

## Community Detection, Topic, and Sentiment Analysis of the Palestine-Israel Issue on Social Media X

### Deteksi Komunitas, Analisis Topik, dan Sentimen Isu Palestina-Israel pada Media Sosial X

Fathnin Nur Azmina\*<sup>1</sup>, Muhaza Liebenlito<sup>2</sup>, Dhea Urfina Zulkifli<sup>3</sup>

<sup>1,2,3</sup>*Department of Mathematics, Faculty of Science and Technology, Syarif Hidayatullah State Islamic University Jakarta*

*Email: <sup>1</sup>fathninnurazmina@gmail.com, <sup>2</sup>muhazaliebenlito@uinjkt.ac.id, <sup>3</sup>dhea.urfina@uinjkt.ac.id*

*\*Corresponding author*

*Received: 8 October 2024, revised: 2 December 2024, accepted: 10 December 2024*

#### Abstract

The combination of community detection, topic modeling, and sentiment analysis provides deep insights into conversation data on the social media platform X (formerly Twitter) regarding the Palestine-Israel issue. The data, collected in Indonesian using several keywords, resulted in 108,969 tweets. The analysis process began with community detection using the Leiden algorithm, which identified five communities. The three dominant communities identified are Community 1 comprising 37.13% of users, Community 2 with 26.95%, and Community 3 with 19.76%. Topic modeling using LDA revealed that these communities focused on various aspects of the conflict. Sentiment analysis using the IndoBERT model uncovered that the majority of users expressed negative attitudes such as disappointment and anger. This study provides insights into public opinions and social dynamics surrounding the conflict.

**Keywords:** Community Detection, Leiden, LDA, IndoBERT, Palestine-Israel

#### Abstrak

Kombinasi deteksi komunitas, pemodelan topik, dan analisis sentimen memberikan wawasan mendalam terhadap data percakapan di media sosial X (sebelumnya Twitter) terkait isu Palestina-Israel. Data yang dikumpulkan dalam bahasa Indonesia menggunakan beberapa kata kunci menghasilkan 108.969 cuitan. Proses analisis dimulai dengan deteksi komunitas menggunakan algoritma Leiden, yang mengidentifikasi lima komunitas. Tiga komunitas dominan yang ditemukan adalah komunitas 1 dengan 37,13% pengguna, komunitas 2 dengan 26,95%, dan komunitas 3 dengan 19,76%. Pemodelan topik menggunakan LDA menunjukkan komunitas-komunitas ini fokus pada berbagai aspek konflik. Analisis sentimen menggunakan model IndoBERT mengungkapkan mayoritas pengguna memiliki sikap negatif seperti kekecewaan dan kemarahan. Penelitian ini memberikan wawasan tentang pandangan publik dan dinamika sosial terkait konflik tersebut.

**Kata kunci:** Deteksi Komunitas, Leiden, LDA, IndoBERT, Palestina-Israel



## 1. PENDAHULUAN

Menurut data *Palestinian Central Bureau of Statistics* (PCBS), tercatat lebih dari 36 ribu korban tewas di Palestina dan 86 ribu lainnya terluka sejak konflik Palestina-Israel yang dimulai pada 7 Oktober 2023<sup>1</sup>. Konflik ini menarik perhatian dunia dan memicu berbagai diskusi di media sosial, yang dapat memberikan wawasan tentang pandangan publik dan dinamika sosial terkait isu tersebut. Di media sosial, terutama di aplikasi X (sebelumnya dikenal Twitter), konflik Palestina-Israel menjadi topik yang ramai dibahas. Tagar '#FreePalestine' kembali *trending* dunia dengan lebih dari 221.000 sebutan, bersama dengan banyak kata kunci lain terkait isu tersebut<sup>23</sup>.

*Social Network Analysis* (SNA) adalah cabang ilmu komputer yang meneliti pola dan hubungan dalam jaringan sosial [10]. Jaringan sosial menggambarkan struktur yang menghubungkan individu atau entitas melalui berbagai jenis relasi. Interaksi antar pengguna di aplikasi X, misalnya melalui *mention*, merupakan salah satu contoh jaringan sosial. Dalam konteks ini, SNA menjadi alat yang sangat penting untuk menganalisis interaksi, mengidentifikasi komunitas, serta memahami penyebaran informasi dan opini [1]. Pengguna yang terhubung melalui *mention* dapat membentuk komunitas dengan topik perbincangan yang sering kali serupa [2]. Untuk mengevaluasi kesesuaian topik dengan suatu isu, pemodelan topik dapat digunakan, sementara analisis sentimen dapat mengungkap sikap atau sentimen dalam cuitan. Kombinasi deteksi komunitas, pemodelan topik, dan analisis sentimen menjadi tantangan utama dalam penelitian ini, terutama karena melibatkan data skala besar terkait isu Palestina-Israel di aplikasi X.

Berbagai metode dapat digunakan untuk mendeteksi komunitas dalam jaringan sosial, salah satunya adalah algoritma Louvain, yang dikenal efektif dalam mendeteksi komunitas pada jaringan sosial berskala besar. Namun, algoritma ini memiliki kelemahan, yaitu potensi terputusnya hubungan antar komunitas jika dijalankan berulang kali. Untuk mengatasi masalah ini, algoritma Leiden dikembangkan sebagai penyempurnaan dari Louvain, menawarkan deteksi komunitas yang lebih stabil dan akurat [17]. Beberapa penelitian yang membandingkan Leiden dan Louvain telah dilakukan dan hasilnya menunjukkan keunggulan Leiden. Penelitian-penelitian tersebut mengatakan bahwa Leiden lebih stabil, lebih cepat, dan mampu menghasilkan komunitas yang lebih berkualitas [7], [16]. Selain itu, algoritma Leiden memiliki keunggulan dalam metrik modularitas dan waktu berjalan, serta dapat berfungsi dengan baik pada berbagai ukuran jaringan, baik kecil, sedang, maupun besar [6], [7]. Berdasarkan keunggulan-keunggulan tersebut, penelitian ini memilih algoritma Leiden untuk deteksi komunitas.

*Latent Dirichlet Allocation* (LDA) adalah metode umum untuk mengidentifikasi topik laten dalam kumpulan dokumen dan banyak digunakan karena efektif, mudah dipahami, serta dapat diterapkan dalam berbagai bidang. LDA telah terbukti efektif dalam menganalisis data dari aplikasi X. Misalnya, dalam penelitian E. Sanandres (2020), metode ini berhasil menghasilkan 12 topik dari 45.000 cuitan [14]. Patmawati (2021) menggunakan LDA untuk menganalisis topik politik dan berhasil mengungkap fokus diskusinya [11]. Selain itu, F. Rasyif (2021) berhasil mengidentifikasi topik diskusi utama mengenai Covid-19 menggunakan LDA [12]. T. D. Dikiyanti (2021) juga menerapkan LDA pada data Twitter terkait "BPJS Kesehatan" dan menemukan 5 topik dengan nilai koherensi mencapai 0,82% [4]. Studi-studi ini menyoroti fleksibilitas LDA dalam memproses data Twitter dalam jumlah besar dan efektivitasnya dalam mengungkap topik yang bermakna di berbagai domain, menjadikannya alat yang berharga untuk analisis media sosial. Berdasarkan bukti ini, LDA dipilih sebagai metode pemodelan topik dalam penelitian ini.

Banyak penelitian telah diterapkan untuk analisis sentimen terhadap teks-teks informal, seperti cuitan dari aplikasi media sosial Twitter. *Bidirectional Encoder Representations from Transformers*

<sup>1</sup> Lihat lebih lanjut di *Palestinian Central Bureau of Statistics* (PCBS): <https://www.pcbs.gov.ps/Default.aspx>, diakses pada 14 Juni 2024.

<sup>2</sup> Lihat lebih lanjut di *Twitter Trending Archive*: <https://archive.twitter-trending.com/>, diakses pada 14 Juli 2024.

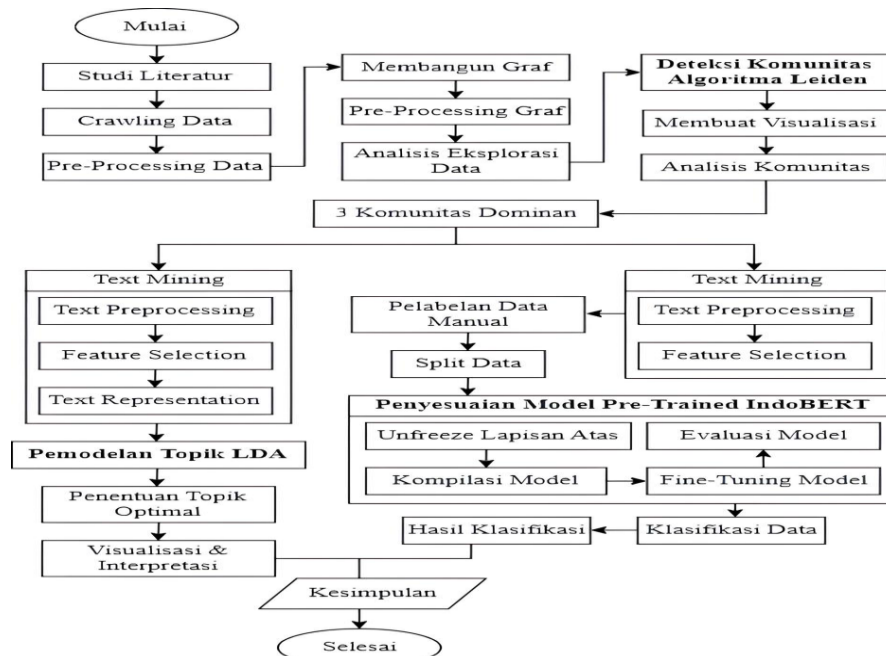
<sup>3</sup> D diambil dari Detikinet: <https://inet.detik.com/cyberlife/d-6980568/seruan-freepalestine-menggema-di-twitter-x-dunia>, diakses pada 15 Juni 2024.

(BERT) adalah model transformasi populer dalam pemrosesan bahasa alami (NLP) yang efektif dalam analisis sentimen teks informal seperti cuitan, karena kemampuannya memahami konteks kata secara *bidirectional* [3]. Meskipun awalnya dikembangkan untuk bahasa Inggris, IndoBERT diperkenalkan pada 2020 sebagai versi khusus untuk bahasa Indonesia [19]. Karena data penelitian ini menggunakan bahasa Indonesia, IndoBERT akan digunakan untuk analisis sentimen. Beberapa penelitian analisis sentimen menggunakan metode ini telah dilakukan, seperti oleh H. Imaduddin dan timnya pada komentar aplikasi kesehatan dengan nilai akurasi 96% [8], serta oleh D. R. Wijaya dan rekan-rekannya pada data Twitter dengan akurasi sebesar 82,5% [18]. S. Saadah dan timnya juga membandingkan beberapa metode analisis sentimen pada data cuitan dan menemukan bahwa IndoBERT memiliki nilai akurasi terbaik [13]. Penelitian-penelitian tersebut memperkuat keputusan untuk memilih IndoBERT sebagai metode analisis sentimen pada penelitian ini.

Penelitian ini mengacu pada beberapa referensi utama: pertama, artikel oleh V. A. Traag et al. (2019) berjudul “*From Louvain to Leiden: Guaranteeing Well-Connected Communities*” [17]; kedua, penelitian T. D. Dikiyanti et al. (2021) yang berjudul “*Sentiment analysis and topic modeling of BPJS Kesehatan based on twitter crawling data using Indonesian Sentiment Lexicon and Latent Dirichlet Allocation algorithm.*” [4]. Penelitian ini berbeda karena fokus pada isu Palestina-Israel di aplikasi X (Twitter) dari 7 Oktober 2023 hingga 25 Juli 2024, serta menerapkan kombinasi deteksi komunitas dengan algoritma Leiden, pemodelan topik dengan LDA, dan analisis sentimen dengan IndoBERT.

## 2. METODELOGI PENELITIAN

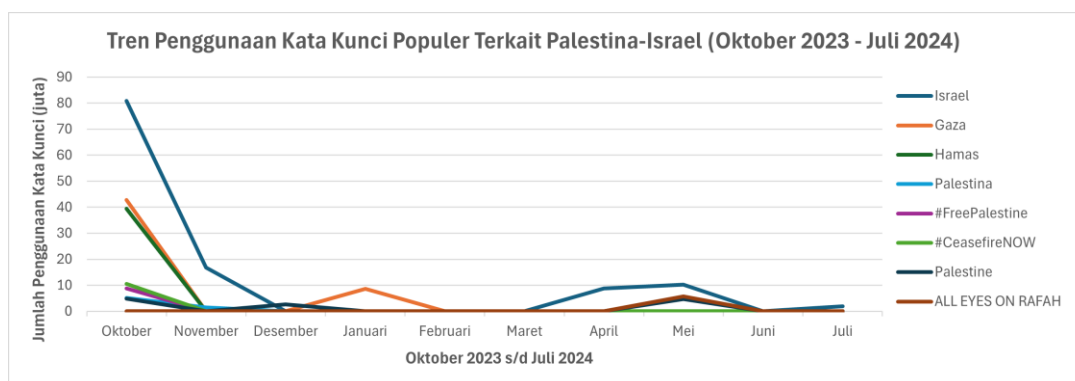
Penelitian ini dilakukan melalui beberapa tahapan utama, mulai dari pengumpulan data cuitan, pra-pemrosesan teks, analisis jaringan untuk mendeteksi komunitas, hingga analisis topik dan sentimen. Gambar 2.1. menjelaskan langkah-langkah sistematis yang ditempuh untuk mencapai tujuan penelitian, di mana setiap tahap saling terkait untuk memastikan hasil yang akurat dan relevan.



**Gambar 2.1.** Diagram Alur Penelitian

## 2.1. Sumber Data

Penelitian ini menggunakan data dari aplikasi X terkait isu Palestina-Israel, yang dikumpulkan dengan menggunakan bahasa Indonesia dan beberapa kata kunci, seperti 'Israel', 'Gaza', 'Hamis', 'Palestina', '#FreePalestine', '#CeasefireNOW', 'Palestine', dan 'ALL EYES ON RAFAH'. Kata-kata ini dipilih karena mencakup entitas kunci yang terlibat langsung dalam konflik, simbol gerakan global, desakan perdamaian, serta peristiwa penting di Rafah, sebuah kota yang menjadi target utama agresi Israel. Pemilihan kata kunci diperkuat dengan seringnya mereka menjadi *trending* di kalangan pengguna X Indonesia (lihat Gambar 2.2.).



Gambar 2.2. Tren Penggunaan Kata Kunci Populer dalam Isu Palestina-Israel<sup>4</sup>

Data tersebut diambil dari aplikasi X sejak awal gencatan senjata Palestina-Israel pada 7 Oktober 2023 hingga waktu pengambilan pada 25 Juli 2024. Pengambilan data dilakukan menggunakan alat *crawling* bernama *Tweet-Harvest* yang memanfaatkan API untuk mengakses data. Hasilnya, terkumpul sebanyak 108.969 data.

## 2.2. Pra-Pemrosesan Data

Pra-Pemrosesan data adalah langkah untuk mempersiapkan data mentah agar siap digunakan dalam analisis lanjutan, dengan tujuan meningkatkan kualitas data. Dalam penelitian ini, *preprocessing* mencakup pemilihan variabel relevan, dan pembersihan data dari nilai hilang dan duplikat. Selain itu, *preprocessing* juga diterapkan pada graf, dengan menghapus simpul yang memiliki derajat nol atau sangat rendah, untuk memastikan komunitas yang terbentuk lebih representatif dan berkualitas [20].

## 2.3. Penambangan Teks

Penambangan teks adalah proses mengubah teks tidak terstruktur menjadi informasi yang dapat digunakan untuk pengambilan keputusan yang lebih baik. Dalam penelitian ini, *text mining* diterapkan pada data cuitan untuk pemodelan topik dan analisis sentimen, namun untuk analisis sentimen hanya dilakukan hingga tahap *feature selection*. Tahapan *text mining* meliputi [9]:

- Text preprocessing*, mengubah teks menjadi format terstruktur melalui langkah-langkah seperti *case folding*, penghapusan URL, entitas HTML, *mention*, angka, tanda baca, *emoticon*, spasi berlebihan, dan normalisasi kata.
- Feature Selection*, menghapus kata-kata tidak relevan untuk meningkatkan efektivitas pengelompokan melalui *stopwords removal*, *stemming/lemmatization*, dan *tokenizing*.

<sup>4</sup> Lihat lebih lanjut di *Twitter Trending Archive*: <https://archive.twitter-trending.com/>, diakses pada 14 Juli 2024.

- c. *Text Representation*, mengubah teks menjadi format numerik, seperti *Document Term Matrix* (DTM), di mana pembobotan kata dilakukan menggunakan metode *Term-Frequency* (TF).

## 2.4. Analisis Eksplorasi Data

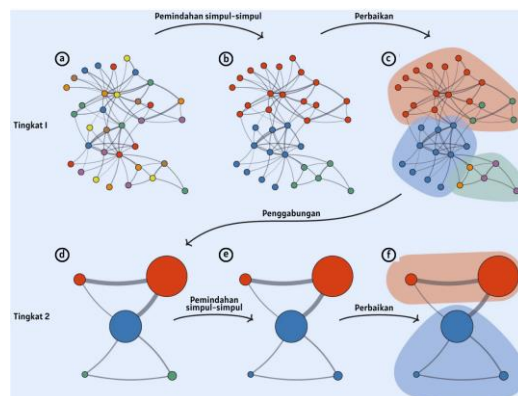
Analisis eksplorasi data adalah proses untuk menemukan pola, anomali, dan hubungan dalam data menggunakan teknik statistik dan visualisasi, bertujuan memahami struktur dasar dataset dan mendapatkan wawasan awal. Dalam penelitian ini, analisis eksplorasi data fokus pada analisis komunitas, dengan memeriksa informasi graf yang terbentuk seperti jumlah simpul, sisi, derajat masuk dan keluar, serta statistik sederhana seperti rerata derajat, diameter jaringan, dan kepadatan graf (*density graph*).

## 2.5. Algoritma Leiden

Algoritma Leiden, dikembangkan oleh Vincent Traag pada tahun 2019, merupakan perbaikan dari algoritma Louvain untuk meningkatkan akurasi dan efisiensi dalam mendeteksi komunitas pada jaringan besar [17]. Algoritma ini memaksimalkan modularitas, yang mengukur kualitas partisi jaringan, dengan nilai modularitas berkisar dari -1 hingga 1. Nilai mendekati 1 menunjukkan struktur komunitas kuat, yang berarti komunitas lebih terhubung dan stabil. Berikut adalah formula modularitas ( $Q$ ) didefinisikan sebagai berikut.

$$Q = \frac{1}{2m} \sum_{i,j \in V} \left( A_{ij} - \gamma \frac{k_i k_j}{2m} \right) \delta(C_i, C_j) \quad (1.1)$$

Pada Persamaan 1.1 di atas,  $m$  adalah jumlah total bobot sisi pada graf.  $A_{ij}$  menunjukkan jumlah bobot sisi antara simpul  $i$  dan  $j$ , dengan  $2m = \sum_{i,j} A_{ij}$ .  $k_i$  adalah jumlah bobot semua sisi yang melekat pada simpul  $i$ .  $C_i$  adalah komunitas tempat simpul  $i$  ditugaskan, dimana  $\delta(C_i, C_j) = 1$  jika  $C_i = C_j$  dan 0 jika tidak. Selain itu,  $\gamma$  merupakan parameter resolusi; jika nilainya  $\gamma > 1$ , hasilnya adalah lebih banyak komunitas yang lebih kecil dan lebih terhubung dengan baik, sedangkan  $\gamma < 1$  menghasilkan lebih sedikit komunitas yang lebih besar dan kurang terhubung.



**Gambar 2.3.** Tahapan Algoritma Leiden [17]

Algoritma Leiden bekerja serupa dengan algoritma Louvain, tetapi menambahkan fase penyempurnaan untuk memastikan keterhubungan yang baik di setiap komunitas. Algoritma ini dimulai dengan partisi singleton, di mana setiap simpul berada dalam komunitasnya sendiri, dan melalui tiga fase utama (lihat Gambar 2.3.):

- a. Fase Pertama: Optimasi Modularitas Cepat

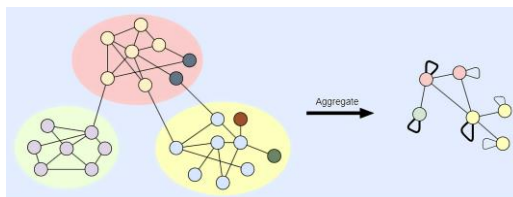
Pada fase ini, algoritma Leiden mengunjungi setiap simpul sekali dan hanya simpul yang lingkungannya berubah yang diproses lebih lanjut. Simpul dipindahkan ke komunitas yang meningkatkan modularitas maksimum atau tetap di komunitas asal jika tidak ada peningkatan. Fase ini menghasilkan partisi  $P$  dari graf.

b. Fase Kedua: Penyempurnaan

Fase ini bertujuan memastikan semua komunitas terhubung dengan baik. Setiap komunitas dievaluasi, dan simpul-simpul yang tidak terhubung dengan baik dapat dipindahkan ke komunitas lain, mempertimbangkan peningkatan modularitas dengan tingkat keacakan tertentu untuk memungkinkan eksplorasi partisi yang lebih luas.

c. Fase Ketiga: Agregasi Komunitas

Graf agregat dibentuk berdasarkan  $P_{refined}$ , tetapi partisi didasarkan pada  $P$ , mirip dengan metode Louvain. Setelah fase ketiga selesai, tahap ini diulang pada graf agregat hingga tidak ada perubahan dan modularitas maksimum tercapai (lihat Gambar 2.4.).



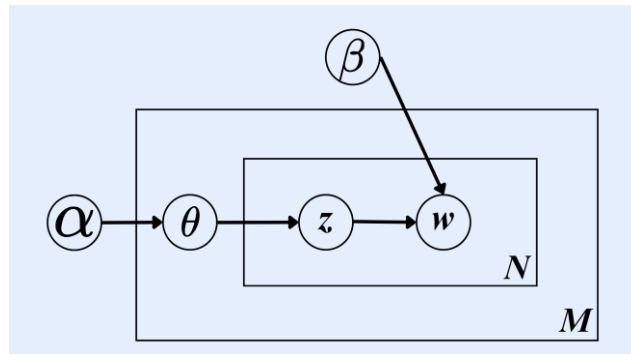
Gambar 2.4. Ilustrasi Agregasi Komunitas [17]

Dalam penelitian ini, deteksi komunitas dilakukan menggunakan algoritma Leiden melalui aplikasi Gephi. Gephi adalah perangkat lunak sumber terbuka yang dirancang untuk visualisasi dan analisis jaringan, menawarkan visualisasi waktu nyata serta antarmuka pengguna yang intuitif [19]. Dengan menggunakan Gephi, peneliti dapat mempelajari hubungan yang kompleks dan mengungkap struktur sistem yang mendasari di berbagai disiplin ilmu [19]. Selain itu, perangkat lunak ini juga memungkinkan analisis perilaku pengguna di platform media sosial melalui pengelompokan graf. Kemampuan Gephi dalam pengoptimalan modularitas sangat bermanfaat untuk mendeteksi dan memvisualisasikan struktur komunitas dalam jaringan yang kompleks, sehingga memberikan wawasan yang lebih mendalam dibandingkan dengan analisis jaringan tradisional [5].

## 2.6. Latent Dirichlet Allocation (LDA)

*Latent Dirichlet Allocation* (LDA) adalah algoritma pemodelan topik yang mendeteksi topik tersembunyi dalam korpus teks. LDA mengasumsikan bahwa setiap dokumen terdiri dari campuran beberapa topik, dan setiap topik terdiri dari sekelompok kata yang saling terkait. Istilah penting dalam LDA meliputi [9]: kata (unit terkecil dalam data)  $\{1, \dots, V\}$ ; dokumen (rangkaian  $N$  kata)  $w = (w_1, w_2, \dots, w_N)$ ; dan korpus (kumpulan  $M$  dokumen)  $D = \{w_1, w_2, \dots, w_M\}$ .

LDA mengasumsikan adanya beberapa proses untuk setiap dokumen  $w$  dalam korpus  $D$ . Pertama, dipilih nilai  $N$  yang mengikuti distribusi Poisson ( $\xi$ ). Kedua, dipilih parameter  $\theta$  yang mengikuti distribusi Dirichlet ( $\alpha$ ). Selanjutnya, untuk setiap kata  $N$  dalam dokumen  $w_n$ , dilakukan dua langkah, yaitu memilih topik  $z_n$  mengikuti distribusi Multinomial ( $\theta$ ), dan memilih kata  $w_n$  berdasarkan peluang multinomial  $P(w_n | z_n, \beta)$ , yang ditentukan oleh topik.



Gambar 2.5. Model Grafis LDA

*Latent Dirichlet Allocation* (LDA) adalah model grafis probabilistik dengan tiga tingkatan: dokumen ( $M$ ), topik ( $\theta$ ), dan kata ( $N$ ) (lihat Gambar 2.5.). Parameter  $\alpha$  dan  $\beta$  mengatur distribusi topik dalam dokumen dan kata dalam topik. Semakin tinggi  $\alpha$ , semakin banyak topik dalam dokumen; semakin tinggi  $\beta$ , semakin banyak kata dalam topik. Variabel  $\theta$  menunjukkan campuran topik dalam dokumen, sementara  $z$  dan  $w$  masing-masing mewakili topik terkait dengan kata dan kata terkait dengan topik. LDA menguraikan *Document Term Matrix* (DTM) menjadi dua matriks: *document-topic matrix* dan *topic-word matrix*, untuk menghasilkan distribusi dokumen-topik dan topik-kata yang optimal.

Proses *Latent Dirichlet Allocation* (LDA) dimulai dengan input berupa *Document Term Matrix* (DTM), yang mencatat frekuensi kata dalam dokumen, dengan baris mewakili dokumen dan kolom mewakili kata unik. Selanjutnya, dilakukan inisialisasi parameter untuk menentukan jumlah topik ( $k$ ) dan hyperparameter  $\alpha$  serta  $\beta$ . Setelah itu, *sampling Dirichlet* dan *Multinomial* dilakukan untuk mendistribusikan topik pada dokumen dan kata. *Gibbs Sampling* digunakan untuk memperkirakan distribusi posterior topik dan kata, dengan iterasi untuk memperbarui probabilitas topik berdasarkan hitungan kata. Terakhir, estimasi akhir menghasilkan *Document-Topic Matrix* dan *Topic-Word Matrix*, yang menunjukkan distribusi probabilitas topik pada dokumen dan kata. Kualitas topik kemudian dinilai menggunakan koherensi topik varian UCI, yang mengukur seberapa bermakna topik dengan melihat seberapa sering kata-kata dalam topik muncul bersama dalam dokumen.

## 2.7. IndoBERT

IndoBERT adalah model bahasa berbasis *transformer* yang dioptimalkan untuk bahasa Indonesia, dibangun dari arsitektur BERT dan dilatih menggunakan korpus teks besar. Model ini efektif untuk berbagai tugas NLP, termasuk analisis sentimen. Dalam penelitian ini, IndoBERT digunakan untuk melabeli cuitan berbahasa Indonesia yang belum berlabel, dengan melatih model pada sebagian kecil data berlabel secara manual dan *unfreeze* lapisan terluar untuk meningkatkan performa.

Sebelum diproses dengan IndoBERT, data harus melalui tahap *preprocessing* dan *feature selection*, lalu diubah menjadi format yang sesuai untuk model, termasuk penambahan token khusus seperti [CLS] dan [SEP]. Data diekode dengan *tokenizer* IndoBERT menjadi indeks berdasarkan kosakata model. IndoBERT menerima input hingga 128 token; kalimat yang lebih pendek diberi padding, sementara yang lebih panjang dipotong. Kata yang tidak ada dalam kosakata model dipecah menjadi sub-kata dengan simbol ##, seperti "konflik" menjadi "kon" dan "##flik". Contoh penyesuaian input yang dapat diterima oleh IndoBERT akan disajikan dalam tabel berikut.

Tabel 2.1. Proses Tokenisasi IndoBERT

Kalimat	konflik di Palestina-Israel semakin memanas
---------	---

Hasil Tokenisasi	'konflik', 'di', 'Palestina', '-', 'Israel', 'semakin', 'memanas'
Penambahan Token Khusus	'[CLS]', 'konflik', 'di', 'Palestina', '-', 'Israel', 'semakin', 'memanas', '[SEP]'
Encoding	2, 1824, 12, 30000, 45, 15345, 23, 987, 3
Attention Mask	1, 1, 1, 1, 1, 1, 1, 1, 1

Setelah tokenisasi, token khusus seperti [CLS] dan [SEP] ditambahkan dan dikodekan menjadi indeks sesuai kosakata model. *Attention mask* membedakan token kata dari *padding*, dengan token *padding* [PAD] diberi nilai 0 dan token kata diberi nilai 1 (lihat Tabel 2.1.). Data loader dibuat untuk mempermudah proses tokenisasi dan inferensi. Input yang disesuaikan dimasukkan ke dalam jaringan IndoBERT, yang terdiri dari 12 lapisan *Transformer Encoder*. Hasil akhir dari setiap token diperoleh setelah melewati semua lapisan, tetapi hanya vektor output dari token [CLS] yang digunakan sebagai representasi kalimat dan input untuk *classifier sentiment*.

*Fine-tuning* IndoBERT untuk klasifikasi dilakukan dengan menambahkan lapisan klasifikasi menggunakan kelas *BertForSequenceClassification* dari perpustakaan *Transformers*. *Output pooler* menghitung *logits* yang diproses dengan fungsi *softmax* untuk menghasilkan prediksi. Model yang digunakan adalah IndoBERTBASE fase 1, dengan 12 lapisan *encoder*, 768 *hidden nodes*, dan 124 juta parameter. *Hyperparameter* optimal yang direkomendasikan meliputi *batch size* 16 atau 32 dan jumlah *epoch* 2, 3, atau 4, dengan kombinasi pengaturan yang memiliki akurasi tertinggi akan digunakan untuk klasifikasi sentimen pada data yang belum berlabel.

### 3. HASIL DAN PEMBAHASAN

Penelitian ini mengidentifikasi tiga komunitas dominan dari data aplikasi X terkait isu Palestina-Israel menggunakan algoritma Leiden untuk deteksi komunitas, pemodelan topik LDA, dan analisis sentimen IndoBERT, memberikan wawasan tentang pandangan publik dan dinamika sosial terkait isu tersebut.

#### 3.1. Pra-Pemrosesan Data

Dari data mentah, hanya variabel 'username' dan 'full\_text' yang akan diolah. Sebelum dilakukan pembersihan (*cleaning*), jumlah data sebanyak 108.966, dan setelah proses pembersihan, jumlah data berkurang menjadi 96.946. Tabel 2.2. menampilkan data sebelum dan sesudah proses pembersihan.

**Tabel 2.2.** Data Sebelum dan Sesudah Proses Pembersihan

Sebelum	
username;;;;;;;;;;	full_text
luqmanha14;;;;;;;;;;	@rgone360 Kena beli sorg lg wing back january ...
AsySyaiful01;;;;;;;;;;	@yogi_bintoro @yo2thok Kemungkinan dapet pas d...
haenesarajawane;;;;;;;;;;	@kompascom Bangsaatt .... Hamas itu bukan tero...
Sesudah	
username	full_text
luqmanha14	@rgone360 Kena beli sorg lg wing back january ...
AsySyaiful01	@yogi_bintoro @yo2thok Kemungkinan dapet pas d...
haenesarajawane	@kompascom Bangsaatt .... Hamas itu bukan tero...

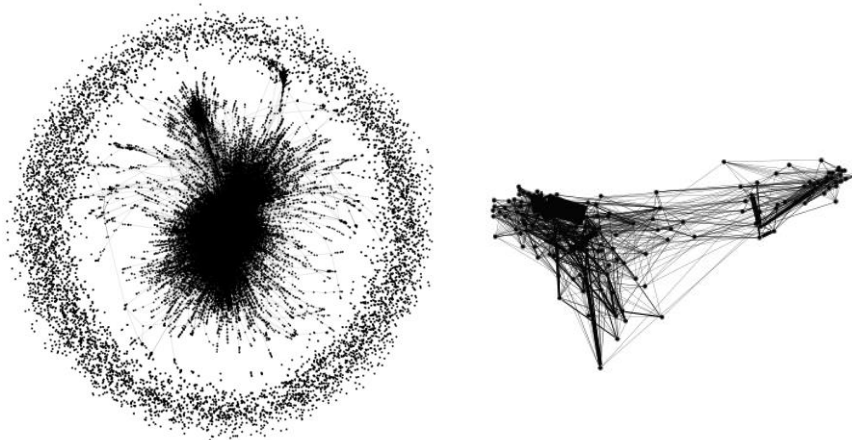
#### 3.2. Membangun Graf



Graf dibangun dengan simpul yang merepresentasikan pengguna X, dan sisi yang menghubungkan simpul-simpul tersebut merepresentasikan interaksi berupa *mention*. Graf berarah dipilih karena mampu memberikan representasi yang lebih realistis dari hubungan *mention* di aplikasi X, dengan menangkap arah interaksi dan pola komunikasi secara lebih detail. Berdasarkan data yang tersedia, graf ini terbentuk dari 54.104 simpul dan 71.130 sisi.

### 3.3. Pra-Pemrosesan Graf

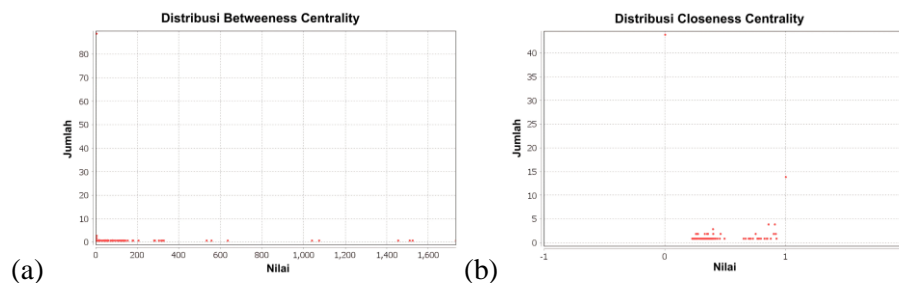
Pra-pemrosesan graf dilakukan untuk menghasilkan komunitas yang lebih representatif dan berkualitas. Beberapa langkah yang diambil meliputi penghapusan simpul dengan derajat 0 hingga 8 serta penghapusan simpul yang memiliki nilai *betweenness centrality* rendah. Tindakan ini bertujuan untuk menyederhanakan jaringan tanpa mengorbankan struktur utamanya. Setelah proses ini dilakukan, graf yang tersisa terdiri dari 167 simpul dan 1.244 sisi (lihat Gambar 3.1.).



**Gambar 3.1.** Graf Sebelum dan Sesudah Pra-Pemrosesan

### 3.4. Analisis Data Eksplorasi

Jaringan atau graf yang telah diproses memiliki rata-rata derajat sebesar 7,449, yang menunjukkan jumlah koneksi rata-rata per simpul. Jaringan ini juga memiliki diameter yang relatif kecil, yaitu 8, dan panjang jalur rata-rata 3,02, yang mencerminkan keterhubungan yang baik antar simpul.



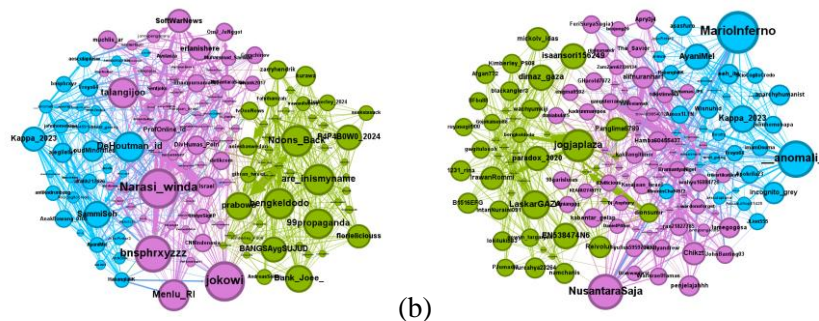
**Gambar 3.2.** Distribusi *Betweenness Centrality* (a), Distribusi *Closeness Centrality* (b)

Gambar 3.2.(a) menunjukkan *betweenness centrality* yang bervariasi dari 0 hingga lebih dari 1600, dengan sebagian besar simpul memiliki nilai rendah dan beberapa simpul berperan sebagai

penghubung penting. Dan Gambar 3.2.(b) menunjukkan *closeness centrality* yang cenderung rendah dalam rentang 0 hingga 1, menunjukkan simpul-simpul relatif dekat satu sama lain. Dengan densitas 0,045, jaringan ini memiliki keterhubungan rendah, menandakan interaksi terbatas dan terfokus pada simpul-simpul utama.

### 3.5. Deteksi Komunitas

Deteksi komunitas dilakukan menggunakan algoritma Leiden dengan parameter resolusi sebesar 1,0, jumlah iterasi sebanyak 15, dan pengulangan (*restart*) sebanyak 3 kali. Kombinasi parameter ini dipilih untuk mencapai keseimbangan antara akurasi deteksi komunitas dan efisiensi komputasi, serta memastikan hasil yang stabil dan representatif. Hasil deteksi komunitas pada graf yang telah melalui pra pemrosesan menunjukkan adanya 5 kluster dengan nilai modularitas sebesar 0,523, yang dapat dianggap cukup baik. Dari kelima kluster tersebut, tiga komunitas dominan dipilih untuk analisis lebih lanjut.



**Gambar 3.3.** Tiga Komunitas Dominan dengan Diameter Simpul sebagai: (A) Jumlah Derajat Masuk dan (B) Jumlah Derajat Keluar

Gambar 3.3. menampilkan 3 komunitas dominan, dimana warna ungu, hijau, dan biru masing-masing mewakili komunitas dominan 1, 2, dan 3. Komunitas-komunitas dominan memiliki persentase anggota sebesar 37,13%, 26,95%, dan 19,76% berturut-turut. Pada gambar tersebut, ukuran simpul memberikan gambaran umum mengenai jumlah derajat masuk (a) dan derajat keluar (b). Hal ini menunjukkan jumlah *mention* yang diterima (derajat masuk) dan dikirimkan (derajat keluar) oleh setiap pengguna, sehingga memberikan gambaran umum mengenai pola interaksi di antara para pengguna.

**Tabel 2.3.** Informasi Komunitas Dominan

Komunitas Dominan	Jumlah Anggota	Jumlah Cuitan
1	62 pengguna	3.111
2	45 pengguna	641
3	33 pengguna	389

Tabel 2.3. menampilkan jumlah anggota dan jumlah cuitan dari setiap komunitas dominan. Terlihat bahwa semakin banyak anggota dalam suatu komunitas, semakin banyak pula jumlah cuitan yang dihasilkan.

**Tabel 2.4.** Tiga Nama Pengguna Paling Banyak Menerima dan Mengeluarkan *Mention*

Menerima <i>Mention</i>	Mengeluarkan <i>Mention</i>
-------------------------	-----------------------------

Komunitas Dominan	Nama Pengguna	Frekuensi	Nama Pengguna	Frekuensi
1	Narasi_winda	35	NusantaraSaja	26
	jokowi	35	alifnuranhar1	15
	bnsphrxyzzz	32	Chikzt	14
2	Ndons_Back	27	jogjaplaza	20
	bengkeldodo	26	LaskarGAZA	17
	are_inismyname	26	dimaz_gaza	17
3	DeHoutman_id	25	__anomali__	32
	SammiSoh	21	MarioInferno	30
	Kappa_2023	19	Kappa_2023	28

Selanjutnya, Tabel 2.4. menyajikan tiga pengguna dengan jumlah *mention* terbanyak, baik sebagai penerima maupun pengirim, dari masing-masing komunitas dominan. Pengguna yang paling banyak menerima *mention* dapat dianggap sebagai akun yang memiliki pengaruh besar dalam komunitas. Sementara itu, pengguna yang paling banyak mengirimkan *mention* bisa diartikan sebagai akun yang aktif dalam diskusi, meskipun dapat diindikasikan sebagai akun *buzzer* atau bahkan *spammer*. Akun-akun ini dapat berperan sebagai alat penyebaran opini, pendukung kampanye, atau pengendali narasi dalam komunitas tertentu.

Dari analisis akun-akun dengan jumlah penerima dan pengirim *mention* terbanyak setiap komunitas dominan, ditemukan bahwa dalam komunitas dominan 1, akun Narasi\_winda merupakan akun yang paling berpengaruh. Akun ini sering membahas isu-isu sosial, budaya, atau politik melalui narasi gambar dan kata-kata, dengan fokus pada kebebasan berekspresi, serta sering memulai percakapan dalam komunitas. Selain itu, akun NusantaraSaja menjadi akun yang paling aktif dengan aktivitas yang secara jelas mendukung negara Israel.

Selanjutnya, dalam komunitas dominan 2, akun Ndons\_Back adalah akun yang paling berpengaruh. Akun ini sering membahas narasi dan berita terkait isu sosial, budaya, atau politik di Indonesia, serta menyuarakan perlindungan anak. Sementara itu, akun jogjaplaza menjadi akun yang paling aktif, terutama dalam mendukung negara Palestina.

Terakhir, dalam komunitas dominan 3, akun paling berpengaruh adalah DeHoutman\_id. Meskipun telah ditangguhkan oleh aplikasi X, sebelumnya akun ini sering mengkritik isu politik dan tetap dianggap sebagai akun yang berpengaruh. Sementara itu, akun \_\_anomali\_\_ menjadi akun yang paling aktif, dikenal dengan pandangan keagamaannya yang kuat dan sering berbagi pandangan alternatif terkait berbagai isu. Yang menarik dalam komunitas ini adalah akun Kappa\_2023, yang memiliki jumlah *mention* masuk dan keluar yang tergolong banyak. Setelah dianalisis lebih lanjut, akun ini bukanlah akun yang sangat berpengaruh, melainkan lebih aktif mengkritik dan berbagi pandangan mengenai isu keagamaan.

### 3.6. Penambangan Teks

Data yang melalui proses *text mining* adalah data yang akan digunakan untuk pemodelan topik dan analisis sentimen, yaitu semua cuitan pada komunitas dominan. Namun, data untuk analisis sentimen hanya sampai tahap *feature selection*. Tabel 2.5. menampilkan perbedaan data sebelum dan setelah melalui tahap *text preprocessing* dan *feature selection*.

**Tabel 2.5.** Data Sebelum dan Setelah Tahap *Text Preprocessing* dan *Feature Selection*.

## JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Fathnin Nur Azmina, Muhaza Liebenlito, Dhea Urfina Zulkifli

Sebelum	Sesudah
<p>@TheGoatNunezIdn @athok_delonge @yo2thok Bukanya thn 1946 daerah itu masih jadi jajahan inggris ya? untuk tahun 1947 itu batas 2 negara yg ditentukan PBB. tahun 1949-1967 tanah yg didapatkan pasca perang 6 hari lawan negara arab yg gak setuju pendirian ne</p> <p>@MahjriProperty @Paltiwest Palestina sekarang apa sama dengan Palestina dulu??? Ini aja dulu jawab nanti kita bahas lg setelah kamu jawab?</p> <p>@zoelfick seharusnya pejuang hamas cari medan perang kurusetra pak. duel jantan. bukan sembunyi di rumah2 penduduk atau di masjid sehabis membantai warga sipil israel di kampung halamannya.</p>	<p>[buka, tahun, daerah, jajah, inggris, batas, negara, tentu, pbb, tanah, dapat, pasca, perang, lawan, negara, arab, tuju, diri, ne]</p> <p>[palestina, palestina, bahas, lagi]</p> <p>[juang, hamas, cari, medan, perang, kurusetra, duel, jantan, sembunyi, rumah, duduk, masjid, habis, bantai, warga, sipil, israel, kampung, halaman]</p>

### 3.7. Deteksi Topik

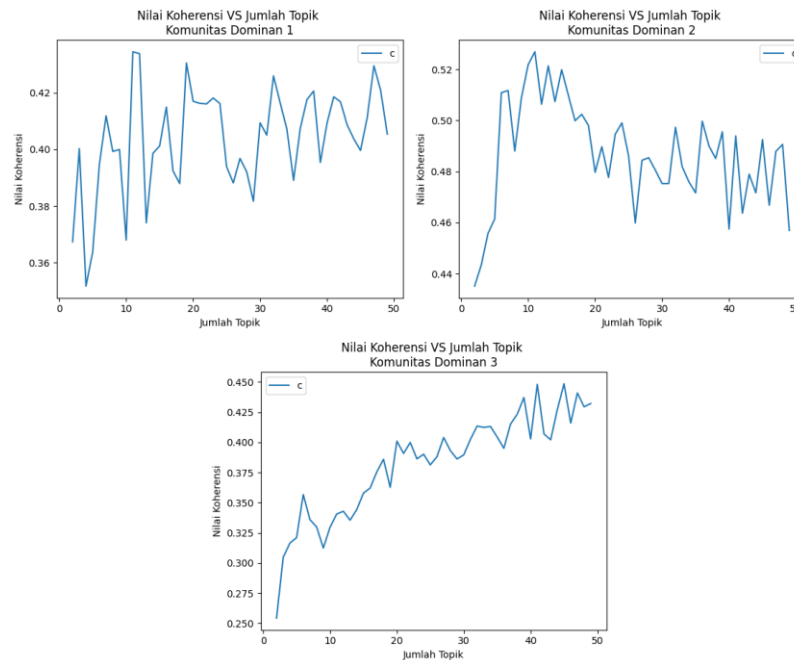


**Gambar 3.4.** Word Cloud dari Semua Cuitan dalam Tiap Komunitas Dominan

Langkah ini bertujuan untuk mengidentifikasi kata-kata yang paling sering digunakan oleh pengguna X dalam setiap komunitas dominan, dengan menggunakan *word cloud*. Gambar 3.4. menampilkan gambaran umum dari kata-kata yang sering dipakai dalam tiap komunitas dominan, di mana ukuran kata yang lebih besar menunjukkan frekuensi pemakaian yang lebih tinggi. Dari gambar tersebut, terlihat bahwa dalam komunitas dominan 1, tiga kata terpopuler adalah "israel," "gaza," dan "palestina." Dalam komunitas dominan 2, tiga kata teratas adalah "palestina," "prabowo," dan "orang." Sedangkan dalam komunitas dominan 3, kata-kata yang paling sering digunakan adalah "palestina," "hamas," dan "israel." Secara keseluruhan, ketiga komunitas dominan ini menunjukkan pola yang konsisten dengan diskusi terkait konflik Palestina-Israel, memperlihatkan bagaimana pengguna aplikasi X berfokus pada isu-isu yang berkaitan dengan konflik tersebut.

### 3.8. Pemodelan Topik

Pemodelan topik menggunakan *Latent Dirichlet Allocation* (LDA) dilakukan untuk mengidentifikasi topik-topik dalam tiga komunitas dominan. Model LDA yang paling stabil dan representatif dipilih berdasarkan nilai koherensi tertinggi, yang digunakan untuk mengukur kualitas topik yang dihasilkan.



**Gambar 3.5.** Grafik Nilai Koherensi Tiap Komunitas Dominan

Gambar 3.5. menampilkan grafik nilai koherensi dari model tiap komunitas dominan, menunjukkan kinerja model pada berbagai jumlah topik. Hasil analisis menunjukkan bahwa jumlah topik optimal yang dihasilkan dari LDA untuk masing-masing komunitas adalah 11 topik untuk komunitas dominan 1 dan 2, serta 45 topik untuk komunitas dominan 3. Jumlah topik ini mencerminkan variasi diskusi dan luasnya isu yang dibahas dalam tiap komunitas dominan, terutama dalam komunitas dominan 3 yang memiliki lebih banyak topik, mungkin karena lebih banyak sub-tema yang muncul dalam diskusi mengenai konflik Palestina-Israel.

**Tabel 2.6.** Sepuluh Kata Populer dalam 5 Topik Teratas

	<b>Topik 1</b>	<b>Topik 2</b>	<b>Topik 3</b>	<b>Topik 4</b>	<b>Topik 5</b>
<b>Komunitas Dominan 1</b>	gaza	israel	israel	palestina	hamas
	israel	gaza	hizbullah	israel	gaza
	tentara	senjata	serang	arab	israel
	jalur	as	lebanon	negara	warga
	zionis	warga	perang	orang	teroris
	palestina	bantu	zionis	bangsa	anak
	tewas	palestina	lawan	yahudi	idf
	serang	hamas	militer	tanah	buruh
	pasu	amerika	utara	tidak	palestina
	brigade	perang	tentara	anda	tolol

	Topik 1	Topik 2	Topik 3	Topik 4	Topik 5
<b>Komunitas Dominan 2</b>	palestina	palestina	palestina	rakyat	israel
	rakyat	sudah	gaza	bagaimana	pakai
	perang	beliau	akibat	makan	palestina
	kamu	israel	indonesia	israel	maksud
	manusia	salah	orang	sampai	obat
	presiden	rip	israel	hamas	gas
	israel	golkar	tunggu	negara	adil
	prabowo	legend	bohong	dukung	bukti
	dewan	ruben	salah	oknum	negara
	produk	salak	rakyat	pro	kalah
	Topik 1	Topik 2	Topik 3	Topik 4	Topik 5
<b>Komunitas Dominan 3</b>	Israel	israel	hamas	israel	israel
	palestina	hamas	palestina	zionis	tipu
	warga	lindung	indonesia	dukung	palestina
	gaza	palestina	israel	bahawa	hamas
	zionis	warga	air	pesek	gaza
	hamas	kecut	yahudi	yahudi	holokous
	bantu	tentara	gaza	aka	dunia
	arab	zionis	terowongan	narasi	perang
	manusia	warga	tidak	salah	teroris
	serang	arab	anda	goblok	arab

Tabel 2.6. menyajikan daftar kata-kata populer dalam lima topik teratas dari tiga komunitas dominan, yang mengarah pada diskusi seputar isu Palestina-Israel, dengan fokus yang berbeda di setiap komunitas.

Kata-kata populer dalam lima topik teratas dari tiga komunitas dominan dapat memberikan wawasan berharga mengenai pembahasan isu Palestina-Israel. Dengan menganalisis kata-kata tersebut, peneliti dapat memastikan apakah komunitas tersebut benar-benar membahas isu Palestina-Israel. Lebih jauh lagi, analisis ini juga dapat mengungkap aspek-aspek spesifik yang menjadi fokus diskusi mereka terkait isu tersebut.

Dalam komunitas dominan 1, Topik 1 membahas kekerasan fisik dan militer di jalur Gaza, dengan kata-kata seperti “gaza”, “israel”, “tentara”, “jalur”, “palestina”, “tewas”, “serang”, “pasu”, dan “brigade”. Topik 2 membahas bantuan persenjataan Israel dari Amerika Serikat dengan kata-kata seperti “israel”, “senjata”, “as”, “bantu”, “amerika”, dan “perang”. Topik 3 mengulas perlawanan hizbullah Lebanon terhadap Israel dan serangan Israel terhadap Lebanon dengan kata-kata seperti “israel”, “hizbullah”, “serang”, “lebanon”, dan “lawan”. Topik 4 menyangkut asal usul konflik Palestina dan Israel dengan kata-kata seperti “palestina”, “israel”, “arab”, “negara”, “bangsa”, “yahudi”, dan “tanah”. dan Topik 5 menyoroti korban anak-anak di Gaza dan mengkritik tindakan Israel dengan kata-kata seperti “hamas”, “gaza”, “israel”, “warga”, “teroris”, “anak”, “idf”, “palestina”, dan “tolol”. Terlihat bahwa komunitas dominan 1 ini memiliki beragam topik yang membahas isu Palestina-Israel.

Selanjutnya, dalam komunitas dominan 2, meskipun isu Palestina-Israel tetap dibahas dengan kata-kata seperti "Palestina", "perang", "Israel", "Gaza", dan "Hamis", topik-topik lain juga muncul, seperti "presiden", "Prabowo", "Golkar", dan sebagainya. Hal ini menunjukkan bahwa komunitas tersebut tidak sepenuhnya terfokus pada konflik Palestina-Israel.

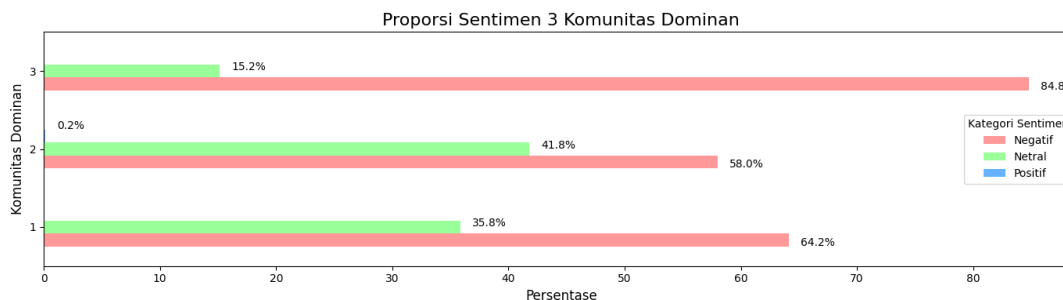
Terakhir, dalam komunitas dominan 3, Topik 1 berfokus pada sejarah dan konflik Palestina-Israel dengan kata-kata seperti “israel”, “palestina”, “gaza”, “zionis”, “hamas”, “bantu”, “arab”, dan

“serang”; Topik 2 mengulas kritik terhadap Hamas dan Israel dengan kata-kata seperti “israel”, “hamas”, “lindung”, “palestina”, “warga”, dan “kecut”; Topik 3 menyoroti perdebatan mengenai terowongan Hamas serta partisipasi Indonesia dengan kata-kata seperti “hamas”, “palestina”, “indonesia”, “israel”, “air”, dan “terowongan”; Topik 4 menyinggung narasi sejarah Yahudi dan Israel dengan kata-kata seperti “israel”, “zionis”, “dukung”, “pesek”, “yahudi”, “narasi”, “salah”, dan “goblok”; dan Topik 5 membahas pandangan dunia tentang konflik, termasuk holokaus dan terorisme dengan kata-kata seperti “israel”, “tipu”, “palestina”, “hamas”, “gaza”, “holokaus”, “dunia”, “perang”, dan “teroris”. Terlihat bahwa dalam komunitas dominan 3 ini terdapat beragam topik terkait isu Palestina-Israel.

Secara keseluruhan, ketiga komunitas ini mencerminkan sudut pandang berbeda terkait konflik Palestina-Israel, mencakup kekerasan, politik, dan dampak kemanusiaan, serta memperlihatkan kompleksitas konflik ini dari berbagai perspektif.

### 3.9. Analisis Sentimen

Analisis sentimen menggunakan model *pre-trained* IndoBERT yang telah dilatih kembali dengan data terkait isu Palestina-Israel yang memiliki akurasi tes sebesar 73,4%. Tujuan analisis ini adalah untuk memahami sentimen masyarakat Indonesia terhadap konflik Palestina-Israel.



**Gambar 3.6.** Diagram Proporsi Kategori Sentimen Cuitan Tiap Komunitas Dominan

Gambar 3.6. menunjukkan proporsi tiap sentimen dari semua cuitan mengenai isu Palestina-Israel dalam setiap komunitas dominan. Warna merah, hijau, dan biru berturut merepresentasikan sentimen negatif, netral, dan positif. Terlihat bahwa ketiga komunitas memiliki pola yang sama, yaitu mayoritas cuitan memiliki sentimen negatif, diikuti sentimen netral, dan hanya sedikit yang memiliki sentimen positif bahkan ada yang tidak ada sama sekali.

Cuitan yang teridentifikasi memiliki sentimen positif, negatif, atau netral dapat mencerminkan pandangan para pengguna yang mempostingnya. Sentimen positif menunjukkan pandangan optimis atau dukungan terhadap salah satu pihak dalam konflik. Sebaliknya, sentimen negatif menggambarkan ketidakpuasan atau kritik tajam terhadap situasi yang terjadi. Sementara itu, sentimen netral menunjukkan bahwa pengguna mungkin hanya menyampaikan informasi tanpa mengekspresikan emosi atau opini yang jelas.

Ketiga komunitas dominan menunjukkan pola proporsi sentimen yang serupa, yaitu didominasi oleh sentimen negatif, diikuti oleh sentimen netral, dan terakhir sentimen positif. Pola ini mengindikasikan bahwa mayoritas pengguna memiliki ketidakpuasan atau pandangan negatif yang kuat terhadap situasi Palestina-Israel. Selain itu, terdapat sejumlah besar pengguna yang hanya menyampaikan informasi atau berita tanpa mengekspresikan emosi atau opini yang jelas. Terakhir, hanya sedikit sekali pengguna yang menunjukkan dukungan atau pandangan positif terkait konflik ini.

**Tabel 2.7.** Contoh Cuitan yang Bersentimen Negatif

<b>Cuitan</b>
teriak palestina teroris hamas teroris orang-orang tinas lawan teriak israel teroris jahat wanita anak palestina
orang israel palestina manusia tunduk hukum internasional ujar benci lanjut bawa bencana
tukang bohong hina hukum taurat usaha tutup busuk zionis israel orang seperti kau hina hukum taurat ludahin sama zionis israel lagak bela zionis ko benci islam

Tabel 2.7. menampilkan contoh-contoh cuitan dengan sentimen negatif. Dari tabel tersebut, terlihat bahwa cuitan yang bersifat negatif umumnya menyampaikan ketidakpuasan atau pandangan yang kritis terhadap situasi yang dibahas. Cuitan-cuitan ini cenderung menunjukkan emosi yang kuat, seperti kekecewaan, kemarahan, atau penentangan, serta mengandung opini atau sikap negatif dari pengguna terkait isu tersebut.

#### **4. KESIMPULAN**

Jaringan hasil *crawling* data isu Palestina-Israel di aplikasi X menggunakan algoritma Leiden mendeteksi 5 komunitas dengan modularitas 0,523, menunjukkan struktur komunitas yang solid. Tiga komunitas dominan terdiri dari 62, 45, dan 33 pengguna, dengan pengguna berpengaruh seperti Narasi\_winda, Ndons\_Back, dan DeHoutman\_id, serta pengguna paling aktif seperti NusantaraSaja, jogjaplaza, dan \_\_anomali\_\_. Pemodelan topik dengan LDA menunjukkan diskusi berfokus pada berbagai aspek konflik Palestina-Israel, seperti kekerasan, politik, dan kemanusiaan. Analisis sentimen mengungkap mayoritas pandangan negatif, mencerminkan kritik terhadap situasi konflik. Kombinasi deteksi komunitas, pemodelan topik, dan analisis sentimen memberikan analisis mendalam tentang aktor utama, topik, dan sentimen publik, sehingga menghasilkan wawasan yang komprehensif dan holistik.

Saran untuk penelitian selanjutnya adalah memperluas jenis interaksi yang digunakan dalam membangun graf. Dalam penelitian ini, graf dibangun berdasarkan interaksi *mention*, namun di aplikasi X (Twitter) terdapat berbagai bentuk interaksi lain, seperti *retweet*, *like*, dan *reply*, yang juga dapat memberikan wawasan tambahan. Dengan memasukkan berbagai jenis interaksi ini, analisis yang dihasilkan diharapkan dapat menjadi lebih mendalam dan komprehensif, sehingga dapat memberikan gambaran yang lebih kaya mengenai dinamika yang terjadi dalam komunitas.

#### **DAFTAR PUSTAKA**

- [1] Ahmad Rifai, Eka Putri Rachmawati, & Dinar Anggit Wicaksono., 2022. Social Network Analysis dalam Mengukur Keaktifan Promosi Universitas di Media Sosial Twitter. *The Indonesian Journal of Computer Science*, 11(3). <https://doi.org/10.33022/ijcs.v11i3.3097>
- [2] D. Wood, I., Glover, J., & Buitelaar, P., 2020. Community Topic Usage in Online Social Media. *ACM Transactions on Social Computing*, 3(3). <https://doi.org/10.1145/3403616>
- [3] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (arXiv:1810.04805). arXiv. <http://arxiv.org/abs/1810.04805>
- [4] Dikiyanti, T. D., Rukmi, A. M., & Irawan, M. I., 2021. Sentiment analysis and topic modeling of BPJS Kesehatan based on twitter crawling data using Indonesian Sentiment Lexicon and



- Latent Dirichlet Allocation algorithm. *Journal of Physics: Conference Series*, 1821(1), 012054. <https://doi.org/10.1088/1742-6596/1821/1/012054>
- [5] Gautam, V., & Gupta, S., 2023. Envisaging Modularity Detecting Communities in Networks: Gephi Visuals. *Evergreen*, 10(4), 2390–2397. <https://doi.org/10.5109/7160929>
- [6] Hairol Anuar, S. H., Abal Abas, Z., Md Yunos, N., Mukhtar, M. F., Setiadi, T., & Shibghatullah, A. S., 2024. Identifying Communities with Modularity Metric Using Louvain and Leiden Algorithms. *Pertanika Journal of Science and Technology*, 32(3), 1285–1300. <https://doi.org/10.47836/pjst.32.3.16>
- [7] Hairol Anuar, S. H., Abas, Z. A., Yunos, N. M., Mohd Zaki, N. H., Hashim, N. A., Mokhtar, M. F., Asmai, S. A., Abidin, Z. Z., & Nizam, A. F., 2021. Comparison between Louvain and Leiden Algorithm for Network Structure: A Review. *Journal of Physics: Conference Series*, 2129(1), 012028. <https://doi.org/10.1088/1742-6596/2129/1/012028>
- [8] Imaduddin, H., A'la, F. Y., & Nugroho, Y. S., 2023. Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach. *International Journal of Advanced Computer Science and Applications*, 14(8). <https://doi.org/10.14569/IJACSA.2023.0140813>
- [9] Kannitha, D. Z. T., Mustafid, M., & Kartikasari, P., 2022. Pemodelan Topik pada Keluhan Pelanggan Menggunakan Algoritma Latent Dirichlet Allocation dalam Media Sosial Twitter. *Jurnal Gaussian*, 11(2), 266–277. <https://doi.org/10.14710/j.gauss.v11i2.35474>
- [10] Kurniawan, A., Nugroho, A., & Al Azam, M. N., 2019. Analisis Jejaring Sosial Tokoh Publik Menggunakan Metode GrapML. *JOINTECS (Journal of Information Technology and Computer Science)*, 4(3), 123. <https://doi.org/10.31328/jointecs.v4i3.1208>
- [11] Patmawati, P., & Yusuf, M., 2021. Analisis Topik Modelling Terhadap Penggunaan Sosial Media Twitter oleh Pejabat Negara. *Building of Informatics, Technology and Science (BITS)*, 3(3), 122–129. <https://doi.org/10.47065/bits.v3i3.1012>
- [12] Rashif, F., Nirvana, G. I. P., Noor, M. A., & Aini, N., 2021. Implementasi LDA untuk Pengelompokan Topik Tweet Akun Bot Twitter bertagar #Covid-19. 7.
- [13] Saadah, S., Kaenova Mahendra Auditama, Ananda Affan Fattahila, Fendi Irfan Amorokhman, Annisa Aditsania, & Aniq Atiqi Rohmawati, 2022. Implementation of BERT, IndoBERT, and CNN-LSTM in Classifying Public Opinion about COVID-19 Vaccine in Indonesia. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 6(4), 648–655. <https://doi.org/10.29207/resti.v6i4.4215>
- [14] Sanandres, E., Lianos, R. A., & Alberto Madariaga Orozco, C., 2020. Topic Modeling of Twitter Conversations: The Case of the National University of Colombia. *Text Analytics, Advances and Challenges*, 241–251. [https://doi.org/10.1007/978-3-030-52680-1\\_19](https://doi.org/10.1007/978-3-030-52680-1_19)
- [15] Schneider, K., & Mühlhaus, T., 2019. FSharpGephiStreamer: An idiomatic bridge between F# and network visualization. *Journal of Open Source Software*, 4(38), 1445. <https://doi.org/10.21105/joss.01445>
- [16] Sliwa, K., Kusen, E., & Strembeck, M., 2024. A Case Study Comparing Twitter Communities Detected by the Louvain and Leiden Algorithms During the 2022 War in Ukraine. 1376–1381. <https://doi.org/10.1145/3589335.3651892>

- [17] Traag, V. A., Waltman, L., & Van Eck, N. J., 2019. From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, 9(1), 5233. <https://doi.org/10.1038/s41598-019-41695-z>
- [18] Wijaya, D. R., Sasmita, G. M. A., & Vihikan, W. O., 2024. Sentiment Analysis of Indonesian Citizens on Electric Vehicle Using FastText and BERT Method. 6(3).
- [19] Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A., 2020. IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding (arXiv:2009.05387). arXiv. <http://arxiv.org/abs/2009.05387>
- [20] Yan, S., Jia, Y., & Wang, X., 2021. Overlapping Community Detection in Temporal Text Networks (arXiv:2101.05137). arXiv. <http://arxiv.org/abs/2101.05137>