



Pemetaan Topik Percakapan Terkait Fenomena “#KaburAjaDulu” di Twitter Menggunakan IndoBERT dan Klasterisasi Hirarki

Andini¹, I Putu Krishna Deny Suadnyana², Garudo Puji Sandiyudho³

^{1,2,3}Data Science, Telkom University Surabaya, Indonesia

¹galuhandini@student.telkomuniversity.ac.id

²putukrishnaaaaa@student.telkomuniversity.ac.id

³Garudosandiyudho@student.telkomuniversity.ac.id

Corresponding author email: galuhandini@student.telkomuniversity.ac.id

Abstract: The “#KaburAjaDulu” phenomenon on Twitter reflects public responses to perceived social, economic, and political pressures. This study aims to map the dominant discussion topics expressed through this hashtag as a form of collective digital discourse. A total of 855 tweets were collected through Twitter crawling, then preprocessed and converted into vector representations using the IndoBERT model. Subsequently, hierarchical clustering was applied using cosine distance and average linkage. The results reveal three major clusters that reflect various interpretations of the hashtag, ranging from emotional expressions to rational considerations and socio-political criticism. These findings highlight the role of social media as a space for articulating public unease and demonstrate the effectiveness of IndoBERT-based NLP approaches in uncovering latent discourse patterns in online conversations.

Keywords: IndoBERT, Twitter, hierarchical clustering, “#KaburAjaDulu”, public discourse

Abstrak: Fenomena “#KaburAjaDulu” di Twitter mencerminkan respons masyarakat terhadap tekanan sosial, ekonomi, dan politik yang dirasakan dalam kehidupan sehari-hari. Penelitian ini bertujuan untuk memetakan topik percakapan publik yang muncul melalui tagar tersebut sebagai bentuk wacana kolektif digital. Sebanyak 855 tweet dikumpulkan melalui proses crawling dari Twitter, kemudian diproses dengan tahapan text preprocessing dan dikonversi menjadi representasi vektor menggunakan model IndoBERT. Selanjutnya, dilakukan klasterisasi menggunakan metode Hierarchical Clustering berbasis cosine distance dan average linkage. Hasil penelitian menunjukkan tiga klaster utama yang mencerminkan variasi makna dalam penggunaan tagar, mulai dari ungkapan emosional, pertimbangan rasional, hingga kritik sosial-politik. Temuan ini menunjukkan bahwa media sosial berperan sebagai ruang artikulasi keresahan publik dan bahwa pendekatan NLP berbasis IndoBERT efektif dalam mengungkap pola wacana yang tersembunyi dalam percakapan daring.

Kata kunci: IndoBERT, Twitter, klasterisasi hirarki, “#KaburAjaDulu”, wacana publik

I. PENDAHULUAN

Media sosial telah menjadi sarana utama bagi masyarakat untuk mengekspresikan opini, emosi, dan reaksi terhadap berbagai isu sosial dan politik [1]. Twitter, sebagai salah satu platform populer, memungkinkan penyebaran informasi dan opini melalui penggunaan tagar (*hashtag*) yang dapat mencerminkan sentimen atau fenomena sosial tertentu [2]. Dalam konteks dinamika opini publik, *hashtag* “#KaburAjaDulu” muncul sebagai ekspresi kolektif pengguna yang ingin “menghindar” dari tekanan sosial atau situasi tidak menyenangkan, sehingga menarik untuk dianalisis [3]. Fenomena ini penting untuk dikaji karena mencerminkan reaksi masyarakat terhadap berbagai tekanan seperti kondisi politik, sosial, dan ekonomi yang memengaruhi kehidupan sehari-hari. Pemetaan topik percakapan melalui *hashtag* dapat mengungkap isu-isu dominan dan struktur diskursus public yang berkembang di media sosial [4].

Analisis topik pada teks berbahasa Indonesia membutuhkan model pemrosesan Bahasa alami yang dapat menangkap konteks dan makna kalimat dengan baik. IndoBERT adalah sebuah model berbasis *Transformer* yang dikembangkan khusus untuk Bahasa Indonesia yang mampu merepresentasikan makna semantic secara efektif sehingga cocok untuk tugas-tugas *NLP* seperti klasifikasi dan ekstraksi topik [5]. Untuk mengelompokkan percakapan ke dalam tema utama, metode klasterisasi hirarki menjadi pilihan yang efektif karena mampu menggambarkan hubungan bertingkat antar kelompok data [6].

Penelitian ini bertujuan untuk memetakan topik percakapan terkait fenomena “#KaburAjaDulu” di Twitter dengan menggunakan representasi teks dari IndoBERT dan Teknik klasterisasi hirarki,

sehingga memberikan pemahaman yang lebih komprehensif tentang dinamika opini public di media sosial Indonesia.

II. KAJIAN PUSTAKA

2.1. Penelitian Terkait

Beberapa penelitian telah dilakukan dalam analisis topik dan sentiment di media sosial. Raess dan Fazilat (2024) menerapkan metode berbasis leksikon untuk analisis sentiment pada data Twitter, menunjukkan bahwa pendekatan ini efektif dalam mengklasifikasikan teks ke dalam kategori positif, negative, dan netral, dengan evaluasi menggunakan berbagai model klasifikasi seperti *Naïve Bayes* dan *Random Forest* [7]. Sementara itu, Indahyanti et al. (2024) menggunakan metode *Latent Dirichlet Allocation* (LDA) untuk menganalisis topik dalam tweet terkait tragedi “Kanjuruhan”, berhasil mengidentifikasi lima topik utama yang mencerminkan reaksi public terhadap peristiwa tersebut [8].

2.2. Analisis Topik

Analisis topik bertujuan untuk mengidentifikasi tema-tema utama dalam kumpulan data teks. Dalam konteks media sosial, pendekatan ini membantu memahami isu-isu yang sedang tren dan opini public. Metode seperti *Latent Dirichlet Allocation* (LDA) sering digunakan untuk tujuan ini, memungkinkan identifikasi topik-topik tersembunyi dalam data teks yang besar [9].

2.3. Text Mining dan Preprocessing

Text mining adalah Teknik mengekstrak informasi berharga dari teks tidak terstruktur. Proses ini dimulai dengan *preprocessing* teks, *case folding* (mengubah huruf menjadi huruf kecil), *tokenizing* (memecah kalimat menjadi kata), *stopword removal* (menghapus kata-kata umum), *stemming* (mengubah kata ke bentuk dasarnya), dan juga *embedding* (mengubah teks menjadi vektor) [10].

2.4. Ekstraksi Fitur dengan IndoBERT

IndoBERT adalah model Bahasa berbasis *Transformer* yang dilatih khusus untuk Bahasa Indonesia. Model ini mampu menghasilkan representasi vektor dari teks yang mempertimbangkan konteks kata, sehingga cocok untuk tugas-tugas seperti klasifikasi teks dan analisis topik [11]. Dalam penelitian ini, IndoBERT digunakan untuk mengubah tweet menjadi representasi vektor sebelum dilakukan klusterisasi.

2.5. Klusterisasi Hirarki

Klusterisasi hirarki adalah metode pengelompokan data yang membentuk hierarki kluster. Metode ini dapat dibagi menjadi dua pendekatan yaitu *agglomerative* (penggabungan) dan *divisive* (pembagian). Dalam pendekatan *agglomerative*, setiap data awalnya dianggap sebagai kluster tersendiri, kemudian secara iteratif digabungkan berdasarkan kemiripan hingga membentuk satu kluster besar [12]. Rumus umum untuk menghitung jarak antar kluster dalam klusterisasi *agglomerative* adalah:

$$D(A, B) = \min_{a \in A, b \in B} d(a, b) \quad (1)$$

Dimana $D(A, B)$ adalah jarak antar kluster A dan kluster B , dan $d(a, b)$ adalah jarak antara elemen a dan elemen b dari masing-masing kluster.

Selain menggunakan metode *single linkage* pada rumus (1), pendekatan lain dalam klusterisasi hirarki adalah *average linkage*, yang menghitung jarak rata-rata antara semua pasangan elemen dari dua kluster berbeda. Pendekatan ini dinilai lebih stabil karena tidak terlalu dipengaruhi oleh *outlier* dibandingkan *single linkage* (menggunakan jarak minimum) atau *complete linkage* (menggunakan jarak maksimum [13]. Rumus *average linkage* adalah:

$$D(A, B) = \frac{1}{|A||B|} \sum_{i \in A} \sum_{j \in B} d(i, j) \quad (2)$$

Keterangan:

- $D(A, B)$: jarak rata-rata antara kluster A dan B;
- $d(i, j)$: jarak antara elemen $i \in A$ dan $j \in B$;
- $|A|$ dan $|B|$: jumlah elemen dalam kluster A dan B.

Pemilihan metode *linkage* sangat memengaruhi bentuk dendrogram yang dihasilkan dan dapat disesuaikan berdasarkan karakteristik data dan tujuan analisis [14].

2.6. Cosine Distance

Dalam pemrosesan teks seperti data twitter, representasi data umumnya dalam bentuk vektor berdimensi tinggi. Untuk mengukur kemiripan antara dua dokumen atau tweet, *cosine distance* sering digunakan karena mampu menangkap perbedaan arah berbasis teks seperti topik percakapan Twitter (X) karena tidak terpengaruh oleh panjang dokumen [15]. Berikut adalah rumus umum *cosine similarity*:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} \quad (3)$$

Dan untuk menghitung *cosine distane*, digunakan rumus:

$$\text{cosine Distane} = 1 - \cos(\theta) \quad (4)$$

Keterangan:

- A dan B : dua vektor teks;
- $A \cdot B$: perkalian antara vektor A dan B;
- $\|A\|$: panjang dari vektor A;
- $\|B\|$: panjang dari vektor B.

Cosine distance bernilai antara 0 (sangat mirip) hingga 1 (tidak mirip), dan banyak digunakan dalam metode klusterisasi teks seperti hierarki *agglomerative* berbasis TF-IDF atau *embeddings* dari BERT [16].

III. METODE PENELITIAN

3.1. Sumber Data dan Variabel Penelitian

Sumber data dalam penelitian ini berasal dari platform media sosial Twitter (X), dengan fokus pada tweet yang mengandung tagar “#KaburAjaDulu” sebagai topik utama. Data dikumpulkan melalui proses *crawling* menggunakan *XAPI* dengan rentang waktu pengambilan data mulai dari 11 maret 2025 hingga 17 maret 2025. Jumlah data yang berhasil dikumpulkan dan digunakan dalam penelitian ini sebanyak 855 tweet. Variable utama dalam penelitian ini adalah teks tweet yang telah melalui tahapan *preprocessing*, termasuk translasi, pembersihan teks, dan normalisasi. Variabel ini kemudian digunakan untuk menghasilkan representasi vektor berbasis *contextual embedding* menggunakan model IndoBERT. Selain itu, variabel hasil klusterisasi berupa label kluster juga menjadi bagian dari analisis yang digunakan untuk mengelompokkan tweet berdasarkan kesamaan makna kontekstual.

3.2. Tahapan Penelitian



Adapun alur tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut:



Gambar 1. Flowchart penelitian

Berikut adalah penjelasan dari alur penelitian di atas:

- Studi literatur untuk memperoleh pemahaman terkait metode klasterisasi hierarki dan terkait IndoBERT;
- Pengumpulan data, dimana data dikumpulkan dari media sosial Twitter (X) menggunakan metode *crawling* berdasarkan kata kunci dan rentang waktu tertentu;
- Preprocessing* teks, dengan beberapa tahapan, yaitu *case folding*, normalisasi *slang dictionary*, *tokenizing*, *stopword removal*, dan juga *stemming*;
- Ekstraksi Fitur dengan IndoBERT, tweet yang telah dibersihkan dikonversi menjadi *embedding* vektor menggunakan IndoBERT (indobenchmark/indobert-base-p1);
- Klasterisasi dilakukan menggunakan algoritma Hierarchical Clustering, di mana proses dimulai dengan menghitung embedding dari setiap tweet menggunakan model IndoBERT. Embedding ini berupa representasi vektor dari tiap tweet di ruang berdimensi tinggi. Kemudian, untuk mengukur kesamaan antar tweet, digunakan jarak kosinus (*cosine distance*), yaitu $1 - \cos(\theta)$, dengan θ sebagai sudut antara dua vektor embedding. Setelah seluruh pasangan jarak antar tweet dihitung, algoritma *Hierarchical Clustering* dimulai dengan menganggap setiap tweet sebagai satu kluster tersendiri. Kluster-kluster tersebut lalu digabungkan secara bertahap berdasarkan jarak antar kluster yang paling kecil, mengikuti metode linkage *average*. Proses penggabungan ini divisualisasikan dalam bentuk dendogram, yaitu diagram pohon yang menunjukkan urutan penggabungan antar kluster;
- Visualisasi dan interpretasi, kluster ditampilkan dalam bentuk dendogram dan *WordCloud*, disertai analisis naratif untung masing-masing kluster berdasarkan tweet di dalamnya.

IV. HASIL DAN DISKUSI

4.1. Hasil Pengumpulan Data

Data dalam penelitian ini diperoleh melalui proses *crawling* menggunakan *Tweet Harvester* dengan autentikasi token (*auth token*) dari Twitter API v2. Focus pengambilan data adalah pada tweet yang mengandung *hashtag* “#KaburAjaDulu”. Periode data dimulai pada tanggal 11 maret 2025 hingga 17 maret 2025. Data yang terkumpul kemudian disimpan dalam format CSV untuk dianalisis lebih lanjut. Total jumlah tweet yang berhasil dikumpulkan adalah 855 tweet.

4.2. *Preprocessing Text*

Tahapan *preprocessing text* dilakukan untuk mengubah data tersebut ke dalam format yang lebih terstruktur. Berikut adalah tahapan *preprocessing text* pada penelitian ini:

- Case folding*
Pada tahap ini, semua huruf yang ada pada data diubah menjadi huruf kecil atau *lowercase*.
- Cleansing*
Proses *cleaning* dilakukan untuk menghapus karakter yang tidak relevan seperti tanda baca, angka, link, emoji, dan simbol lainnya.

- c. Normalisasi dengan *slang dictionary*
Pada tahap ini, kata-kata tidak baku atau *slang* yang umum digunakan di media sosial diganti dengan kata baku dengan menggunakan kamus *slang*.
- d. *Tokenizing*
Pada tahap ini, kalimat akan dipecah menjadi *unit-unit* kata (token).
- e. *Stopword removal*
Stopword removal digunakan untuk menghapus kata umum yang tidak penting dalam analisis, seperti kata penghubung “dan”, “yang”, dan lainnya.
- f. *Embedding*
Pada tahapan ini, teks diubah menjadi vektor.

Tabel 1. Hasil *preprocessing text*

Text Input	Clean text	Tokens	Stemmed	Stopword	Stemmed_nostop	Embedding
kalau BEI Divisi Komunikasin ya atau Mar...	kalau bursa efek indonesia divisi komunikasiny ...	[kalau, bursa, efek, indonesi a, divisi, komuni ...]	kalau bursa efek indonesia divisi komunika si b...	[bursa, efek, divisi, komunika si, bagus, pasar...]	bursa efek divisi komunikasi bagus pasar sekur...	[0.0574868 2, 0.9045462 6, 0.3...]

4.3. Ekstraksi Fitur menggunakan IndoBERT

Setelah data dibersihkan, dilakukan proses ekstraksi fitur menggunakan model IndoBERT. Setiap tweet diubah menjadi representasi vektor menggunakan metode *mean pooling* terhadap keluaran dari model. Representasi vektor ini memberikan pemetaan semantic yang lebih dalam dan kontekstual terhadap isi dari setiap tweet, sehingga memudahkan dalam proses klasterisasi. Berikut adalah sampel *output* dari hasil *embedding*

Tabel 2. *Output Embedding*

Sampel <i>Output</i> Hasil <i>Embedding</i>			
0.50506192445755,	0.9766284823417664,	0.37497028708457947,	0.22300629317760468,
0.34056174755096436,		
0.783271849155426,	0.1521611511707306,	0.9791715145111084,	0.37094253301620483,
1.0465178489685059,		
0.514991283416748,	1.0732635259628296,	0.6891756057739258,	-0.21098501980304718,
0.10105781257152557,		
0.783271849155426,	0.1521611511707306,	0.9791715145111084,	0.37094253301620483,
1.0465178489685059,		
0.28178778290748596,	0.8233219981193542,	0.3325433135032654,	0.2516762316226959,
0.9397497177124023,	0.22423668205738068,	

4.4. Klasterisasi

Tahapan klasterisasi dilakukan menggunakan metode *Hierarchical Clustering* dengan teknik penggabungan *average linkage* dan metrik jarak *cosine distance*. Input untuk proses ini berupa representasi vektor dari setiap tweet, yang dihasilkan melalui ekstraksi fitur menggunakan model IndoBERT dengan teknik *mean pooling*. Dengan pendekatan ini, setiap tweet direpresentasikan dalam bentuk vektor berdimensi tetap (768 dimensi).

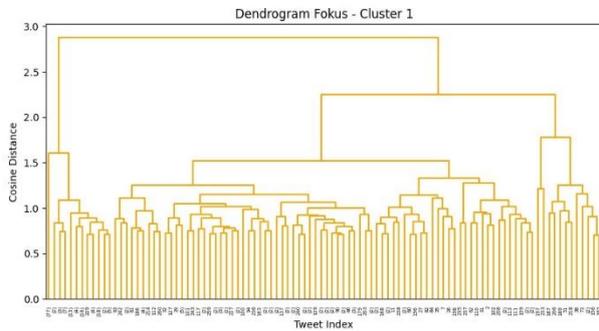


Setelah setiap *tweet* direpresentasikan sebagai vektor numerik melalui ekstraksi fitur menggunakan model IndoBERT, dihitunglah matriks jarak cosine untuk mengukur tingkat kemiripan semantik antar *tweet*. *Cosine distance* dipilih karena mampu menangkap perbedaan arah antar vektor teks tanpa dipengaruhi besar nilainya, dengan nilai berkisar antara 0 (sangat mirip) hingga 1 (sangat berbeda). Matriks ini digunakan sebagai dasar dalam metode klasterisasi hierarki dengan teknik average linkage, di mana dua klaster digabung berdasarkan rata-rata jarak antar anggotanya. Proses ini berlangsung secara iteratif hingga semua *tweet* tergabung dalam struktur pohon hierarkis, yang kemudian divisualisasikan menggunakan dendrogram untuk menunjukkan pola kemiripan antar kelompok *tweet*. Berikut adalah contoh output dari penggabungan *average linkage* dan metrik jarak *cosine distance*.

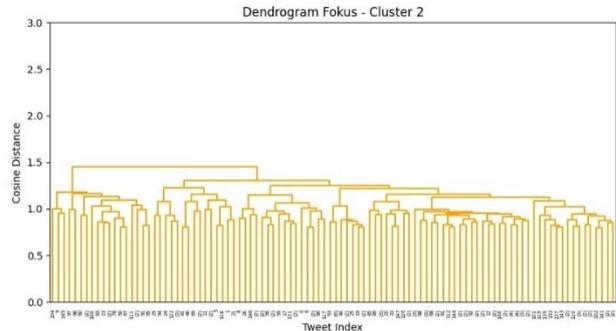
Tabel 3. Table contoh output penggabungan average linkage dan metrik jarak cosine distance.

no	0	1	2	3
1	12.0	13.0	1.6858739404357614e-07	2.0
2	159.0	160.0	3.371747880871523e-07	2.0
3	5	161.0	435.0	3.371747880871523e-07
4	6	162.0	436.0	3.371747880871523e-07
5	7	11.0	434.0	3.816559456248789e-07

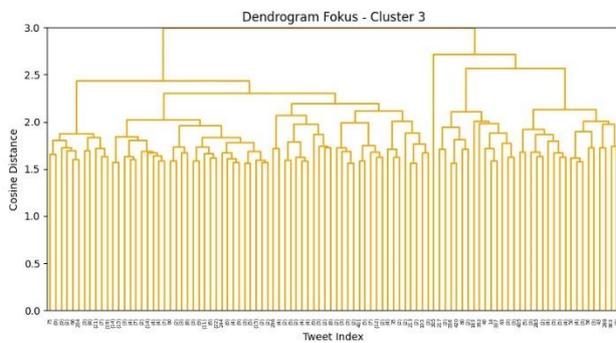
Tabel diatas menunjukkan hasil perhitungan jarak antar representasi vektor teks yang diperoleh melalui perhitungan cosine distance dan metode pengelompokan linkage average. Kolom pertama dan kedua merepresentasikan indeks pasangan data yang dibandingkan, kolom ketiga merupakan nilai jarak kosinus antar vektor, dan kolom keempat menunjukkan label klaster yang terbentuk. Nilai jarak yang sangat kecil (mendekati nol) mengindikasikan bahwa pasangan teks tersebut memiliki kemiripan konteks yang tinggi dan cenderung tergabung dalam klaster yang sama. Untuk memperoleh jumlah klaster yang optimal dan memahami struktur hierarki dari hasil klasterisasi, visualisasi dendrogram dan analisis menggunakan elbow method dilakukan.



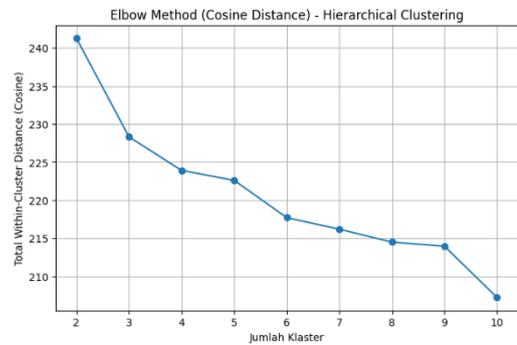
Gambar 2a. Dendrogram Hierarki – *Cosine Distance cluster 1*



Gambar 2b. Dendrogram Hierarki – *Cosine Distance cluster 2*



Gambar 3c. Dendrogram Hierarki – *Cosine Distance cluster 3*

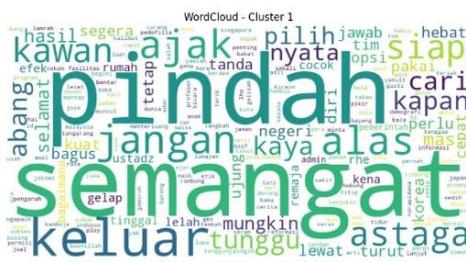


Gambar 2d. Elbow method – *Hierarchical Clustering 3*

Berdasarkan eksplorasi visual melalui dendrogram (Gambar 2a–2c) serta pengujian jumlah kluster yang optimal menggunakan elbow method (Gambar 2d), ditentukan bahwa data akan dikelompokkan ke dalam 3 kluster. Penentuan ini didasarkan pada titik siku (elbow) pada grafik, yang menunjukkan penurunan nilai total within-cluster distance secara signifikan hingga jumlah kluster ke-3. Tiap kluster yang terbentuk merepresentasikan kelompok topik atau sentimen yang berbeda berdasarkan kemiripan semantik antar tweet.

4.5. Visualisasi dan Interpretasi

Untuk memahami distribusi kata dan pola tematik dalam setiap kelompok hasil klusterisasi, dilakukan visualisasi menggunakan word cloud dan tabel data acak untuk memahami karakteristik setiap kluster. Word cloud merupakan teknik visualisasi teks yang menampilkan kata-kata paling sering muncul dalam bentuk grafis, di mana ukuran huruf mencerminkan frekuensi atau bobot relatif sebuah kata. Visualisasi dan tabel ini membantu mengidentifikasi kata-kata kunci dan tema dominan secara cepat dan intuitif. Berikut merupakan word cloud dan tabel kluster yang dihasilkan untuk masing-masing kluster:





Tabel 2. Tabel data *cluster 1*

Full Text	Clean Text
Lihat negara lain makmur kaya gini apa ga makin pengen #KaburAjaDulu	lihat negara lain makmur kaya gini apa tidak makin ingin
@Key392305 Oh ya? Mau #KaburAjaDulu ke Jepang? Wkwk	oh ya ingin ke jepang tertawa
Semakin memantapkan diri untuk #KaburAjaDulu	semakin meyakinkan diri untuk
@ridwanhr Pindah ke Malaysia aja yuk #KaburAjaDulu #AyoKeMalaysia	pindah ke malaysia saja yuk
pengen ikutan #KaburAjaDulu soon ya	ingin ikut segera ya

Tabel di atas menampilkan beberapa contoh kalimat yang mewakili narasi dominan dalam kluster 1 yang terdiri dari 275 tweet. Tweet-tweet dalam kluster ini mencerminkan keinginan kuat untuk meninggalkan negara dan mencari kehidupan yang dianggap lebih layak di luar negeri. Narasi yang muncul cenderung bersifat personal dan reflektif, dengan nada yang campuran antara harapan dan keputusasaan. Banyak tweet membandingkan kondisi negara sendiri dengan negara lain yang lebih maju, mengungkapkan kekaguman, dan sekaligus kekecewaan terhadap situasi di dalam negeri. Ajakan seperti “pindah ke Malaysia”, “ingin ikut segera”, atau “oh ya ingin ke Jepang” menjadi gambaran umum dari aspirasi untuk “kabur” demi mencari masa depan yang lebih baik. Kluster ini memperlihatkan bagaimana persepsi terhadap kemajuan negara lain bisa mendorong keinginan untuk migrasi sebagai bentuk protes diam terhadap keadaan yang ada.



Gambar 4. Cluster 2

Tabel 3. Tabel data *cluster 2*

Full Text	Clean Text
pengen #KaburAjaDulu tapi aku masih pengangguran di negara carut marut yg penuh korupsi ini	ingin tapi aku masih pengangguran di negara kacau acak acakan yang penuh korupsi ini
tunjukkan lah aku jalan yg lurus dan tunjukkanlah aku jalan untuk #kaburajadulu juga yaAllah	tunjukkan ya aku jalan yang lurus dan tunjukkanlah aku jalan untuk juga ya allah
@333salsaa Ya sebenarnya gak sia-sia banget karena akhirnya dia tau realita dunia kerja di negara kita yang sekarang mending emang #KaburAjaDulu	ya sebenarnya tidak sia sia sangat karena akhirnya dia tahu kenyataan dunia kerja di negara kita yang sekarang lebih baik memang
negara ini makin lama makin bobrok anjir. Dari #KaburAjaDulu jadi #KaburAjaYuk yukk	negara ini makin lama makin bobrok astaga dari jadi ayo



@somexthread Mau ngemis malu mau korupsi ingin mengemis malu ingin korupsi bukan bukan pejabat. #KaburAjaDulu lah pejabat ya

Tabel di atas menampilkan beberapa contoh kalimat yang mewakili narasi dominan dalam kluster 2 yang terdiri dari 149 tweet. Tweet-tweet dalam kluster ini mengekspresikan kekecewaan dan frustrasi terhadap kondisi sosial-politik di Indonesia, terutama terkait isu korupsi, kepemimpinan, dan kesenjangan sosial. Narasi yang muncul cenderung kritis dan emosional, dengan penekanan pada ketidakadilan serta ketidakpercayaan terhadap institusi negara. Banyak pengguna menggunakan tagar #KaburAjaDulu sebagai bentuk sindiran atau pelampiasan keputusan terhadap situasi yang dianggap semakin memburuk. Bahasa yang digunakan dalam tweet sering kali menyoroti persoalan pengangguran, sulitnya mencari pekerjaan, serta keresahan generasi muda yang merasa terjebak dalam sistem yang tidak berpihak pada mereka.



Gambar 4. Cluster 3

Tabel 4. Tabel data cluster 3

Full Text	Clean Text
Lapangan kerja di NKRI harga Mati ini semakin tak jelas dan kian suram [hampir gelap] CPNS pun	lapangan kerja di nkri harga mati ini semakin tak jelas dan kian suram hampir gelap cpns juga ditunda dan Pegawai Pemerintah dengan Perjanjian Kerja hingga tahun ke depan phk massal serentak menjelang hari raya idul fitri perasaan siapa yang tak akan sedih melihat fakta hidup kian kelabu wajarlah
@CNNIndonesia Terserah lah bikin aturan sendiri aja tergantung pemimpin nya... Biar tambah kacau negara ini. apalagi PHK dimana-mana investor kabur penerimaan negara anjlok rupiah tertekan. #KaburAjaDulu	terserah ya bikin aturan sendiri saja tergantung pemimpin nya biar tambah berantakan negara ini apalagi phk di mana mana investor kabur penerimaan negara turun drastis rupiah tertekan
@pasyaaza indonesia makin ga relevan buat banyak orang. politik kotor polisi ga bisa dipercaya korupsi merajalela infrastruktur & kesejahteraan ga berkembang. sementara negara lain udh lebih stabil & nyaman buat ditinggalin. wajar aja banyak yg lebih milih cabut & pake tag #kaburajadulu	indonesia makin tidak relevan untuk banyak orang politik kotor polisi tidak bisa dipercaya korupsi merajalela infrastruktur dan kesejahteraan tidak berkembang sementara negara lain sudah lebih stabil dan nyaman untuk ditinggalin wajar saja banyak yang lebih memilih pergi dan pakai menandai
@mas_veel Tau gak China banyak buka full beasiswa kpd mahasiswa baru yg pintar. Dikasih uang saku dan penginapan. Yg pintar pd #kaburajadulu bljr gratis. Banyak yg tidak mau pulang krn di fasilitasi bag R&D. Krn di sini ga dpt. Malah di sini suruh pinjol byr UKT. PIP BOS aja di korup.	serius aku tu aslinya hanya bertanya tapi dianggap jualan sampai dititipin ya tidak apa apa jasa titip takjil tidak ini kakak bagaimana y buka puasa pakai omongan manis membatalkan enggak takutnya bukan puasanya yang batal tapi hubungannya



@NESVERLAND Dah gaji minimum nyari kerja nya susah kerja nya juga pada gak ngotak ditambah kelakuan pejabatnya pada maling semua udah paling bener #KaburAjaDulu harga mati

sudah tahu sekarang itu lg meningkatnya saya saja gemetar gelisah tingkat tidak kebagian gara gara itu makanya sudah tinggal terus di situs web nya mempersiapkan pas foto di galeri jadi urutan pertama

Tabel di atas menyajikan beberapa kalimat representatif dari Klaster 3, yang merupakan klaster terbesar dengan 431 tweet. Klaster ini merupakan yang terbesar, mencerminkan diskursus kritis terhadap berbagai institusi dan sistem di Indonesia, seperti pemerintahan, ekonomi, dan layanan publik. Tweet-tweet di dalamnya menyuarkan keluhan yang lebih konkret dan terarah, mulai dari rendahnya gaji, ketidakpastian lapangan kerja, hingga kualitas pendidikan bagi anak-anak. Penggunaan narasi "kabur" dalam konteks ini bukan sekadar wacana emosional, tetapi dilandasi oleh pertimbangan rasional dan pengalaman hidup sehari-hari. Klaster ini memperlihatkan bagaimana keresahan masyarakat terakumulasi menjadi argumen logis bahwa berpindah negara dianggap sebagai solusi untuk mencapai stabilitas hidup yang lebih baik.

Visualisasi word cloud pada masing-masing klaster memperlihatkan adanya perbedaan nuansa tematik dalam percakapan seputar "#KaburAjaDulu". Klaster pertama didominasi oleh kata-kata bernada optimis dan ajakan kolektif, menunjukkan adanya dorongan emosional untuk "keluar" dari situasi yang dianggap tidak ideal. Klaster kedua cenderung merepresentasikan pertimbangan rasional dan perencanaan yang lebih matang mengenai tujuan atau arah perpindahan. Sementara itu, klaster ketiga memuat narasi yang lebih kritis dan reflektif, memperlihatkan kekecewaan terhadap situasi sosial-politik dan kecenderungan membandingkan dengan kondisi luar negeri.

V. KESIMPULAN

Penelitian ini menganalisis percakapan di media sosial Twitter yang menggunakan tagar "#KaburAjaDulu" untuk memahami dinamika wacana publik terkait keresahan sosial, aspirasi migrasi, serta respons masyarakat terhadap tekanan kehidupan. Dengan menggunakan pendekatan Hierarchical Clustering berbasis cosine distance terhadap representasi vektor teks hasil ekstraksi dari model IndoBERT, sebanyak 855 tweet berhasil dikelompokkan ke dalam tiga klaster utama yang menunjukkan kemiripan semantik antar percakapan.

Hasil klasterisasi menunjukkan bahwa masing-masing klaster merepresentasikan dimensi narasi yang berbeda namun saling berkaitan. Klaster 1 (275 tweet) didominasi oleh keinginan pribadi untuk mencari kehidupan yang lebih baik di luar negeri. Tweet-tweet dalam klaster ini banyak mengekspresikan harapan, kekaguman terhadap negara lain, serta rasa kecewa terhadap kondisi dalam negeri, yang membuat migrasi tampak sebagai pilihan rasional dan emosional. Klaster 2 (149 tweet) mencerminkan kekecewaan mendalam terhadap kondisi sosial-politik di Indonesia, dengan banyak ekspresi frustrasi terhadap isu seperti korupsi, kepemimpinan, dan kesenjangan sosial. Di sini, frasa "kabur aja dulu" digunakan sebagai bentuk sindiran terhadap situasi yang dinilai tidak adil. Klaster 3 (431 tweet) menampilkan diskursus yang lebih terstruktur dan logis terkait kritik terhadap sistem pemerintahan, ekonomi, dan layanan publik. Narasi dalam klaster ini muncul dari pengalaman konkret, seperti rendahnya pendapatan atau buruknya akses pendidikan, yang mendorong pemikiran rasional untuk mencari stabilitas hidup di tempat lain.

Secara keseluruhan, penggunaan frasa "kabur aja dulu" bukan sekadar ekspresi spontan atau lelucon semata, tetapi merefleksikan spektrum emosi yang luas: dari dorongan emosional untuk menjauh, pertimbangan rasional untuk migrasi, hingga kritik terhadap struktur sosial-politik. Visualisasi melalui word cloud turut memperkuat identifikasi tema dominan pada tiap klaster dan memperjelas pola narasi publik.

Temuan ini menegaskan bahwa media sosial merupakan ruang penting dalam artikulasi opini publik dan bentuk partisipasi digital masyarakat. Kombinasi model IndoBERT dengan metode klasterisasi hirarki terbukti efektif dalam memetakan wacana kolektif secara lebih kontekstual dan mendalam. Ke depan, pendekatan ini dapat dikembangkan lebih lanjut dengan menambahkan analisis sentimen untuk memperkaya pemahaman terhadap intensi dan sikap pengguna.



VI. REFERENSI

- [1] J. H. Kietzmann, K. Hermkens, I. P. McCarthy, and B. S. Silvestre, “Social media? Get serious! Understanding the functional building blocks of social media,” *Bus Horiz*, vol. 54, no. 3, pp. 241–251, May 2011, doi: 10.1016/j.bushor.2011.01.005.
- [2] A. Bruns and J. Burgess, “Twitter Hashtags from Ad Hoc to Calculated Publics,” *Soc Media Soc*, vol. 1, no. 2, Aug. 2015, [Online]. Available: <http://eprints.qut.edu.au/46515/>.
- [3] N. Bahar and R. Latief, “Cyberactivism Sebagai Budaya...(Nurfadhilah,Rahmawati) CYBERACTIVISM SEBAGAI BUDAYA POPULER: ANALISIS WACANA HASHTAG #PERCUMALAPORPOLISI DI TWITTER”.
- [4] D. Putri Sitompul, Y. Sitorus, E. Gracia Br Sibuea, and S. Destemi Elsi, “PERAN MEDIA SOSIAL DALAM MEMPENGARUHI PERILAKU PEMILIH PEMULA,” *Journal of Law, Administration, and Social Science*, vol. 4, no. 5, 2024.
- [5] F. Koto, J. H. Lau, and T. Baldwin, “IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization,” *Soc Media Soc*, vol. 1, no. 2, Sep. 2021, [Online]. Available: <http://arxiv.org/abs/2109.04607>
- [6] C. C. Sujadi, Y. Sibaroni, and A. F. Ihsan, “Analysis Content Type and Emotion of the Presidential Election Users Tweets using Agglomerative Hierarchical Clustering,” *Sinkron : Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 3, pp. 1230–1237, Jul. 2023, doi: 10.33395/sinkron.v8i3.12616.
- [7] M. Raees and S. Fazilat, “Lexicon-Based Sentiment Analysis on Text Polarities with Evaluation of Classification Models,” Sep. 2024, [Online]. Available: <http://arxiv.org/abs/2409.12840>
- [8] U. Indahyanti, Y. Findawati, A. Ariansyah, and E. Asmawati, “Twitter Topic Modelling Using Latent Dirichlet Allocation Approach,” *CENTRAL ASIAN JOURNAL OF THEORETICAL AND APPLIED SCIENCES*, vol. 4, no. 9, Sep. 2023, [Online]. Available: <https://cajotas.centralasianstudies.org>
- [9] B. Helmi Puspita, M. Muhajir, and H. Aliady, “Topic Modeling Using Latent Dirichlet Allocation (LDA) and Sentiment Analysis for Marketing Planning Tiket.com Kilometers 2 PT. Global Tiket Network (Tiket.com),” in *Proceedings of the 2nd International Seminar on Science and Technology (ISSTEC 2019)*, ATLANTIS PRESS, 2019. [Online]. Available: <https://t.co/Zdjh2A0WLj>
- [10] A. Cahya Kamilla *et al.*, “ANALISIS SENTIMEN FILM AGAK LAEN DENGAN KECERDASAN BUATAN: TEXT MINING METODE NAÏVE BAYES CLASSIFIER,” *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 3, pp. 2923–2928, Jun. 2024.



- [11] Y. A. Singgalen, “Performance Analysis of IndoBERT for Sentiment Classification in Indonesian Hotel Review Data,” *Journal of Information System Research*, vol. 6, no. 2, pp. 976–986, Jan. 2025, doi: 10.47065/josh.v6i2.6505.
- [12] A. Krisna Suyono, M. Nor Hayati, M. Siringoringo, S. Prangga, and M. Fathurahman, “Penerapan Algoritma Divisive Analysis dalam Pengelompokan Provinsi di Indonesia Berdasarkan Prevalensi Stunting Application of the Divisive Analysis Algorithm in Clustering Provinces in Indonesia Based on Stunting Prevalence,” *Jurnal EKSPONENSIAL*, vol. 15, no. 2, Nov. 2024, doi: 10.30872/eksponensial.v15i2.1341.
- [13] S. Indah Pratiwi, T. Widiharih, and A. Rachman Hakim, “ANALISIS KLASSTER METODE WARD DAN AVERAGE LINKAGE DENGAN VALIDASI DUNN INDEX DAN KOEFISIEN KORELASI COPHENETIC (Studi Kasus: Kecelakaan Lalu Lintas Berdasarkan Jenis Kendaraan Tiap Kabupaten/Kota di Jawa Tengah Tahun 2018),” *JURNAL GAUSSIAN*, vol. 8, no. 4, pp. 486–495, 2019, Accessed: Jun. 09, 2025. [Online]. Available: <http://ejournal3.undip.ac.id/index.php/gaussian>
- [14] D. Widyadhana, R. B. Hastuti, I. Kharisudin, and F. Fauzi, “Perbandingan Analisis Klaster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah,” *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 4, pp. 584–594, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [15] B. M. Kumarahadi, H. Pratiwi, and S. Subanti, “Penerapan Metode Hierarchical Clustering Untuk Pengelompokan Kota/Kabupaten di Indonesia Berdasarkan Indikator Kemiskinan,” *Jurnal Teknologi Informasi dan Komunikasi (TIKOMSiN)*, vol. 11, no. 2, p. 13, Oct. 2023, doi: 10.30646/tikomsin.v11i2.754.
- [16] A. F. Dewi and K. Ahadiyah, “Agglomerative Hierarchy Clustering Pada Penentuan Kelompok Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Pendidikan,” *Zeta - Math Journal*, vol. 7, no. 2, pp. 57–63, Nov. 2022, doi: 10.31102/zeta.2022.7.2.57-63.