasus adalah hasil dari proses pengelolaan *integrated corporate and marketing communication* melalui media-media *multichannel*, salah satunya adalah memanfaatkan *publicity channel* yang terdiri dari *corporate-owned media* (situs Maranatha News) dan *external media* (media massa online dan situs lainnya). *Brand name* universitas dipaparkan melalui artikel-artikel yang dipublikasikan pada dua kategori *publicity channel* tersebut, selain pada *channel-channel non-publicity* lainnya.

Salah satu bentuk *brand exposure monitoring* yang dilakukan di universitas studi kasus adalah dengan cara memonitor isu-isu berkaitan dengan *brand name* Universitas yang dipaparkan melalui kedua kategori *publicity channel*. Isu-isu yang dipaparkan kepada publik berpengaruh pada *brand recognition* dan dapat menimbulkan asosiasi antara *brand name* Universitas dengan topik-topik tertentu (bisa berupa aktivitas, kejadian, tokoh, prestasi, dll.) yang bisa bersifat temporer ataupun jangka panjang. Asosiasi temporer cenderung bersifat tren jangka pendek, sedangkan asosiasi jangka panjang dapat bersifat permanen dan berdampak pada citra dan reputasi Universitas.

Secara periodik tim *monitoring* mencari dan melakukan pendataan terhadap artikel yang diterbitkan pada dua kategori media. Proses *monitoring* melibatkan *web searching (branded search)* dan membaca hasil pencarian untuk mendapatkan wawasan mengenai topik-topik informasi yang dipaparkan kepada publik. Hasil *monitoring* ini menjadi referensi bagi tim pengelola media & komunikasi institusi untuk melakukan evaluasi dan membantu pengambilan keputusan strategis terkait pengelolaan media & komunikasi di tingkat yang lebih tinggi. Proses pemantauan manual tersebut memiliki beberapa kelemahan, yaitu faktor subjektivitas yang tinggi karena mengandalkan kemampuan, kecakapan, dan kecermatan staf pemonitor. Proses konvensional tersebut juga tidak dapat menampilkan metrik yang konsisten, karena kesimpulan bergantung analisis masing-masing staf pemonitor. Berdasarkan beberapa kelemahan tersebut, rumusan masalah (*problem questions*) dalam penelitian ini adalah 1) Bagaimana *brand exposure monitoring* yang dilakukan melalui mesin pencari dapat didukung dan ditingkatkan agar dapat menghasilkan informasi yang mudah dicerna oleh *stakeholder* pengambil keputusan? 2) Bagaimana cara mendapatkan informasi topik *exposure* terhadap *brand* dengan cara yang lebih mudah dan objektif (*data driven*)? 3) Apakah topik pada *corporate website* dapat menjadi referensi informasi yang berguna dalam proses *brand exposure monitoring* bila dihubungkan dengan *website external media*? 4) Indikator apa yang dapat digunakan untuk menghubungkan *brand exposure* pada *corporate website* dengan *external media*?

Penelitian ini bertujuan 1) merancang model kerangka kerja *brand exposure monitoring* untuk mendukung dan meningkatkan kinerja tim pemonitor di institusi studi kasus, yang dapat menghasilkan *report* yang mudah dipahami pengambil keputusan; 2) mengolah informasi yang dihasilkan oleh *search engine* menggunakan metode-metode *machine learning* dan *natural language processing (NLP)* yang sesuai dengan kebutuhan kerangka kerja; 3) mengekstraksi informasi hasil pencarian *corporate website* dan *external media* untuk menemukan topik yang terkandung di dalamnya; dan 4) menentukan indikator dan metrik yang dibutuhkan sebagai salah satu bentuk output dari kerangka kerja.

Ruang lingkup *brand exposure monitoring* pada penelitian ini mencakup *exposure* pada media internal, media massa dan media online lainnya yang mengandung *brand mention* Universitas. Batasan dan asumsi pada penelitian ini adalah 1) *Search engine* yang digunakan adalah Google Search Engine dengan pertimbangan a) merupakan *search engine* terpopuler yang diasumsikan juga digunakan oleh publik untuk mendapatkan informasi mengenai Universitas; b) merupakan *search engine* yang saat ini digunakan oleh tim *monitoring*. 2) *Scraping* dilakukan pada seluruh halaman *organic SERP* (tidak termasuk *sponsored SERP, location-based results, personalized results*, dan tidak hanya pada *top SERP*), sehingga data yang didapatkan tidak terpengaruh oleh posisi *ranking SERP*. 3) Pencarian dibatasi hanya di wilayah Indonesia dan dalam bahasa

Indonesia. 4) Metode pemodelan topik yang dieksplorasi adalah *Latent Dirichlet Allocation* (LDA). Keseluruhan proses pengujian dalam penelitian ini dijalankan menggunakan bahasa pemrograman Python (3.10.12) dan *library-library* terkait.

Penelitian-penelitian sebelumnya memanfaatkan metode pemodelan topik untuk ekstraksi informasi dari sumber-sumber data bervariasi, yaitu artikel berita, *web search query*, media sosial, kuesioner *open-ended*, ulasan pengguna, dan arsip dokumen abstrak (Tabel 1). Kebaruan penelitian ini terletak pada 1) tujuan pemodelan topik dalam konteks kerangka kerja *brand exposure monitoring*; 2) jenis *dataset* berupa *SERP snippet*; 3) implementasi LDA tidak hanya untuk interpretasi topik, tetapi juga untuk mengukur metrik *brand exposure*; dan 4) langkah dan parameter *preprocessing* dan *pipeline* keseluruhan yang *domain-specific*. Penelitian ini berkontribusi memberikan wawasan baru dan mengisi *research gap* dalam hal 1) implementasi *LDA topic modeling* untuk mengekstraksi informasi dari *dataset* berupa *SERP snippet*, dan 2) metode *text preprocessing* dan validasi yang perlu dilakukan pada studi kasus dalam bidang terapan *brand exposure monitoring*.

TABEL 1
RINGKASAN PENELITIAN SEBELUMNYA MENGENAI PEMODELAN TOPIK MENGGUNAKAN METODE LDA

Peneliti	Tahun	Tujuan	Variabel		Metode		Outcome/temuan
			Input	Output	Preprocess- ing spesifik	Processing & evaluasi model	
Liu <i>et al.</i> [1]	2020	Analisis berita Covid-19	Berita <i>online</i> dari <i>database</i> WiseSearch	Interpretasi 20 topik, 9 kelompok tema; <i>time series reports</i>	Vektorisasi TF-IDF	LDA; Gibbs <i>sampling</i> ; <i>coherence</i> ; pyLDAvis	Indikasi peran & fokus pemberitaan pada masa awal pandemi di China
Abebe <i>et al.</i> [2]	2019	Ekstraksi informasi data kesehatan	Teks <i>search engine queries</i> pengguna Bing di Afrika	100 topik HIV/AIDS, 50 malaria, 50 TBC; penge-lompokan lanjutan	EDA menggunakan Spearman <i>correlation</i>	LDA (Mallet); <i>coherence</i> ; interpretasi pakar	Pola informasi kesehatan yang dibutuhkan oleh kelompok usia, gender, dan negara
Hidayatullah <i>et al.</i> [3]	2019	Analisis teks kondisi cuaca dan iklim	Teks media sosial Twitter BMKG	Interpretasi 5 topik	Model <i>bigram</i>	LDA; pyLDAvis	Tren cuaca dan bencana alam di Pulau Jawa
Karnalim <i>et al.</i> [4]	2023	Ekstraksi informasi mitigasi plagiarisme	Isian kuesioner survei <i>open-ended</i> (6 model pada 6 pertanyaan)	2 & 3 topik (2 pertanyaan); 5 & 6 topik (4 pertanyaan)	Model <i>bigram</i>	LDA; <i>coherence</i> ; <i>topic relation graph</i>	<i>Topic centrality</i> konsep jawaban atas pertanyaan <i>open-ended</i>
Wahyudin [5]	2019	Analisis pemberitaan PSBB	Berita dari 9 portal berita <i>online</i> Indonesia	20 topik, 4 kelompok	EDA menggunakan <i>time series graphs</i>	LDA; <i>topic coherence</i> ; <i>prevalence</i> ; <i>dendrogram</i>	Jenis pemberitaan selama PSBB April 2020
Santoso <i>et al.</i> [6]	2022	Analisis topik tagar <i>covid-indonesia</i>	<i>Caption</i> media sosial Instagram	Interpretasi 5 topik	<i>Manual stopwords</i>	LDA; <i>perplexity</i> ; <i>topic coherence</i>	Topik pembicaraan dengan <i>hashtag covidindonesia</i>
Sagala & Toba [7]	2021	Analisis aspek-aspek produk <i>smartphone</i>	Ulasan pengguna <i>e-commerce</i> Amazon	3 aspek terpenting berdasarkan 10 skenario	<i>Bigram</i> ; <i>lemmatization</i>	LDA; <i>sentiment analysis</i> (VADER)	Aspek terpenting dari 4 <i>brand smartphone</i>
Kannitha <i>et al.</i> [8]	2022	Analisis keluhan pelanggan	Teks media sosial Twitter (2 kata kunci nama ISP)	10 & 11 topik; probabilitas tertinggi	<i>Stemming</i> ; <i>manual stopwords</i>	LDA; Gibbs <i>sampling</i> ; <i>log-likelihood</i>	Topik keluhan pengguna 2 provider internet (ISP)
Kharisudin & Masri'an [9]	2022	Analisis produk aplikasi <i>mobile</i>	Ulasan pengguna Whats-App dari Google Play	31 topik dengan 4 topik teratas	<i>Lemmatization</i>	LDA; <i>perplexity</i> ; <i>cross-validation</i>	Topik yang paling sering dibahas oleh pengguna aplikasi
Novarian <i>et al.</i> [10]	2023	Analisis topik tugas akhir mahasiswa	Teks abstrak dari <i>website</i> repositori	3 <i>top terms</i> pada 8 topik	Vektorisasi TF-IDF	LDA; <i>coherence</i> ; pyLDAvis	Tren topik penelitian tugas akhir mahasiswa

II. METODE PENELITIAN

A. Hipotesis

Hipotesis yang diajukan pada penelitian ini adalah: 1) Proses *monitoring* dapat ditingkatkan menggunakan *query* berupa *branded search* dengan parameter kondisional dan/atau *query terms* spesifik untuk menghasilkan SERP yang relevan; 2) SERP *snippet* mengandung data yang cukup representatif untuk proses ekstraksi informasi menggunakan metode pemodelan topik; 3) Teknik ekstraksi menggunakan *topic modeling* dapat diaplikasikan pada sumber berita media eksternal dan *corporate website* untuk mendeteksi topik-topik yang diekspos pada periode *monitoring*; 4) Topik-topik yang dihasilkan dari proses ekstraksi informasi pada kedua kategori media tersebut dapat dibandingkan dan dilakukan penghitungan untuk analisis pengambilan keputusan.

B. Konsep dan Teori

Brand atau merek adalah aset *intangible* paling berharga dari sebuah institusi [11]. Salah satu komponen *brand* adalah *brand name*. Bagi institusi perguruan tinggi, nama institusi merupakan *brand name* yang menjadi pembeda produk atau jasa antara perguruan tinggi satu dengan yang lainnya. *Brand equity* atau ekuitas merek adalah nilai yang dimiliki oleh sebuah *brand*, yang dapat memberikan nilai tambah bagi produk atau jasa yang ditawarkan oleh sebuah *brand* [11][12]. *Brand equity* juga dapat mengurangi nilai dari sebuah *brand*, contohnya pada kasus *brand image* atau citra institusi yang buruk. Salah satu komponen *brand asset* pembentuk *brand equity* adalah *name awareness* [12]. Shimp & Andrews [13] menyebutkan bahwa *brand equity* dari perspektif konsumen dapat dipandang dari seberapa familier sebuah *brand* dikenali dan diminati oleh konsumen. Dengan demikian, *brand equity* berkaitan dengan *brand knowledge* yang dibentuk dari dua unsur yaitu *brand awareness* dan *brand image*. *Brand knowledge* juga berkaitan dengan dimensi *brand image* yang merepresentasikan *brand association* [14].

Brand equity dapat dibangun dan ditingkatkan melalui aktivitas komunikasi, yaitu mengirimkan atau memaparkan pesan positif mengenai *brand* kepada khalayak. *Marketing communication* berperan penting dalam penciptaan *brand equity*, tetapi tidak terbatas memanfaatkan bentuk-bentuk komunikasi pemasaran konvensional saja, misalnya melalui iklan [13]. *Integrated marketing communication* (IMC) adalah proses mengelola *marketing communication* secara terpadu untuk mengkoordinasikan semua aspek komunikasi yang melibatkan berbagai jenis sarana, pesan, media, dan khalayak [11]. Salah satu tujuan strategis dalam perencanaan IMC adalah meningkatkan *brand exposure* untuk membentuk *brand awareness*. Hubungan antara *brand exposure*, *brand awareness*, dan *brand equity* pada beberapa sektor bisnis dikonfirmasi oleh beberapa peneliti, antara lain oleh Cornwell *et al.* [15], Pitts & Slattery [16], Arthana *et al.* [17], Pomalaa *et al.* [18], Khoa [19], Pinar [20], Perwito *et al.* [21].

Beberapa literatur (Gatfield *et al.* [22]; Gray *et al.* [23]; Pinar *et al.* [24]; Sujchaphong [25]; Noor *et al.* [26]; de Heer [27]; Khoa [19]; Pinar [20]) menyebutkan bahwa pengelolaan *brand* pada institusi perguruan tinggi tidak sama dengan pengelolaan *brand* pada institusi komersial pada umumnya. Pendapat dan temuan para peneliti tersebut serupa dengan pengalaman empiris di universitas studi kasus. Faktor yang paling mendasar adalah perbedaan proses bisnis dari kedua jenis institusi. Proses bisnis perguruan tinggi khususnya di Indonesia berpusat pada aktivitas tridharma, yaitu terdiri dari pengajaran, penelitian, dan pengabdian masyarakat. Sedangkan proses bisnis perusahaan komersial pada umumnya berpusat pada penjualan barang atau jasa. Dengan demikian, praktik-praktik komunikasi *branding* yang umumnya dilakukan di perusahaan komersial tidak sepenuhnya dapat langsung diimplementasikan di institusi perguruan tinggi. Aspek penting dari institusi perguruan tinggi yang melekat erat pada identitasnya adalah aspek citra dan reputasi. Citra dan reputasi perguruan tinggi memegang peranan sangat penting bagi keberadaannya. Oleh karena itu, citra dan reputasi adalah elemen *brand asset* pembentuk *brand equity* yang sangat penting bagi institusi perguruan tinggi. Hal ini menjadi kajian pada literatur dan penelitian yang dilakukan oleh Binnie [28], Chapleo [29], Pinar [20], dan Gutiérrez [30].

Ada banyak faktor yang mempengaruhi citra dan reputasi perguruan tinggi, antara lain adalah akreditasi, sejarah, rekam jejak institusi, kualitas dosen, kelengkapan sarana penunjang pendidikan, dan semua aktivitas akademik serta tridharma yang dilakukan. Upaya pembentukan citra dan reputasi perguruan tinggi dilakukan sebagai bagian dari strategi *corporate branding* dalam kerangka yang lebih besar, yaitu mencakup *integrated corporate & marketing communication*. Dalam konteks yang komprehensif ini, *branding* tidak hanya masuk dalam dimensi *marketing communication*, tetapi juga merupakan dimensi *corporate communication* yang cakupannya lebih luas. *Corporate identity* sebagai *brand asset* memiliki perbedaan dengan *brand identity* dalam konteks *brand marketing* tradisional, karena *corporate identity* berkaitan dengan semua *stakeholder* institusi, tidak hanya terbatas pada khalayak konsumen saja [31]. *Corporate identity* dipaparkan kepada khalayak atau *stakeholder* melalui saluran-saluran komunikasi yang multifaset, membentuk *corporate image* dan *corporate reputation* [31]. *Corporate image* dan *corporate reputation* sebagai bagian dari dimensi *corporate identity* secara strategis menjadi *competitive advantages* bagi institusi.

Media monitoring adalah aktivitas pemantauan media yang bertujuan untuk mencari informasi spesifik mengenai suatu hal atau tren yang ingin diamati. *Media monitoring* dilakukan oleh institusi untuk memantau isu yang beredar, yang

berhubungan dengan citra dan reputasi institusi. *Media monitoring* dalam konteks yang lebih luas digunakan untuk memantau isu yang lebih besar. Contohnya pada kasus pemantauan berita mengenai Covid-19 yang bertujuan untuk mendeteksi *hoax* yang beredar di masyarakat [32], dan pemantauan berita mengenai Covid-19 pada masa awal pandemi merebak di Tiongkok, untuk mengetahui efektivitas media massa dalam menyediakan informasi terbaru mengenai Covid-19 kepada masyarakat [1].

Media monitoring paling sederhana dan tradisional dilakukan dengan cara melakukan klipring berita (Jefkins [33]; Friedel & Lukman [34]; Habiba [35]; Astiti, *et al.* [36]). Hasil klipring kemudian dianalisis untuk menemukan informasi spesifik sesuai kebutuhan. Analisis pada umumnya bertujuan menemukan sentimen atau *tonality* pemberitaan positif, negatif, atau netral. Dalam konteks komunikasi pemasaran, *media monitoring* dilakukan tidak hanya untuk mendeteksi sentimen atau tren, tetapi juga dapat digunakan untuk mendeteksi *brand mention* yang beredar di media massa. Hal ini merupakan salah satu aspek yang dipantau oleh tim *monitoring* di universitas studi kasus, sebagai bagian dari unsur *brand exposure* melalui *publicity channel*. Aktivitas *media monitoring* dengan cara klipring berita pada dasarnya adalah aktivitas *information retrieval* yang dimulai dari proses penelusuran informasi. Saat ini penelusuran informasi berita dan informasi apa pun sebagian besar dilakukan secara *online* pada World Wide Web (*web search*) memanfaatkan *search engine* sebagai sarana *information retrieval* [37]. Dengan banyaknya informasi yang tersedia di web, atau dengan banyaknya media *online* sebagai sumber penyedia informasi, proses *information retrieval* perlu dilakukan dengan metode yang tepat, agar informasi yang dicari dapat ditemukan secara efektif dan akurat.

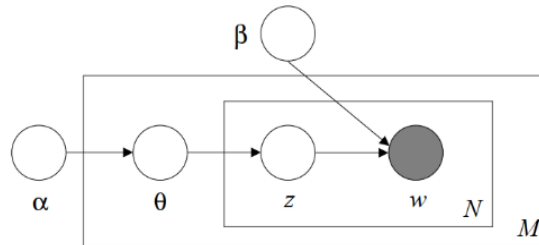
Web search engine pada umumnya terdiri dari beberapa komponen dan tahapan pemrosesan, yaitu 1) *Query processing*; 2) *Document retrieval*; 3) *Result ranking*; 4) *Snippet generation*; 5) *Result aggregation and displaying* [38]. Proses pencarian dimulai dari pengguna memasukkan *query terms* atau kata kunci pencarian, dan diakhiri dengan *search engine* menampilkan hasil pencarian (SERP atau *Search Engine Results Page*). SERP berisi daftar hasil pencarian yang paling relevan beserta deskripsi singkat (*snippet*) mengenai isi dokumen atau halaman web yang dicari. Hasil pencarian pada SERP disusun berdasarkan peringkat dan relevansi. Metode dan algoritma yang digunakan oleh *search engine* berperan dalam faktor akurasi *search engine* tersebut untuk memberikan hasil pencarian yang relevan. Selain itu, *search engine* dapat menghasilkan informasi yang relevan dengan *query* pencarian, apabila *query terms* yang dimasukkan sebagai input pada *search engine* tepat. Diperlukan strategi pencarian dan formulasi *query* yang tepat agar *search engine* dapat melakukan penelusuran dengan efektif dan menemukan informasi yang relevan dengan yang diharapkan. *Boolean search* atau *boolean retrieval model* adalah salah satu metode pencarian informasi menggunakan *search engine* yang efisien dan dapat memberikan hasil akurat [37]. Akurasi dan relevansi hasil pencarian menggunakan metode *boolean search* dapat ditingkatkan dengan menyusun *search query* yang kompleks menggunakan *boolean operators*. Beberapa peneliti (Aula [39]; Russel-Rose & Shokraneh [40]; Hambarde & Proença [41]; Sachin & Patel [42]) melakukan kajian dan mengembangkan metode-metode yang lebih *advanced* untuk membantu proses *query formulation* agar *search engine* dapat melakukan pencarian dengan lebih efektif dan akurat.

Dalam kasus *media monitoring*, dokumen pilihan didapatkan dari proses klipring, atau bila *monitoring* dilakukan menggunakan *web search*, maka hasil “klipring” adalah berupa SERP. Langkah selanjutnya adalah memilah hasil klipring atau SERP untuk mengambil informasi spesifik yang sedang dipantau. Langkah ini melibatkan metode ekstraksi informasi (*information extraction*), baik secara tradisional maupun memanfaatkan sarana yang lebih canggih seperti NLP yang melibatkan *machine learning* dan *artificial intelligence* (AI). Beberapa metode *machine learning* yang populer digunakan untuk ekstraksi dan identifikasi informasi antara lain adalah *sentiment analysis*, *topic modeling*, *named entity recognition*, *text summarization* [43].

Pemodelan topik *Latent Dirichlet Allocation* (LDA) adalah metode analisis teks secara statistik untuk menemukan tema dan hubungan antar-tema dalam kumpulan dokumen teks tidak terstruktur [44]. LDA menemukan topik dari sekumpulan dokumen dengan cara melakukan komputasi pada dokumen sebagai objek observasi, dan melakukan inferensi struktur topik yang tersembunyi (*latent variables*) dari pola kata yang muncul dalam dokumen. *LDA topic modeling* mengasumsikan sebuah topik terdiri dari sekumpulan kata-kata (*vocabulary of terms*); dan sebuah dokumen terdiri dari beberapa topik yang terdistribusi secara probabilistik (distribusi Dirichlet). Konsep dasar *LDA topic modeling* adalah *nested concepts* dari *corpus*, dokumen, dan *terms* [45]. *Corpus* atau *corpora* adalah koleksi teks yang akan dimodelkan; *corpus* terdiri dari dokumen; dokumen terdiri dari kata-kata (*terms*). Dokumen direpresentasikan sebagai gabungan random dari topik laten, dengan masing-masing topik merupakan distribusi dari kata-kata [46]. LDA merupakan *generative probabilistic modeling* yang bekerja berdasarkan *joint probability distributions*, untuk menghitung *conditional distribution* (*posterior distribution*) dari *hidden variables* berdasarkan *observed variables* [44]. *Joint distribution* dari sebuah distribusi topik (θ), sebuah himpunan N topik (z), dan sebuah himpunan N kata-kata (w) didefinisikan dengan Persamaan 1. Parameter α adalah Dirichlet *prior parameter* yang mempengaruhi distribusi topik dalam setiap dokumen, sedangkan β adalah Dirichlet *prior parameter* yang mempengaruhi distribusi kata dalam setiap topik.

$$p(\theta, z, w|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^N p(z_n|\theta) p(w_n|z_n, \beta) \quad (1)$$

Proses generatif LDA direpresentasikan dengan model grafis berupa diagram *plate notation* (Gambar 1), dengan langkah: 1) untuk setiap dokumen, sebuah distribusi topik (θ) dicuplik dari distribusi Dirichlet (α); 2) Kemudian, untuk setiap kata dalam dokumen, sebuah topik (z) ditugaskan dari distribusi topik dalam dokumen, dan sebuah kata (w) ditugaskan dari distribusi kata dalam topik (β) berdasarkan topik yang ditugaskan (z). *Plate N* merepresentasikan kumpulan kata dalam sebuah dokumen; *plate M* merepresentasikan kumpulan dokumen pada *corpus*.



Gambar 1. Diagram plate notation model LDA [46]

Algoritma LDA bekerja dengan cara melakukan inisiasi setiap kata w secara random ke seluruh topik berjumlah K ; dan untuk setiap kata w dalam dokumen d akan dihitung 1) $p(\text{topic } k \mid \text{document } d)$ yaitu proporsi topik k yang ditugaskan ke dalam dokumen d (*topics probabilities to documents*), dan 2) $p(\text{word } w \mid \text{topic } k)$ yaitu proporsi kata yang ditugaskan ke topik k (*term probabilities to topics*) dalam dokumen d . Kemudian algoritma akan memaksimalkan *joint likelihood* dengan iterasi adaptasi nilai *word-topic distribution matrix* β dan *document-topic distribution matrix* θ . Model LDA menghasilkan output berupa *word-topic matrix* yaitu *vocabulary* kata yang terkoneksi dengan masing-masing topik, dan *document-topic matrix* yaitu daftar kumpulan topik yang terkoneksi dengan masing-masing dokumen. Evaluasi terhadap model LDA dilakukan untuk memastikan bahwa model dapat menghasilkan output berupa K topik dan *vocabulary* kata dengan baik. Sebuah model LDA yang baik dapat membentuk daftar kata optimal yang dapat merepresentasikan topik dengan jelas dan dapat diinterpretasikan oleh manusia dengan relatif mudah [47]. Salah satu metode evaluasi model LDA adalah menggunakan *coherence score*, yaitu mengukur tingkat koherensi topik dengan mengecek seberapa sering *top token* atau *top term* (daftar kata teratas) muncul dalam dokumen pada setiap topik. Iterasi dilakukan untuk mendapatkan nilai *coherence* terbaik.

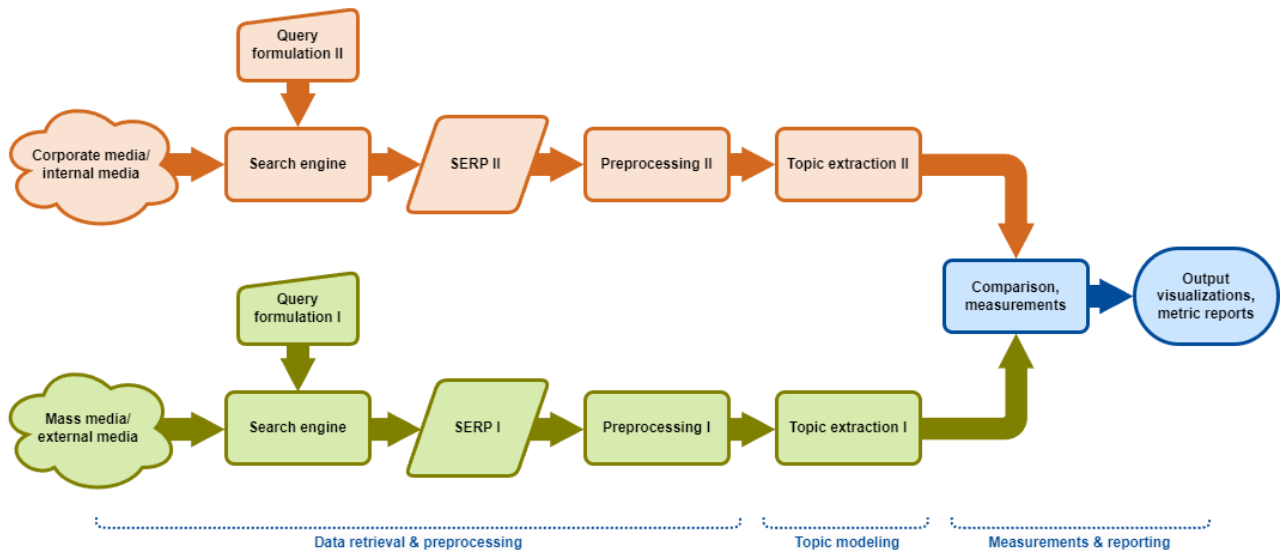
Similaritas antara dua *dataset* dapat dihitung menggunakan metode *Jaccard score* atau *Jaccard index* [48] atau *Tanimoto index* [49] dan *cosine similarity* [37][50]. *Jaccard score* (J) antara himpunan A dan B dihitung berdasarkan irisan anggota kedua himpunan (Persamaan 2). Sedangkan *cosine similarity* ($\cos \theta$) antara vektor A (terdiri dari *terms* $a_{1,\dots,n}$) dan B (terdiri dari *terms* $b_{1,\dots,n}$) dihitung berdasarkan sudut vektor yang terbentuk (Persamaan 3). Implementasi *cosine similarity* pada penelitian ini menggunakan *library* Scikit-learn [51].

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

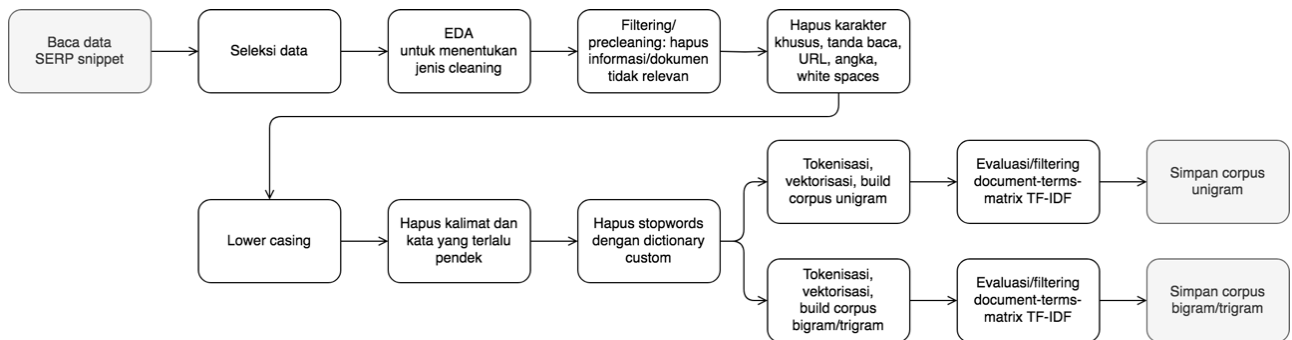
$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n a_i b_i}{\sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}} \quad (3)$$

C. Desain Penelitian

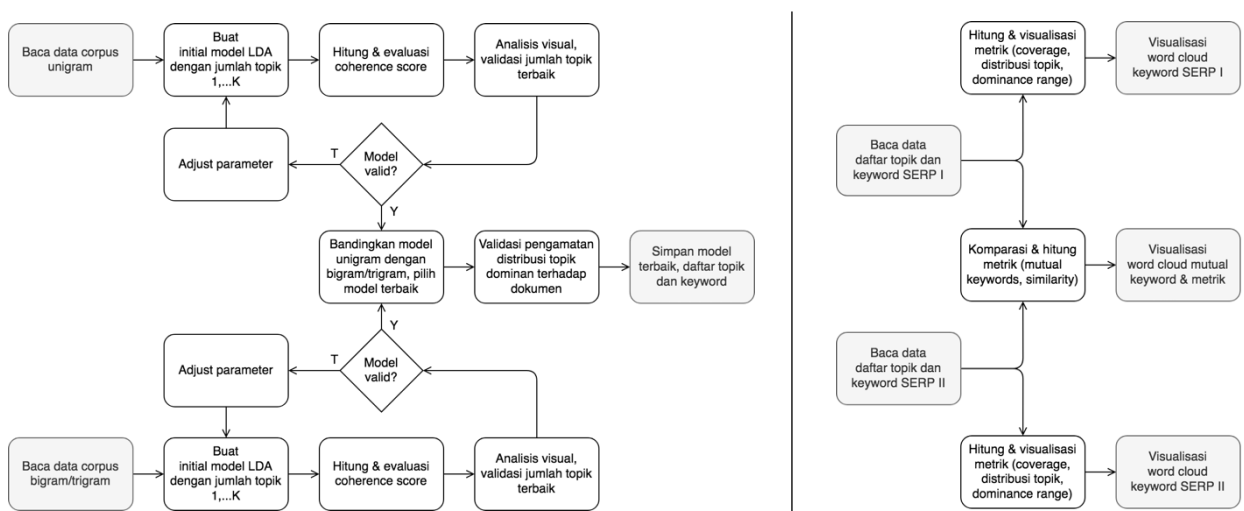
Penelitian ini berpusat pada perancangan dan pengujian kerangka kerja “*Focused Brand Exposure Monitoring*” (FBxM) yang digambarkan pada diagram blok Gambar 2. Terdapat tiga kelompok proses utama dalam kerangka kerja, yaitu 1) *data retrieval & preprocessing*; 2) *topic modeling*; 3) *measurements & reporting*. Proses *data retrieval & preprocessing* (diagram Gambar 3) dan *topic modeling* (diagram Gambar 4 kiri) dilakukan dua kali, yaitu pada media massa (*external media*) dan media korporat Universitas. *Data collection* dilakukan menggunakan Google Search Engine, kemudian pada *snippet* hasil pencarian (SERP I dan SERP II) dilakukan *scraping* menggunakan *browser extension* SERP Snippet Extractor v2.2 [52], menghasilkan data *tabular*. *Browser* yang digunakan adalah LibreWolf 107.0.1-1. Proses berikutnya adalah *preprocessing* untuk membersihkan data dari *noise*. Dalam penelitian ini akan dilakukan perbandingan performa model topik antara *corpus unigram* dan *bigram/trigram*, maka data bersih setelah tahap *preprocessing* tersebut selanjutnya dilakukan tokenisasi, vektorisasi, dan pembentukan *corpus unigram* dan *bigram/trigram*. Evaluasi terhadap *document-term matrix* TF-IDF (*term frequency - inverse document frequency*) dilakukan untuk melihat distribusi kata berdasarkan bobot TF-IDF [38], untuk memperoleh wawasan indikasi tingkat kepentingan atau signifikansi kata-kata pada seluruh dokumen.



Gambar 2. Diagram blok rancangan kerangka kerja FBxM



Gambar 3. Tahapan langkah data retrieval & preprocessing



Gambar 4. Tahapan langkah proses pemodelan topik LDA (kiri), measurements & reporting (kanan)

Proses selanjutnya adalah pemodelan topik LDA (menggunakan *library* Gensim [53]), dan visualisasi untuk *validasi* memanfaatkan *library* pyLDAvis [54] sebagai alat bantu untuk pengamatan *intertopic distance map*. Proses validasi selanjutnya dilakukan dengan membuat tabel validasi untuk melakukan analisis perbandingan topik dominan terhadap dokumen aslinya (yaitu SERP *snippet* yang sudah melewati tahap pembersihan). Tabel validasi berguna untuk melihat distribusi topik dominan terhadap dokumen; tingkat *coverage* pada keseluruhan topik; *mutual terms* pada masing-masing topik; distribusi Jaccard *score* pada setiap pasangan topik dan dokumen.

Tahap berikutnya adalah *measurements & reporting* (diagram Gambar 4 kanan) berupa proses komparasi dan output visualisasi. Masing-masing topik yang dihasilkan oleh model topik SERP I dan SERP II divisualkan menggunakan *word cloud* untuk mempermudah *stakeholder* mengamati *keywords* yang terbentuk. Besarnya ukuran kata merepresentasikan bobot *keyword*. Output visualisasi berikutnya pada masing-masing model topik menampilkan *metrics report* berisi a) distribusi topik; b) persentase *topic coverage* dan kesesuaian maksimal; c) rentang *topic dominance* minimal, rata-rata, dan maksimal. Berikutnya, komparasi antara model topik SERP I dan SERP II dilakukan dengan menghitung *mutual top keywords count* menggunakan Jaccard *score* pada hasil topik SERP I dan SERP II, setelah gabungan kata *bigram/trigram* dikembalikan ke bentuk semula yaitu *unigram*. *Overall topics similarity* dihitung menggunakan *cosine similarity* antara gabungan semua topik yang terbentuk pada SERP I dan SERP II.

D. Pengujian Data Retrieval & Preprocessing SERP I

Tahapan langkah *data retrieval* dan *preprocessing* dilakukan berdasarkan diagram proses pada Gambar 3. Sampel data SERP I diambil dari *external website* (media massa dan situs lainnya) dengan cara melakukan *branded search* (pencarian dengan menggunakan *brand name* sebagai kata kunci utama [55][56]). Parameter *query* pencarian terdiri dari periode waktu *monitoring* (2023-09-30 s.d. 2023-11-03) dan beberapa parameter kondisional lainnya, termasuk parameter untuk memfilter situs-situs yang tidak relevan untuk analisis topik, antara lain situs media sosial dan *e-commerce*, berdasarkan format *query formulation I* (Tabel 2).

TABEL 2
FORMAT DASAR QUERY FORMULATION I

Brand Name	Operator	Search Parameter 1	Operator	Search Parameter 2	Operator	Periode Tanggal Awal	Operator	Periode Tanggal Akhir
“brand name”		-site:[URL 1]		-intitle:[text 1]				
OR		AND (+)		AND (+)				
“alias 1”	AND (+)	-site:[URL 2]	AND (+)	-intitle:[text 2]	AND (+)	after: yyyy-mm-dd	AND (+)	before: yyyy-mm-dd
OR		AND (+)		AND (+)				
“alias 2”		-site:[URL 3]		-intitle:[text 3]				

Proses *scraping* SERP I menghasilkan 134 baris data *tabular*, kemudian diseleksi dengan mengambil kolom *Title* dan *Description*, menghasilkan sampel data mentah (Gambar 5). Langkah berikutnya adalah melakukan *filtering* dan *cleaning* untuk menghapus *noise* berupa data yang tidak diperlukan, antara lain data halaman yang berisi *tag* navigasi dan halaman indeks; menghapus tanda baca, karakter khusus, angka, penunjuk waktu, dan *stopwords* menggunakan *custom dictionary* yang ditentukan berdasarkan hasil *exploratory data analysis* (EDA); dan dokumen yang tidak relevan, antara lain indikasi jurnal akademik dan indikasi artikel lainnya yang tidak diperlukan. Langkah *preprocessing* ini menghasilkan data bersih pada Gambar 6. Distribusi kata berdasarkan frekuensi kemunculan *terms* dan bobot TF-IDF setelah pembentukan *corpus bigram/trigram* dapat dilihat pada Gambar 7 dan 8.

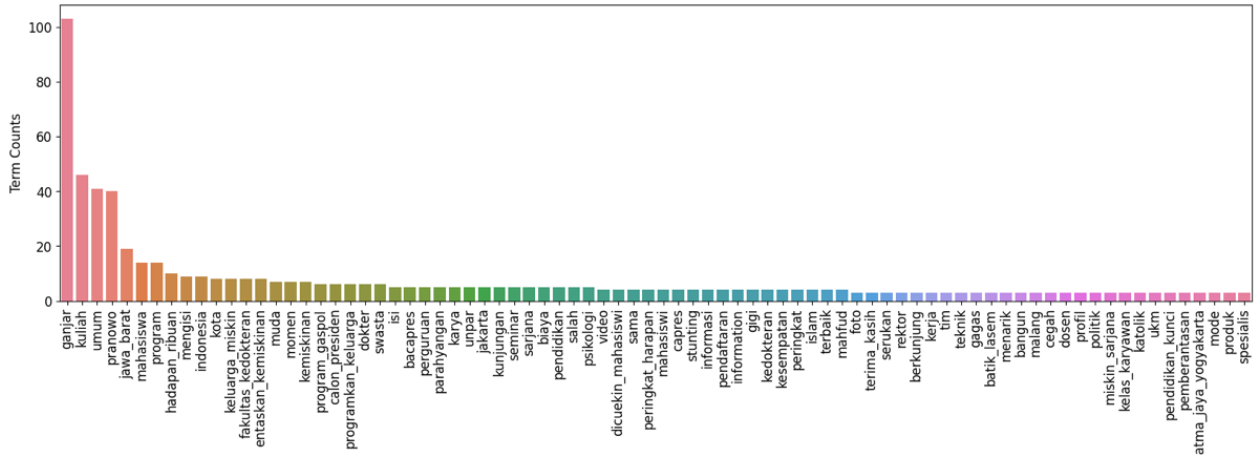
12 Siswa Teknik Otomasi Industri (TOI) SMK Negeri 1 Cimahi ... - 5 days ago - ... Universitas Kristen Maranatha, bekerja sama dengan Rexroth Bosch Company. Kompetisi yang berlangsung pada tanggal 25 ...
 13 Ganjar Kenalkan Program Gaspol Saat Isi Kuliah Umum di ... - Oct 13, 2023 - Jakarta, Gesuri.id - Calon presiden dari PDI Perjuangan, Ganjar Pranowo mengenalkan program Gaspol saat mengisi kuliah u...
 14 UKDW Terima Kunjungan dari Universitas Kristen ... - Oct 1, 2023 - Bertempat di Ruang Seminar Pdt. Dr. Harun Hadiwijono, Universitas Kristen Duta Wacana (UKDW) menerima kunjungan dari Universitas ...
 15 AdminSAAC, Author at SAAC - Universitas Nusa Putra - Oct 17, 2023 - Program Studi Teknik Elektro Universitas Kristen Maranatha bersama Rexroth Bosch Company Mengadakan HaKathon Competition 2023. D...
 16 Ganjar Gagasan Program Satu Keluarga Miskin Satu Sarjana - Oct 12, 2023 - Gagasan itu disampaikan Ganjar saat dirinya mengisi kuliah umum di Universitas Kristen Maranatha, Bandung, Jawa Barat, Rabu,...
 17 Batik Lasem UK Maranatha Raih Penghargaan Seoul ... - 4 days ago - "The New Face of Batik Lasem Industry" karya tim UK Maranatha, Indonesia menempati posisi pemenang ketiga, bersama dengan lima ka...

Gambar 5. Sampel hasil seleksi data mentah SERP I

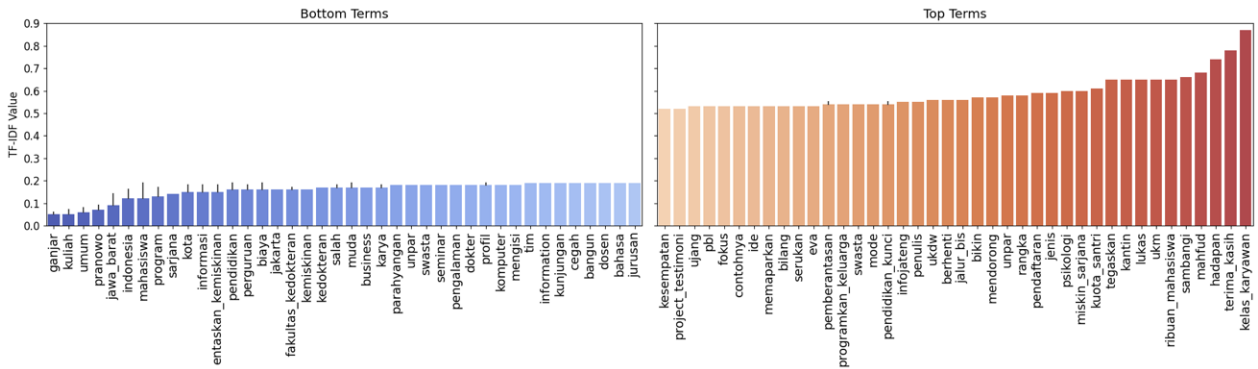
4
5
6
7
8
9
10
11
12
13
14
15

LEI AMB KASAH LEI AMB KASAH LEI AMB KASAH peribnyngin karya wabala epasube delabu memaparakni keyalaha
isi kuliah umum ganjar serukan isi kuliah umum ganjar serukan perempuan nikah rektor prof sri widiyantoro
ganjar serukan nikah muda pemateri parahyangan unpar ganjar pranowo materi kuliahnya
momen ganjar pranowo dicuekin mahasiswa viral media sosial ganjar pranowo dicuekin mahasiswa berkunjung video beragam respon
kota stasiun lokal berjarak meter berjalan jalur bis berhenti jalur bis berhenti
widyatama borong penghargaan kerja sama peringkat harapan peringkat harapan institut teknologi nasional hasil penilaian tim juri kategori kerja sama
siswa teknik otomasi industri toi smk negeri cimahi sama rexroth bosch company kompetisi
ganjar kenalkan program gaspol isi kuliah umum jakarta gesuri calon presiden pdi perjuangan ganjar pranowo mengenalkan program gaspol mengisi kuliah umum
ukdw terima kunjungan seminar pdt harun hadiwijono duta wacana ukdw menerima kunjungan
adminsaac author saac nusa putra program studi teknik elektro rexroth bosch company haxathon competition dilaksanakan
ganjar gagasan program keluarga miskin sarjana gagasan ganjar mengisi kuliah umum jawa barat ganjar
batik lasem raih penghargaan seoul new face batik lasem industry karya tim indonesia menempati posisi pemenang ketiga karya india brasil

Gambar 6. Sampel data bersih setelah preprocessing SERP I



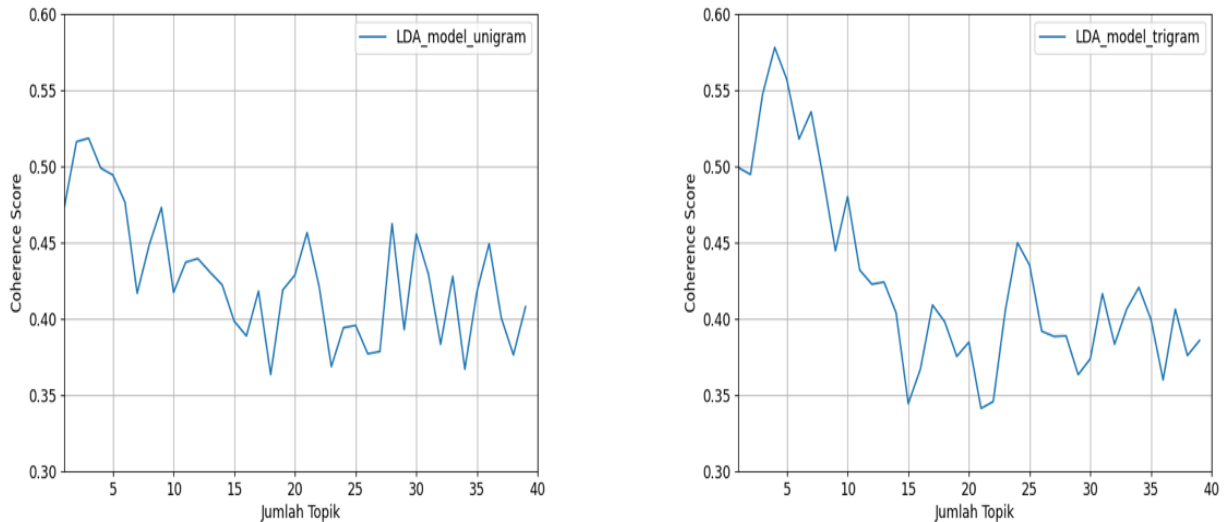
Gambar 7. Distribusi kata teratas pada corpus bigram/trigram SERP I



Gambar 8. Distribusi kata terbawah dan teratas berdasarkan bobot TF-IDF SERP I

E. Pengujian Pemodelan Topik SERP I

Tahapan langkah pemodelan topik SERP I dilakukan berdasarkan diagram proses pada Gambar 4 (kiri). Dari *corpus unigram* dan *bigram/trigram* yang telah dihasilkan pada langkah sebelumnya, dibentuk masing-masing 40 model LDA (jumlah topik $K=1$ s.d. 40) untuk dievaluasi berdasarkan perhitungan *coherence score* (Gambar 9). Pada rentang jumlah topik yang diharapkan, yaitu antara 5 s.d. 15, terlihat bahwa model LDA *bigram/trigram* dapat menghasilkan *coherence score* (cv) lebih tinggi. Kandidat jumlah topik yang dipilih adalah $K=5$ ($cv=0,5568$), $K=7$ ($cv=0,5357$), dan $K=10$ ($cv=0,48$).

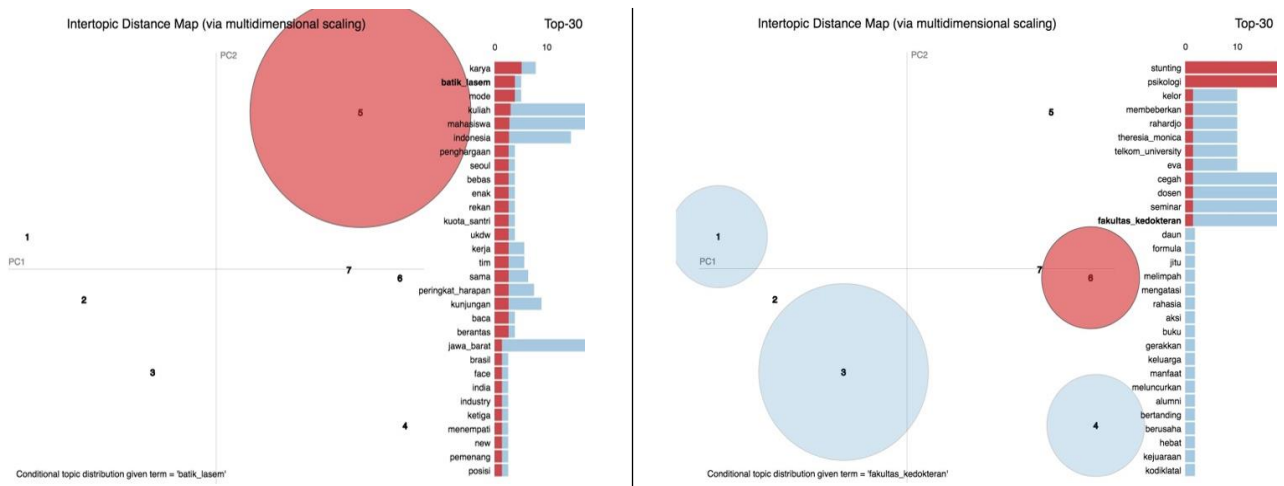


Gambar 9. Grafik *coherence score* model LDA *unigram* (kiri) dan *bigram/trigram* (kanan).

Untuk menentukan jumlah topik terbaik, selanjutnya dilakukan visualisasi terhadap ketiga kandidat model tersebut. Dari diagram *intertopic distance map* (Gambar 10) terlihat bahwa model dengan jumlah topik $K=7$ (diagram tengah) dapat dipilih sebagai model terbaik dengan kriteria masing-masing kelompok topik (direpresentasikan dalam bentuk lingkaran) dapat terpisahkan dengan baik; tidak terdapat banyak lingkaran yang saling beririsan; dan distribusi kata (*terms*) yang terbentuk dapat mengelompok dengan baik (Gambar 11). Daftar 7 topik yang dihasilkan oleh model terbaik beserta *top terms* dan bobot *terms* penyusun masing-masing topik ditampilkan pada Gambar 12. Tabel validasi model topik SERP I ditampilkan pada Gambar 13.



Gambar 10. Analisis visual untuk melihat keberhasilan model memisahkan topik (kiri ke kanan: $K=5, 7, 10$).



Gambar 11. Distribusi term “batik_lasem” (kiri) dan “fakultas_kedokteran” (kanan) pada topik

```
[0,
'0.078*ganjar" + 0.049*umum" + 0.049*kuliah" + 0.026*pranowo" + 0.014*program_gaspol" + 0.014*isi" + 0.011*mengisi" + 0.010*mahasiswa" + 0.008*calon_presiden" + 0.008*jakarta'),
(1,
'0.072*ganjar" + 0.036*pranowo" + 0.022*kuliah" + 0.019*umum" + 0.015*muda" + 0.015*swasta" + 0.015*fakultas_kedokteran" + 0.011*terima_kasih" + 0.008*ajak" + 0.008*politik'),
(2,
'0.020*karya" + 0.015*mode" + 0.015*batik_lasem" + 0.012*kuliah" + 0.011*mahasiswa" + 0.011*indonesia" + 0.010*tim" + 0.010*penghargaan" + 0.010*seoul" + 0.010*sam'),
(3,
'0.012*lukas" + 0.012*jenis" + 0.012*pendaftaran" + 0.009*program" + 0.009*kuliah" + 0.008*fasilitas" + 0.008* LENGKAP" + 0.008*hukum" + 0.008*pengalaman" + 0.008*information'),
(4,
'0.015*tpa_berhitung" + 0.015*bahasa_Inggris" + 0.008*ips" + 0.008*osc" + 0.008*online" + 0.008*matematika" + 0.008*test" + 0.008*cek" + 0.008*diujikan" + 0.008*ipa'),
(5,
'0.021*psikologi" + 0.021*stunting" + 0.011*rahardjo" + 0.011*dosen" + 0.011*fakultas_kedokteran" + 0.011*membeberkan" + 0.011*kelor" + 0.011*theresia_monica" + 0.011*cegah" +
0.011*telkom_university'),
(6,
'0.106*ganjar" + 0.040*pranowo" + 0.036*jawa_barat" + 0.035*umum" + 0.035*kuliah" + 0.018*program" + 0.014*mahasiswa" + 0.014*momen" + 0.012*keluarga_miskin" + 0.012*kemiskinan')]
```

Gambar 12. Daftar 7 topik yang dihasilkan oleh model beserta top terms dan bobotnya

Doc	Document_Text_Content	Keywords	Topic	Dominance	Mutual_Keywords	Mutual_Count	J_Score
0	[isi, 'kuliah', 'umum', 'ganjar', 'bacapres', 'ganjar', 'pranowo', 'mengisi', 'kuliah', 'umum', 'perguruan', 'kota']	ganjar, umum, kuliah, pranowo, program_gaspol, isi, mengisi, mahasiswa, calon_presiden, jakarta	0	0.9858	{ganjar, isi, mengisi, umum, pranowo, kuliah}	6	46.15
1	[ganjar, 'pranowo', 'program_gaspol', 'hadapan', 'ganjar', 'pranowo', 'program_gaspol', 'hadapan', 'mahasiswa', 'jawa_barat', 'foto']	ganjar, umum, kuliah, pranowo, program_gaspol, isi, mengisi, mahasiswa, calon_presiden, jakarta	0	0.7707	{program_gaspol, ganjar, mahasiswa, pranowo}	4	30.77
2	[kunjungi, 'ganjar', 'pranowo', 'calon_presiden', 'bacapres', 'ganjar', 'pranowo', 'mengenalkan', 'program_gaspol', 'mengisi', 'kuliah', 'umum', 'bertajuk']	ganjar, umum, kuliah, pranowo, program_gaspol, isi, mengisi, mahasiswa, calon_presiden, jakarta	0	0.9869	{program_gaspol, ganjar, kuliah, mengisi, umum, pranowo, calon_presiden}	7	50.00
3	[isi, 'kuliah', 'umum', 'ganjar', 'ingatkan', 'baca', 'faktor', 'ganjar', 'pemilu', 'video', 'rekomendasi', 'tpst', 'bantargetbang']	karya, mode, batik_lasem, kuliah, mahasiswa, indonesia, tim, penghargaan, seoul, sama	2	0.6273	{kuliah}	1	4.76
4	[terima_kasih, 'terima_kasih', 'terima_kasih', 'parahyangan', 'karya', 'wisata', 'episode', 'belliau', 'memaparkan', 'kegiatan']	ganjar, pranowo, kuliah, umum, muda, swasta, fakultas_kedokteran, terima_kasih, ajak, politik	1	0.9819	{terima_kasih}	1	5.88

Gambar 13. Tabel validasi pemodelan topik SERP I

F. Pengujian Data Retrieval & Preprocessing SERP II

Tahapan *pipeline data retrieval* dan *preprocessing* SERP II dilakukan berdasarkan diagram proses pada Gambar 3. Sampel data diambil dari *corporate website* dengan cara melakukan *web search* menggunakan parameter *query* pencarian yaitu periode waktu *monitoring* (2023-06-30 s.d. 2023-11-03), berdasarkan format *query formulation* II (Tabel 3). Seleksi data hasil *scraping* SERP II (berisi 100 baris data *tabular*) dilakukan dengan mengambil kolom *Title* dan *Description*, kemudian dilakukan *filtering* dan *cleaning* untuk menghapus *noise* berupa data yang tidak diperlukan, antara lain data halaman yang berisi *tag navigasi* dan halaman indeks; menghapus tanda baca, karakter khusus, angka, penunjuk waktu, dan *stopwords* menggunakan *custom dictionary* yang ditentukan berdasarkan EDA SERP II. Langkah *preprocessing* ini menghasilkan data bersih (Gambar 14).

TABEL 3
FORMAT DASAR QUERY FORMULATION II

Search Parameter	Operator	Periode Tanggal Awal	Operator	Periode Tanggal Akhir
site:[URL]	AND (+)	after:yyyy-mm-dd	AND (+)	before:yyyy-mm-dd

```

0          belajar ganjar ganjar persis mengupayakan keadilan kesejahteraan tantangan harapan regulasi
1          semnas road unggul foto video materi seminar nasional road unggul berkolaborasi lembaga layanan pendidikan lldikti wilayah ban
2          kedokteran terakreditasi internasional asiin program sarjana kedokteran profesi dokter menerima sertifikat akreditasi internasional asiin tertanggal
3          pelantikan mahasiswa sidang terbuka senat rangka pelantikan mahasiswa pembukaan akademik
4          semnas kepemimpinan transformasional ganjar pranowo seminar nasional kepemimpinan transformasional menyejahterakan wong cilik pembicara ganjar
5          penelitian dosen raih hibah kemendikbudristek sepuluh penelitian dosen terpilih penerima hibah kemendikbudristek hibah diterima sembilan orang
6          jalan kerja sama pemkot samarinda menandatangani memorandum understanding mou pemerintah kota pemkot samarinda penandatanganan
7          mengenal jurusan teknik industri prospek kerjanya sales engineer lulusan teknik industri bertanggung menjual produk teknis kemampuan teknis miliki
8          bangkitnya batik wajah indonesia hadapan dunia batik kebanggaan budaya indonesia diakui dunia unesco menetapkan batik warisan budaya nonbendawi batik
9          prospek kerja teknik elektro rencana kariermu prospek kerja teknik elektro bidang manufaktur ahli teknik elektro sektor manufaktur profesional bertanggung
10         menkumham yasonna laoly tantangan wujudkan yasonna indonesia emas impian tujuan diwujudkan mencapai tujuan
11         pengukuhan guru besar ilmu hukum ilmu biologi guru besar dikukuhkan guru besar prof pan lindawaty suherman sewu
12         jurusan kuliah bidang bisnis beserta prospek kerjanya jurusan kuliah bisnis prospek kerjanya kewirausahaan akuntansi bisnis berkelanjutan administrasi bisnis manajemen bisnis
13         batik lasem fsrd raih penghargaan seoul satunya karya indonesia berjudul new face batik lasem industry resmi menerima penghargaan pemenang ketiga seoul design award
  
```

Gambar 14. Sampel data bersih setelah preprocessing SERP II

G. Pengujian Pemodelan Topik SERP II

Tahapan langkah pemodelan topik SERP II dilakukan berdasarkan diagram proses pada Gambar 4 (kiri). Langkah-langkah kerja evaluasi dan validasi untuk menentukan jumlah topik terbaik sama dengan langkah kerja pada pemodelan topik SERP I yang telah dipaparkan pada bagian E. Model terbaik yang dipilih adalah dengan jumlah topik $K=6$. Daftar 6 topik beserta *top terms* dan bobot *terms* penyusun masing-masing topik ditampilkan pada Gambar 15. Tabel validasi model topik SERP II ditampilkan pada Gambar 16.

```

[[0,
  '0.017*elektro" + 0.017*seminar" + 0.017*pembicara" + 0.017*kepemimpinan" + 0.017*ganjar_pranowo" + 0.017*hibah" + 0.017*teknik" + 0.011*kemendikbudristek" + 0.011*kerja" +
  0.011*penelitian'),
  (1,
  '0.024*mahasiswa" + 0.014*studi" + 0.014*juara" + 0.011*mou" + 0.011*seni_rupa" + 0.011*kerja_sama" + 0.011*desain" + 0.011*kabupaten" + 0.011*program" + 0.011*pemerintah'),
  (2,
  '0.021*fakultas" + 0.017*dies_natalis" + 0.013*indonesia" + 0.013*rangkaiannya" + 0.009*start" + 0.009*perayaan" + 0.009*kedokteran" + 0.009*layoff" + 0.009*wiratha_fest" +
  0.009*abdimas'),
  (3,
  '0.023*mahasiswa" + 0.023*program" + 0.014*sarjana" + 0.012*generasi" + 0.012*akreditasi" + 0.012*teknik" + 0.012*indonesia" + 0.010*wiratha_fest" + 0.010*internasional" +
  0.010*batik'),
  (4,
  '0.011*batik_lasem" + 0.011*indonesia" + 0.011*penghargaan" + 0.011*menerima" + 0.011*seharis" + 0.011*seminar" + 0.011*exposure" + 0.011*kunjungan" + 0.006*pemenang" + 0.006*seoul'),
  (5,
  '0.029*bisnis" + 0.011*fakultas" + 0.011*kuliah" + 0.011*manajemen" + 0.011*dekan" + 0.008*jurusan" + 0.008*bantu" + 0.008*lomba" + 0.008*prospek_kerjanya" + 0.008*alat')]]
  
```

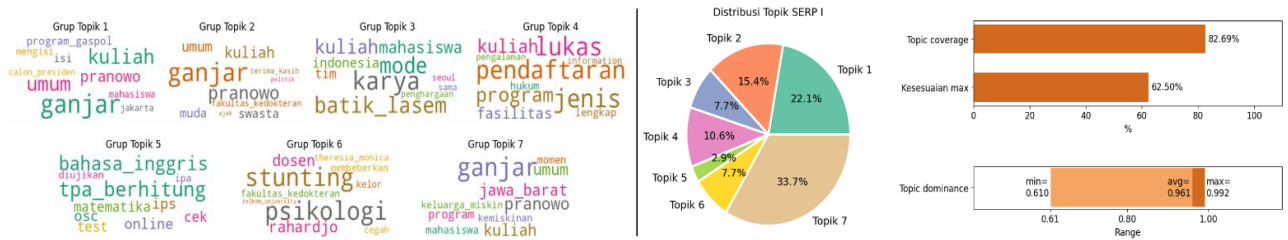
Gambar 15. Daftar 6 topik yang dihasilkan oleh model SERP II

Doc	Document_Text_Content	Keywords	Topic	Dominance	Mutual_Keywords	Mutual_Count	J_Score	
0	0	[belajar, ganjar, ganjar, persis, mengupayakan, keadilan, kesejahteraan, tantangan, harapan, regulasi]	bisnis, fakultas, kuliah, manajemen, dekan, jurusan, bantu, lomba, prospek_kerjanya, alat	5	0.9878	0	0.00	
1	1	[semnas, road, unggul, foto, video, materi, seminar, nasional, road, unggul, berkolaborasi, lembaga, tayanari, pendidikan, lldikti, wilayah, bani]	fakultas, dies_natalis, indonesia, rangkaian, start, perayaan, kedokteran, layoff, wiratha_fest, abdimas	2	0.9927	0	0.00	
2	2	[kedokteran, terakreditasi, internasional, asiin, program, sarjana, kedokteran, profesi, dokter, menerima, sertifikat, akreditasi, internasional, asiin, tertanggal]	mahasiswa, program, sarjana, generasi, akreditasi, teknik, indonesia, wiratha_fest, internasional, batik	3	0.9929	{sarjana, internasional, program, akreditasi}	4	22.22
3	3	[pelantikan, mahasiswa, sidang_terbuka, senat_rangka, pelantikan, mahasiswa, pembukaan, akademik]	mahasiswa, program, sarjana, generasi, akreditasi, teknik, indonesia, wiratha_fest, internasional, batik	3	0.9867	{mahasiswa}	1	6.67
4	4	[semnas, kepemimpinan, transformasional, ganjar_pranowo, seminar, nasional, kepemimpinan, transformasional, menyejahterakan, wong, cilik, pembicara, ganjar]	elektro, seminar, pembicara, kepemimpinan, ganjar_pranowo, hibah, teknik, kemendikbudristek, kerja, penelitian	0	0.9901	{seminar, kepemimpinan, ganjar_pranowo, pembicara}	4	23.53
...	

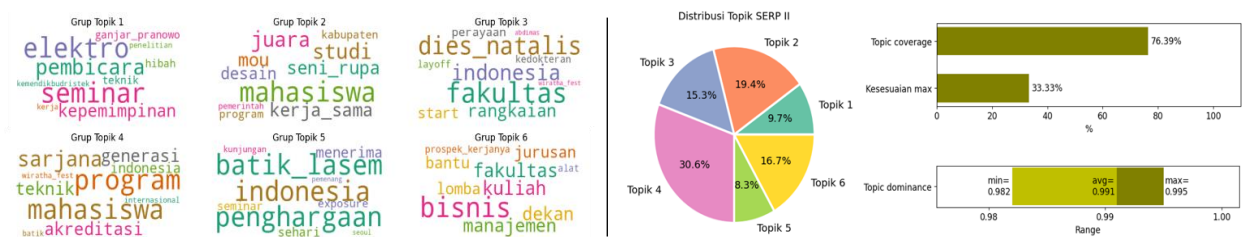
Gambar 16. Tabel validasi pemodelan topik SERP II

H. Pengujian Komparasi dan Output Visualisasi

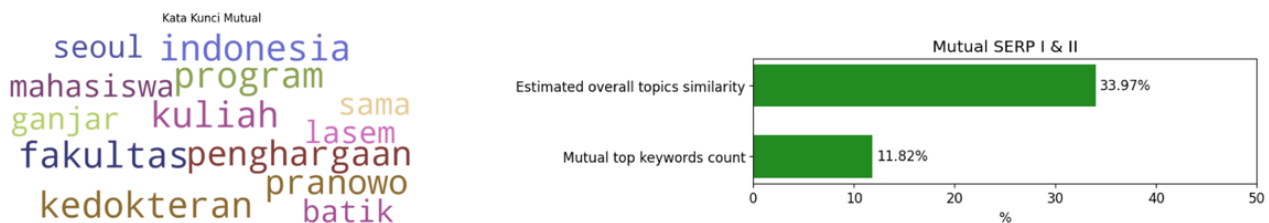
Tahapan langkah proses komparasi dan output visualisasi dilakukan berdasarkan diagram proses pada Gambar 4 (kanan). Masing-masing topik yang dihasilkan oleh model topik SERP I dan SERP II divisualkan menggunakan *word cloud* dan *metrics report* (Gambar 17 & 18). Hasil komparasi ditampilkan berupa visualisasi *word cloud mutual keywords* (tanpa bobot) dan *metrics report* (Gambar 19).



Gambar 17. Visualisasi word cloud topik dan metrics report SERP I



Gambar 18. Visualisasi word cloud topik dan metrics report SERP II



Gambar 19. Visualisasi word cloud mutual keywords dan metrics report komparasi SERP I & SERP II

III. HASIL DAN PEMBAHASAN

A. Analisis & Interpretasi Topik SERP I

Memperhatikan hasil pemodelan topik SERP I (Gambar 12 & 17), dapat diinterpretasikan 7 kelompok topik dominan yang diuraikan pada Tabel 4. *Topic coverage* SERP I mencapai 82,69% (metrik Gambar 17 kanan), berarti topik-topik yang dibentuk oleh model dapat mencakup 82,69% dari keseluruhan dokumen yang terdapat pada SERP I; sisanya (17,31%) adalah topik yang tidak tercakup, yaitu dokumen yang gagal diekstraksi, salah penempatan topik, dan dokumen yang frekuensinya sangat rendah. Kesesuaian maksimal 62,5% mengindikasikan adanya topik yang dapat merepresentasikan dokumen aslinya hingga sebesar 62,5% dari keseluruhan *terms* gabungan. Semakin besar angka kesesuaian, topik yang terbentuk cenderung semakin homogen (tidak terdiri dari banyak subtopik). Tingkat dominasi cenderung tinggi (rata-rata 0,961 dari skala 1) mengindikasikan lebih banyak dokumen yang direpresentasikan oleh topik yang dominan (berbobot tinggi).

TABEL 4
INTERPRETASI TOPIK DOMINAN PADA SERP I

Grup Topik	Interpretasi Topik
Topik 1	kuliah umum Ganjar Pranowo mengenai program gaspol
Topik 2	kuliah umum Ganjar Pranowo berkaitan dengan politik dan generasi muda; kuliah fakultas kedokteran; perguruan tinggi swasta
Topik 3	penghargaan batik Lasem; penghargaan karya mahasiswa; Program Desain Mode
Topik 4	pendaftaran kuliah; informasi fasilitas kampus; informasi Fakultas Hukum
Topik 5	materi ujian online
Topik 6	dosen Fakultas Kedokteran membahas pencegahan <i>stunting</i> ; informasi mengenai Fakultas Psikologi
Topik 7	kuliah umum Ganjar Pranowo mengenai program keluarga miskin di Jawa Barat

Hasil analisis terhadap model topik SERP I: 1) Topik-topik yang dihasilkan secara umum dapat menggambarkan isu-isu dengan *brand mention* Universitas yang dipublikasikan pada situs media massa dan situs lainnya; 2) Terdapat topik yang sangat dominan dan sangat homogen mengerucut ke satu topik khusus (topik 1, 5, 7), dan terdapat topik yang heterogen (antara lain topik 2 & 6); 3) Topik-topik yang heterogen terdiri dari beberapa subtopik yang tidak dapat dipisahkan menjadi topik tersendiri; 4) Beberapa topik terdiri dari *terms* yang sangat khas dan tidak dapat diinterpretasikan dengan gamblang; 5) Beberapa *keyword* spesifik dapat menjadi kandidat formulasi kueri untuk penelusuran lanjutan, bila dibutuhkan oleh tim pemonitor.

B. Analisis & Interpretasi Topik SERP II

Memperhatikan hasil pemodelan topik SERP II (Gambar 15 & 18), dapat diinterpretasikan 6 kelompok topik dominan yang diuraikan pada Tabel 5. *Topic coverage* SERP II mencapai 76,39% (metrik Gambar 18 kanan), berarti topik-topik yang dibentuk oleh model dapat mencakup 76,39% dari keseluruhan dokumen pada SERP II; sisanya (23,61%) merupakan topik yang tidak tercakup, yaitu dokumen yang gagal diekstraksi, salah penempatan topik, dan dokumen yang frekuensinya sangat rendah. Kesesuaian maksimal sebesar 33,33% mengindikasikan adanya topik yang dapat merepresentasikan dokumen aslinya hingga sebesar 33,33% dari keseluruhan *terms* gabungan. Semakin kecil angka kesesuaian, topik yang terbentuk cenderung semakin heterogen (terdiri dari banyak subtopik). Tingkat dominansi tinggi (0,982 sampai 0,995 dari skala 1) mengindikasikan hampir semua dokumen direpresentasikan oleh topik yang sangat dominan (berbobot tinggi).

TABEL 5
INTERPRETASI TOPIK DOMINAN PADA SERP II

Grup Topik	Interpretasi Topik
Topik 1	seminar kepemimpinan dengan pembicara Ganjar Pranowo; peluang kerja teknik elektro; kegiatan penelitian hibah Kemendikbudristek
Topik 2	aktivitas mahasiswa Fakultas Seni Rupa dan Desain; perolehan juara; kerja sama dengan pemerintah
Topik 3	rangkaian dies natalis dan Wiratha Fest; kegiatan nasional; kegiatan pengabdian masyarakat; kegiatan Fakultas Kedokteran
Topik 4	kegiatan mahasiswa program sarjana; program akreditasi; tema mengenai Indonesia
Topik 5	karya batik Lasem menerima penghargaan di Seoul; kegiatan seminar dan kunjungan
Topik 6	prospek kerja dan kegiatan Fakultas Bisnis dan jurusan manajemen; kegiatan melibatkan dekan; kegiatan lomba

Hasil analisis terhadap model topik SERP II: 1) Topik-topik yang dihasilkan secara umum dapat menggambarkan garis besar isu yang dipublikasikan pada situs Maranatha News; 2) Topik-topik ini membentuk kaitan asosiasi *brand* dengan isu-isu yang diterbitkan pada periode *monitoring*; 3) Masing-masing topik terdiri dari beberapa subtopik yang tidak dapat dipisahkan menjadi topik tersendiri; 4) Daftar *terms* pada 6 topik gabungan cukup menggambarkan hal-hal penting yang diekspos publikasi.

C. Analisis Komparasi Topik SERP I & SERP II

Hasil analisis komparasi dengan memperhatikan *word cloud* SERP I (Gambar 17) dan SERP II (Gambar 18), juga *mutual keywords* (Gambar 19): 1) Terdapat kesamaan isu yang dibahas pada situs Maranatha News dengan situs media massa dan situs lainnya, yaitu sebesar 33,97%, dan terdapat *mutual terms* sebanyak 11,82%; 2) Kesamaan topik paling dominan adalah mengenai Ganjar Pranowo, batik Lasem, dan Fakultas Kedokteran, mengindikasikan *brand exposure* yang lebih luas dari topik-topik ini, dengan paparan mencakup kedua kategori media; 3) Topik mengenai Ganjar Pranowo homogen pada topik 7 SERP I dengan distribusi tertinggi (33,7%), dan pada 2 topik lainnya; sedangkan di SERP II hanya ditampilkan pada topik 1 dengan bobot rendah, mengindikasikan jumlah artikel Maranatha News yang membahas topik ini sangat sedikit; 4) Ketidaksamaan di luar 33,97% mengindikasikan masing-masing kategori media memiliki topik-topik khas yang berlainan.

D. Temuan & Diskusi

Beberapa temuan eksperimen dan pengujian dalam penelitian ini dari aspek teknis diuraikan sebagai berikut. 1) Proses LDA *topic modeling* pada data berupa SERP *snippet* secara garis besar dapat berfungsi sebagai *dimension reduction*, dapat mereduksi banyak data menjadi beberapa topik tanpa menghilangkan informasi penting yang terkandung. 2) Topik yang dihasilkan tidak sepenuhnya mudah untuk diinterpretasikan secara gamblang, dikarenakan adanya campuran subtopik. 3) Keterbatasan tersebut tidak membatasi performa dan kebermanfaatan kerangka kerja, karena tujuan yang ingin dicapai adalah mengekstraksi informasi yang paling dominan secara garis besar, bukan untuk mengklasifikasi informasi. 4) Topik yang relatif mudah diinterpretasikan adalah topik yang cenderung homogen, atau yang mengandung *terms* dominan (berbobot tinggi). 5) Topik yang terdiri dari banyak *terms* granular membentuk campuran subtopik, menghasilkan satu topik yang heterogen. 6) Karakter data input menentukan kemudahan interpretasi dan kualitas topik yang dihasilkan. Proses *filtering* dan

cleaning sangat penting dalam pemodelan topik LDA, untuk memperoleh data yang “*balance*” tanpa banyak *noise*, tetapi tidak terlalu steril sehingga menyulitkan interpretasi topik. 7) *Dictionary* untuk melakukan *filtering* pada penelitian ini masih terbatas, dan perlu diperluas lagi seiring implementasi kerangka kerja pada periode-periode monitoring selanjutnya, untuk mengakomodasi pola-pola *noise* yang belum ditemukan. 8) Pembentukan *corpus bigram/trigram* sebagai metode *feature engineering* dapat membantu meningkatkan performa model, dan dapat membantu mempermudah interpretasi topik.

Sedangkan dari aspek kebermanfaatan, output kerangka kerja berupa visualisasi *word cloud* dan *metrics report* dapat mempermudah tim *monitoring* dan *stakeholder* untuk mencerna informasi yang terkandung dalam masing-masing SERP. Daftar kelompok topik yang dihasilkan cukup memberikan gambaran garis besar mengenai isu-isu yang dapat dikaitkan dengan *brand mention* dan *brand exposure* institusi pada periode pemantauan. Metrik komparasi dapat menjadi indikator penting bagi tim pemonitor, yang menunjukkan adanya kesamaan dan perbedaan pandangan mengenai isu-isu berkaitan dengan *brand* institusi yang dianggap penting oleh institusi dan media eksternal.

IV. SIMPULAN

Berdasarkan eksperimen dan pengujian yang telah dilakukan, rancangan kerangka kerja FBxM dapat memenuhi ekspektasi untuk membantu proses *brand exposure monitoring*, dan selanjutnya dapat digunakan untuk analisis pengambilan keputusan pengelolaan media dan komunikasi di universitas studi kasus. Keempat hipotesis yang diajukan dapat teruji dengan baik, dan tujuan penelitian dapat tercapai. Masing-masing hipotesis terbukti dengan 1) dihasilkannya SERP I dari *web search* berdasarkan *query formulation* I yang berhasil digunakan sebagai input kerangka kerja; 2) dihasilkannya sejumlah topik hasil proses ekstraksi informasi dari SERP I dan SERP II yang *interpretable*; 3) output yang dihasilkan oleh proses pemodelan topik pada SERP I dan SERP II, masing-masing dapat merepresentasikan garis besar topik yang diekspos pada masing-masing jenis media pada periode tertentu; dan 4) didapatkannya hasil komparasi berupa *mutual keywords* dan metrik *estimated overall topics similarity* sebagai indikator tingkat kesamaan dan kekhasan isu yang dipaparkan pada kedua jenis media. Metode pemodelan topik yang digunakan dalam kerangka kerja pada penelitian ini dapat digunakan sebagai *baseline* untuk penelitian-penelitian selanjutnya, yaitu menggunakan metode pemodelan topik lainnya, guna memperoleh performa yang lebih baik. Penelitian selanjutnya dapat dilakukan untuk menemukan metode pemodelan topik dengan kinerja yang lebih baik dan dapat mengatasi keterbatasan pada metode yang telah diterapkan, sehingga kebermanfaatan kerangka kerja dapat lebih ditingkatkan. Pengembangan juga dapat dilakukan dengan merancang otomasi tahapan proses dalam kerangka kerja, misalnya dengan membuat fungsi untuk menjalankan *pipeline* antarmodul secara *sequential* atau berdasarkan kondisi-kondisi tertentu. Kemudian dapat dibuat antarmuka yang *user friendly* untuk memungkinkan implementasi praktis, sehingga kerangka kerja dapat dimanfaatkan oleh pemonitor tanpa perlu kemampuan pemrograman.

DAFTAR PUSTAKA

- [1] Q. Liu et al., “Health Communication Through News Media During the Early Stage of the COVID-19 Outbreak in China: Digital Topic Modeling Approach,” *J. Med. Internet Res.*, vol. 22, no. 4, p. e19118, 2020.
- [2] R. Abebe, S. Hill, J. W. Vaughan, P. M. Small, and H. A. Schwartz, “Using Search Queries to Understand Health Information Needs in Africa.” *ICWSM*, vol. 13, no. 1, pp. 3–14, 2019.
- [3] A. F. Hidayatullah, S. K. Aditya, Karimah, and S. T. Gardini, “Topic modeling of weather and climate condition on twitter using latent dirichlet allocation (LDA),” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 482, no. 1, p. 012033, 2019.
- [4] O. Kamalim, M. Ayub, and H. Toba, “Mitigating plagiarism and collusion in K-12 as initial knowledge for programming instructors in higher education,” *World Trans. Eng. Technol. Educ.*, vol. 21, no. 1, pp. 44–49, 2023.
- [5] W. Wahyudin, “Aplikasi Topic Modeling pada Pemberitaan Portal Berita Online Selama Masa PSBB Pertama,” *Semin. Nas. Off. Stat.*, vol. 2020, no. 1, pp. 309-318, 2021.
- [6] K. R. A. P. Santoso, A. Husna, N. W. Putri, and N. A. Rakhmawati, “Analisis Topik Tagar Covidindonesia pada Instagram Menggunakan Latent Dirichlet Allocation,” *JISKA*, vol. 7, no. 1, pp. 1–9, 2022.
- [7] H. Sagala and H. Toba, “Penentuan Aspek yang Berpengaruh Terhadap Produk Smartphone Berdasarkan Ulasan Berbasis Teksual,” *JuTISI*, vol. 7, no. 1, pp. 287–295, 2021.
- [8] D. Z. T. Kannitha, M. Mustafid, and P. Kartikasari, “Pemodelan Topik pada Keluhan Pelanggan Menggunakan Algoritma Latent Dirichlet Allocation dalam Media Sosial Twitter,” *J. Gaussian*, vol. 11, no. 2, pp. 266–277, 2022.
- [9] I. Kharisudin and H. Masri’an, “Topic Modeling on WhatsApp User Reviews Using Latent Dirichlet Allocation,” *Sci. J. Inform.*, vol. 9, no. 1, pp. 51–62, 2022.
- [10] N. Novarian, S. Khomsah, and A. B. Arifa, “Topic Modeling Tugas Akhir Mahasiswa Fakultas Informatika Institut Teknologi Telkom Purwokerto Menggunakan Metode Latent Dirichlet Allocation,” *LEDGER J. Inform. Inf. Technol.*, vol. 2, no. 1, pp. 14–27, 2023.
- [11] P. Kotler and K. L. Keller, *Marketing Management*, 15th ed., Boston: Pearson, 2016.
- [12] D. A. Aaker, *Managing Brand Equity: Capitalizing on the Value of a Brand Name*, DIANE Publishing Company, 2000.
- [13] T. A. Shimp and J. C. Andrews, *Advertising, Promotion, and Other Aspects of Integrated Marketing Communications*, 9th ed., Cengage Learning, 2013.
- [14] K. L. Keller, “Conceptualizing, Measuring, and Managing Customer-Based Brand Equity,” *J. Mark.*, vol. 57, no. 1, pp. 1–22, 1993.
- [15] T. B. Cornwell, G. E. Relyea, R. L. Irwin, and I. S. J. Maignan, “Understanding Long-Term Effects of Sports Sponsorship: Role of Experience, Involvement, Enthusiasm and Clutter,” *Int. J. Sports Mark. Spons.*, vol. 2, no. 2, pp. 127–144, 2000.
- [16] B. Pitts and J. Slattery, “An examination of the effects of time on sponsorship awareness levels,” *Sport Mark. Q.*, vol. 13 no. 1, pp. 43–54, 2004.
- [17] R. Arthana, Noermijati, and C. Susilowati, “Pengaruh Brand Exposure dan Brand Experience Terhadap Brand Trust dan Brand Recall (Studi pada Produk Smartphone di Wilayah Kota Malang),” *Ekon. Bisnis*, vol. 21, no. 2, pp. 201–213, 2016.

- [18] L. Pomalaa, Y. T. Mursityo, and A. D. Herlambang, "Analisis Faktor Brand Awareness, Brand Exposure, Customer Engagement, dan Electronic Word-of-Mouth dalam Pemasaran Melalui Media Sosial pada The Body Shop Indonesia", *J-PTIHK*, vol. 2, no. 10, pp. 4082–4091, 2018.
- [19] K. T. Tran, P. V. Nguyen, H. T. S. Do, and L. T. Nguyen, "University students' insight on brand equity," *Manag. Sci. Lett.*, vol. 10, pp. 2053–2062, 2020.
- [20] M. Pinar, "University Branding and Measuring Brand Equity: Conceptual Framework and Empirical Studies," *J. Int. Sci. Res.*, vol. 5, pp. 1–5, 2020.
- [21] Perwito, A. Rahayu, and H. Hendrayati, "Integrated Marketing Communication Analysis and Its Effect Towards Brand Equity," in *Proceedings of the 5th Global Conference on Business, Management and Entrepreneurship (GCBME 2020)*, Atlantis Press, 2021, pp. 306–310.
- [22] T. Gatfield, M. Barker, and P. Graham, "Measuring communication impact for university advertising materials," *Corp. Commun. Int. J.*, vol. 4, no. 2, pp. 73–79, 1999.
- [23] B. Gray, V. Llanes, and K. Fam, "Branding universities in Asian markets," *Journal of Product and Brand Management*, vol. 12, no. 2, pp. 108–118, 2003.
- [24] M. Pinar, P. Trapp, T. Girard, and T. E. Boyt, "University brand equity: An empirical investigation of its dimensions," *Int. J. Educ. Manag.*, vol. 28, no. 6, pp. 616–634, 2014.
- [25] N. Sujchaphong, B. Nguyen, and T. Melewar, "Towards a branding oriented higher education sector: An overview of the four perspectives on university marketing studies," *Mark. Rev.*, vol. 17, no. 1, pp. 87–116, 2017.
- [26] S. Noor, K. Abdul Manan, and H. M. Abdul Kuthoos, "Assessing Corporate Brand Equity of Public Universities," *J. Komun. Malays. J. Commun.*, vol. 35, no. 3, pp. 283–299, 2019.
- [27] F. de Heer, "Exploring the Understanding of University Brand Equity: Perspectives of Public Relations and Marketing Directors," *IOSR J. Bus. Manag. IOSR-JBM*, vol. 22, no. 7, pp. 49–57, 2020.
- [28] C. Binnie, "Constructing University Brands Through University Research Magazines," University of Ottawa, 2008.
- [29] C. Chapleo, "What defines 'successful' university brands?," *Int. J. Public Sect. Manag.*, vol. 23, no. 2, pp. 169–183, 2010.
- [30] B. Gutiérrez-Villar, P. Alcaide-Pulido, and M. Carbonero-Ruz, "Measuring a University's Image: Is Reputation an Influential Factor?," *Educ. Sci.*, vol. 12, no. 1, p. 19, 2022.
- [31] J. M. T. Balmer and E. R. Gray, "Corporate identity and corporate communications: creating a competitive advantage," *Ind. Commer. Train.*, vol. 32, no. 7, pp. 256–262, 2000.
- [32] M. S. Dewi and I. I. Wahyuni, "Strategi Media Monitoring pada Pemberitaan Covid-19 di Media Daring," *MEDIALOG J. Ilmu Komun.*, vol. 5, no. 1, pp. 35–43, 2022.
- [33] F. Jeffkins, *Public Relations*, 4th ed., Intl Ideas, 1992.
- [34] S. Lukman and T. Friedel, "Program Kliping Media Sebagai Kegiatan Media Monitoring Divisi Humas di PT Perusahaan Listrik Negara (PLN) Persero Distribusi Jawa Barat," *JRK J. Ris. Komun.*, vol. 10, no. 2, pp. 181–194, 2020.
- [35] H. A. Habiba, "Peran Media Monitoring dalam Penanganan Pemberitaan Limbah Pabrik pada Bagian Informasi dan Komunikasi Departemen Humas PT Pusri Palembang," Universitas Sriwijaya, 2020.
- [36] I. A. P. Astiti, E. Rusdiyana, and D. Beywiarno, "Aktivitas Media Monitoring di PT. Bisnis Indonesia Konsultan (Bisnis Indonesia Intelligence Unit)," *SENASTER Semin. Nas. Ris. Teknol. Terap.*, vol. 1, no. 1, 2020.
- [37] W. B. Croft, D. Metzler, and T. Strohman, *Search Engines - Information Retrieval in Practice*. Pearson Education, Inc., 2015.
- [38] T. Yang and A. Gerasoulis, "Web search engines: Practice and experience," *Comput. Handb. Two-Vol. Set*. CRC Press, 2022. pp. 1–24.
- [39] A. Aula, "Query Formulation in Web Information Search," *Proceedings of the IADIS International Conference WWW/Internet*, 2003. pp. 403–410.
- [40] T. Russell-Rose and F. Shokraneh, "Designing the Structured Search Experience: Rethinking the Query-Builder Paradigm," *Weave J. Libr. User Exp.*, vol. 3, no. 1, 2020.
- [41] K. A. Hambarde and H. Proenca, "Information Retrieval: Recent Advances and Beyond," *IEEE Access*, vol. 11, pp. 76581–76604, 2023.
- [42] Sachin and D. Patel, "Event-centric Query Suggestion for Online News." arXiv:2201.04716.
- [43] P. Ghasiya and K. Okamura, "Investigating COVID-19 News Across Four Nations: A Topic Modeling and Sentiment Analysis Approach," *IEEE Access*, vol. 9, pp. 36645–36656, 2021.
- [44] D. M. Blei, "Probabilistic topic models," *Commun. ACM*, vol. 55, no. 4, pp. 77–84, 2012.
- [45] D. Maier et al., "Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology," *Commun. Methods Meas.*, vol. 12, no. 2–3, pp. 93–118, 2018.
- [46] D. M. Blei, "Latent Dirichlet Allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.
- [47] S. Bhatia, J. H. Lau, and T. Baldwin, "An Automatic Approach for Document-level Topic Model Evaluation," in *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, R. Levy and L. Specia, Eds., Vancouver, Canada: Association for Computational Linguistics, 2017. pp. 206–215.
- [48] P. Jaccard, "The Distribution of the Flora in the Alpine Zone.1," *New Phytol.*, vol. 11, no. 2, pp. 37–50, 1912.
- [49] T. T. Tanimoto, "An Elementary Mathematical Theory of Classification and Prediction," *International Business Machines Corp.*, 1958.
- [50] Scikit-learn developers, "6.8. Pairwise metrics, Affinities and Kernels," scikit-learn. [Online]. Available: <https://scikit-learn/stable/modules/metrics.html>. [Accessed 29 December 2023].
- [51] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [52] S. Czysch, "SERP Snippet Extractor browser extension - getIndexed." [Online]. Available: <https://www.getindexed.io/en/serp-snippet-extractor/>. [Accessed 29 December 2023].
- [53] R. Řehůřek, "Gensim: topic modelling for humans." [Online]. Available: https://radimrehurek.com/gensim/auto_examples/tutorials/run_lda.html#sphx-glr-auto-examples-tutorials-run-lda-py. [Accessed 29 December 2023].
- [54] B. Mabey, "pyLDavis: Interactive topic model visualization. Port of the R package." [Online]. Available: <https://github.com/bmabey/pyLDavis>. [Accessed 29 December 2023].
- [55] I. C. Drivas, D. P. Sakas, and G. A. Giannakopoulos, "Display Advertising and Brand Awareness in Search Engines: Predicting the Engagement of Branded Search Traffic Visitors," in *Business Intelligence and Modelling*, D. P. Sakas, D. K. Nasiopoulos, and Y. Taratuhina, Eds., in *Springer Proceedings in Business and Economics*. Cham: Springer International Publishing, 2021. pp. 3–15.
- [56] V. Tunuguntla, K. Rakshit, and P. Basu, "Bidding for an optimal portfolio of keywords in sponsored search advertising: From generic to branded keywords," *Eur. J. Oper. Res.*, vol. 307, no. 3, pp. 1424–1440, 2023.