

ANALISIS TOPIK PERCAKAPAN DAN SENTIMEN PUBLIK PADA *TWEET* TERKAIT PEMILIHAN UMUM PRESIDEN DI INDONESIA TAHUN 2024

Oleh:

M.Alif Al Hakim

Rakha Abid Bangsawan

Tengku Laras Malahayati



SEMIFINAL
BIG DATA CHALLENGE
2024



Puspresnas
Pusat Prestasi Nasional

BPTI
Badan Pengembangan Telekomunikasi Indonesia

Telkom
University

Kampus
Merdeka
INDONESIA JAYA

ANALISIS TOPIK PERCAKAPAN DAN SENTIMEN PUBLIK PADA *TWEET* TERKAIT PEMILIHAN UMUM PRESIDEN DI INDONESIA TAHUN 2024

Abstrak

Perkembangan media sosial menandai dinamika proses komunikasi, termasuk peran media sosial seperti Twitter dalam aktivitas selama gelaran pemilu presiden Indonesia tahun 2024. Politisi, calon presiden dan wakil presiden, serta masyarakat bertemu dalam ruang komunikasi guna menyalurkan aspirasi, pandangan, dan pendapat tentang berbagai isu terkait ajang pemilu presiden 2024. Penelitian ini merupakan analisis topik percakapan dan sentimen publik pada *tweet* terkait pemilu presiden di Indonesia tahun 2024 yang dilakukan dengan menggunakan metode ekstraksi topik, analisis sentimen, dan analisis jaringan sosial. Ekstraksi topik dilakukan dengan menggunakan *framework* BERTopic. Sementara itu, analisis sentimen dilakukan dengan menggunakan *fine tuning* model *Transformers*. Selain itu, jaringan sosial dianalisis berdasarkan pengguna yang di-*retweet* dan topik pembicaraan. Hasil penelitian ini menunjukkan bahwa debat antar pasangan kandidat merupakan momen yang paling memengaruhi opini dan tanggapan masyarakat, serta teridentifikasinya berbagai pola penyebaran informasi dari setiap pasangan calon. Penelitian ini menyimpulkan pentingnya fokus pada penyebaran program kerja dan pencegahan terlibat pada isu-isu sensitif bagi setiap pasangan calon.

Kata kunci: Analisis Sentimen, Analisis Jaringan Sosial, Ekstraksi Topik, Pemilu Presiden Indonesia tahun 2024

I. PENDAHULUAN

A. Latar Belakang Penelitian

Dalam beberapa tahun terakhir, media sosial telah mengalami pertumbuhan pesat dalam jumlah pengguna dan mengubah cara penyebaran informasi dalam kehidupan sehari-hari (Wigand et al., 2010). Sebelum adanya perkembangan media sosial, penyebaran informasi sering terhambat oleh biaya infrastruktur yang diperlukan untuk mencapai audiens yang luas. Saat ini, dengan luasnya akses terhadap Internet, hambatan ini sebagian besar telah teratasi melalui penggunaan situs jejaring sosial (*Social Network Services/SNS*) (Kaplan & Haenlein, 2010). Selain itu, seiring dengan perkembangan media

sosial, proses pencarian, pengumpulan informasi, dan kemampuan untuk mengumpulkan serta berbagi pendapat dan ide dalam berbagai topik telah berubah secara signifikan (Agrawal et al., 2011). Akibatnya, peningkatan relevansi komunikasi di media sosial menandai perubahan dalam proses komunikasi publik yang sebelumnya dikendalikan oleh aktor-aktor tertentu, seperti politisi, perusahaan, dan jurnalis (Chadwick & Howard, 2009). Fenomena ini menjadi fokus perhatian para ahli dari berbagai disiplin ilmu untuk memahami bagaimana komunikasi digital terjadi, termasuk dalam pembentukan opini di media sosial.

Penggunaan media sosial yang pesat di berbagai aspek kehidupan terlihat pula dalam dunia politik, baik oleh masyarakat umum maupun institusi politik seperti politisi dan partai politik. Media sosial menjadi alat dan sumber informasi yang dianggap ideal untuk menyuarakan aspirasi politik terkait kebijakan dan posisi politik dari para politisi serta untuk membangun dukungan masyarakat terhadap kandidat yang mencalonkan diri untuk jabatan publik (Zeng et al., 2010). Beberapa penelitian telah menunjukkan adanya kebutuhan untuk terus mengumpulkan, memantau, menganalisis, meringkas, dan memvisualisasikan informasi politik yang relevan dari media sosial guna meningkatkan komunikasi antara kandidat dengan masyarakat umum dan pemilih (Zeng et al., 2010; Kavanaugh et al., 2011; Stieglitz & Dang-Xuan, 2013). Sebagai contoh, penting untuk mengidentifikasi jenis pengguna atau opini pemimpin politik agar dapat mengetahui tren diskusi yang terjadi di kalangan mereka. Selain itu, identifikasi isu-isu, tren, serta memprediksi topik-topik yang berpotensi meningkat di kalangan masyarakat juga merupakan langkah penting yang perlu dilakukan.

Salah satu platform media sosial yang kerap digunakan untuk komunikasi politik, khususnya selama kampanye pemilu, adalah Twitter, yang sekarang dikenal dengan nama X. Penggunaan Twitter dalam kampanye politik mulai mencuri perhatian publik saat Barack Obama berhasil memanfaatkannya secara efektif dalam pemilihan presiden Amerika Serikat tahun 2008 (Wattal et al., 2010). Kesuksesan Obama ini dianggap sebagai inspirasi bagi politisi di seluruh dunia untuk mempertimbangkan penggunaan Twitter sebagai bagian dari

strategi kampanye mereka. Sementara itu di Indonesia, Twitter digunakan dalam pemilihan umum tahun 2014 dan 2019 antara calon presiden Joko Widodo dan Prabowo Subianto, serta dalam pemilihan gubernur DKI Jakarta tahun 2017 antara Anies Baswedan dan Basuki Tjahaja Purnama.

Penggunaan media sosial sebagai sarana untuk menginformasikan berbagai hal, termasuk di bidang politik, telah memicu munculnya kajian mengenai analisis sosial media (*Social Media Analytics*). Dari pengamatan dan penelusuran kepustakaan, tampaknya bahasan tentang analisis sosial media masih terbatas. Beberapa penelitian yang telah melakukan analisis sosial media diantaranya adalah Akbar et al. (2021), Tumasjan et al. (2010), serta Allcott dan Gentzkow (2017). Penelitian Akbar et al. (2021) mengkaji penggunaan Twitter sebagai media kampanye dalam pemilu presiden Amerika Serikat tahun 2020, dengan fokus pada analisis isu politik, SARA, serta tema dan sentimen kampanye yang berkaitan dengan kandidat calon presiden saat itu, yaitu Donald Trump dan Joe Biden. Penelitian Tumasjan et al. (2010) menyelidiki apakah Twitter digunakan sebagai forum untuk pertimbangan politik dan apakah topik bahasan pada *tweet* di Twitter mencerminkan situasi politik selama pemilu federal di Jerman tahun 2009. Penelitian Allcott dan Gentzkow (2017) mengkaji pengaruh penyebaran berita hoaks selama pemilu presiden Amerika Serikat tahun 2016. Dari beberapa kajian tentang analisis sosial media tersebut terdapat kesenjangan penelitian dalam hal analisis sentimen dan *social network* dalam penggunaan media sosial Twitter dimana ketiganya menampilkan penggunaan dalam konteks di luar negeri. Kajian tentang analisis sosial media dalam dunia politik di Indonesia masih terbatas, meskipun bila dilihat dari sisi jumlah pengguna, Indonesia tergolong sebagai negara dengan jumlah pengguna media sosial terbanyak di dunia (We Are Social, 2024). Pada posisi inilah penelitian ini dilakukan dengan berfokus pada platform kampanye pemilu presiden di Indonesia pada pemilu presiden tahun 2024.

Pada saat Indonesia menggelar pemilu presiden di tahun 2024, Twitter menjadi platform yang kerap digunakan oleh masyarakat Indonesia untuk memberikan opini tentang pasangan calon presiden dan wakil presiden. Masyarakat Indonesia aktif menyampaikan pandangan mereka mengenai isu-isu

politik yang relevan, serta untuk mengulas kinerja dan program dari setiap pasangan calon presiden. Perilaku pemilih, komunikasi antara pasangan calon presiden dan wakil presiden dengan calon pemilih, serta para politisi dalam gelaran Pemilu Presiden Indonesia tahun 2024 yang terungkap dalam media sosial Twitter menjadi hal yang meramaikan penggunaan sosial media di Indonesia.

Pemanfaatan teknologi ekstraksi topik dan analisis sentimen menggunakan *Artificial Intelligence* dapat membantu pihak tim kampanye ataupun pasangan calon itu sendiri dalam menentukan strategi yang cocok dalam menghadapi Pemilu Presiden tahun 2024. Dengan mengembangkan kerangka kerja untuk menganalisis topik dan sentimen untuk keperluan Pemilu, pihak terkait yang membutuhkan dapat menghemat waktu dan biaya dalam proses pembuatan strategi atau keputusan.

B. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah yang akan tim peneliti ajukan adalah sebagai berikut.

1. Apa topik pembicaraan yang sering dibicarakan oleh pengguna Twitter terkait Pemilu Presiden Indonesia tahun 2024?
2. Momen atau peristiwa seperti apa yang mendapat banyak respon dari masyarakat selama Pemilu Presiden Indonesia tahun 2024?
3. Topik pembicaraan seperti apa yang cenderung mendapatkan respon positif dari masyarakat selama Pemilu Presiden Indonesia tahun 2024?
4. Apakah jumlah *followers* berpengaruh besar terhadap penyebaran informasi suatu topik?
5. Apakah ada pengguna yang menjadi aktor kunci dalam penyebaran informasi mengenai suatu topik?

C. Tujuan Penelitian

Penelitian ini dilakukan untuk mencapai beberapa tujuan sebagai berikut.

1. Mengetahui topik pembicaraan yang sering dibicarakan oleh pengguna Twitter terkait Pemilu Presiden Indonesia tahun 2024.
2. Mengetahui momen apa yang mendapat banyak respon dari masyarakat selama Pemilu Presiden Indonesia tahun 2024.

3. Mengetahui jenis topik pembicaraan yang cenderung mendapatkan respon positif dari masyarakat selama Pemilu Presiden Indonesia tahun 2024.
4. Mengetahui pengaruh jumlah *followers* terhadap penyebaran informasi suatu topik.
5. Mengetahui apakah ada pengguna yang menjadi aktor kunci dalam penyebaran informasi suatu topik.

D. Manfaat Penelitian

Penelitian yang dilakukan memiliki manfaat sebagai berikut.

1. Mendukung pengembangan keilmuan Ilmu Komputer dalam memahami berbagai bidang kehidupan masyarakat termasuk bidang sosial politik. Dengan demikian, Ilmu Komputer tidak hanya ditelisik dari sisi pengembangan aplikasi semata, tetapi juga perlu dikembangkan untuk mendukung kehidupan manusia secara luas.
2. Membuka potensi pengembangan kajian tentang *Social Media Analytics*, khususnya untuk pemilu presiden.
3. Mendorong pengembangan analisa berpikir dan kajian tentang penggunaan media sosial dalam kampanye pemilu presiden.
4. Menjadi sumber aspirasi bagi politisi dalam menyiapkan kampanye politik dengan memanfaatkan *Social Media Analytics*.

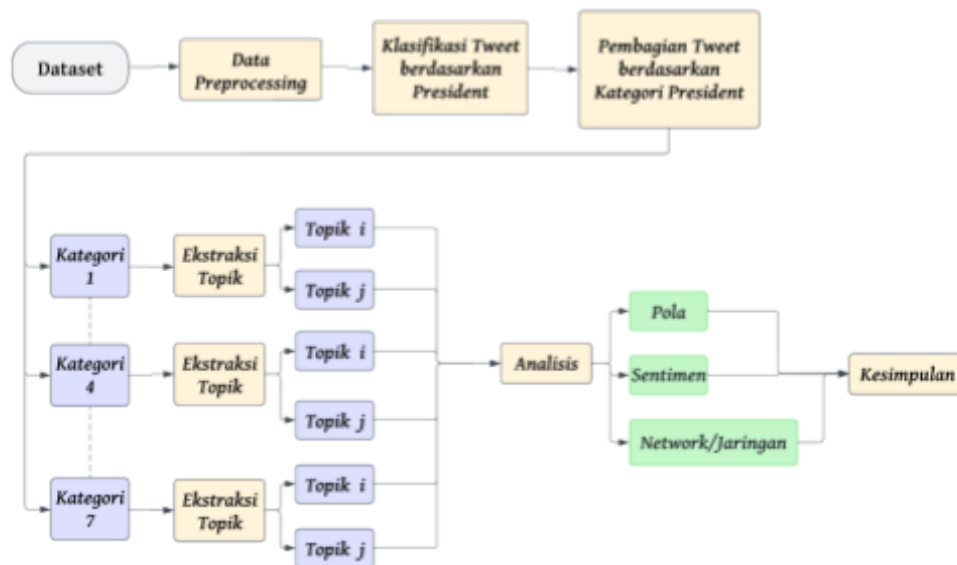
II. METODOLOGI

A. Metode Eksperimen

Penelitian ini menggunakan sebuah *dataset* yang berisi kumpulan interaksi masyarakat Indonesia di media sosial Twitter, yang sekarang dikenal dengan nama X, selama masa kampanye Pemilu Presiden Indonesia tahun 2024. *Dataset* tersebut berisi *tweet* yang berkaitan dengan tiga pasangan calon presiden dan wakil presiden pada Pemilu Presiden Indonesia tahun 2024. Ketiga pasangan tersebut adalah:

- Pasangan nomor urut 01: Anies Baswedan dan Muhaimin Iskandar
- Pasangan nomor urut 02: Prabowo Subianto dan Gibran Rakabuming Raka
- Pasangan nomor urut 03: Ganjar Pranowo dan Mahfud MD

Berikut adalah gambaran alur penelitian yang dilakukan oleh tim peneliti.



Gambar 1: Alur Penelitian

Dalam penelitian ini, akan dilakukan analisis pola pada topik, sentimen, serta struktur jaringan interaksi yang terdapat pada setiap topik pembahasan. Tujuan analisis ini adalah untuk menggali informasi mengenai topik-topik yang ramai dibicarakan, mengidentifikasi reaksi mayoritas masyarakat Indonesia terhadap topik tersebut, dan mengenali pengguna yang berperan penting dalam penyebaran informasi terkait suatu topik.

Secara umum, penelitian dimulai dengan melakukan prapemrosesan data pada *dataset* awal. Tahapan prapemrosesan data mencakup proses pembersihan teks *tweet*, penghapusan kolom, dan penanganan data *tweet* yang memiliki *missing values*. Setiap *tweet* kemudian diklasifikasikan berdasarkan pasangan calon yang sedang dibahas. Proses klasifikasi melibatkan pemodelan yang menggunakan kata-kata kunci yang berkaitan dengan masing-masing pasangan

calon. Hasil klasifikasi yang diperoleh mengelompokkan setiap *tweet* menjadi tujuh kategori berdasarkan pasangan calon yang dibahas dalam *tweet* tersebut. Berikut adalah gambaran pembagian kategori tersebut.

	Kategori 1	Kategori 2	Kategori 3	Kategori 4	Kategori 5	Kategori 6	Kategori 7
Pasangan Calon 1	✓			✓	✓		✓
Pasangan Calon 2		✓		✓		✓	✓
Pasangan Calon 3			✓		✓	✓	✓

Tabel 1: Pembagian Kategori Dataset

Pembagian kategori ini dilakukan untuk membantu menemukan pola-pola yang sifatnya lebih terperinci dan menghindari temuan yang sifatnya global atau terlalu umum.

Setiap kategori kemudian diterapkan ekstraksi topik untuk menemukan pola-pola percakapan masyarakat saat membicarakan pasangan calon tertentu. Proses ekstraksi topik dilakukan dengan menggunakan *framework* BERTopic.¹ Selanjutnya, setiap topik akan dianalisis polanya serta hubungannya dengan fitur lain dalam *dataset*, seperti fitur *follower*, jumlah *retweet*, *timestamp*, dan lainnya. Selain itu, juga akan dilakukan analisis jaringan sosial dan analisis sentimen pada setiap topik untuk menemukan wawasan tambahan.

B. Dataset

1. Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini adalah kumpulan interaksi dari media sosial Twitter terkait tiga pasangan calon presiden selama kampanye Pemilu Presiden Indonesia tahun 2024, yang disediakan oleh Media Kernels Indonesia.² Data ini diperoleh melalui teknik *scraping* pada media sosial Twitter. Hasil *scraping* ini menghasilkan *dataset* besar yang terdiri dari beberapa kolom, antara lain:

¹ <https://maartengr.github.io/BERTopic/>

² Deskripsi Dataset pada Laman *Website* Satria Data

Nama Kolom	Deskripsi Kolom
created_at	waktu <i>tweet</i> dibuat
username	<i>username</i> Twitter yang sudah dienkrpsi
content	teks <i>tweet</i>
interaction_type	jenis interaksi pengguna (<i>retweet</i> , <i>mention</i> , atau <i>reply</i>)
following_count	jumlah <i>following</i> (teman)
followers_count	jumlah <i>followers</i> (pengikut)
tweet_count	jumlah <i>tweet</i> yang telah dibuat oleh pengguna
location	lokasi yang dicantumkan di profil pengguna
language	bahasa yang terdeteksi dari konten <i>tweet</i>

Tabel 2: Deskripsi Dataset

2. Konstruksi Data

Data Understanding. Pada penelitian ini, tim peneliti diberikan *dataset* dengan total 9,8 juta atau 9.817.355 baris konten *tweet* terkait tiga pasangan calon presiden pada masa kampanye pemilihan presiden 2024. *Dataset* yang akan dianalisis memiliki 12 kolom seperti yang sudah dijelaskan pada bagian Deskripsi Dataset.

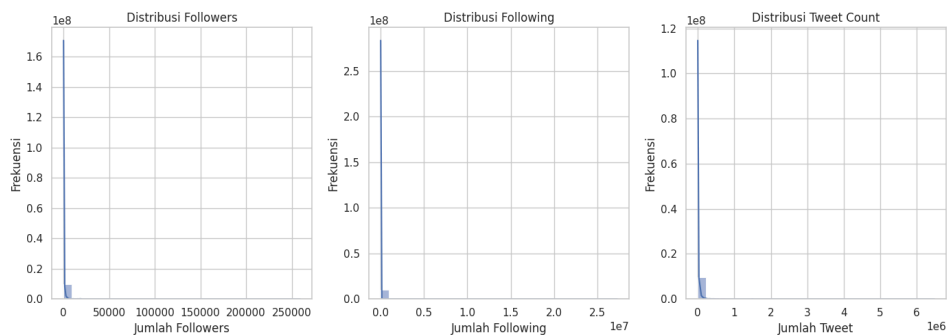
Kolom *username* Twitter tidak dapat tim peneliti lakukan analisis lebih lanjut karena *username* tersebut telah dienkrpsi dan tim peneliti tidak diberikan informasi mengenai algoritma enkripsi atau *key* sehingga tim peneliti tidak dapat melakukan dekripsi untuk mengetahui *username* Twitter yang melakukan interaksi.³

Tim peneliti memutuskan untuk menghapus kolom “type”, “lang”, dan “loc” dari *dataset* karena beberapa alasan. Kolom “type” hanya memiliki satu nilai, yaitu “twit” sehingga tidak memberikan

³ Is it theoretically possible to decrypt a file without the key/password? - Information Security Stack Exchange

variasi data dan tidak memiliki *missing value* karena jumlahnya sama dengan jumlah baris data (9.817.355). Kolom “lang” menunjukkan 13 bahasa berbeda sesuai standar ISO 639-1⁴, tetapi analisis lebih lanjut menunjukkan banyaknya kesalahan deteksi bahasa sehingga data tersebut tidak akurat seperti pada Gambar 2. Terakhir, kolom loc memiliki banyak *missing value*, sekitar 60% dari total data, dan banyak data noise yang dihasilkan dari pengguna yang meng-*custom* informasi lokasi mereka sehingga tidak memberikan informasi yang berguna.

Gambar 3 merupakan visualisasi persebaran data dari jumlah *following* dan *followers* yang dimiliki oleh pengguna, serta *tweet* yang dibuat oleh pengguna.



Gambar 3: Distribusi Data *Following*, *Followers*, dan Jumlah *Tweet* Pengguna

Berdasarkan gambar 3 dan setelah dianalisis lebih lanjut, tim peneliti mendapatkan informasi bahwa dataset yang diberikan berisi berbagai tipe pengguna, mulai dari yang terkenal (dilihat dari banyak *followers*) sampai tidak terkenal. Tim peneliti menduga bahwa pengguna yang memiliki *followers* yang banyak akan berdampak pada suatu topik yang sedang dibahas. Contoh dampak yang dimaksud adalah dapat mempengaruhi atau menggiring opini terkait topik yang sedang dibahas.⁵

Data Cleaning. Sebelum melakukan tahap pemrosesan lebih lanjut, dataset yang diberikan perlu dibersihkan terlebih dahulu. Proses prapemrosesan ini melibatkan beberapa langkah sebagai berikut.

⁴ ISO 639-1 - Wikipedia

⁵ <https://www.mdpi.com/2076-0760/12/7/402>

- 1) **Memperbaiki atau menormalisasi teks dalam *tweet*.** Terdapat data teks yang memiliki masalah seperti ada karakter Unicode yang tidak valid. Untuk memperbaiki masalah tersebut, tim peneliti menggunakan library *ftfy* (*fixes text for you*) sehingga teks yang dihasilkan lebih konsisten dan dapat diinterpretasikan dengan baik oleh algoritma atau tahapan berikutnya.
- 2) **Menangani data yang hilang.** Data yang hilang akan dihapus dari dataset.
- 3) **Menangani data *noise*.** Data *noise* seperti pada kolom *loc* akan dihapus dari dataset.
- 4) **Menggunakan metode *Regex* dan *method* dari *python*.** Metode terbaik untuk membuat data teks pada kolom *content* lebih konsisten adalah dengan metode *Regex* dan *method* dari *python* seperti *lower()* dan *replace()* yang dapat dilihat pada Tabel 3.

Salah satu contoh hasil dari proses prapemrosesan di atas dapat dilihat pada Tabel 4.

C. Pemodelan

1. Model Klasifikasi Presiden

Untuk mengklasifikasikan pasangan calon presiden dan wakil presiden yang disebut atau dibahas dalam suatu *tweet*, dibangunlah sebuah model algoritma pencocokan frasa. Model ini dirancang untuk mengenali berbagai variasi penyebutan terkait pasangan calon, termasuk nama lengkap, nama panggilan, singkatan, dan kombinasi nama pasangan. Pengembangan model dimulai dengan pembangunan kamus yang mengaitkan setiap pasangan calon dengan daftar kata kunci yang terkait.

Proses klasifikasi teks melibatkan dua langkah utama. Pertama, model mencoba mencocokkan frasa secara tepat dengan daftar kata kunci. Jika tidak ditemukan, model beralih ke pencocokan dengan algoritma *fuzzy*. Pencocokan secara *fuzzy* dapat membantu menemukan *string* yang mirip walaupun tidak identik. Teknik pencocokan secara

fuzzy yang digunakan dalam penelitian ini adalah algoritma rasio parsial dengan ambang batas kemiripan sebesar 80 persen. Kemudian, indikator biner dibuat untuk setiap pasangan calon agar dapat menunjukkan apakah suatu *tweet* menyebutkan pasangan calon tertentu atau tidak. Indikator ini kemudian digunakan untuk membagi *dataset* menjadi tujuh kategori berbeda seperti pada Tabel 1.

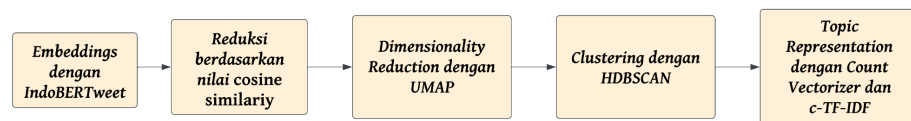
2. Ekstraksi Topik

Topik diekstrak menggunakan *framework BERTopic*. Namun, sebelum itu, *dataset* untuk masing-masing kategori dilakukan prapemrosesan berupa penghapusan *tweet* yang duplikat. Hal ini perlu dilakukan agar model tidak bias terhadap data yang berulang dan mencegah model gagal menggeneralisir suatu topik. Kemudian, setiap *tweet* diterapkan ekstraksi fitur-fiturnya menggunakan model *Transformers IndoBERTweet*. Model ini dipilih karena telah dilatih pada data Twitter dalam Bahasa Indonesia sehingga memiliki kamus yang lebih sesuai dan lebih besar untuk data teks yang akan dianalisis. Hasil *embeddings* kemudian dilakukan perhitungan *cosine similarity* antar setiap *tweet* untuk menemukan *tweet* yang serupa antara satu *tweet* dan *tweet* lain.. Hal ini perlu dilakukan karena terdapat *tweet* yang hanya berbeda satu atau dua kata, baik karena sengaja dibuat oleh pengguna maupun akibat enkripsi *username* pada *mention*. Sebagai contoh, *tweet* “@QOS7XYPB Pasangan calon 01 punya kebijakan yang bagus” dan “@Mwvr2Fp Pasangan calon 01 punya kebijakan yang bagus” merupakan *tweet* yang mirip setelah dienkripsi meskipun keduanya berasal dari akun Twitter yang berbeda.

Hasil *embeddings* kemudian dilakukan *dimensionality reduction* menggunakan UMAP (Uniform Manifold Approximation and Projection). Hal ini perlu dilakukan agar menghindari terjadi *high dimensionality curse*. Fitur yang telah direduksi kemudian dikelompokkan dalam beberapa klaster menggunakan model HDBSCAN. Model ini digunakan karena model ini tidak memaksakan

setiap data untuk termasuk dalam suatu kategori topik, sehingga menghasilkan kategori yang berisi *outliers*. Hal ini memastikan kemurnian topik yang diekstrak dari *tweet* lain yang sebenarnya tidak berkaitan.

Klaster yang telah dihasilkan kemudian direpresentasikan dengan menggunakan *count vectorizer* dan *c-TF-IDF*. Pada *count vectorizer*, ditambahkan daftar kata *stopwords* agar kata-kata tersebut tidak diperhitungkan menjadi representasi topik. *Stopwords* terdiri dari *stopwords* dalam Bahasa Indonesia, bahasa *gaul*, dan kata-kata umum yang berkaitan dengan Pemilu Presiden Indonesia tahun 2024, seperti nama paslon, 'capres', 'cawapres', dan lainnya. Setiap kata kunci kemudian dihitung bobotnya menggunakan *c-TF-IDF*.



Gambar 4: Alur Ekstraksi Topik

Model kemudian digunakan untuk memprediksi topik keseluruhan *dataset*. Hasil prediksi topik kemudian dianalisis polanya berdasarkan waktu pembuatan *tweet*, pengikut pengguna, jumlah *retweet*, sentimen, dan lainnya.

3. Analisis Sentimen

Analisis sentimen dilakukan dengan menggunakan tiga model *pretrained* yang berbeda, yaitu Indo-RoBERTa,⁶ Indo-BERT,⁷ dan Indo-BERTweet.⁸ Ketiga model ini telah dilatih untuk tugas klasifikasi sentimen dalam bahasa Indonesia dan menunjukkan performa yang baik pada eksperimen-eksperimen terdahulu berdasarkan hasil yang telah

⁶ wllwo/indonesian-roberta-base-sentiment-classifier · Hugging Face

⁷ mdhugol/indonesia-bert-sentiment-classification · Hugging Face

⁸ ridhodaffasyah/sentiment-analysis-indobertweet · Hugging Face

dilakukan pada platform Hugging Face. Setiap topik akan diklasifikasikan sentimen berdasarkan setiap *tweet*-nya menggunakan model-model tersebut.

Penggunaan beberapa model ini bertujuan untuk mencapai triangulasi hasil. Jika ketiga model memberikan hasil yang serupa, hal ini dapat meningkatkan kepercayaan terhadap validitas prediksi sentimen yang dilakukan. Selain itu, perbandingan antara hasil prediksi dari ketiga model dapat memberikan wawasan tambahan mengenai kelebihan dan kekurangan masing-masing model dalam konteks data *tweet* yang dianalisis.

4. Analisis Jaringan Sosial berdasarkan Akun Pengunggah

Pada dataset terdapat kolom *username* dan *mention* pada kolom *content*. Namun, *username* tersebut telah dienkripsi sehingga tidak memungkinkan untuk dilakukan analisis jaringan sosial. Meskipun demikian, setiap *retweet* memiliki kode pada akhir *tweet* seperti berikut “[re nama_akun_pengunggah]”. Tim peneliti memanfaatkan informasi ini untuk melakukan analisis terhadap akun pengunggah terkait topik yang telah diekstrak.

Tujuan dari analisis ini adalah untuk mengidentifikasi akun-akun yang berperan dalam penyebaran informasi mengenai suatu topik. Perlu diingat bahwa informasi paling banyak menyebar melalui *retweet* sehingga akun-akun pengunggah tersebut dapat menjadi aktor kunci dalam penyebaran topik tertentu. Dengan demikian, analisis ini dapat membantu memahami peran penting akun-akun tertentu dalam dinamika penyebaran informasi mengenai Pemilu Presiden Indonesia tahun 2024 di media sosial Twitter.

III. HASIL DAN PEMBAHASAN

Berdasarkan analisis mengenai topik, sentimen, dan jaringan sosial yang telah dilakukan, tim peneliti memperoleh beberapa hasil dan temuan yang menarik. Hasil dan temuan ini juga dikaitkan dengan data pendukung berupa

berita dan informasi yang beredar di berita mengenai masing-masing pasangan calon selama masa kampanye Pemilu Presiden Indonesia tahun 2024.

1. Debat Calon Presiden dan Wakil Presiden

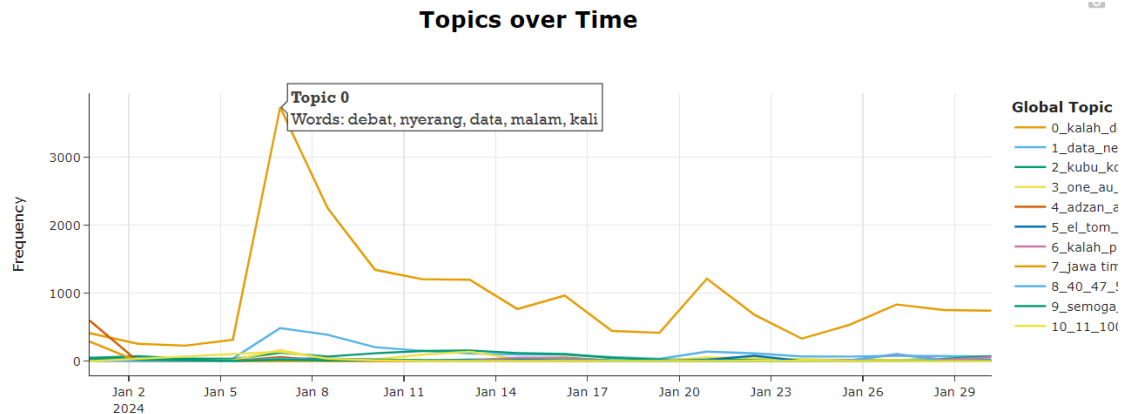
Selama masa kampanye Pemilu Presiden Indonesia tahun 2024, Komisi Pemilihan Umum (KPU) mengadakan debat kandidat pasangan calon presiden dan wakil presiden. Acara ini berperan sebagai ajang untuk menyebarluaskan profil, visi, misi, dan program para pasangan calon kepada pemilih dan masyarakat (Wahyuni, 2024). Selain itu, debat ini memberikan informasi menyeluruh kepada pemilih sebagai salah satu pertimbangan dalam menentukan pilihan, serta menggali dan mengelaborasi lebih dalam setiap tema yang diangkat dalam kampanye pemilu.

Ajang debat ini juga menjadi sarana bagi masyarakat untuk mengetahui pandangan pasangan calon mengenai isu-isu tentang kehidupan berbangsa dan bernegara di Indonesia. Tahun ini, acara debat kandidat diselenggarakan sebanyak lima kali dengan beberapa tema berbeda, yang mencakup hukum dan Hak Asasi Manusia (HAM), pertahanan dan keamanan, ekonomi dan perdagangan, energi dan Sumber Daya Alam (SDA), serta teknologi informasi dan ketenagakerjaan (Badan Kesbangpol Kulon Progo, 2023).

Debat ini menjadi salah satu peristiwa yang mengundang reaksi masyarakat. Hal ini juga dapat terlihat dari hasil analisis *dataset*. Berdasarkan analisis *dataset* kategori 4, 5, 6, dan 7, banyak pengguna Twitter yang membicarakan dan membentuk opini mengenai ketiga pasangan calon. Dari pembicaraan tersebut, tim peneliti memperoleh bahwa debat antara pasangan calon nomor urut satu dan dua merupakan topik bahasan yang kerap dibicarakan masyarakat pada Twitter.

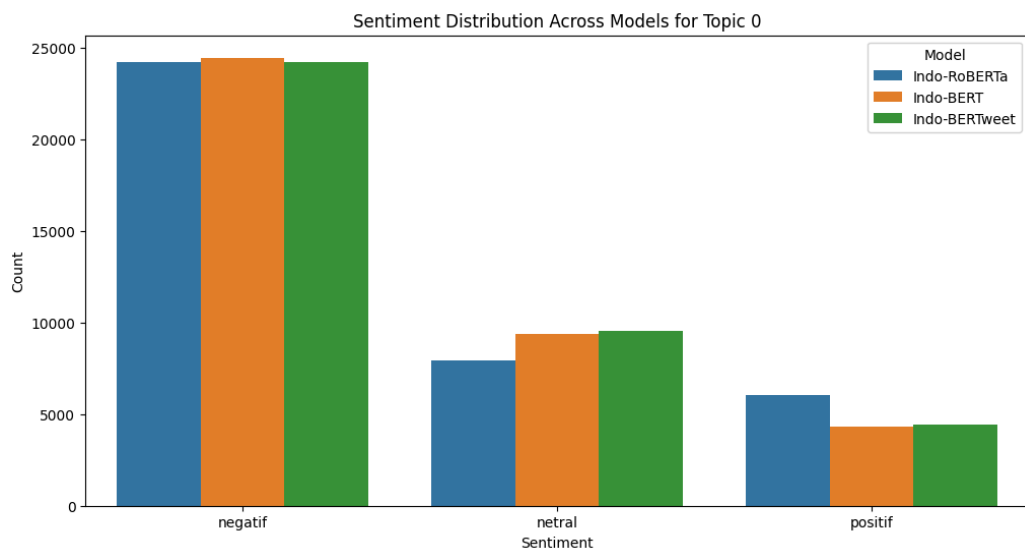
Selain itu, pada gambar 5, terlihat banyak pembicaraan mengenai alutsista negara dan persentase lahan. Kedua topik tersebut merupakan tema acara debat kandidat ketiga dan keempat yang membahas pertahanan dan keamanan serta energi dan SDA. Data pertahanan negara juga menjadi salah satu topik perbincangan masyarakat Indonesia di Twitter. Topik tersebut menjadi

salah satu pembicaraan hangat pada saat itu karena terdapat diskursus mengenai keterbukaan informasi perihal data serta fakta pertahanan dan keamanan negara (Kompas.com, 2024). Adanya lonjakan jumlah *tweet* sekitar tanggal 7 dan 8 Januari 2024 juga menjadi bukti pendukung bahwa debat kandidat menjadi pusat perhatian masyarakat karena debat mengenai pertahanan dan keamanan negara diadakan pada tanggal 7 Januari 2024.



Gambar 6: Perkembangan dan Frekuensi Topik Tiap Waktu Dataset Kategori 4

Berdasarkan topik-topik yang dibahas pada Gambar 5, terlihat bahwa mayoritas *tweet* membahas mengenai pasangan calon nomor urut 2. Hasil klasifikasi sentimen juga menunjukkan bahwa sebagian besar topik tersebut memiliki hasil bersentimen negatif.



Gambar 7: Klasifikasi Sentimen Topik 0 Dataset Kategori 4

Dari uraian dan visualisasi di atas, tim peneliti menyimpulkan bahwa debat kandidat presiden dan wakil presiden memengaruhi opini masyarakat Indonesia dalam media sosial Twitter.

2. Gaya Kampanye Setiap Pasangan Calon

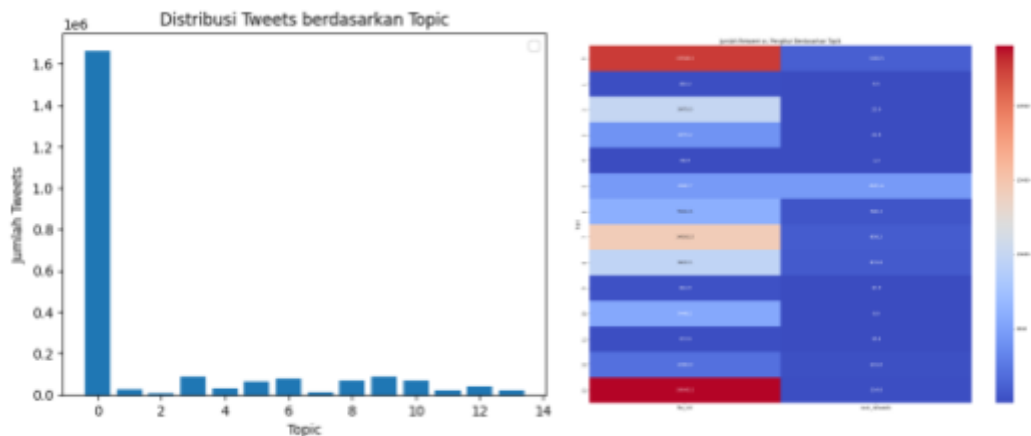
Metode atau gaya kampanye mengenai pasangan calon nomor urut satu menjadi salah satu topik yang sering dibicarakan. Pada gambar 8, terlihat bahwa kata ‘desak’ berperan penting dalam membentuk topik 0 mengenai kampanye pasangan calon nomor urut satu. Hal ini merujuk kepada gaya kampanye pasangan calon nomor urut satu yang mengadakan suatu program bernama Desak Anies. Selain itu, juga terlihat bahwa pasangan calon nomor urut satu dapat menarik perhatian dari kelompok penggemar musik korea atau yang biasa disebut sebagai *kpopers*. Akan tetapi, berdasarkan gambar 9 dan gambar 10, terlihat bahwa masih terdapat banyak perbedaan pendapat ataupun polarisasi terhadap gaya kampanye tersebut. Berdasarkan *tweet* pada *dataset* kategori 5, pasangan calon ini mendapatkan sentimen buruk terkait dengan komentarnya terhadap pengungsi rohingya dan adanya indikasi penggunaan adzan sebagai alat kampanye, seperti yang terlihat pada gambar 11 dan gambar 8, serta hasil analisis sentimennya pada gambar 12 dan gambar 13.

Sementara itu, informasi topik mengenai kampanye pasangan calon nomor urut dua tidak terlalu sering dibicarakan. Akan tetapi, pasangan calon nomor urut dua sering mendapatkan perhatian terkait dengan *tweet* yang membahas tentang kegiatan kemanusiaan dan juga mengenai santri seperti terlihat pada gambar 14. Komentar-komentar pasangan calon ini juga sering menjadi pusat pembicaraan publik. Komentar-komentar tersebut diantaranya dikeluarkan saat debat dan menghasilkan sentimen buruk dari masyarakat.

Pasangan calon nomor urut tiga memiliki *tweet* yang berpusat pada program-program kerja yang dimilikinya. Program-program tersebut berpusat kepada rakyat, seperti petani dan nelayan, serta juga membahas mengenai keamanan digital. Topik-topik mengenai program kerja ini mendapat sentimen yang baik dari masyarakat seperti terlihat pada gambar 15, 16, 17, 18, dan 19.

3. Pengaruh Jumlah *Followers* terhadap Kepopuleran Topik

Pada penelitian ini, tim peneliti menyadari bahwa metode *retweet* menjadi metode yang paling banyak digunakan dalam menyebarkan informasi, seperti pada gambar 20. Namun, setelah dilakukan analisis, tim peneliti menemukan bahwa tidak adanya kaitan antara jumlah *followers* dengan kepopuleran topik. Pengguna dengan *followers* banyak tidak menjamin *tweet*-nya akan di-*retweet* oleh banyak pengguna lain ataupun topik yang sama dengan *tweet*-nya juga dibahas oleh orang lain. Berikut adalah salah satu visualisasi yang menggambarkan keadaan tersebut.



Gambar 21: Visualisasi Heatmap *Followers* dan *Number of Retweets* dan Frekuensi Topik

Terlihat bahwa topik 13 jarang dibahas walaupun rata-rata *follower* yang dimiliki oleh pengguna tergolong tinggi.

4. Analisis Jaringan Interaksi

Berdasarkan interaksi antara akun pengunggah dan topik yang dibahasnya, tim peneliti menemukan beberapa pola sebagai berikut.

1. Akun “mdy_asmara1701” dan “aniesbubble” memiliki peran penting dalam menyebarkan informasi mengenai kampanye pasangan calon nomor urut satu, seperti terlihat pada gambar 22 jaringan interaksi kategori 1 pada lampiran.
2. Akun “pdi_perjuangan” memiliki peran penting dalam menyebarkan informasi mengenai program pasangan calon nomor urut tiga. Hal ini

terlihat pada gambar 23 jaringan interaksi kategori 3 yang terdapat pada lampiran.

3. Akun-akun berita, seperti “tribunnews”, berperan penting dalam menyebarkan informasi mengenai isu-isu faktual yang terkait dengan pasangan calon. Contohnya topik mengenai persentase lahan pada data kategori 4 ataupun isu mengenai koalisi antar pasangan calon pada data kategori 5. Hal ini terlihat pada gambar 24 jaringan interaksi kategori 4 dan gambar 25 kategori 5 pada lampiran.

IV. PENUTUP

1. Kesimpulan

Berdasarkan analisis yang telah dilakukan, tim peneliti menemukan beberapa temuan utama. Pertama, debat presiden dan wakil presiden menjadi momen yang memicu respon besar dari pengguna Twitter di Indonesia. Kedua, mayoritas topik yang dibahas adalah mengenai dukungan terhadap pasangan calon presiden dan wakil presiden tertentu. Selain itu, terdapat diskusi mengenai gaya kampanye masing-masing pasangan calon, dimana pasangan calon nomor urut satu menonjolkan gaya kampanye yang lebih kekinian, pasangan calon nomor urut dua berfokus pada tindakan kemanusiaan, dan pasangan calon nomor urut tiga menekankan program kerjanya. Ketiga, tidak ditemukannya korelasi yang signifikan antara jumlah *followers* dengan popularitas topik pembicaraan. Keempat, aktor-aktor kunci sering kali muncul dalam berbagai topik terkait pasangan calon atau isu tertentu. Terakhir, topik pembicaraan yang berorientasi pada kepentingan rakyat cenderung memiliki sentimen positif, sementara topik yang berkaitan dengan isu-isu sensitif seperti agama atau golongan cenderung memiliki sentimen yang beragam.

2. Saran

Beberapa saran dapat dihasilkan dari hasil analisis yang telah dilakukan. Berikut adalah saran yang dapat tim peneliti berikan.

1. Peneliti lain dapat mengembangkan kajian serupa dengan menggunakan Social Media Analytics guna memperkaya kajian tentang media sosial dalam berbagai bidang kehidupan.
2. Dalam bidang politik, politisi dan orang-orang yang berminat mencalonkan diri dalam ajang pemilihan kepala daerah, presiden, ataupun jabatan penting lainnya di pemerintahan sebaiknya menghindari komentar apapun terhadap isu-isu sensitif karena akan berdampak pada penurunan persepsi masyarakat.
3. Setiap pasangan calon ataupun tim kampanye sebaiknya memfokuskan kepada penyebaran informasi mengenai program-program yang akan dijalankan.
4. Pasangan calon sebaiknya mempersiapkan diri dengan lebih baik dalam menghadapi momen-momen seperti debat karena akan menjadi pusat perhatian masyarakat.
5. Jumlah *followers* seorang pengguna tidak menjamin meningkatkan kepopuleritasan suatu topik. Oleh karena itu, tim kampanye sebaiknya memfokuskan pada penyampaian program yang menarik.
6. Penting bagi setiap pasangan calon dan tim kampanye untuk menjaga *image*-nya di depan pers. Hal ini karena informasi-informasi faktual akan tersebar dengan cepat melalui akun-akun berita atau pers tersebut.

V. REFERENSI

- Wigand, R. T., Wood, J. D., & Mande, D. M. (2010, 8 5). Taming the Social Network Jungle: From Web 2.0 to Social Media. *AMCIS 2010 Proceedings*, 416.
<https://aisel.aisnet.org/amcis2010/416>
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, 53(1), 59-68.
<https://www.sciencedirect.com/science/article/pii/S0007681309001232>
- Agrawal, D., Budak, C., & Abbadi, A. E. (2011). Information diffusion in social networks: observing and influencing societal interests. *Proceedings of the VLDB Endowment*, 4(12), 1512-1513.
- Chadwick, A., & Howard, P. N. (Eds.). (2009). *Routledge Handbook of Internet Politics*. Routledge.
- Zeng, D., Chen, H., Lusch, R., & Li, S.-H. (2010). Social Media Analytics and Intelligence. *IEEE Intelligent Systems*, 25(6), 13-16.
- Wattal, S., Schuff, D., Mandviwalla, M., & Williams, C. B. (2010). Web 2.0 and Politics: The 2008 U.S. Presidential Election and an E-Politics Research Agenda. *MIS Quarterly*, 34(4), 669-688. <http://www.jstor.org/stable/25750700>
- Kavanaugh, A., Fox, E. A., Sheetz, S., Yang, S., Li, L. T., Whalen, T., Shoemaker, D., Natsev, P., & Xie, L. (2011). Social media use by government: From the routine to the critical. In *Proceedings of the 12th annual international digital government research conference: digital government innovation in challenging times* (pp. 121-130).
- Stieglitz, S., & Dang-Xuan, L. (2013). Social media and political communication: a social media analytics framework. *Social network analysis and mining*, 3, 1277-1291.

- Akbar, P., Irawan, B., Taufik, M., Nurmandi, A., & Suswanta. (2021). Social Media in Politic: Political Campaign on United States Election 2020 Between Donald Trump and Joe Biden. In *HCI International 2021 - Late Breaking Posters* (pp. 359-367). Springer International Publishing.
- Tumasjan, A., Sprenger, T., Sandner, P., & Welp, I. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the international AAAI conference on web and social media* (Vol. 4, pp. 178-185).
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2), 211-236.
- Digital 2024: 5 Billion Social Media Users. (2024, 1 31). We Are Social. Retrieved 7 13, 2024, from <https://wearesocial.com/id/blog/2024/01/digital-2024-5-billion-social-media-users/>
- Kompas.com. (2024, January 9). Debat Pilpres Ketiga: Data Pertahanan dan Keamanan Rahasia Negara. Kompas. <https://nasional.kompas.com/read/2024/01/09/07000011/debat-pilpres-ketiga--data-pertahanan-dan-keamanan-rahasia-negara>
- Wahyuni, W. (2024, January 2). Manfaat debat capres dan cawapres bagi calon pemilih. Hukumonline. <https://www.hukumonline.com/berita/a/manfaat-debat-capres-dan-cawapres-bagi-calon-pemilih-lt6593e7553434b/>
- Badan Kesbangpol Kulon Progo. (2023, December 4). Tema debat capres-cawapres Pemilu 2024. Kesbangpol Kulon Progo.

<https://kesbangpol.kulonprogokab.go.id/detil/881/tema-debat-capres-cawapres-pemilu-2024>

VI. LAMPIRAN

Regex dan Method Python	Deskripsi
<code>r'@[^\s]+'</code>	Menghapus mention dari suatu <i>tweet</i> .
<code>r'http\S+ www\S+'</code>	Menghapus URL dari <i>tweet</i> , baik yang dimulai dengan 'http' maupun 'www'.
<code>r'\s+'</code>	Mengganti beberapa spasi dengan satu spasi dan menghapus spasi di awal dan akhir teks.
<code>r'^rt'</code>	Menghapus awalan 'RT' dari suatu <i>tweet</i> .
<code>.replace('\n', ' ')</code>	Mengganti karakter <i>whitespace</i> ('\n') menjadi spasi (' ').
<code>.lower()</code>	Mengubah semua huruf pada <i>tweet</i> menjadi huruf kecil.

Tabel 3: Penggunaan Regex dan *Method* dari Python untuk Mengekstrak Data *Tweet*

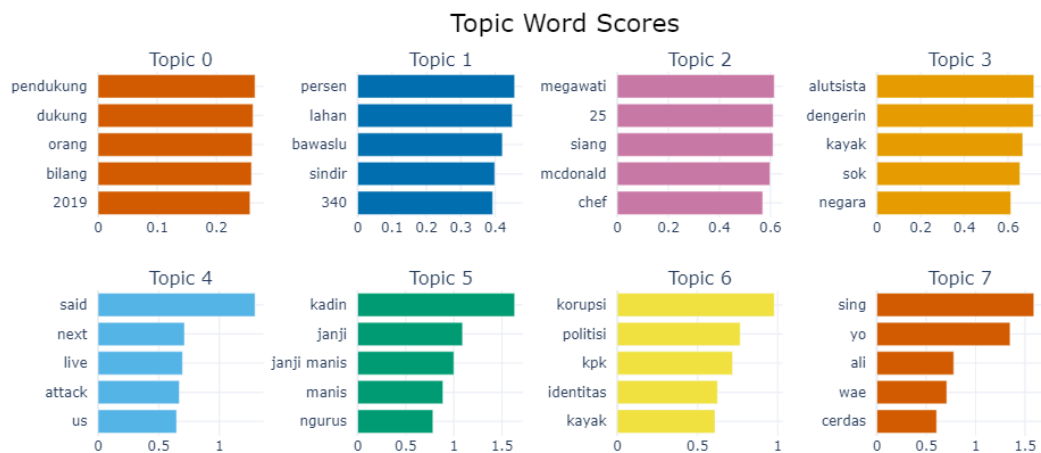
Teks Awal	Teks Setelah Dilakukan <i>Preprocessing</i>
@HUo2P8p6enbSrHhNtJ hdkhUGz+KnIaG9j7IK6 F0uJSg= Sorry yeee.. sorryy yeee.. Prabowo Gibran tidak pernah	sorry yeee.. sorryy yeee.. prabowo gibran tidak pernah...
RT Media Asing Soroti Cara Anies Gaet Pemilih Muda Pakai 'Desak Anies' https://t.co/UYEMKs	media asing soroti cara anies gaet pemilih muda pakai 'desak anies'

4Q10	
@5Tz753nhSrc3MyKUjK PGAQ59HMT5TnaXVJrY BVFXro= Timses AniEs jg ada tofa lemon, ga kalah rasisnya itu	timses anies jg ada tofa lemon, ga kalah rasisnya itu

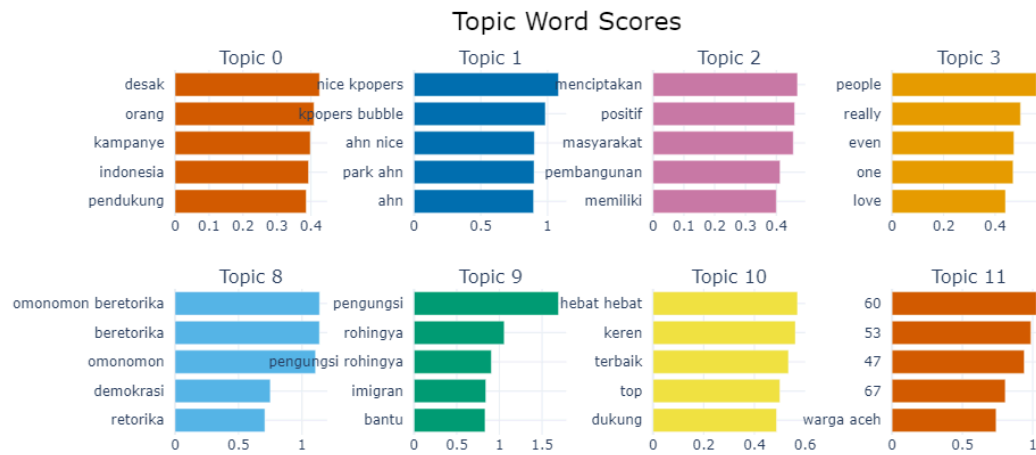
Tabel 4: Contoh Hasil dari Prapemrosesan Data

danish.csv (1.29 kB)				
			en0000rnc0gnv4x00qj oS2BW6Af8ldRmvWdo43D vDBE= Pak Anie...	
		0.0	@w6kZNe5Q4YaSb+H6QJa 81YLUNPxbwaK9zBXdg5i pl4A= Suasana Cilacap-Banyuwangi semakin akrab dengan taglin...	da
		0.0	@5JCA1qShm02hEexpN13 xAuweNSwsuCcjGI4WSGv Lgxxg= Minangkabau peeps tuh tuh sukses banget bikin kampanye...	da

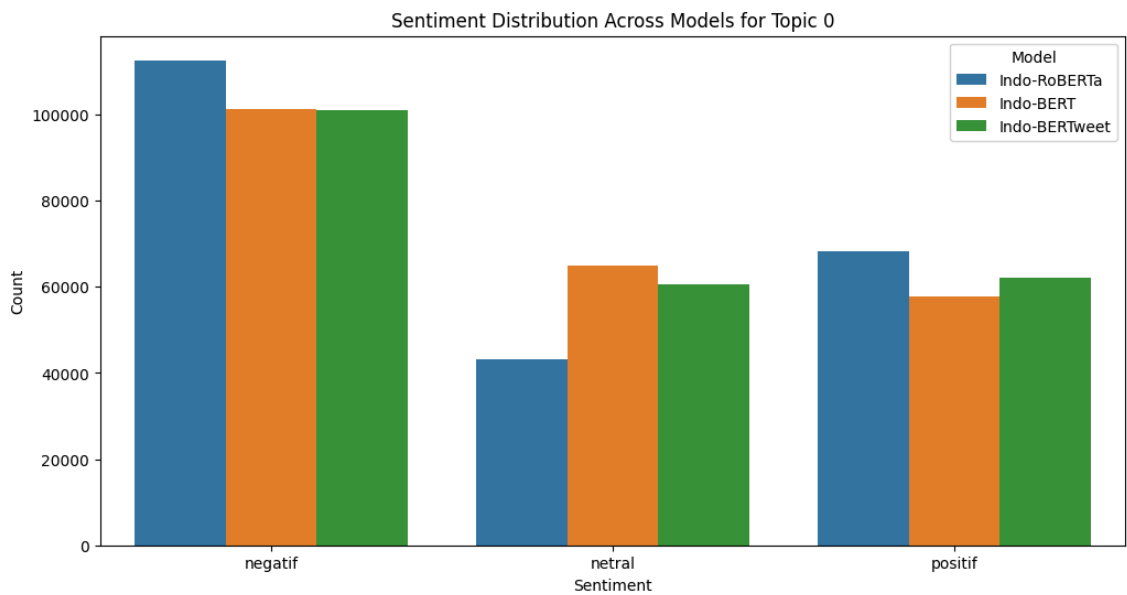
Gambar 2: Contoh Sampel Kesalahan Deteksi Bahasa



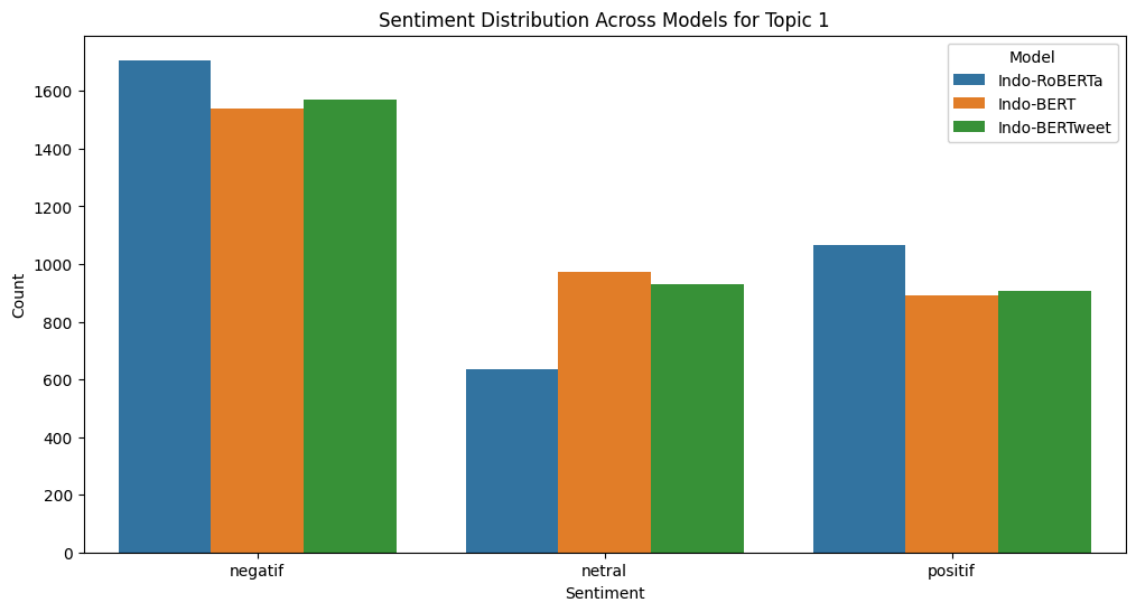
Gambar 5: Daftar Topik Bahasan Mengenai Data Kategori 4



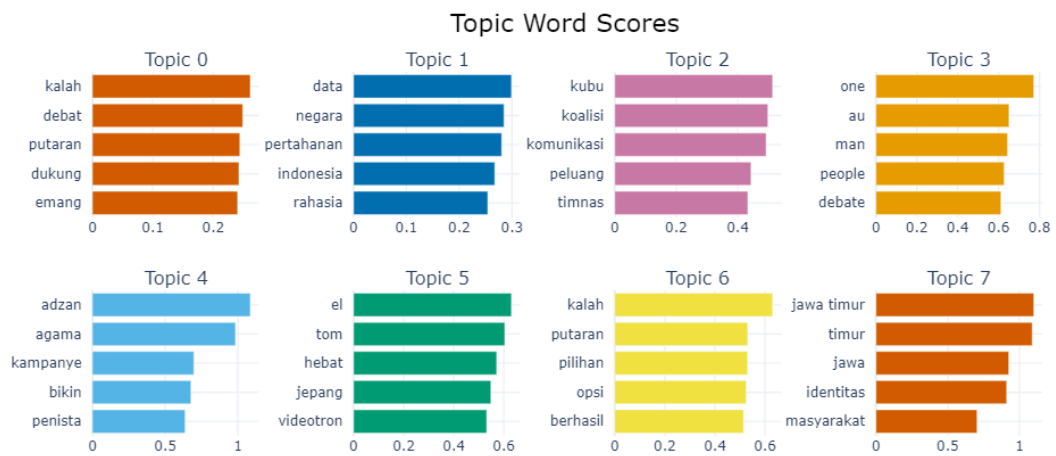
Gambar 8: Daftar Topik Bahasan Mengenai Data Kategori 1



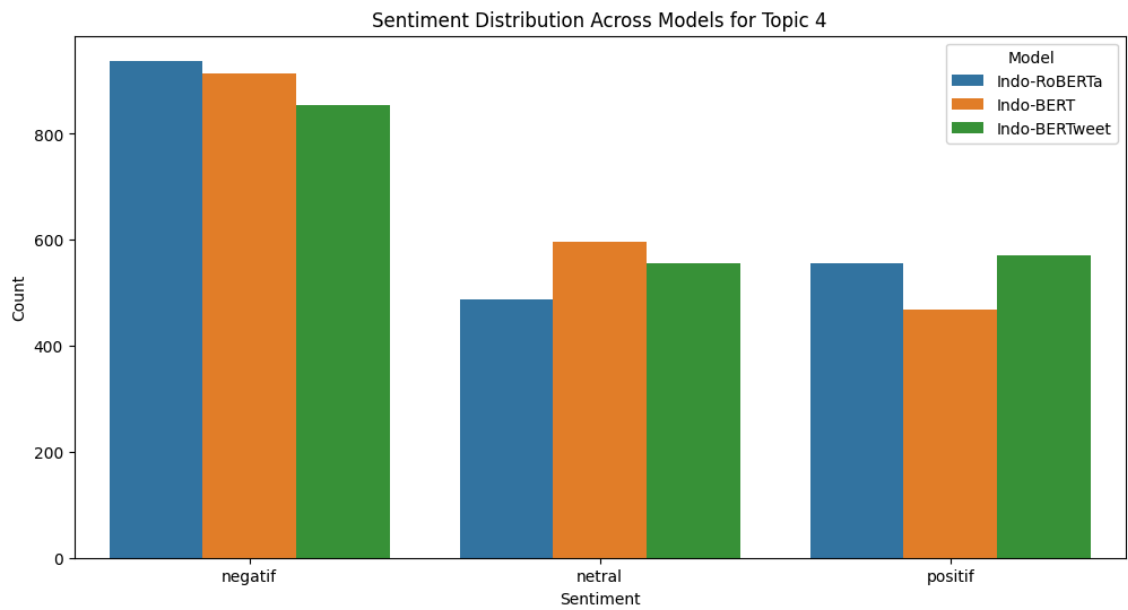
Gambar 9: Distribusi Topik Sentimen Mengenai Data Kategori 1 Tentang Topik 0



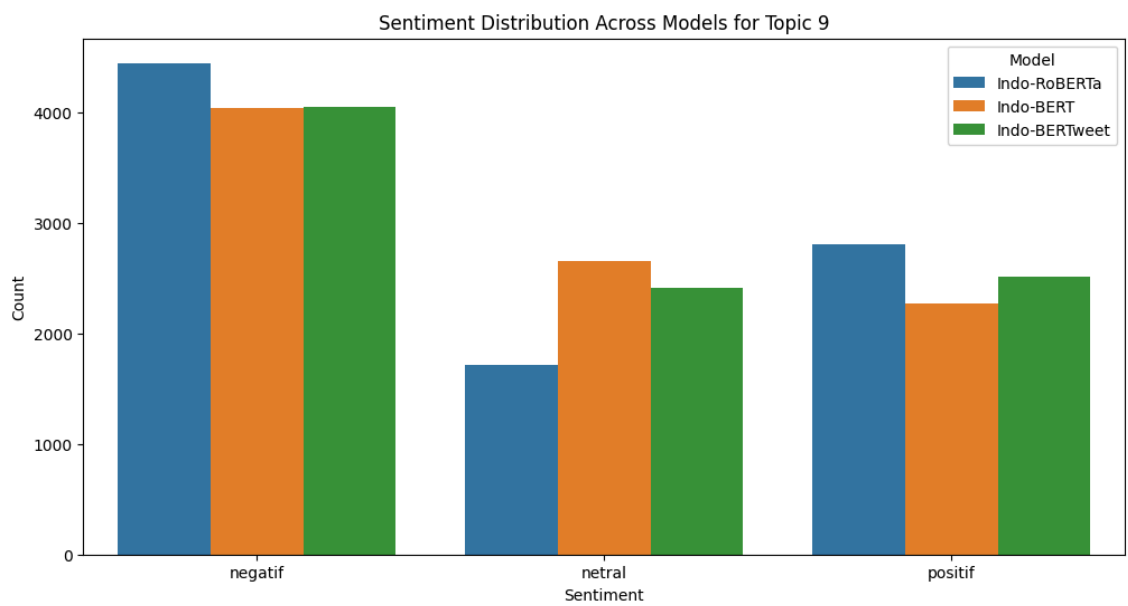
Gambar 10: Distribusi Topik Sentimen Mengenai Data Kategori 1 Tentang Topik 1



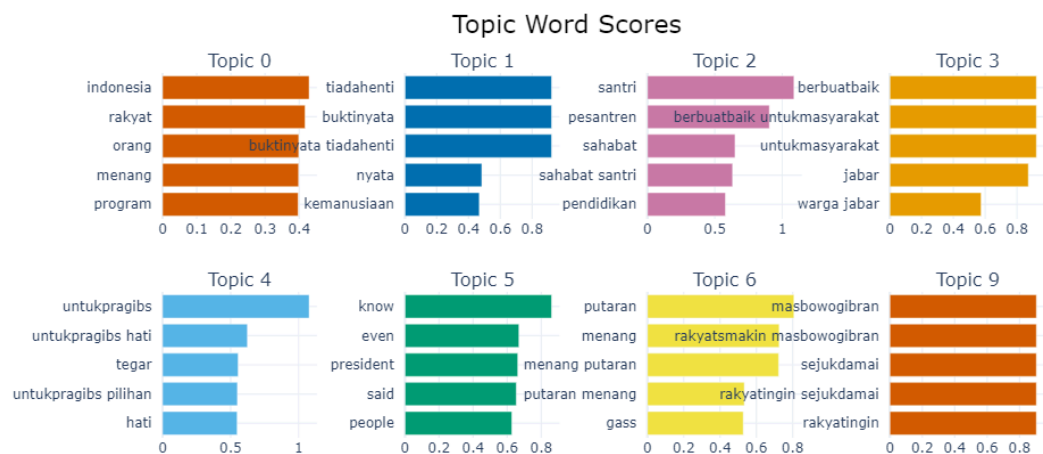
Gambar 11: Daftar Topik Bahasan Mengenai Data Kategori 5



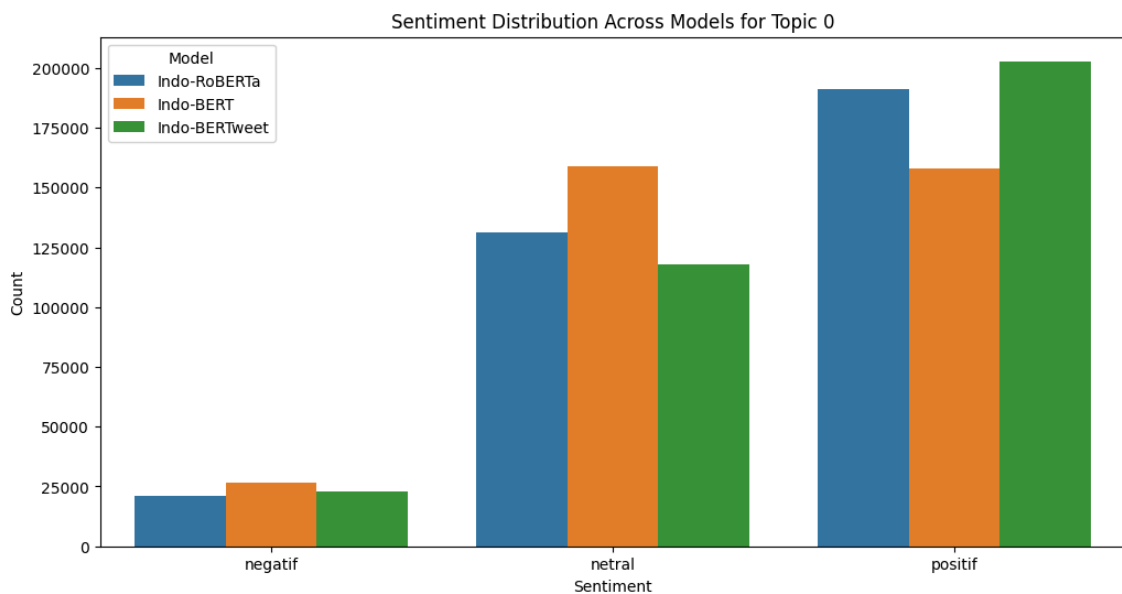
Gambar 12: Distribusi Topik Sentimen Mengenai Data Kategori 5 Tentang Topik 4



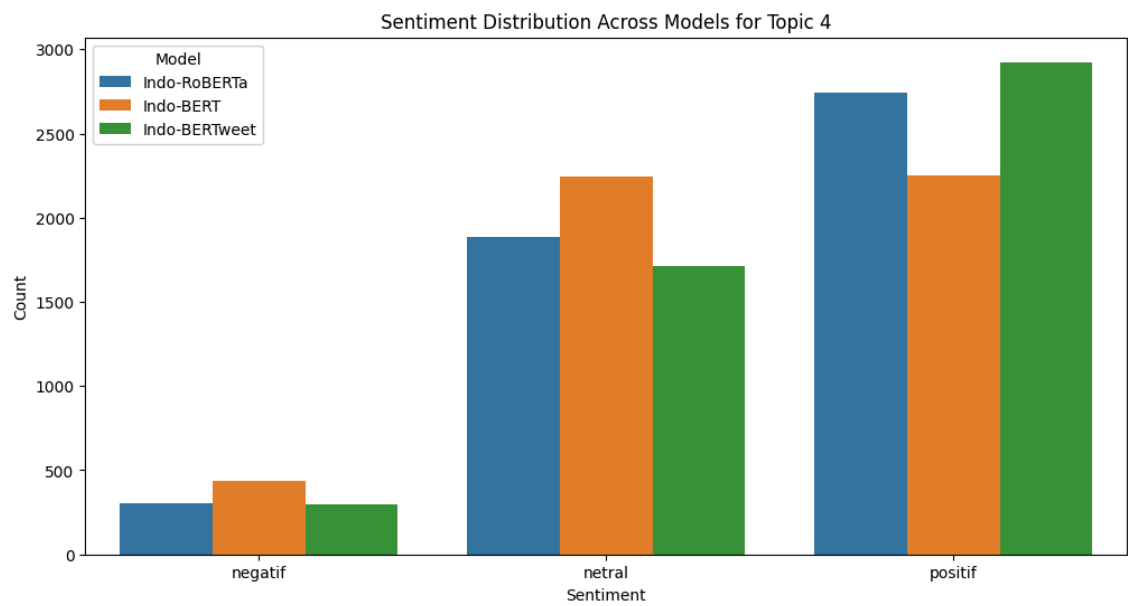
Gambar 13: Distribusi Topik Sentimen Mengenai Data Kategori 1 Tentang Topik 9



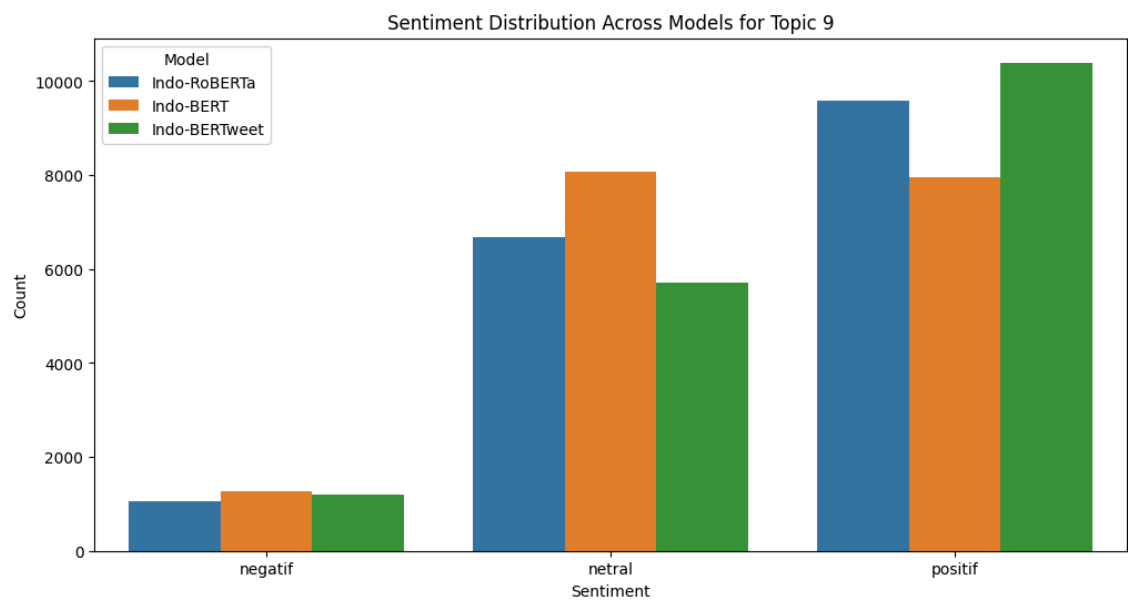
Gambar 14: Daftar Topik Bahasan Mengenai Data Kategori 2



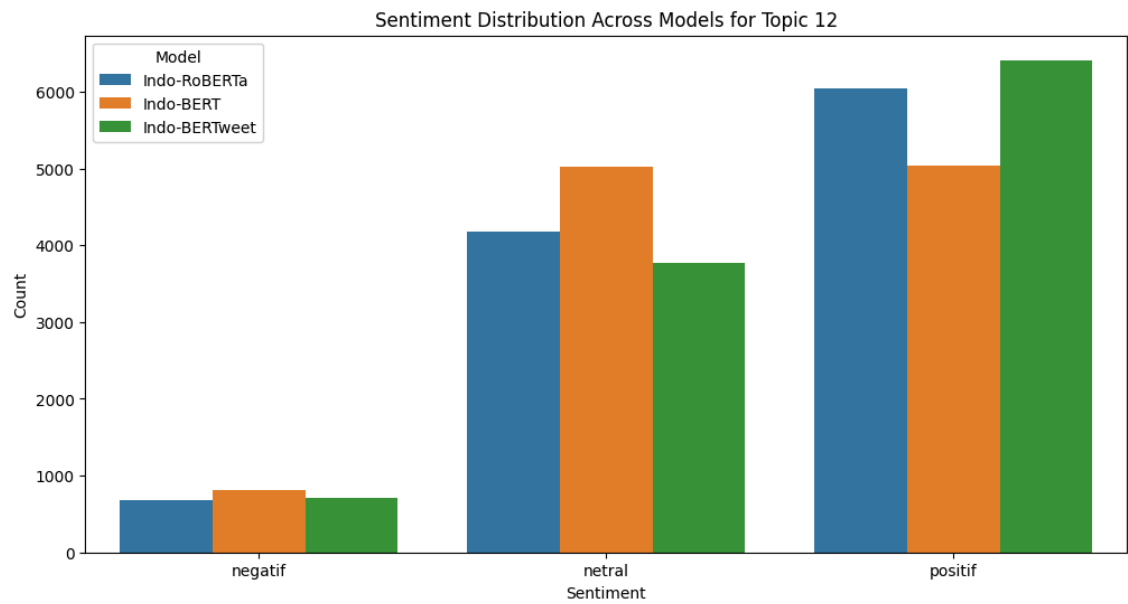
Gambar 15: Distribusi Topik Sentimen Mengenai Data Kategori 3 Tentang Topik 0



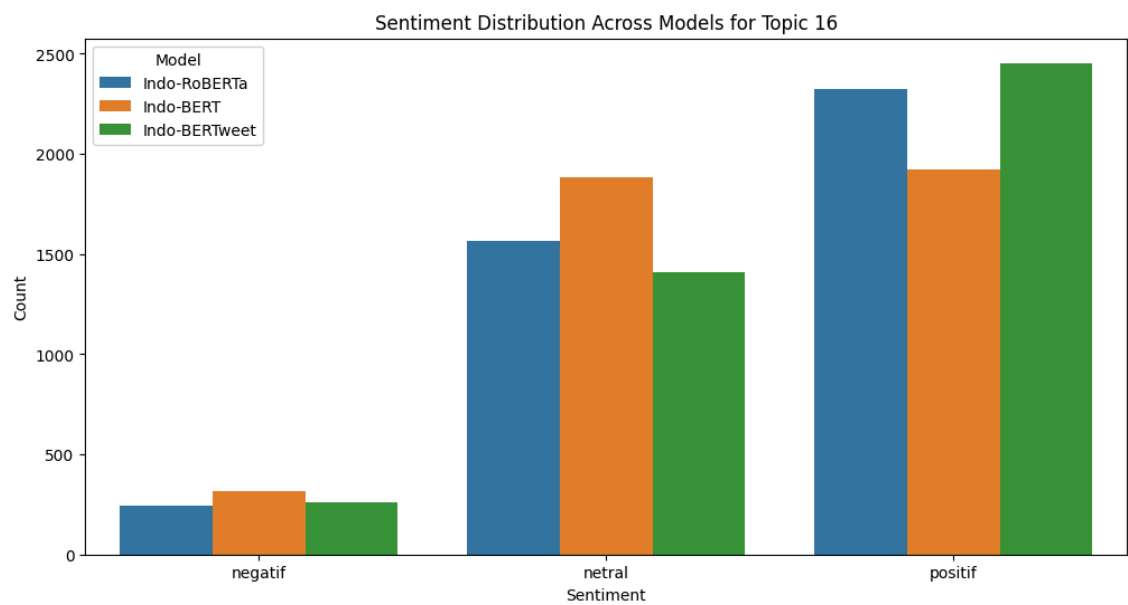
Gambar 16: Distribusi Topik Sentimen Mengenai Data Kategori 3 Tentang Topik 4



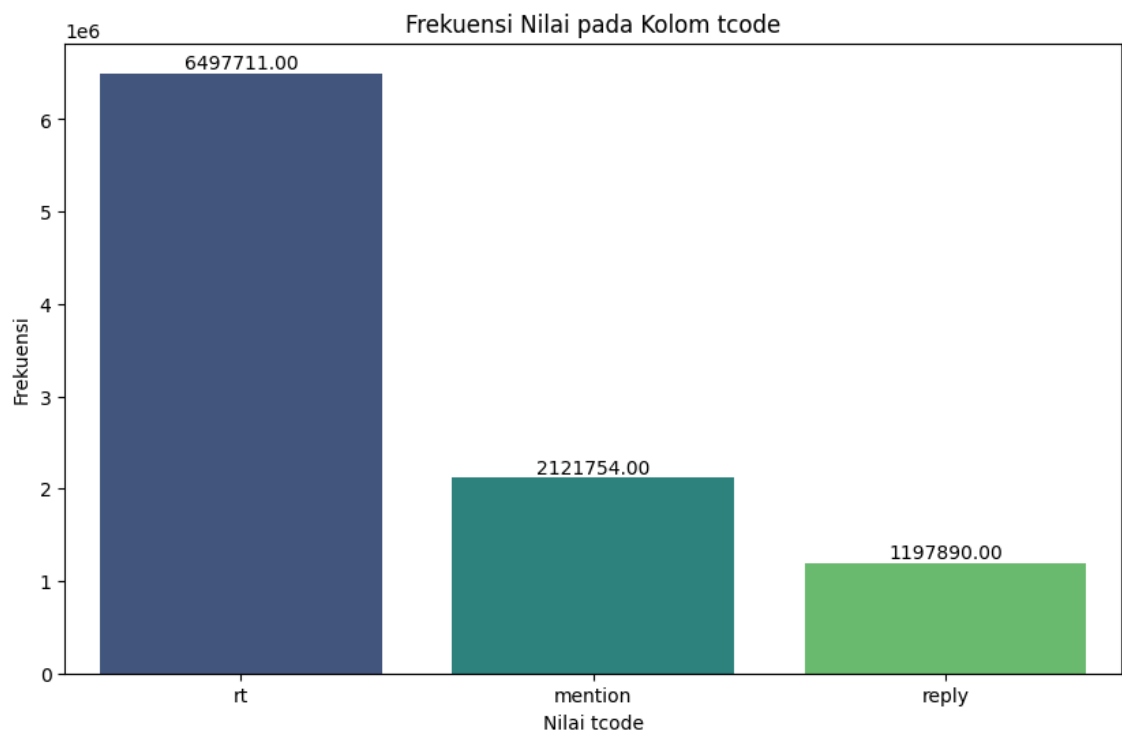
Gambar 17: Distribusi Topik Sentimen Mengenai Data Kategori 3 Tentang Topik 9



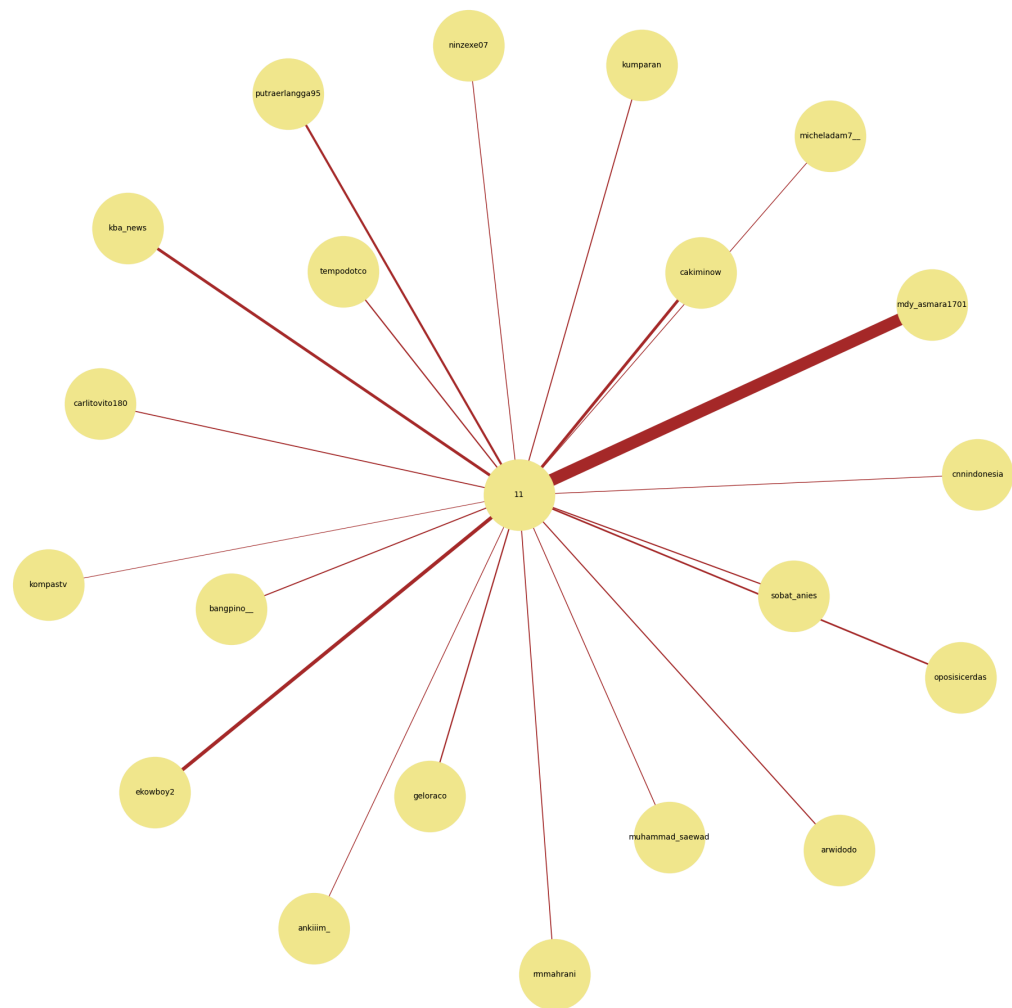
Gambar 18: Distribusi Topik Sentimen Mengenai Data Kategori 3 Tentang Topik 12

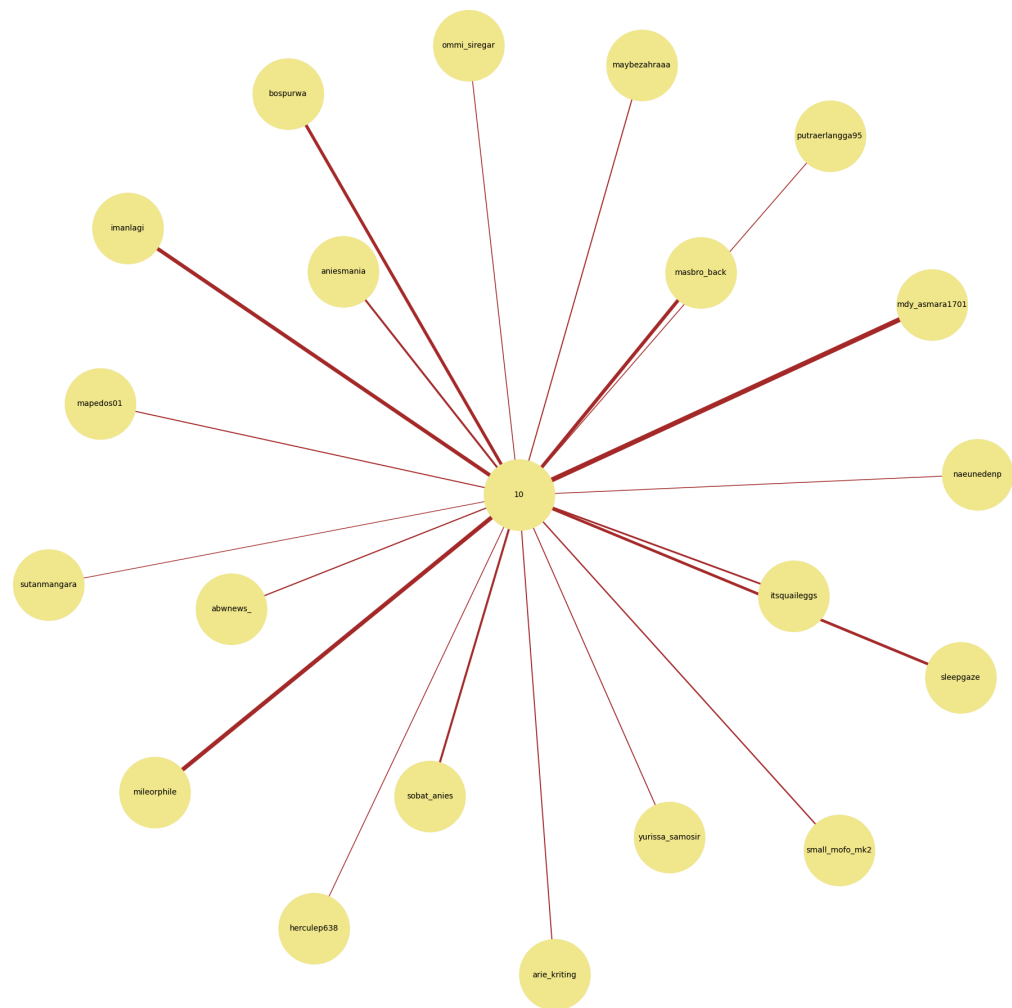


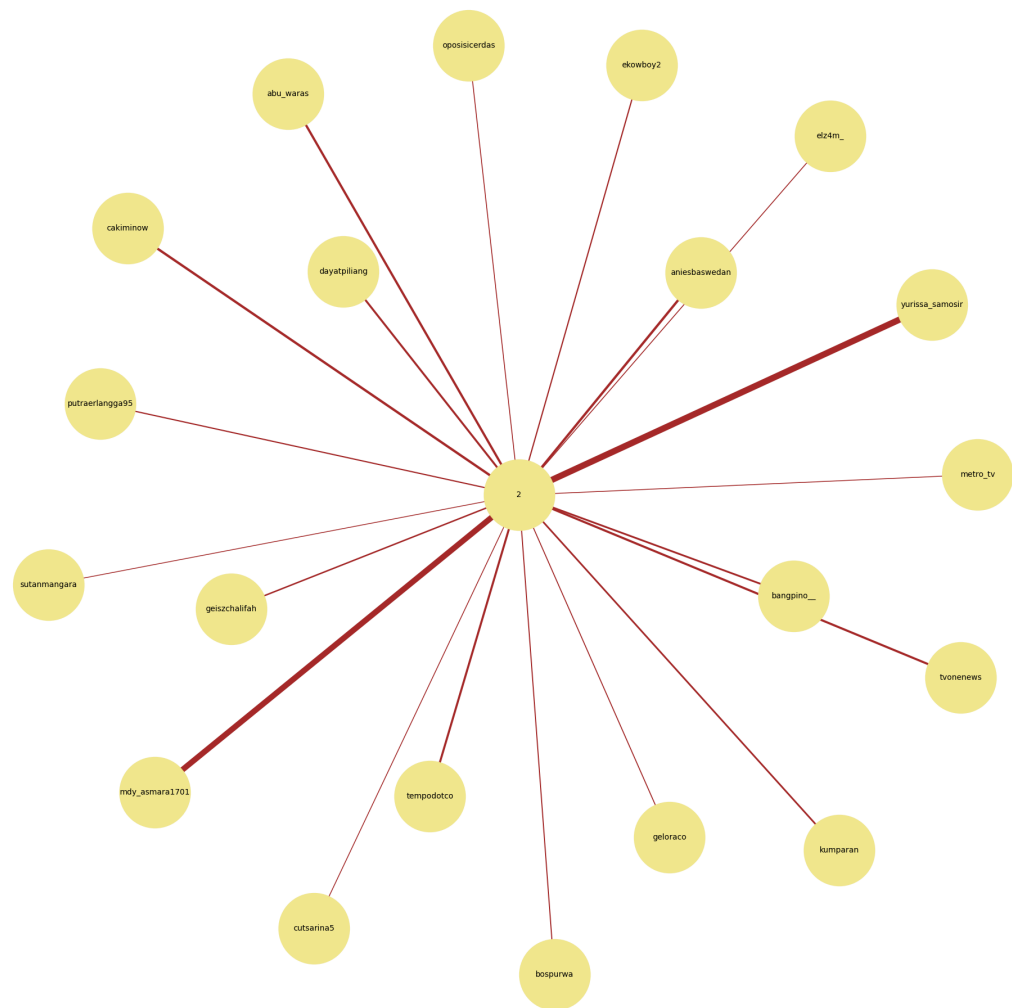
Gambar 19: Distribusi Topik Sentimen Mengenai Data Kategori 3 Tentang Topik 16

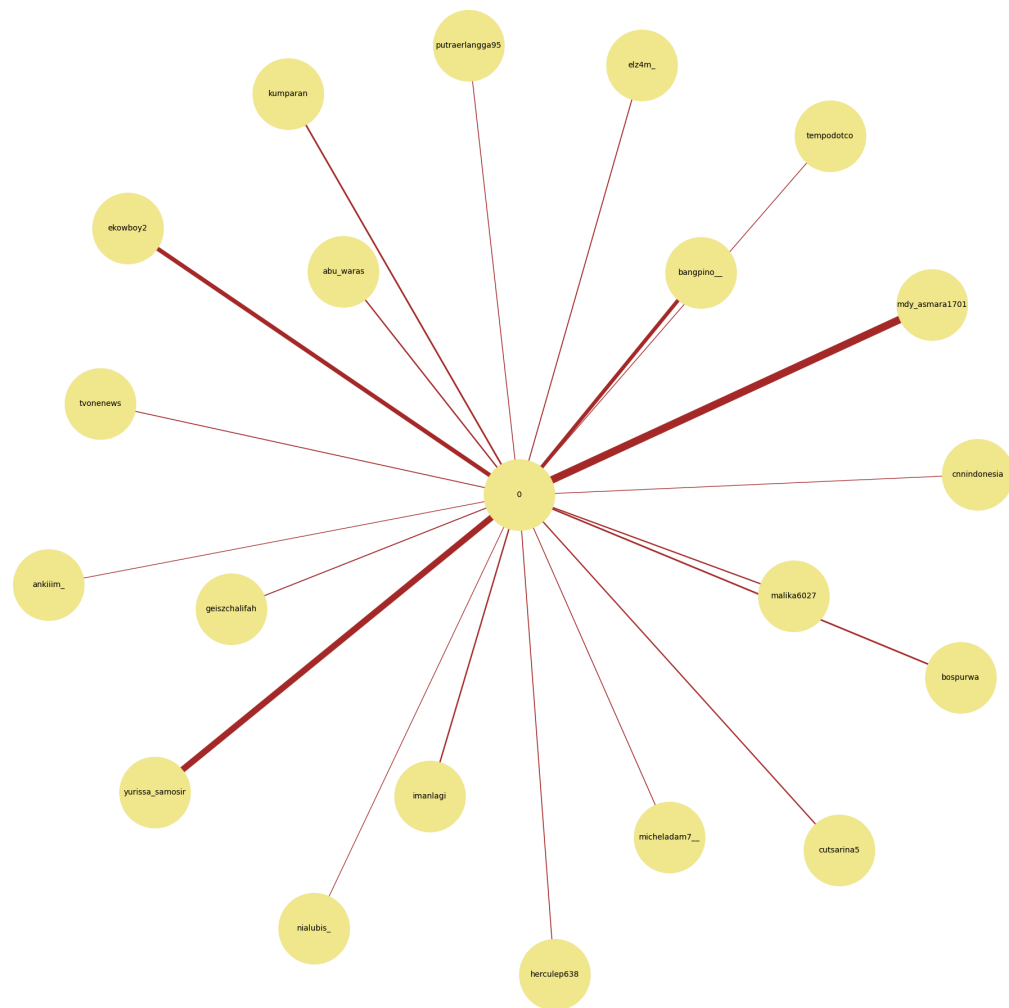


Gambar 20: Frekuensi Nilai pada Kolom tcode

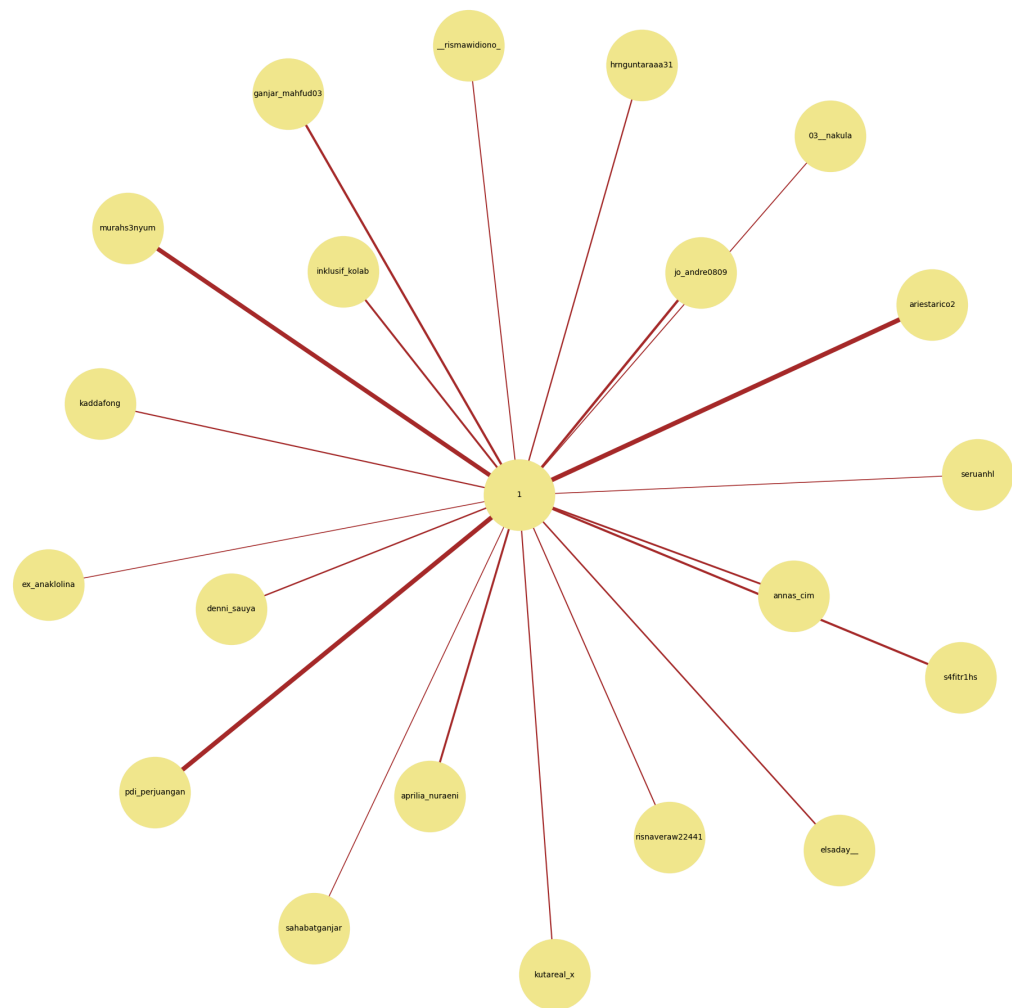


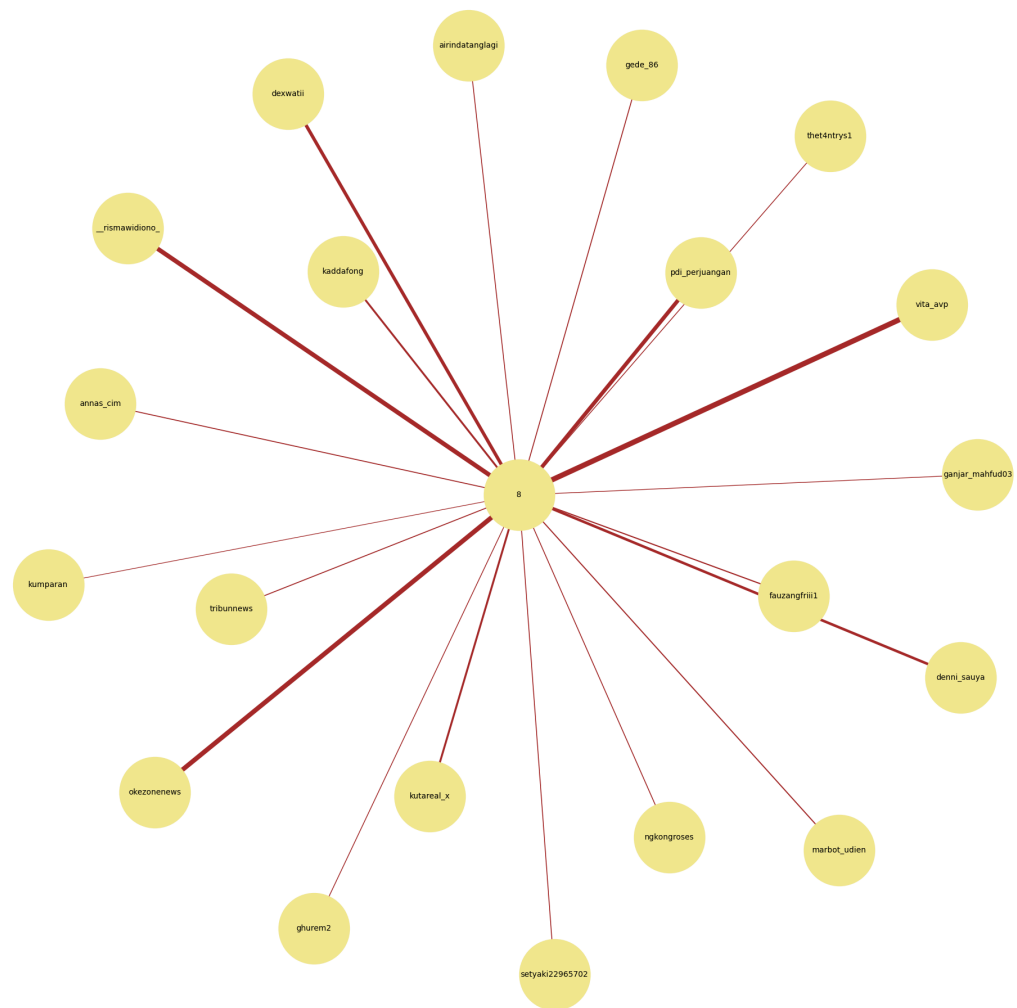


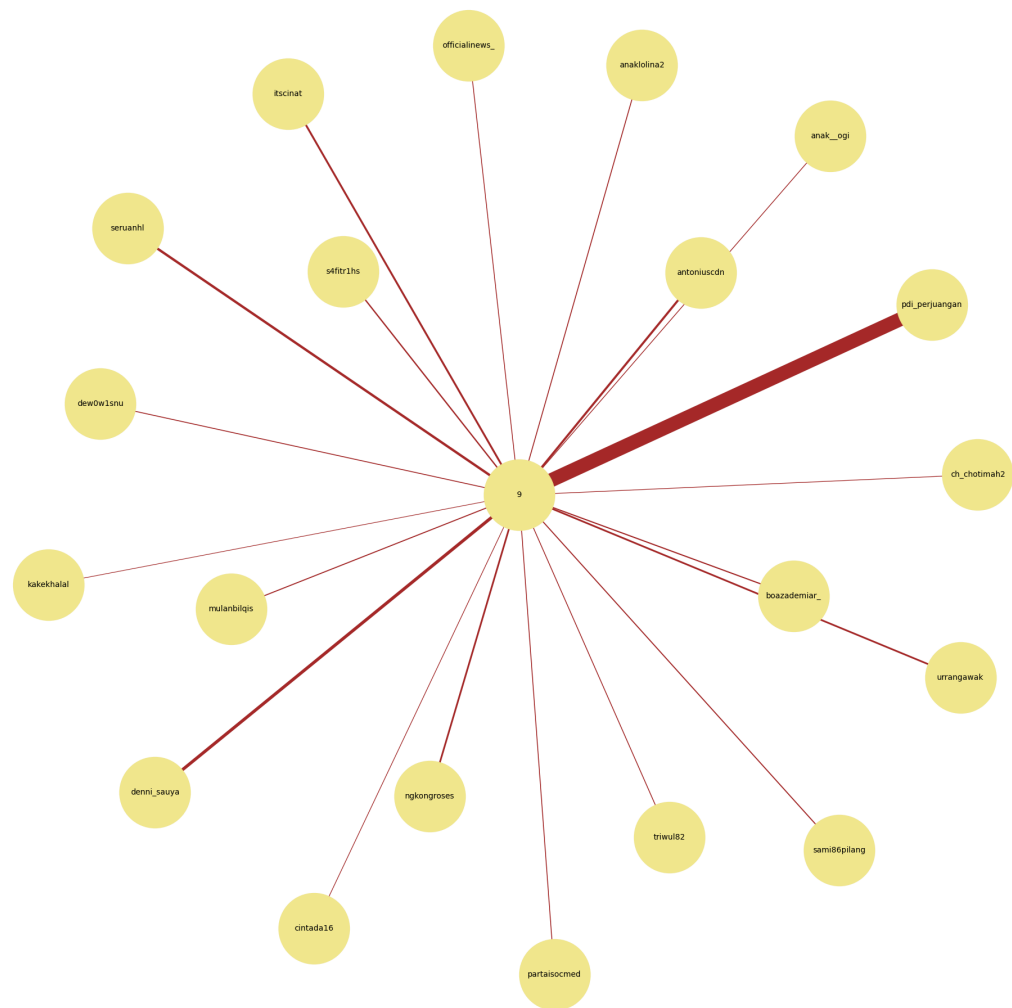


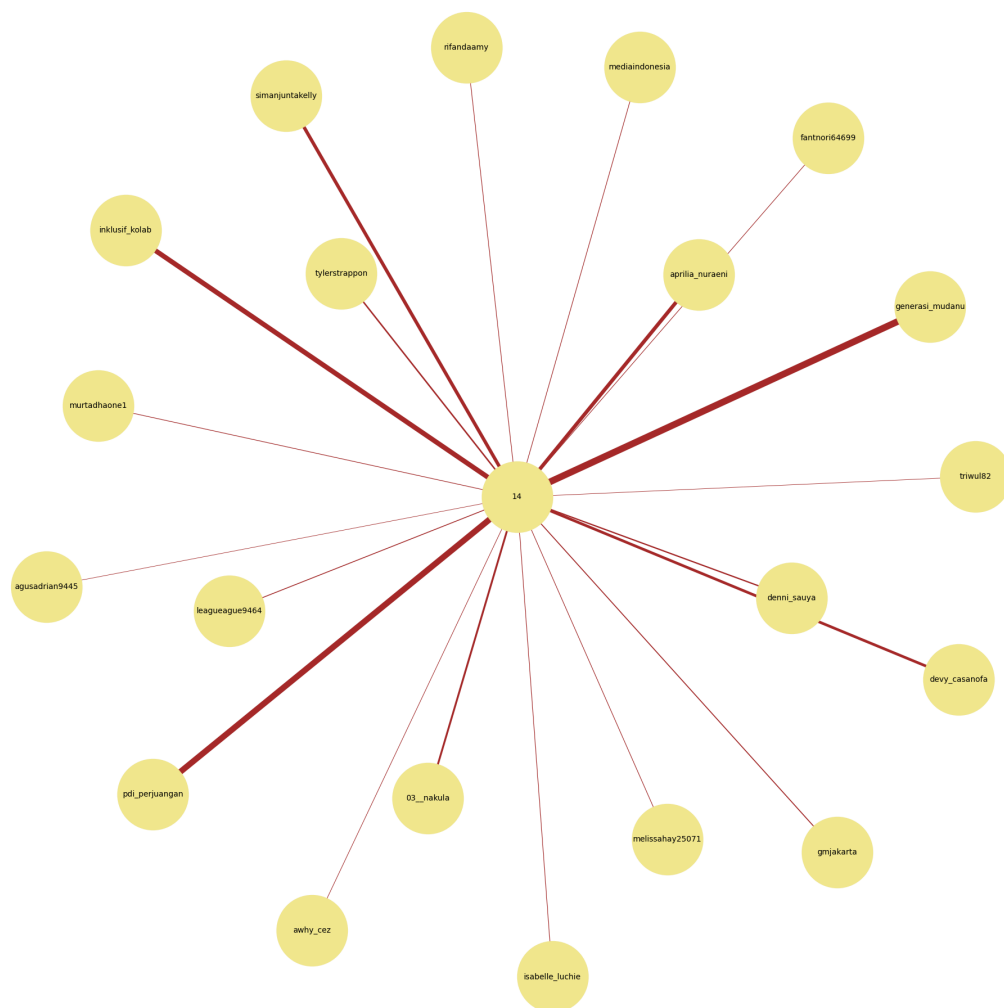


Gambar 22: Kategori 1 Topik 11; 10; 2; 0

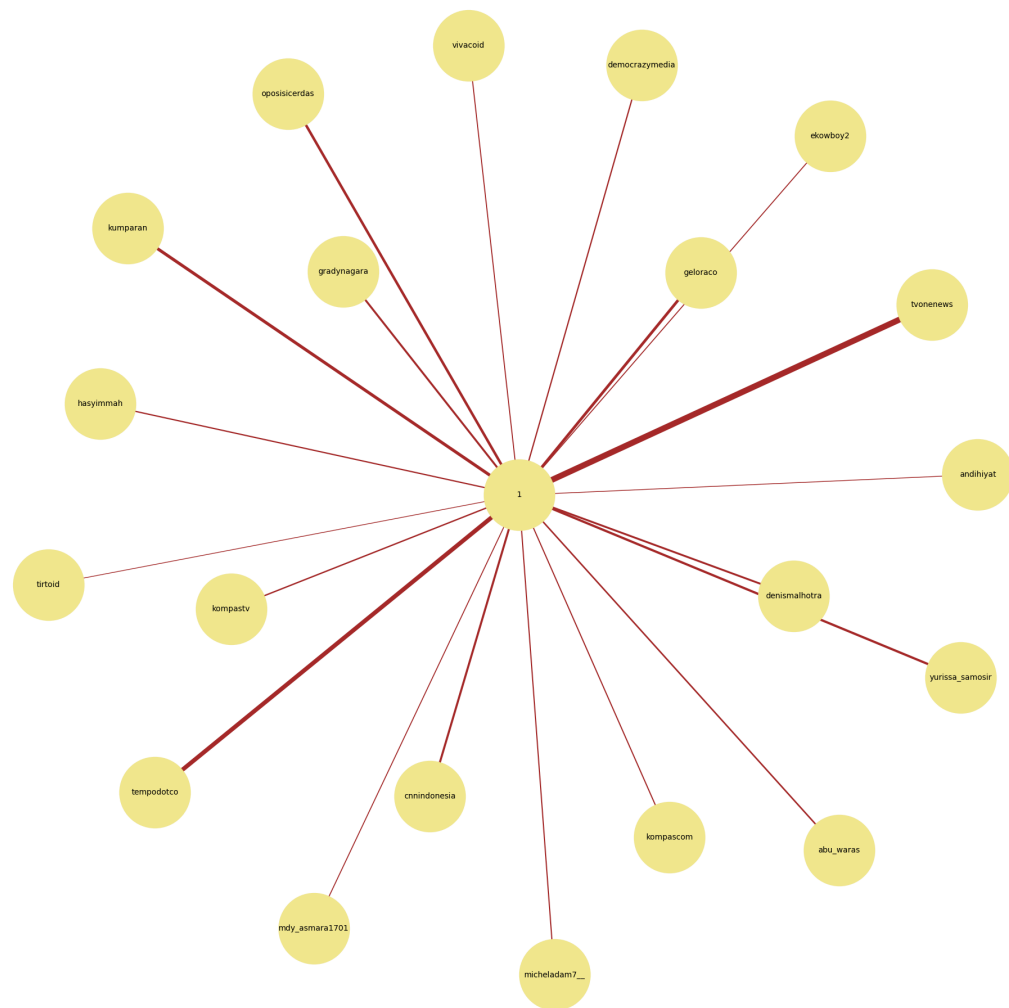




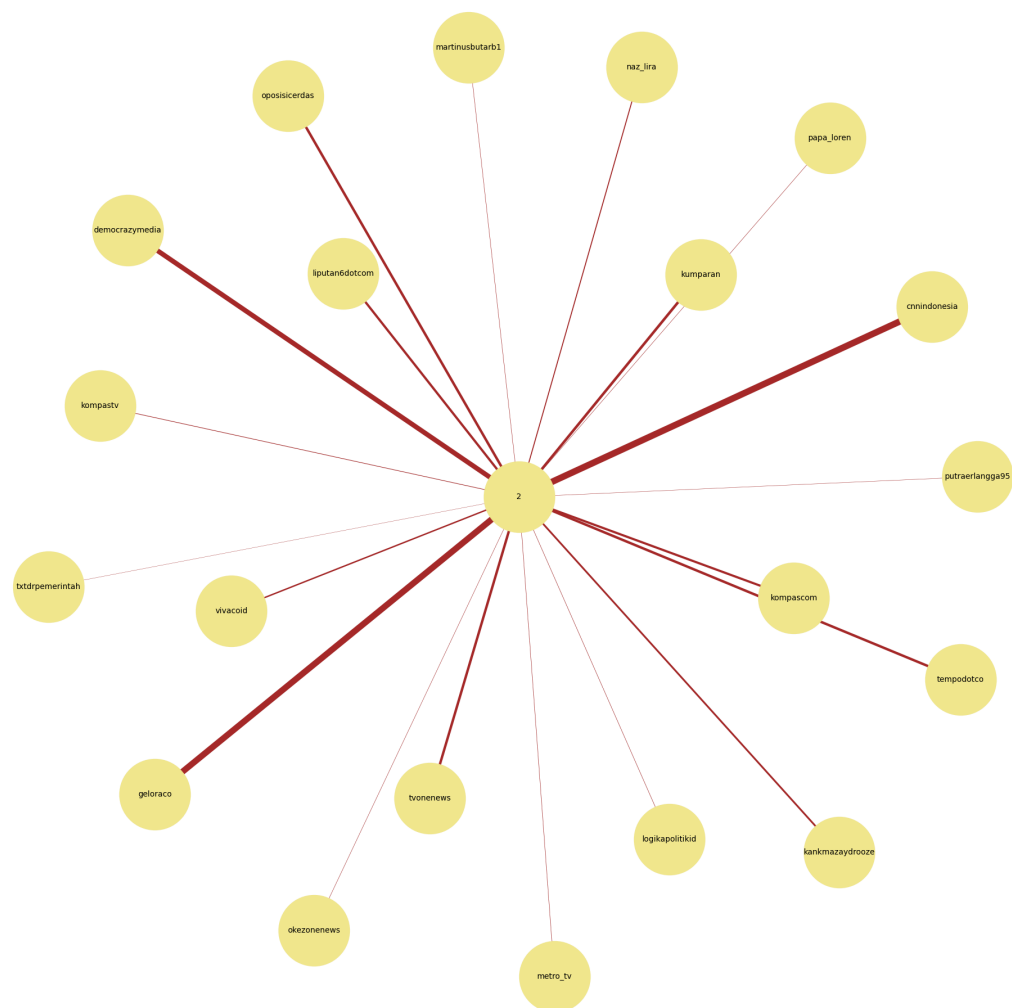




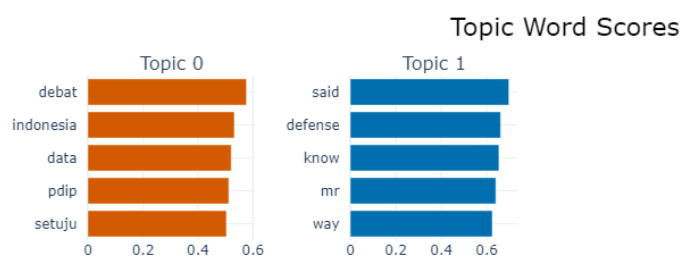
Gambar 23: Kategori 3 Topik 1; 8; 9; 14



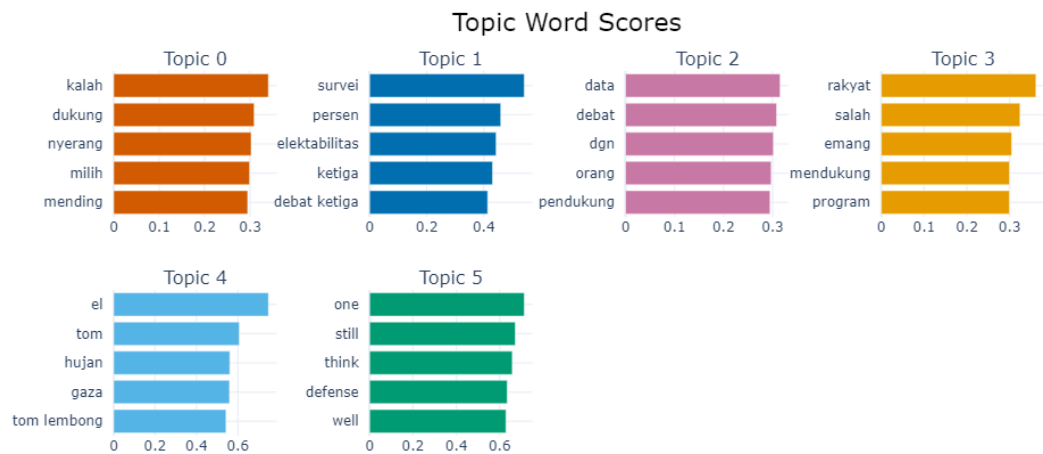
Gambar 24: Kategori 4 Topik 1



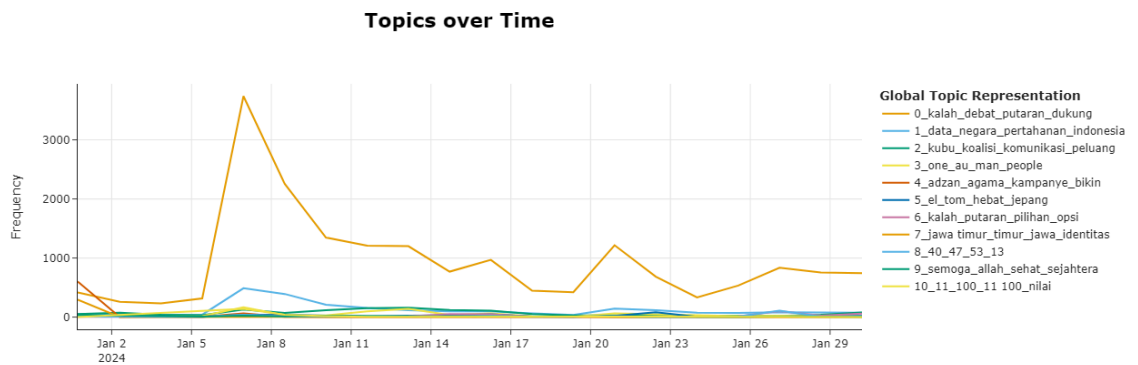
Gambar 25: Kategori 5 Topik 2



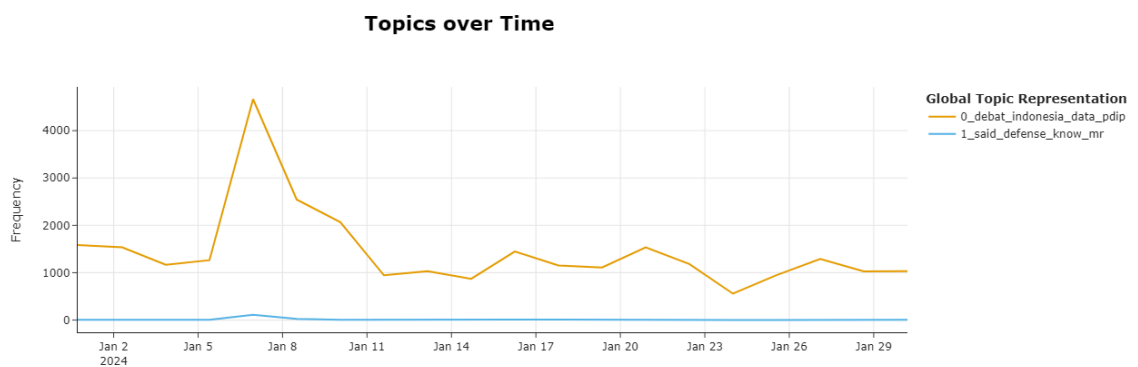
Gambar 26: Daftar Topik Bahasan Mengenai Data Kategori 6



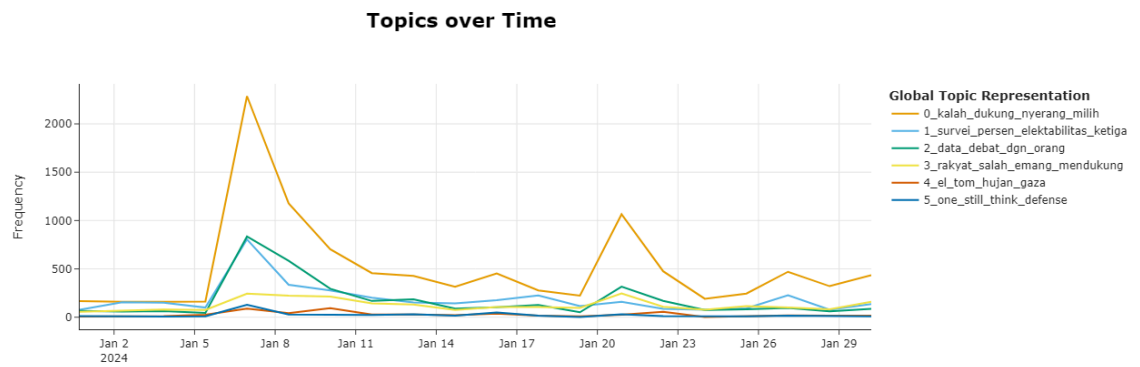
Gambar 27: Daftar Topik Bahasan Mengenai Data Kategori 7



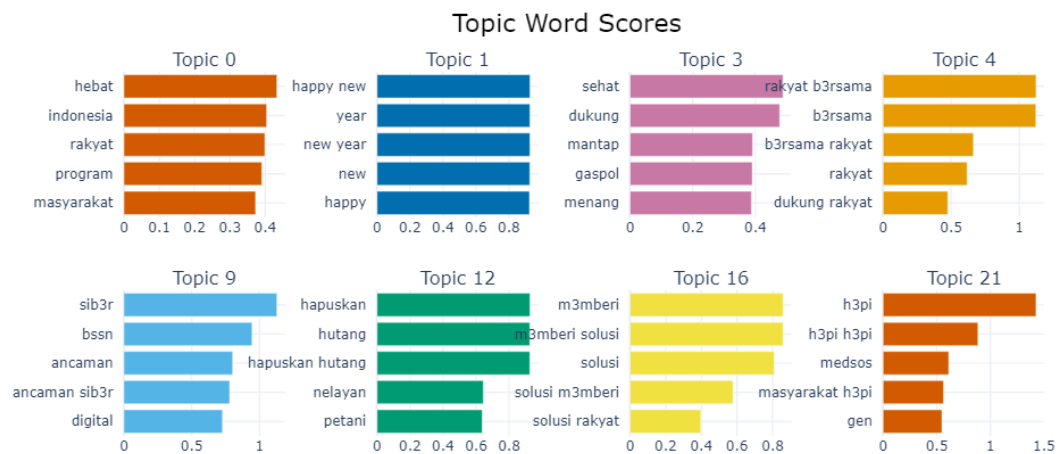
Gambar 28: Grafik Topik Bahasan Mengenai Data Kategori 5 dalam Jangka Waktu Satu Bulan



Gambar 29: Grafik Topik Bahasan Mengenai Data Kategori 6 dalam Jangka Waktu Satu Bulan



Gambar 30: Grafik Topik Bahasan Mengenai Data Kategori 7 dalam Jangka Waktu Satu Bulan



Gambar 31: Daftar Topik Bahasan Mengenai Data Kategori 3