

Implementasi *Graph Neural Network* dalam Analisis Jejaring Sosial Media X pada Pilpres Indonesia 2024

Oleh:

Hajran Azbytama Winarya

Billiartho Sastro Cemerson

Muhammad Goldy Wahyu Haryadi

SEMIFINAL
BIG DATA CHALLENGE
2024



Puspresnas
Pusat Prestasi Nasional

BPTI
Badan Pengembangan Talenta Indonesia

Telkom
University

Kampus
Merdeka
INDONESIA JAYA

JUDUL

Implementasi *Graph Neural Network* dalam Analisis Jejaring Sosial Media X pada Pilpres Indonesia 2024

Abstrak

Media sosial telah menjadi arena penting dalam dinamika politik kontemporer, khususnya selama periode pemilihan umum, dengan membentuk opini publik dan memengaruhi preferensi pemilih. Dalam konteks pemilihan residen (pilpres) Indonesia 2024, pemahaman terhadap pola interaksi, pembentukan komunitas, dan sentimen di media sosial menjadi krusial untuk menganalisis lanskap politik yang berkembang. Penelitian ini bertujuan menganalisis dinamika jejaring sosial yang meliputi struktur komunitas, peran *influencer* dan *bot*, serta sentimen terkait pilpres 2024 di platform X. Deteksi komunitas dilakukan menggunakan model turunan *Graph Neural Network* yaitu GraphSAGE+Louvain dan GraphCL+Louvain, dengan evaluasi berdasarkan *modularity*, *coverage*, dan *conductance*. Analisis sentimen dilakukan dengan model DistilBERT multilingual untuk anotasi, dilanjutkan dengan klasifikasi menggunakan Naïve Bayes dan *Support Vector Machine* (SVM). Hasil analisis menunjukkan algoritma GraphSAGE+Louvain bekerja lebih baik dan berhasil mengidentifikasi enam komunitas utama dengan karakteristik interaksi yang beragam. Individu dengan tingkat pengaruh tertinggi atau *most influence* juga teridentifikasi dalam analisis. Ditemukan juga 53% akun teridentifikasi sebagai *bot*, dengan distribusi bervariasi antar komunitas. Analisis sentimen mengungkapkan dominasi sentimen negatif (68,5%) dibandingkan positif (31,5%), dengan SVM menunjukkan performa lebih baik (akurasi 97,31%) dibandingkan Naïve Bayes (akurasi 85,81%). Penelitian ini memberikan wawasan tentang dinamika opini publik dan polarisasi politik di media sosial, serta menyoroti pentingnya pemahaman terhadap peran *bot* dan *influencer* dalam membentuk diskursus politik online menjelang Pemilu 2024.

Kata Kunci: Pilpres; Jejaring Sosial; GNN; Louvain; Analisis Sentimen

I. PENDAHULUAN

Pemilihan presiden (pilpres) Indonesia tahun 2024 menandai periode penting dalam politik nasional, di mana kandidat-kandidat terkemuka berkompetisi dalam kontestasi elektoral yang intens. Dinamika pilpres ini mencerminkan pergeseran dalam taktik

kampanye dan penggunaan *platform* media sosial yang telah berkembang sejak pilpres 2019 (Harahap, 2020). Fenomena ini menandai pentingnya peran teknologi dan analisis data dalam politik *modern* (Stieglitz & Dang-Xuan, 2013)

Intensitas kampanye di media sosial mengakibatkan fenomena segregasi kelompok berdasarkan afiliasi politik, yang dikenal sebagai *echo chambers*. Studi empiris menunjukkan bahwa fenomena ini berpotensi mempengaruhi opini publik dan preferensi elektoral (Garimella et al., 2018). Lebih lanjut, analisis sentimen pada *platform X* telah terbukti mampu memprediksi preferensi pemilih secara signifikan (Chaudhry et al., 2021), menekankan urgensi pemahaman komprehensif terhadap fenomena ini dalam konteks elektoral Indonesia. Oleh karena itu, analisis sentimen perlu dilengkapi dengan analisis jejaring sosial sehingga dapat menghasilkan informasi lebih dalam.

Penelitian ini mengaplikasikan pendekatan *multi-dimensional* dalam analisis jejaring sosial, meliputi deteksi komunitas menggunakan kombinasi *Graph Neural Networks* (GNN) dan algoritma Louvain, identifikasi *most influence*, serta deteksi *bot*. Metodologi ini dipilih berdasarkan efisiensi komputasional dan kemampuannya dalam menganalisis struktur jejaring kompleks (You et al., 2020). Selanjutnya, analisis sentimen dilakukan menggunakan model *Naïve Bayes* dan *Support Vector Machine* (SVM), yang telah terbukti efektif dalam klasifikasi sentimen teks berbahasa Indonesia (Buntoro, 2017; Widowati & Sadikin, 2021; Sujadi, 2022)

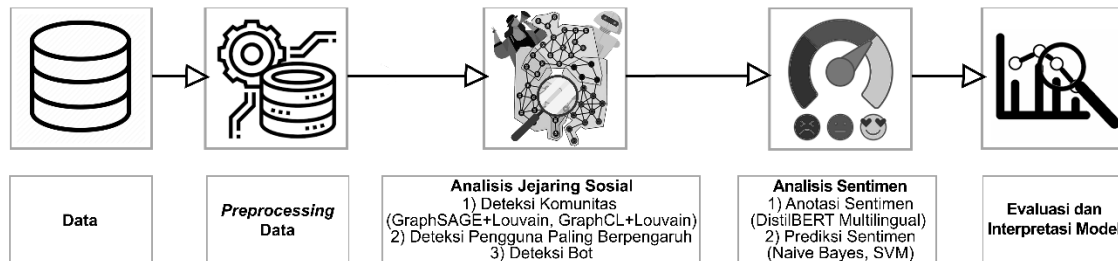
Penelitian ini diharapkan dapat mengisi kesenjangan literatur mengenai analisis jejaring sosial dan sentimen *real-time* dalam konteks elektoral negara berkembang, seperti Indonesia. Dengan fokus pada dinamika pilpres 2024, studi ini bertujuan mengembangkan model komprehensif untuk menganalisis struktur komunitas, peran *influencer* dan *bot*, serta sentimen di *platform X*. Hasil penelitian ini diharapkan dapat memberikan kontribusi signifikan terhadap wawasan akademis dan aplikasi praktis dalam strategi kampanye dan formulasi kebijakan, serta meningkatkan partisipasi politik dan literasi digital masyarakat.

II. METODOLOGI

A. Desain Penelitian

Penelitian ini merupakan studi kuantitatif yang mengintegrasikan metode analisis jejaring sosial (*Social Network Analysis*) dan analisis sentimen. Dalam konteks analisis jejaring sosial, penelitian ini menyajikan hasil deteksi komunitas serta deteksi *most*

influence dan deteksi *bot* dalam komunitas tersebut. Analisis sentimen dilakukan dengan model DistilBERT *multilingual* untuk anotasi sentimen positif-negatif, dilanjutkan dengan pelatihan model prediksi *Naïve Bayes* dan *Support Vector Machine* (SVM). Evaluasi dilakukan pada model deteksi komunitas dan model prediksi analisis sentimen. Tahapan penelitian ini ditampilkan pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

B. Data

Data yang digunakan dalam penelitian ini merupakan kumpulan interaksi dari media sosial X terkait tiga pasangan calon presiden pada masa kampanye pemilihan presiden 2024 yang disediakan oleh Media Kernels Indonesia. Data terdiri atas 9.817.352 baris dan 13 variabel. Variabel yang digunakan dalam penelitian ini dapat dilihat pada Lampiran 1.

C. *Preprocessing* Data

Tahapan *preprocessing* dilakukan untuk menyiapkan data sebelum analisis. Pada tahap awal *preprocessing*, dilakukan penghapusan data dengan *missing value* dan duplikasi. Selanjutnya, *username* dari *tweet* diekstrak untuk membentuk kolom baru yang merepresentasikan interaksi antar pengguna, dan baris dengan hasil ekstraksi kosong dihapus. Selanjutnya dilakukan pemilihan fitur yang relevan untuk analisis. Data dieksplorasi sehingga setiap baris merepresentasikan satu interaksi dengan satu pengguna, dan pengguna dengan interaksi kurang dari dua dihapus.

Untuk analisis sentimen, beberapa *preprocessing* tambahan dilakukan pada konten tweet, seperti *casefolding*, *cleaning*, *tokenizing*, dan *stopword removal*. *Cleaning* dilakukan dengan menghapus kata *rt*, *re*, *username*, tautan, hashtag, karakter emoji, karakter non-alfanumerik, angka, tanda baca, dan *whitespace* berlebih.

D. Analisis Jejaring Sosial

Analisis jejaring sosial dilakukan terhadap data hasil *preprocessing*. Analisis ini bertujuan untuk mengungkap hubungan antar individu yang berperan dalam penyebaran perilaku dan informasi, yang pada akhirnya membentuk sebuah struktur sosial. Metode

ini mengungkap pola tersembunyi, mengevaluasi ukuran sentralitas, serta mengidentifikasi subkelompok koheren atau komunitas dalam jejaring (Theocharis et al., 2017; Wasserman & Faust, 1994). Dalam penelitian ini digunakan tiga metode turunan dari analisis jejaring sosial, yaitu deteksi komunitas, deteksi pengguna paling berpengaruh (*most influence*), dan deteksi *bot*.

1. Deteksi Komunitas

Penelitian ini mengintegrasikan teknik *Graph Neural Network* (GNN) dengan algoritma Louvain untuk deteksi komunitas mengikuti temuan Liu et al., (2024) karena ditunjukkan adanya peningkatan konsisten dalam kinerja model GNN. Studi ini mengadopsi dua inovasi terbaru dalam GNN, yakni *Graph Sample and Aggregate* (GraphSAGE), yang dikembangkan oleh Hamilton et al., (2017), dan *Graph Contrastive Learning* (GraphCL) yang diperkenalkan oleh You et al., (2020). Kombinasi dari kedua model, yaitu GraphSAGE+Louvain dan GraphCL+Louvain, merepresentasikan terobosan dalam deteksi komunitas dan kebaruan dalam penelitian ini.

Pemodelan dilakukan secara terpisah untuk tiga jenis interaksi, yaitu *retweet*, *mention*, dan *reply*, serta satu pemodelan gabungan untuk ketiga interaksi tersebut. Untuk setiap jenis interaksi dan gabungan interaksi, *embedding* dilakukan menggunakan GraphSAGE dan GraphCL untuk menghasilkan representasi vektor dari *node-node* dalam *graph*. Selanjutnya, algoritma Louvain digunakan untuk mendeteksi komunitas berdasarkan *embedding* tersebut.

Algoritma Louvain dimungkinkan menghasilkan komunitas yang terlalu banyak dengan ukuran yang kecil, sehingga sulit diinterpretasikan (Traag et al., 2018). Untuk mengatasi hal tersebut, penelitian ini mengusulkan reduksi komunitas dengan dua pendekatan: pertama, dengan menggabungkan komunitas yang memiliki anggota di bawah *threshold* (T_K) dalam analisis interaksi terpisah; kedua, dengan menggabungkan komunitas di bawah *threshold* (T_G) dalam analisis interaksi gabungan. Dengan demikian, komunitas yang terbentuk memiliki ukuran jumlah anggota yang cukup untuk dianalisis dan relevan terhadap struktur jejaring yang dipelajari.

2. Interpretasi Komunitas

Interpretasi komunitas pada analisis terpisah dan gabungan dilakukan berdasarkan model terbaik, yang dipilih berdasarkan nilai metrik evaluasi yang telah disebutkan sebelumnya. Deteksi pengguna yang paling berpengaruh dalam komunitas dilakukan dengan menerapkan aturan *threshold* sederhana dengan ketentuan: lebih dari 100.000 pengikut, kurang dari 100 *following*, 10.000 daftar publik yang mencantumkan pengguna, atau lebih dari 10 *retweet* yang diterima pada *tweet* tersebut.

Deteksi *bot* dilakukan dengan menerapkan aturan *threshold* sederhana pada variabel jumlah teman (*following*), pengikut (*followers*), dan jumlah status atau *tweet* yang dibuat oleh pengguna. Penelitian ini menggunakan fungsi untuk mengidentifikasi akun sebagai *bot* jika memenuhi salah satu dari tiga kondisi berikut: memiliki lebih dari 5.000 teman tetapi kurang dari 1 pengikut, memiliki lebih dari 10.000 status, atau memiliki rasio teman terhadap pengikut lebih dari 1.000.

E. Analisis Sentimen

Setelah deteksi komunitas, selanjutnya dilakukan analisis sentimen pada konten *tweet*. Beberapa tahapan analisis sentimen yang dilakukan adalah sebagai berikut.

1. Anotasi Sentimen

Anotasi sentimen dalam penelitian ini menggunakan model *pre-trained* DistilBERT *multilingual* yang dikhususkan untuk analisis sentimen. Model yang digunakan adalah "*distilbert-base-multilingual-cased-sentiments-student*" yang dikembangkan oleh Yuan, (2023). Model ini digunakan karena lebih efisien, cepat, dan tidak terbatas pada *lexicon*. Proses anotasi dimulai dengan membagi teks menjadi *chunk-chunk* kecil untuk mengatasi batasan panjang *input* model. Setiap *chunk* dianalisis menggunakan model sentimen, yang mengklasifikasikan teks ke dalam dua kategori: positif dan negatif. Hasil dari seluruh *chunk* kemudian digabungkan; skor sentimen dihitung dengan merata-ratakan skor dari seluruh *chunk*, sementara label sentimen akhir (positif atau negatif) ditentukan berdasarkan label yang paling sering muncul di antara *chunk-chunk* tersebut.

2. TF-IDF (*Term Frequency-Inverse Document Frequency*)

Konten *tweet* yang telah melewati tahap *preprocessing* selanjutnya dilakukan vektorisasi menggunakan metode TF-IDF (*Term Frequency-Inverse Document*

Frequency). TF-IDF menggabungkan dua komponen: *Term Frequency* (TF), yang mengukur frekuensi kemunculan kata dalam dokumen, dan *Inverse Document Frequency* (IDF), yang mengukur seberapa jarang kata tersebut muncul dalam keseluruhan korpus (Kane, 2017)

3. Pembentukan Model Sentimen

Model sentimen yang digunakan adalah *Naïve Bayes* dan *Support Vector Machine* (SVM). *Naïve Bayes* adalah pengklasifikasi terawasi yang sederhana dan efektif berdasarkan penerapan aturan Bayes dengan asumsi independensi bersyarat, yaitu prediktor independen secara kondisional satu sama lain berdasarkan variabel kelas (Tan et al., 2024). Model *Naïve Bayes* yang digunakan dalam analisis ini adalah BinomialNB, yang sangat cocok untuk klasifikasi biner seperti analisis sentimen dengan dua label (positif dan negatif). SVM adalah metode yang memisahkan data menggunakan *hyperplane* untuk memaksimalkan *margin* antara dua kelas data (Hidayat et al., 2021). Dalam analisis ini, SVM digunakan dengan kernel *linear* karena kemampuannya yang lebih unggul dibandingkan kernel lainnya dalam analisis sentimen (Makhtum & Muhajir, 2023).

4. Evaluasi Model

Evaluasi model dilakukan pada analisis deteksi komunitas dan analisis sentimen. Kualitas komunitas yang terbentuk dievaluasi menggunakan nilai *modularity*, *coverage*, dan *conductance*. *Modularity* mengukur kualitas partisi komunitas, dimana nilai yang lebih tinggi menunjukkan struktur komunitas yang lebih baik. *Coverage* mengukur fraksi *edge* yang berada dalam komunitas, dengan nilai yang lebih tinggi menunjukkan bahwa lebih banyak koneksi yang terjadi dalam komunitas daripada antar komunitas. Adapun *conductance* didefinisikan sebagai rasio jumlah *edges* pada batas komunitas atas jumlah derajat *node* di komunitas (Wagenseller & Wang, 2017), dimana nilai yang lebih rendah menunjukkan komunitas yang lebih terisolasi. Nilai *modularity* dan *coverage*, dihitung menggunakan persamaan (1) dan (2), dimana n adalah jumlah *nodes* dalam *graph*, m adalah jumlah *edges* dalam *graph*, k_c adalah jumlah derajat dari *nodes* dalam komunitas c , dan E_c adalah jumlah *edges* dalam komunitas (Setiadi et al., 2023). Sedangkan, nilai *conductance* dihitung menggunakan persamaan (3), dengan E_b adalah jumlah *edges* pada batas komunitas.

$$modularity = \sum_1^n \left[\frac{E_c}{m} - \left(\frac{k_c}{2m} \right)^2 \right] \quad (1)$$

$$coverage = \frac{E_c}{m} \quad (2)$$

$$conductance = \frac{E_b}{k_c} \quad (3)$$

Pada analisis sentimen, evaluasi model menggunakan metrik *accuracy*, *precision*, *recall*, dan *f1-score* (Dalianis, 2018). Model dipilih berdasarkan nilai evaluasi model tertinggi.

III. HASIL DAN PEMBAHASAN

A. Hasil

1. *Preprocessing Data*

Tahap *preprocessing* data dimulai dengan penghapusan data yang mengandung *missing value* dan duplikasi, menghasilkan pengurangan jumlah baris data dari 9.817.354 menjadi 9.817.352. Selanjutnya, dilakukan ekstraksi *username* dari konten *tweet*, menghasilkan kolom baru. Baris dengan hasil ekstraksi kosong dihapus, menyisakan 4.346.024 baris. Data kemudian dieksplorasi untuk memastikan setiap baris merepresentasikan satu interaksi dengan satu pengguna, dan pengguna dengan interaksi kurang dari dua dihapus. Tahap ini menghasilkan dataset akhir dengan 15.822 baris. Informasi *username* yang paling banyak dan paling sedikit muncul dalam konten *tweet* ditampilkan pada Tabel 1.

Tabel 1. Informasi Kemunculan *Username*

Paling Banyak Muncul		Paling Sedikit Muncul	
<i>Username</i>	Jumlah Kemunculan	<i>Username</i>	Jumlah Kemunculan
@aniesbaswedan	4.291	@cakiminnow,	2
@prabowo	1.588	@DevGabriel,	
@ganjarpranowo	1.414	@ronaldy596,	

Berdasarkan Tabel 1, terlihat bahwa @aniesbaswedan menjadi *username* yang paling sering muncul dalam konten *tweet* dengan 4.291 kemunculan, lebih dari dua kali lipat jumlah kemunculan @prabowo dan @ganjarpranowo. Sementara itu, beberapa *username* seperti @ronaldy596, @DevGabriel, dan @cakiminnow hanya muncul 2 kali, yang merupakan jumlah kemunculan minimum yang dipertahankan dalam dataset. Dominasi kemunculan akun-akun calon presiden seperti @aniesbaswedan, @prabowo, dan @ganjarpranowo mengindikasikan tingginya aktivitas dan interaksi seputar figur-figur utama dalam kampanye.

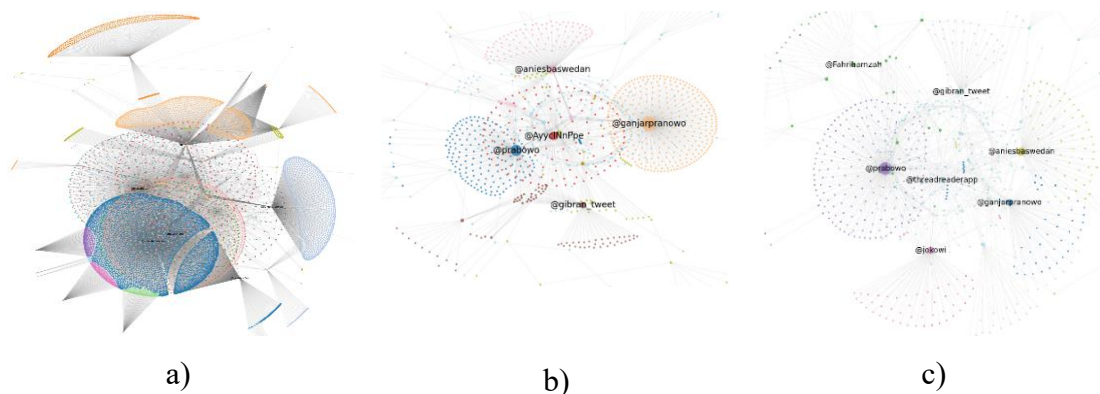
2. Deteksi Komunitas

Hasil evaluasi model GraphSAGE+Louvain dan GraphCL+Louvain dalam deteksi komunitas menggunakan data interaksi terpisah untuk *retweet*, *mention*, dan *reply* ditampilkan pada Tabel 2.

Tabel 2. Evaluasi Model pada Data Interaksi Terpisah

Interaksi	<i>n</i>	Model	<i>Modularity</i>	<i>Coverage</i>	<i>Conductance</i>
<i>Retweet</i>	10.478	GraphSAGE+Louvain	0,7710	0,9454	0,0250
		GraphCL+Louvain	0,7700	0,7564	0,2875
<i>Mention</i>	1.061	GraphSAGE+Louvain	0,6658	0,8992	0,0240
		GraphCL+Louvain	0,7247	0,8205	0,2050
<i>Reply</i>	683	GraphSAGE+Louvain	0,7079	0,9101	0,0131
		GraphCL+Louvain	0,6641	0,7564	0,2875

Hasil evaluasi model pada data interaksi terpisah pada Tabel 2 menunjukkan performa yang bervariasi. Pada interaksi *retweet*, GraphSAGE+Louvain unggul dengan *coverage* lebih tinggi dan *conductance* lebih rendah, mengindikasikan kemampuannya mengidentifikasi komunitas yang lebih kohesif dan terisolasi. Untuk interaksi *mention*, GraphCL+Louvain menunjukkan performa lebih baik dengan *modularity* lebih tinggi, menandakan struktur komunitas yang lebih kuat meski ada *trade-off* dalam *conductance*. Sementara itu, pada interaksi *reply*, GraphSAGE+Louvain unggul di semua metrik. Visualisasi komunitas yang terbentuk untuk setiap interaksi berdasarkan model terbaik ditunjukkan pada Gambar 2 dan Lampiran 2.



Gambar 2. Visualisasi Komunitas Setiap Interaksi a) *Retweet*, b) *Mention*, c) *Reply*

Visualisasi komunitas untuk interaksi *retweet* menunjukkan bahwa model mampu mengidentifikasi komunitas yang kohesif. Komunitas yang terbentuk terlihat rapat dengan garis koneksi yang jelas dan tebal, menunjukkan interaksi yang sering di antara anggota komunitas. Warna-warna berbeda mewakili komunitas yang terisolasi dengan baik, menandakan bahwa *retweet* cenderung terjadi dalam kelompok-kelompok yang memiliki kesamaan minat atau afiliasi. Pada visualisasi

komunitas untuk interaksi *mention*, terlihat bahwa terbentuk komunitas dengan *modularity* yang tinggi, ditunjukkan oleh adanya beberapa pengguna sentral yang terhubung ke banyak komunitas berbeda. Pengguna-pengguna sentral ini sering disebutkan oleh berbagai kelompok, mungkin karena peran mereka sebagai *influencer* atau figur publik. Meskipun ada beberapa *overlap* antara komunitas, struktur keseluruhan menunjukkan adanya koneksi kuat di sekitar pengguna-pengguna kunci.

Visualisasi komunitas untuk interaksi *reply* juga memperlihatkan bahwa model baik dalam mendeteksi komunitas. Setiap komunitas memiliki inti yang kuat, menunjukkan interaksi balasan yang sering di antara anggota-anggotanya, yang mencerminkan diskusi atau percakapan intensif dalam kelompok-kelompok tersebut. Hasil evaluasi kedua model dalam deteksi komunitas menggunakan data semua interaksi ditampilkan pada Tabel 3.

Tabel 3. Evaluasi Model pada Data Semua Interaksi

Model	<i>n</i>	<i>Modularity</i>	<i>Coverage</i>	<i>Conductance</i>
GraphSAGE+Louvain	12.222	0,7822	0,9512	0,0197
GraphCL+Louvain		0,7839	0,6667	0,3528

Berdasarkan Tabel 3, terlihat bahwa kedua model menunjukkan performa yang kompetitif dalam deteksi komunitas. GraphCL+Louvain memiliki nilai *modularity* yang sedikit lebih tinggi, mengindikasikan struktur komunitas yang sedikit lebih kuat. Namun, GraphSAGE+Louvain unggul dalam hal *coverage* dan *conductance*. Hal ini menunjukkan bahwa GraphSAGE+Louvain mampu mengidentifikasi komunitas yang mencakup proporsi *edge* yang lebih besar dan lebih terisolasi dalam konteks keseluruhan interaksi.

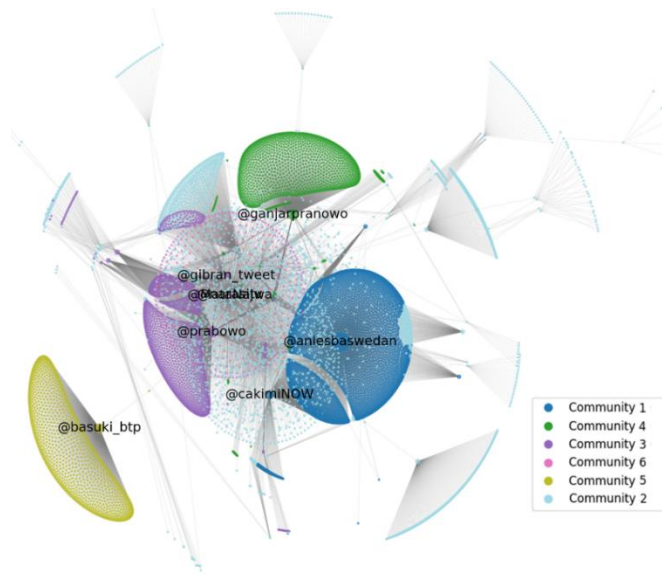
Dalam deteksi komunitas dilakukan reduksi komunitas, dimana komunitas yang berukuran kecil (kurang dari 400 anggota) digabung menjadi satu kategori, dan menghasilkan enam komunitas. Setelah itu, dilakukan pengubahan nama komunitas untuk mempermudah interpretasi. Sebaran anggota pada setiap komunitas ditampilkan pada Tabel 4.

Tabel 4. Sebaran Anggota pada Komunitas

Komunitas	1	2	3	4	5	6
Jumlah	3.963	3.690	1.593	1.416	1.091	469

Pada Gambar 3 Grafik hasil deteksi komunitas menunjukkan enam komunitas dalam jejaring media sosial berdasarkan semua interaksi pengguna. Pengguna

dengan pengaruh besar, seperti @ganjarpranowo, @gibran_tweet, @prabowo, @basuki_btp, @cakimiNOW, dan @aniesbaswedan, ditampilkan dengan ukuran teks lebih besar, menyoroti peran mereka sebagai pusat interaksi.



Gambar 3. Visualisasi Komunitas Semua Interaksi

Setelah terbentuk enam komunitas, selanjutnya dilakukan identifikasi *user* yang paling berpengaruh pada tiap komunitas dan diperlihatkan pada Tabel 5.

Tabel 5. Sampel Data *Most Influence* tiap Komunitas

Komunitas	Username	Followers	Following	List Publik	Retweet
1	@bP+d1TveFxiPs4B/dl0x+oyE2HrUtLxBroNAiykP3T4=	11.058.769	34	11.031	187
2	@9RF507JHRHt5NRIACW0CVktayPYxBN34HA7N6kY6i	11.060.380	34	11032	22
3	@AwilS5kWTBgpz0gtAS1vqiA/P4NBJaDJ4Hrw6QliJn4=	11.060.024	34	11.033	107
4	@GuoAqR7BORqUxd7WQPZfuO2BVmfDszM4euVjUde9	11.061.421	35	11.033	19
5	@z4qviTTe42stKKg6NxjlvIMV2eXY+tQYo7QKlItlsJo=	11.058.303	34	11.029	176
6	@/YUJNigzZOgi8UV26GBhYsHx5icL7QDLfVFocJaq2qE=	11.057.736	33	11.034	87

Pada tabel 5 terlihat bahwa setiap komunitas memiliki pengguna yang sangat berpengaruh dengan karakteristik yang hampir serupa. Hal tersebut menunjukkan bahwa dalam setiap komunitas, ada pengguna-pengguna kunci yang sangat dominan dalam hal popularitas dan interaksi yang kemungkinan besar memainkan peran penting dalam menyebarkan informasi dan membentuk opini dalam komunitas tersebut selama masa kampanye. Dalam analisis ini, dari total 12.086 pengguna,

ditemukan bahwa sebanyak 6.461 akun atau 53% dari total diidentifikasi sebagai *bot*. Sebaran *bot* dalam komunitas ditampilkan pada Tabel 6.

Tabel 6. Sebaran *Bot* pada Komunitas

Komunitas	1	2	3	4	5	6
Jumlah <i>Bot</i>	2.035	1.976	849	793	562	246
Persentase <i>Bot</i>	31,49%	39,59%	13,14%	12,27%	8,70%	3,81%

Tabel 6 menunjukkan bahwa komunitas 1 dan 2 menonjol dengan jumlah *bot* yang signifikan, masing-masing memiliki 2.035 dan 1.976 akun *bot*. Sebaliknya, komunitas 6 menunjukkan jumlah *bot* yang relatif lebih sedikit dengan hanya 246 akun *bot*. Persentase *bot* dalam komunitas 1 dan 2 tinggi, yaitu 31,49% dan 30,59%, sementara komunitas 6 memiliki persentase *bot* terendah sebesar 3,81%.

3. Analisis Sentimen

Analisis sentimen dimulai dengan tahapan *preprocessing* konten *tweet*. Setelah *preprocessing*, konten *tweet* yang tidak memiliki isi dihapus, menyebabkan pengurangan jumlah data dari 12.222 menjadi 12.208 baris. Tabel 7 menampilkan sampel data sebelum dan sesudah *preprocessing*.

Tabel 7. Sampel Data Sebelum dan Sesudah *Preprocessing*

Sebelum <i>Preprocessing</i>	Sesudah <i>Preprocessing</i>
RT Prabowo Dekat Dengan Ulama, Fadli Zon: Sejak Letkol Sudah Datang ke Kiai NU dan Muhammadiyah â @prabowoââ https://t.co/Q5S7eNR5W4 [RE fadlizon]	prabowo ulama fadli zon letkol kiai nu muhammadiyah

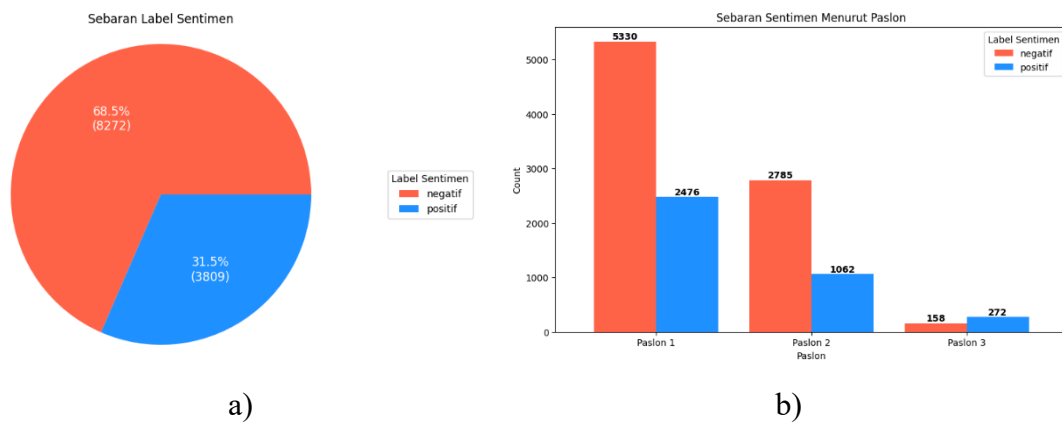
Pelabelan paslon dilakukan dengan pendekatan *keyword-based search* yang berkaitan dengan masing-masing paslon. *Keyword* yang digunakan dalam penentuan label paslon ditampilkan pada Lampiran 3. Dalam proses ini, 122 baris konten *tweet* yang tidak terkait dengan ketiga paslon dihapus. Hasil akhir pelabelan paslon ditampilkan pada Tabel 8 yang menunjukkan distribusi yang tidak merata antara ketiga paslon, dengan paslon 1 memiliki jumlah *tweet* terbanyak, diikuti oleh paslon 2, dan paslon 3 dengan jumlah *tweet* paling sedikit.

Tabel 8. Sebaran Label Paslon

Label Paslon	Paslon 1	Paslon 2	Paslon 3
Jumlah	7.808	3.848	430

Hasil akhir dari proses pelabelan sentimen menggunakan model DistilBERT *multilingual* ditampilkan pada Gambar 4 (a). Sebaran sentimen untuk setiap paslon

ditampilkan pada Gambar 4 (b). Lebih lanjut Tabel 10 menyajikan sebaran sentimen setiap paslon pada masing-masing komunitas.



Gambar 4. Visualisasi Label Sentimen (a) Seluruh *tweet* (b) *Tweet* per paslon

Berdasarkan Gambar 4 (a), terlihat bahwa secara keseluruhan, konten *tweet* memiliki jumlah label negatif lebih banyak sebanyak 8.273, dibandingkan label positif yang berjumlah 3.810. Pada Gambar 4 (b), terlihat bahwa Paslon 1 mendapatkan jumlah sentimen negatif terbesar dengan 5.330 *tweet*, diikuti oleh Paslon 2 dengan 2.785 *tweet*, dan Paslon 3 dengan 158 *tweet*. Paslon 1 juga mendapatkan jumlah sentimen positif terbesar dengan 2.476 *tweet*, diikuti oleh Paslon 2 dengan 1.062 *tweet*, dan Paslon 3 dengan 272 *tweet*. Hal ini menunjukkan bahwa Paslon 1 memiliki eksposur yang lebih tinggi baik dalam sentimen negatif maupun positif dibandingkan dengan dua paslon lainnya.

Tabel 9. Sebaran Sentimen Paslon pada Komunitas

Komunitas	Paslon 1		Paslon 2		Paslon 3	
	Negatif	Positif	Negatif	Positif	Negatif	Positif
1	2.259	1.706	0	0	0	0
2	1.188	586	1.286	249	70	193
3	199	180	689	523	0	0
4	128	22	806	290	88	79
5	1.090	0	1	0	0	0
6	466	0	3	0	0	0

Berdasarkan Tabel 9, paslon 1 menunjukkan jumlah sentimen negatif yang tinggi secara keseluruhan, dengan kritik yang signifikan terhadap mereka. Namun, mereka juga mendapat dukungan positif yang kuat di komunitas 1 dan 2. Di komunitas 5 dan 6, paslon ini lebih banyak mendapat kritik negatif tanpa dukungan positif. Paslon 2 juga menghadapi banyak sentimen negatif, terutama di komunitas 2 dan 4, namun mereka juga mendapat dukungan positif di komunitas 3 dan 4. Secara keseluruhan,

paslon 2 mendapat respon yang beragam dari berbagai komunitas. Paslon 3 memiliki jumlah sentimen negatif yang paling sedikit, namun dukungan positif yang lebih rendah. Komunitas 2 menunjukkan dukungan yang signifikan, sementara di komunitas 4, terdapat keseimbangan antara sentimen positif dan negatif. Hal ini menunjukkan bahwa paslon 3 mungkin memiliki basis dukungan yang lebih seimbang dibandingkan dengan dua paslon lainnya.

Selanjutnya dilakukan pelatihan dan prediksi analisis sentimen menggunakan model *Naïve Bayes* dan SVM. Setelah dilakukan *encoding* label sentimen, dataset dibagi menjadi bagian data latih (*train*) dan data uji (*test*) dengan proporsi 80:20. Dilakukan vektorisasi data dengan TF-IDF, dan penanganan kelas tidak seimbang dengan metode *Synthetic Minority Over-sampling Technique* (SMOTE). Hasil evaluasi metode *Naïve Bayes* dan SVM ditampilkan pada Tabel 11.

Tabel 10. Hasil Evaluasi Model Analisis Sentimen

Model	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
<i>Naïve Bayes</i>	0,8581	0,8432	0,8923	0,8499
SVM	0,9731	0,9720	0,9656	0,9687

Hasil evaluasi model *Naïve Bayes* dan SVM pada Tabel 11 untuk analisis sentimen menunjukkan performa yang baik. Model *Naïve Bayes* mencapai akurasi 85,81%, dengan *precision* 84,32%, *recall* 89,23%, dan *f1-score* 84,99%, menunjukkan kemampuan yang memadai dalam mengklasifikasikan sentimen dari konten *tweet*. Di sisi lain, SVM menunjukkan performa yang lebih tinggi dengan akurasi 97,31%, *precision* 97,20%, *recall* 96,56%, dan *f1-score* 96,87%. Hasil ini menunjukkan bahwa SVM memiliki keunggulan dalam klasifikasi sentimen pada dataset ini.

IV. PENUTUP

Penelitian ini mengeksplorasi interaksi dan sentimen terkait pemilihan umum presiden Indonesia tahun 2024 di platform media sosial X. Model GraphSAGE+Louvain unggul dalam deteksi komunitas pada interaksi *retweet* dan *reply* serta GraphCL+Louvain menunjukkan performa yang baik pada interaksi *mention*. Pada analisis sentimen Model SVM lebih unggul dari *Naïve Bayes* dalam mengklasifikasikan teks. Penelitian ini juga berhasil mengidentifikasi pola-pola komunitas, *bot*, dan *user/akun* yang paling berpengaruh.

Dalam diskusi di X tentang pilpres 2024, paslon 1 (@aniesbaswedan) dan @cakimiNOW mendapatkan perhatian paling banyak, diikuti oleh paslon 2 (@prabowo) @gibran_tweet dan paslon 3 (@ganjarpranowo). Komunitas 1, 2, 5, dan 6 sering membahas paslon 1, sedangkan pada komunitas 3 dan 4 sering membahas paslon 2. Sentimen negatif lebih dominan dalam diskusi di X tentang pilpres 2024. Tingginya keberadaan *bot* dan *user* paling berpengaruh pada setiap komunitas memungkinkan adanya kontribusi pada tingginya sentimen negatif.

Rekomendasi ke Pemerintahan

Bot dan *influencer* memungkinkan penguasaan sosmed X, maka pemerintah perlu melakukan pengawasan dan regulasi secara ketat terkait penggunaan *bot* maupun *influencer*. Melalui akses dan kewenangan yang dimiliki, Pemerintah dapat melakukan intervensi terhadap konten yang sensitif, berbahaya, atau memuat informasi palsu agar tidak disalahgunakan. Pengawasan dan monitoring yang ketat pada sosial media juga harus tetap dilakukan tetapi tetap menerapkan memperhatikan hak privasi dan kebebasan berekspresi warga agar terjaga privasinya.

Rekomendasi ke Kandidat dan Tim Kampanye

Tim kampanye paslon 2 dan 3 harus lebih mengencangkan upayanya dikarenakan paslon 1 sering dibahas tetapi harus tetap mendorong pendukung nyata untuk lebih aktif dalam menyebarkan pesan-pesan positif dan faktual. Strategi yang dapat dilakukan oleh kedua paslon tersebut antara lain pemberdayaan komunitas, program relawan, *engage influencer*, dan pemanfaatan *platform* lain seperti tiktok atau instagram untuk meningkatkan *engagement*. Untuk tim kampanye paslon 1, diperlukan *bot management* dan konten edukatif untuk mengurangi dampak sentimen negatif. Kampanye tatap muka seperti dialog publik secara terbuka, tur daerah, dan kampanye relawan perlu dilakukan untuk mengimbangi sentimen negatif yang berkembang di media sosial, sehingga dapat memperbaiki citra baik dan membangun kepercayaan.

Rekomendasi ke Masyarakat

Masyarakat harus lebih selektif dalam merespon konten agar tidak termakan *hoax* dan konten yang mengarahkan ke hal negatif sehingga merespon buruk, dikarenakan dimungkinkan konten tersebut dibuat oleh *bot* atau *user* berpengaruh yang tidak bertanggung jawab.

V. REFERENSI

- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *INTEGER: Journal of Information Technology*, 2(1). <https://doi.org/10.31284/j.integer.2017.v2i1.95>
- Chaudhry, H. N., Javed, Y., Kulsoom, F., Mehmood, Z., Khan, Z. I., Shoaib, U., & Janjua, S. H. (2021). Sentiment analysis of before and after elections: Twitter data of U.S. election 2020. *Electronics (Switzerland)*, 10(17). <https://doi.org/10.3390/electronics10172082>
- Dalianis, H. (2018). Clinical text mining: Secondary use of electronic patient records. In *Clinical Text Mining: Secondary Use of Electronic Patient Records*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-78503-5>
- Garimella, K., De Francisci Morales, G., Gionis, A., & Mathioudakis, M. (2018). Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. *The Web Conference 2018 - Proceedings of the World Wide Web Conference, WWW 2018*, 913–922. <https://doi.org/10.1145/3178876.3186139>
- Hamilton, W. L., Ying, R., & Leskovec, J. (2017). *Inductive Representation Learning on Large Graphs*. <http://arxiv.org/abs/1706.02216>
- Harahap, I. H. (2020). Kampanye Pilpres 2019 Melalui Media Sosial dan Pengaruhnya Terhadap Demokrasi Indonesia. *Maret*, 17(1), 1.
- Hidayat, T. H. J., Ruldeviyani, Y., Aditama, A. R., Madya, G. R., Nugraha, A. W., & Adisaputra, M. W. (2021). Sentiment analysis of twitter data related to Rinca Island development using Doc2Vec and SVM and logistic regression as classifier. *Procedia Computer Science*, 197, 660–667. <https://doi.org/10.1016/j.procs.2021.12.187>
- Kane, F. (2017). *Hands-on data science and Python machine learning : perform data mining and machine learning efficiently using Python and Spark*. Packt Publishing.
- Liu, C., Han, Y., Xu, H., Yang, S., Wang, K., & Su, Y. (2024). *A Community Detection and Graph Neural Network Based Link Prediction Approach for Scientific Literature*. <http://arxiv.org/abs/2401.02542>
- Makhtum, A. R., & Muhajir, M. (2023). Sentiment Analysis of Omnibus Law Using Support Vector Machine (SVM) with Linear Kernel. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 17(4), 2197–2206. <https://doi.org/10.30598/barekengvol17iss4pp2197-2206>
- Setiadi, T., Yaakub, M. R., & Bakar, A. A. (2023). Community Detection Methods in Library's Books and Borrowers Social Network Segmentation. *Journal of Advances in Information Technology*, 14(6), 1177–1185. <https://doi.org/10.12720/jait.14.6.1177-1185>

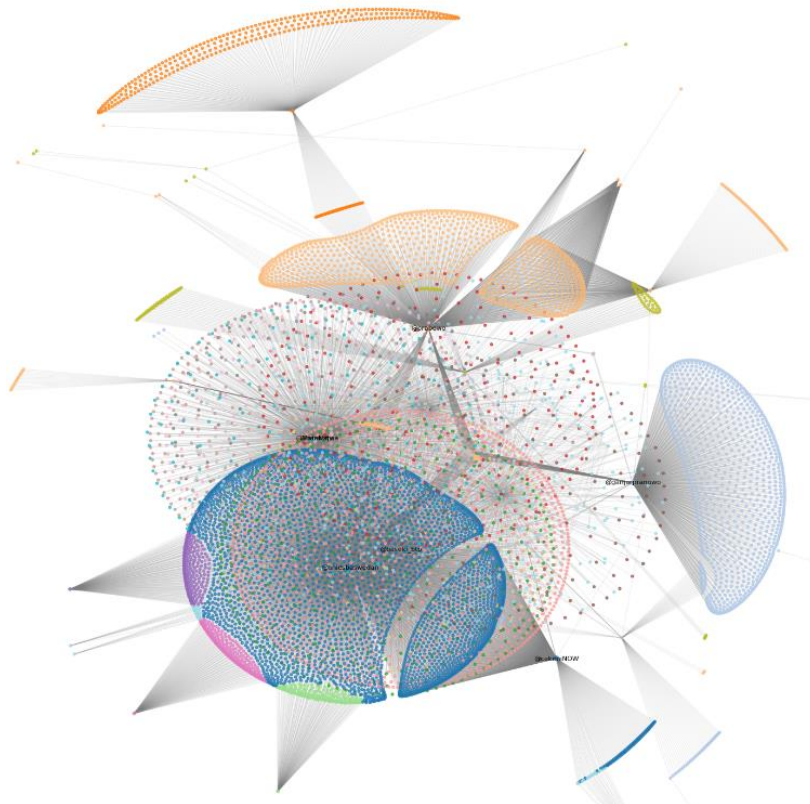
- Stieglitz, S., & Dang-Xuan, L. (2013). Social media and political communication: a social media analytics framework. *Social Network Analysis and Mining*, 3(4), 1277–1291. <https://doi.org/10.1007/s13278-012-0079-3>
- Sujadi, H. (2022). Analisis Sentimen Pengguna Media Sosial Twitter Terhadap Wabah COVID-19 Dengan Metode Naive Bayes Classifier dan Support Vector Machine. *INFOTECH Journal*, 8(1), 22–27. <https://doi.org/10.31949/infotech.v8i1.1883>
- Tan, Y., Sherwood, B., & Shenoy, P. P. (2024). A naïve Bayes regularized logistic regression estimator for low-dimensional classification. *International Journal of Approximate Reasoning*, 172. <https://doi.org/10.1016/j.ijar.2024.109239>
- Theocharis, D., Bekiari, A., & Koustelios, A. (2017). Exploration of Determinants of Verbal Aggressiveness and Leadership through Network Analysis and Conventional Statistics: Using School Class as an Illustration. *Sociology Mind*, 07(02), 27–43. <https://doi.org/10.4236/sm.2017.72003>
- Traag, V., Waltman, L., & van Eck, N. J. (2018). *From Louvain to Leiden: guaranteeing well-connected communities*. <https://doi.org/10.1038/s41598-019-41695-z>
- Wagenseller, P., & Wang, F. (2017). *Size Matters: A Comparative Analysis of Community Detection Algorithms*. <http://arxiv.org/abs/1712.01690>
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511815478>
- Widowati, T. T., & Sadikin, M. (2021). Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 11(2), 626–636. <https://doi.org/10.24176/simet.v11i2.4568>
- You, Y., Chen, T., Sui, Y., Chen, T., Wang, Z., & Shen, Y. (2020). *Graph Contrastive Learning with Augmentations*. <http://arxiv.org/abs/2010.13902>
- Yuan, L. (2023). *distilbert-base-multilingual-cased-sentiments-student*. Hugging Face. <https://doi.org/https://doi.org/10.57967/hf/1422>

VI. LAMPIRAN

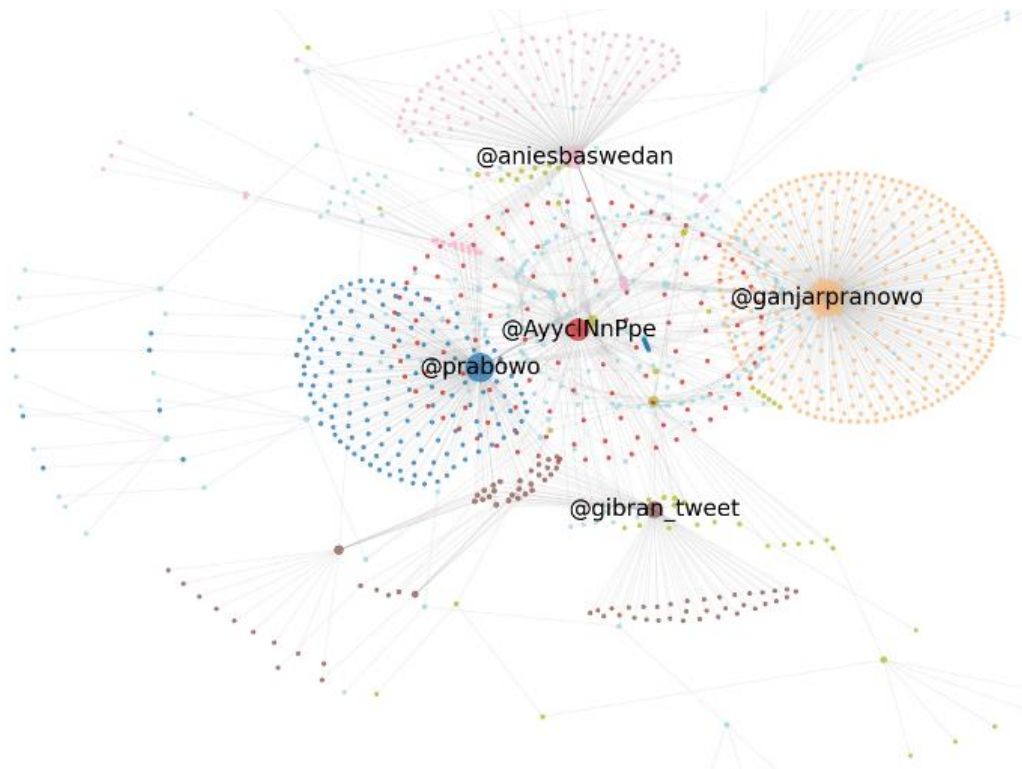
Lampiran 1. Informasi Variabel Penelitian

Variabel	Keterangan
username	Username atau nama pengguna Twitter yang membuat tweet. Dalam data ini, username dienkripsi.
tcode	Kode unik yang menunjukkan jenis tweet (rt = retweet, mention = menyebutkan username lain, reply = membalas tweet).
num_retweets	Jumlah retweet yang diterima oleh tweet tersebut.
frn_cnt	Jumlah teman (following) yang dimiliki oleh pengguna.
flw_cnt	Jumlah pengikut (followers) yang dimiliki oleh pengguna.
sts_cnt	Jumlah status atau tweet yang telah dibuat oleh pengguna.
lst_cnt	Jumlah daftar (lists) publik yang mencantumkan pengguna.
content	Isi atau konten dari tweet. Dalam beberapa kasus, konten tweet mengandung informasi tambahan seperti URL, hashtag, atau username yang disebut (mention).

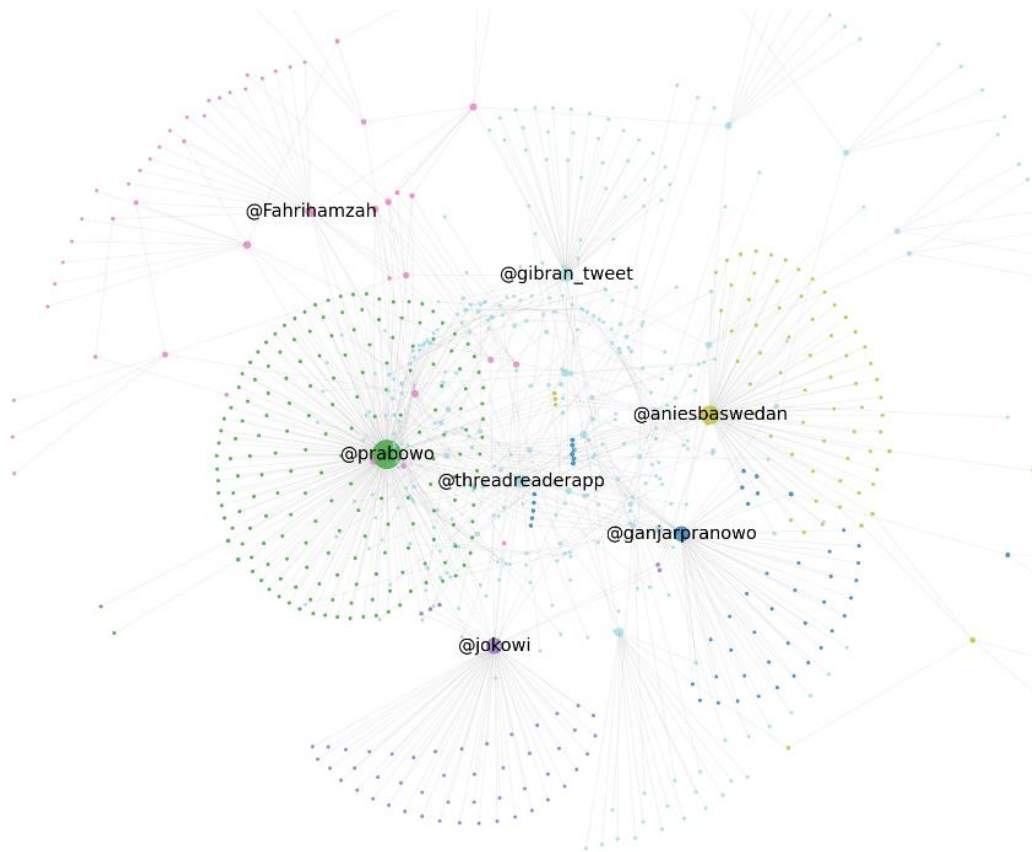
Lampiran 2. Visualisasi Komunitas Menurut Interaksi



Interaksi *Retweet*



Interaksi *Mention*



Interaksi *Reply*

Lampiran 3. Informasi *Keyword* Pelabelan Paslon

Paslon	<i>Keyword</i>
Paslon 1	['anies', 'muhammad', 'amin', 'nasdem', 'pks', 'pkb', '@aniesbaswedan', '@cakiminow', '@nasdem', '@tomlembong', '01', 'paslon 1']
Paslon 2	['prabowo', 'subianto', 'gibran', 'gerindra', 'golkar', 'pan', 'ppp', 'demokrat', '@prabowo', '@gibran_tweet', '@gerindra', '@muzanigerindra', '@fraksi_gerindra', '@bobbynasution_', '02', 'menhan', 'paslon 2', 'gemoy', 'omon']
Paslon 3	['ganjar', 'mahfud', 'pdi-p', 'pdi perjuangan', 'pdip', 'hanura', 'perindo', '@ganjarpranowo', '@pdi_perjuangan', '03', 'paslon 3']

Lampiran 4. Tautan Kode Penelitian

https://drive.google.com/file/d/1w3MRc_OuNGOa6rnLXLkKpPTM_7m2QfDd/view?usp=sharing