PROJECT A

Analisis Sentimen, *POS Tagging*, dan *Named Entity Recognition* pada Artikel Berita Danantara menggunakan *Library newspaper3k*, *spaCy-UDPipe*, dan *transformers*

DISUSUN OLEH:

Diva Ardelia Alyadrus	5026221029
Shof Watun Niswah	5026221043
Muhammad Daffa Alvinoer Rahman	5026221180

PENGOLAHAN BAHASA ALAMI (A)

DEPARTEMEN SISTEM INFORMASI
FAKULTAS TEKNOLOGI ELEKTRO DAN INFORMATIKA CERDAS
INSTITUT TEKNOLOGI SEPULUH NOPEMBER

SEMESTER GENAP 2025

DAFTAR ISI

ABSTRAK

Dalam era digital saat ini, informasi publik tentang kebijakan negara menyebar dengan cepat melalui berbagai media daring. Penelitian ini mengkaji pemberitaan mengenai kebijakan BPI Danantara dengan menggunakan pendekatan Natural Language Processing (NLP), mencakup analisis sentimen, Part-of-Speech (POS) tagging, dan Named Entity Recognition (NER). Data diperoleh dari 226 artikel berita hasil scraping menggunakan library newspaper3k, lalu dianalisis menggunakan TextBlob dan spaCy-UDPipe. Hasil menunjukkan bahwa sebagian besar artikel cenderung bersentimen positif atau netral-positif, dengan fokus utama pada isu ekonomi, pengelolaan aset, dan keterlibatan tokoh politik. Analisis TF-IDF dan frekuensi kata mengungkap bahwa kata-kata seperti danantara, bumn, dan prabowo mendominasi teks artikel. Pendekatan NLP ini terbukti efektif dalam menangkap kecenderungan narasi media serta mengidentifikasi aktor dan topik dominan dalam pemberitaan. Temuan ini diharapkan dapat menjadi landasan dalam memahami opini publik terhadap kebijakan nasional.

Kata kunci: Danantara, *Natural Language Processing,* Analisis sentimen, POS tagging, Named Entity Recognition, berita daring, opini publik, TF-IDF, TextBlob, spaCy.

1. PENDAHULUAN

1.1 Latar Belakang

Dalam era digital yang berkembang pesat dan di tengah memanasnya dinamika politik saat ini, arus informasi mengalir dengan sangat cepat, terutama melalui media daring seperti portal berita. Banyaknya sumber informasi yang beredar memunculkan tantangan baru, termasuk perbedaan respons dan interpretasi publik terhadap isu-isu yang diberitakan. Salah satu tantangan utama adalah bagaimana memahami persepsi publik terhadap suatu kebijakan melalui pemberitaan yang tersebar di berbagai media. Kebijakan Danantara sebagai salah satu kebijakan strategis menjadi sorotan di berbagai platform berita. Berbagai artikel yang membahas kebijakan ini tidak hanya menyampaikan fakta, tetapi juga memuat opini, kritik, maupun dukungan yang narasinya menjadi ruang demokrasi baru bagi masyarakat untuk berpartisipasi dalam kebijakan dan keputusan politik yang dihasilkan pemerintah (Zempi CN et. al, 2023).

Survei dan audiensi publik tidak selalu dapat menjangkau seluruh lapisan masyarakat dan mungkin tidak mewakili opini publik secara keseluruhan. Oleh karena itu, dibutuhkan metode baru untuk menganalisis sentimen masyarakat terhadap kebijakan pemerintah secara real-time dan komprehensif (Abdurrohim & Rahman, 2024). Untuk memahami persepsi yang terkandung dalam berita, diperlukan analisis mendalam terhadap isi artikel tersebut. Salah satu pendekatan yang dapat dilakukan adalah analisis sentimen untuk mengidentifikasi kecenderungan opini dalam teks. Selain itu, analisis lanjutan seperti *Part-of-Speech* (POS) tagging dan *Named Entity Recognition* (NER) juga penting dilakukan guna mengenali struktur kata dan entitas yang terdapat dalam teks berita. POS *tagging* membantu mengetahui peran atau fungsi kata dalam kalimat, sedangkan NER berperan dalam menemukan entitas penting seperti nama tokoh, organisasi, maupun lokasi yang disebutkan. Informasi ini sangat berguna untuk memetakan isu, aktor, dan konteks kebijakan yang sedang dibahas.

Dengan kemampuannya mengolah data teks secara efisien, NLP dapat memberikan wawasan yang berharga bagi pembuat kebijakan dalam merancang dan memperbaiki kebijakan yang sesuai dengan kebutuhan dan aspirasi masyarakat (Abdurrohim & Rahman, 2024). Untuk mendukung proses analisis ini, digunakan *library newspaper3k*, yaitu sebuah pustaka Python yang mampu melakukan ekstraksi konten berita secara otomatis, mulai dari pengambilan tautan artikel, metadata, hingga isi teks

utama. Penggunaan *library* ini mempercepat dan mempermudah proses pengumpulan data dari berbagai sumber berita daring. Melalui penerapan analisis sentimen, POS tagging, dan NER terhadap artikel berita mengenai kebijakan Danantara, diharapkan dapat diperoleh gambaran yang lebih objektif terkait persepsi media, aktor yang terlibat, serta membantu proses pengambilan keputusan atau evaluasi terhadap kebijakan tersebut di masa mendatang.

1.2 Rumusan Masalah

1. Bagaimana tren pemberitaan mengenai kebijakan Danantara ditinjau dari sumber portal berita dan periode waktu tertentu?
2. Apa saja kata atau istilah yang paling sering muncul dalam berita tentang kebijakan Danantara?
3. Bagaimana tingkat **sentiment polarity** dan **subjectivity** yang terdapat dalam artikel berita Danantara?
4. Bagaimana hasil POS dan NER ini membantu mendukung analisis linguistik atau pemrosesan data teks?

2. TINJAUAN PUSTAKA

2.1 Scraping

Web scraping adalah proses ekstraksi data atau informasi dari berbagai situs web secara otomatis (Thota & Elmasri, 2021). Proses ini memungkinkan pengumpulan data dalam jumlah besar tanpa perlu melakukan penyalinan manual. Web scraping juga dikenal dengan istilah lain seperti *web data extraction*, *web harvesting,* atau *screen scraping*. Secara umum, *scraping* dilakukan untuk memperoleh data yang umumnya bersifat *unstructured* di halaman web, kemudian diubah menjadi format *structured* seperti CSV, *spreadsheet*, atau database agar lebih mudah dianalisis. Menurut Thota & Elmasri (2021), proses *web scraping* terdiri dari tiga tahapan utama, yaitu (1) *Fetching*, mengambil halaman web dengan mengirimkan permintaan HTTP ke server dan menerima dokumen HTML sebagai respons. (2) *Extracting*, mengekstrak informasi yang relevan dari dokumen HTML dengan teknik seperti HTML parsing, DOM parsing, XPath, atau pattern matching. (3) *Transforming*, mengubah data yang telah diekstrak menjadi format terstruktur seperti CSV atau spreadsheet untuk penyimpanan dan analisis lebih lanjut.

2.2 Sentiment Polarity dan Sentiment Subjectivity

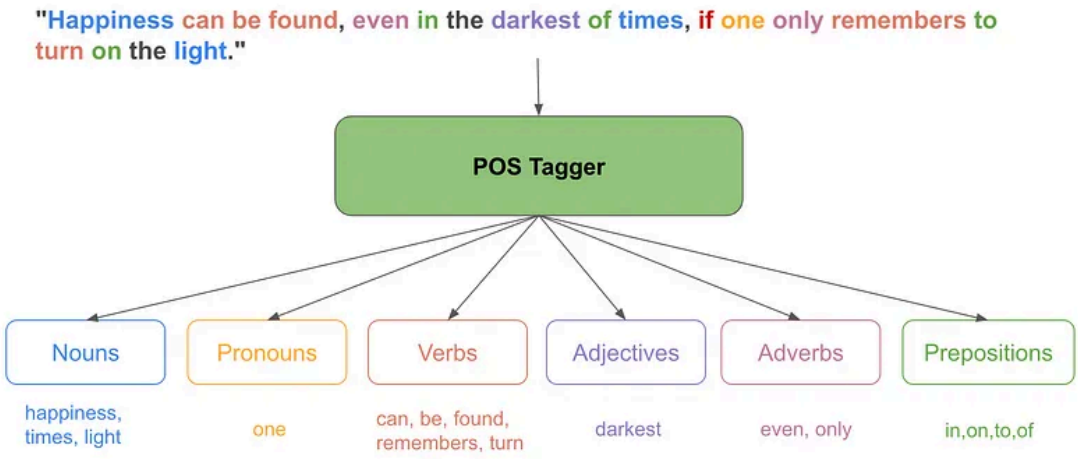
Analisis sentimen merupakan teknik yang digunakan untuk mengidentifikasi dan mengekstrak opini atau emosi yang terkandung dalam suatu teks. Salah satu aspek penting dalam analisis sentimen adalah *sentiment polarity*, yaitu ukuran yang digunakan untuk menentukan arah atau sifat emosi dalam teks, apakah bersifat positif, negatif, atau netral (Nafees et al., 2021). Misalnya, pernyataan seperti "Aplikasi ini sangat bermanfaat" memiliki *polarity* positif, sedangkan "Saya kecewa dengan layanan aplikasi ini" memiliki *polarity* negatif. Penentuan *polarity* memungkinkan peneliti memahami kecenderungan sikap atau persepsi publik terhadap suatu isu, produk, atau kebijakan.

Selain *polarity*, analisis sentimen juga mempertimbangkan aspek *sentiment subjectivity*, yaitu ukuran yang menunjukkan sejauh mana suatu teks mengandung opini pribadi (subjektif) atau hanya menyampaikan informasi faktual (objektif) (Pang & Lee, 2008). Semakin subjektif suatu teks, semakin banyak opini, perasaan, atau pandangan pribadi yang diungkapkan; sedangkan semakin objektif, teks tersebut lebih berfokus pada penyampaian fakta tanpa opini. Contohnya, kalimat "Aplikasi ini dirilis pada tahun 2021" bersifat objektif, sedangkan "Saya merasa aplikasi ini kurang bermanfaat" bersifat subjektif.

Kedua konsep ini saling melengkapi dalam analisis sentimen. Sentiment polarity menjawab pertanyaan "apakah sikapnya positif, negatif, atau netral?", sedangkan sentiment subjectivity menjawab "apakah pernyataannya berupa opini atau fakta?". Dengan menganalisis *polarity* dan *subjectivity* secara bersamaan, peneliti dapat memperoleh gambaran yang lebih komprehensif mengenai nada emosional dan tingkat keberpihakan suatu teks, baik dalam konteks media sosial, ulasan produk, maupun pemberitaan media.

2.3 Part-of-Speech

Part-of-Speech (POS) *tagging* atau penandaan kelas kata adalah proses otomatis untuk menetapkan kategori gramatikal seperti kata benda, kata kerja, kata sifat, atau kata keterangan pada setiap kata dalam suatu kalimat (Chiche & Yitagesu, 2022). POS *tagging* menjadi salah satu komponen penting dalam bidang *Natural Language Processing* (NLP) karena berfungsi sebagai dasar bagi berbagai aplikasi seperti *machine translation*, *question answering,* dan *information extraction*.



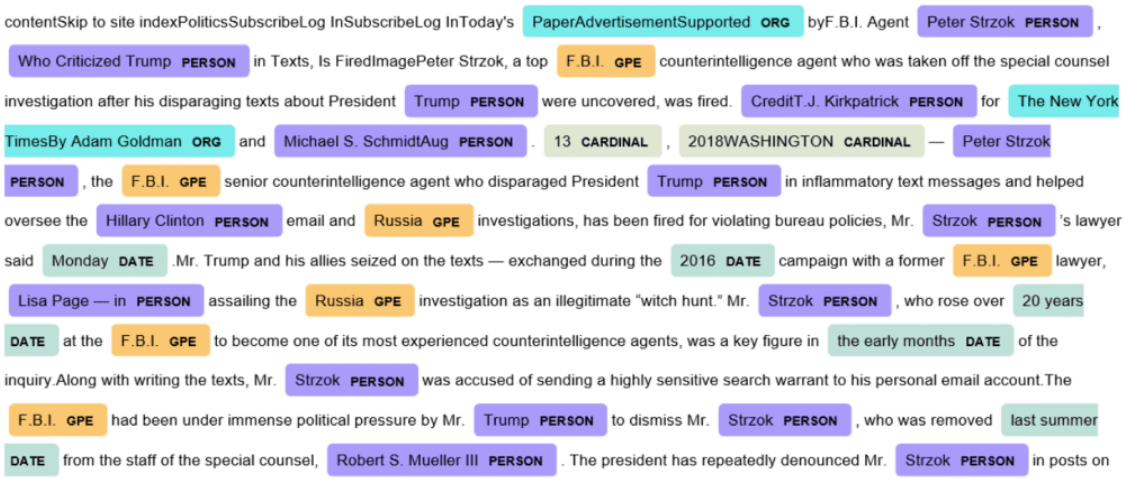
Source: Medium

Gambar 2.1 POS Tagging Example

Penandaan ini tidak hanya berdasarkan bentuk kata, tetapi juga mempertimbangkan konteks kemunculan kata dalam kalimat. Proses ini mengatasi ambiguitas kata yang memiliki kelas kata berbeda tergantung penggunaannya. Implementasi metode *ML/DL* untuk POS *tagging* dinilai mampu meningkatkan akurasi dan mengurangi kesalahan penandaan, meskipun membutuhkan sumber daya komputasi yang lebih besar. Oleh karena itu, perkembangan POS *tagging* berbasis AI menjadi tren penting dalam NLP untuk mendukung berbagai aplikasi analisis teks, termasuk dalam analisis artikel berita dan kebijakan.

2.4 Named Entity Recognition

Named Entity Recognition (NER) adalah salah satu teknik penting dalam Natural Language Processing (NLP) yang digunakan untuk mengidentifikasi dan mengklasifikasikan entitas tertentu seperti nama orang, organisasi, lokasi, tanggal, atau entitas lain dari suatu teks (Naseer et al., 2021). Proses ini bertujuan untvcbuk mengekstrak informasi yang bermakna dari data teks tidak terstruktur sehingga dapat diolah lebih lanjut untuk berbagai keperluan, termasuk analisis informasi, sistem tanya jawab, dan ekstraksi data. Sebagai contoh, kalimat “John membeli 500 saham di Acme Corp. pada tahun 2016” akan diidentifikasi menjadi entitas “John [Person]”, “Acme Corp [Organization]”, dan “2016 [Time]”.



Source: Wisecube AI

Gambar 2.2 NER Tagging Example

Dalam aplikasinya, berbagai library populer seperti *spaCy*, *StanfordNLP*, *TensorFlow*, dan *Apache OpenNLP* telah menyediakan model NER bawaan yang mendukung berbagai bahasa dan domain. Hasil penelitian menunjukkan bahwa *spaCy* memiliki performa terbaik dalam hal akurasi dan kecepatan prediksi dibandingkan *library* lainnya (Naseer et al., 2021). Oleh karena itu, NER menjadi salah satu komponen esensial dalam analisis teks modern, termasuk dalam konteks analisis berita, opini publik, dan dokumentasi kebijakan.

2.5 Library newspaper3k

Newspaper3k adalah sebuah pustaka Python yang dirancang untuk melakukan ekstraksi otomatis terhadap konten artikel berita dari situs web, termasuk mengambil teks artikel, judul, metadata, tanggal publikasi, dan elemen penting lainnya (Aniketh et al., 2025). Pustaka ini memanfaatkan teknik parsing HTML dan algoritma ekstraksi konten untuk mengidentifikasi bagian utama artikel, memisahkannya dari elemen web lain seperti iklan, menu, atau tautan navigasi. Dengan fitur ini, *newspaper3k* mempermudah proses web scraping terhadap berita online, sehingga konten yang diekstrak lebih bersih dan terstruktur. Menurut Aniketh et al. (2025), penggunaan *newspaper3k* dalam sistem pengumpulan data berita terbukti efisien dalam mengekstrak isi artikel dari berbagai sumber portal berita dengan format HTML yang beragam. Library ini juga memiliki keunggulan dalam kecepatan pemrosesan dan kemudahan integrasi dengan pipeline analisis Natural Language Processing (NLP) lainnya. Pada penelitian ini, *newspaper3k* diimplementasikan untuk mengambil konten berita sebelum dilakukan analisis sentimen dan summarization berbasis machine learning, sehingga mendukung alur kerja sistem otomatis untuk klasifikasi dan ringkasan berita.

3. DATA ACQUISITION

3.1 Scrapping Link Artikel Berita

Pada tahap akuisisi data, proses pengumpulan dilakukan dengan cara manual crawling terhadap artikel berita yang relevan. Hal ini dilakukan dengan menelusuri dan mencari artikel berita terkait kebijakan Danantara secara langsung melalui mesin pencari dan portal berita daring. Setiap artikel yang ditemukan kemudian didokumentasikan ke dalam sebuah file spreadsheet (SPS) sebagai dataset awal. Proses manual *crawling* ini dipilih karena sifat spesifik topik kebijakan Danantara yang belum tersedia dalam bentuk dataset siap pakai di platform berita daring atau API publik. Dengan metode ini, peneliti memiliki kontrol penuh terhadap kualitas, relevansi, dan keberagaman sumber berita yang digunakan dalam penelitian. Dataset hasil akuisisi manual ini selanjutnya digunakan sebagai input dalam tahap ekstraksi konten dan analisis lanjutan.

3.2 Pendataan *tag* Artikel Berita

Setelah proses pengumpulan link berita selesai dilakukan, langkah berikutnya adalah pendataan tag atau kategori artikel. Proses ini dilakukan secara manual dengan membaca dan menelaah konten masing-masing artikel untuk menentukan kategori yang sesuai. Setiap artikel diberi tag berdasarkan lima kategori yang telah ditetapkan, yaitu: (1) *Economy*, (2) *Local News*, (3) *Foreign News*, (4) *Opinion*, dan (5) *Education*. Penentuan tag dilakukan dengan mempertimbangkan tema utama yang diangkat dalam artikel, berdasarkan topik dominan, penggunaan istilah kunci, serta konteks isi berita. Sebagai contoh, artikel yang membahas kebijakan ekonomi Danantara dikategorikan ke dalam tag *Economy*, sementara artikel yang menyoroti reaksi masyarakat lokal terhadap kebijakan tersebut dimasukkan ke dalam *Local News*. Pendataan tag ini dicatat langsung dalam kolom tag pada dataset spreadsheet yang telah dibuat sebelumnya, berdampingan dengan kolom link artikel. Proses kategorisasi manual dipilih untuk memastikan akurasi klasifikasi berita, mengingat keterbatasan otomatisasi dalam mendeteksi konteks spesifik topik kebijakan Danantara.

3.3 Scrapping Konten Artikel Berita

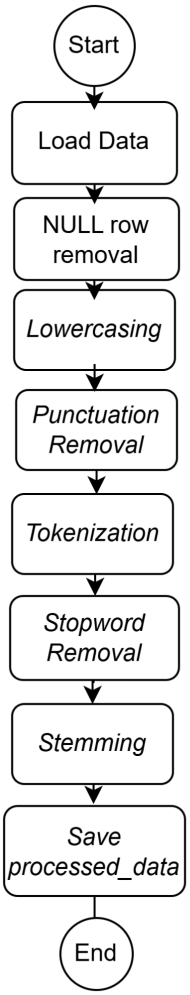
Setelah link artikel dan kategori tag terdokumentasi dalam dataset, langkah selanjutnya adalah pengambilan konten artikel secara otomatis menggunakan library *newspaper3k*. *Newspaper3k* dipilih sebagai alat ekstraksi karena kemampuannya dalam mengambil teks utama artikel, judul, tanggal publikasi, serta metadata lainnya secara cepat dan relatif bersih dari elemen-elemen non-konten seperti iklan, menu, atau tautan navigasi. Dalam tahap ini, setiap link artikel yang tercatat pada dataset diinputkan ke dalam fungsi *Article()* dari library *newspaper3k*. Proses scraping dilakukan dengan

memanfaatkan metode *download()* untuk mengunduh halaman artikel dan *parse()* untuk mengekstrak isi artikel. Hasil scraping ini kemudian disimpan dan ditambahkan ke dataset yang telah ada, sehingga setiap artikel memiliki kolom tambahan berupa judul, isi artikel, dan tanggal publikasi. Proses ini mempermudah persiapan data untuk tahap analisis lebih lanjut, seperti analisis sentimen, POS tagging, dan *Named Entity Recognition* (NER). Penggunaan *newspaper3k* dianggap efektif karena mampu mengekstrak konten dari berbagai struktur HTML portal berita dengan konsistensi yang baik, meskipun masih memerlukan pengecekan manual untuk memastikan kualitas data yang diambil. Dengan pendekatan ini, proses ekstraksi konten berita dapat dilakukan secara otomatis namun tetap terkontrol dari sisi validitas data.

4. DATA PREPROCESSING

4.1 Data Preprocessing Artikel Berita

Praproses dilakukan untuk membersihkan teks dari elemen-elemen yang tidak relevan serta mengubahnya ke dalam bentuk yang lebih terstruktur dan konsisten, sehingga data siap digunakan untuk tahap analisis lanjutan seperti analisis sentimen, frekuensi kata, *Part-of-Speech* (POS) tagging, dan *Named Entity Recognition* (NER). Seluruh tahapan praproses yang dilakukan dijelaskan secara visual melalui skema alur pada gambar berikut untuk memberikan gambaran proses secara menyeluruh dan sistematis.



Gambar 4.1 Data Preprocessing Flowchart

Praprosesing dimulai dengan *Load Data*, yaitu memuat dataset berisi isi artikel hasil *scraping*. Selanjutnya, dilakukan *Lowercasing*, yakni mengubah seluruh huruf menjadi huruf kecil untuk menghindari duplikasi kata akibat kapitalisasi. Setelah itu, dilakukan *Punctuation Removal*, yaitu penghapusan seluruh tanda baca dan karakter non-alfabetik yang tidak relevan untuk analisis linguistik.Langkah berikutnya adalah *Tokenization*, yaitu pemecahan teks menjadi unit kata (token) agar dapat diproses lebih lanjut. Setelah token diperoleh, dilakukan *Stopword Removal*, yakni penghapusan kata-kata umum yang tidak memiliki makna signifikan secara semantik, baik dalam bahasa Indonesia maupun Inggris. Kemudian, dilakukan proses *Stemming* menggunakan *library* Sastrawi untuk mengembalikan kata ke bentuk dasarnya. Terakhir, hasil praproses disimpan dalam bentuk terstruktur berbentuk df, dan siap digunakan untuk berbagai tahap analisis lanjutan.

5. DEFINISI DATASET

5.1 Dataset Link Artikel Berita

Dataset link artikel berita merupakan dataset awal yang digunakan dalam penelitian ini yang merupakan hasil pendataan manual yang disimpan dalam *file* bernama Link_Scraping_Danantara.csv, dengan format Comma-Separated Values (CSV). File ini berisi daftar artikel berita yang dikumpulkan secara manual dan berkaitan dengan kebijakan Danantara. Dataset terdiri atas dua kolom utama, yaitu:

Tabel 5.1 Deskripsi Dataset Link Artikel Berita

Kolom	Deskripsi
link	URL artikel berita yang telah dikurasi secara manual agar sesuai dengan topik penelitian.
tag	klasifikasi atau kategori isi artikel berdasarkan lima jenis tag yang telah ditentukan sebelumnya, yaitu: economy, local news, foreign news, opinion, dan education.

Dataset ini digunakan sebagai input awal dalam tahap scraping konten artikel berita menggunakan library newspaper3k sebagai sumber referensi utama dalam proses scraping konten. Data pada kolom link dimasukkan ke dalam pipeline ekstraksi konten menggunakan library newspaper3k, sementara data pada kolom tag digunakan sebagai label klasifikasi awal untuk keperluan analisis tren, visualisasi distribusi topik, dan evaluasi hasil analisis sentimen berdasarkan kategori.

5.2 Dataset Konten Artikel Berita

Dataset hasil scraping ini disimpan dalam file bernama Scraping_Danantara.csv, yang merupakan keluaran dari proses ekstraksi konten berita menggunakan library newspaper3k. Dataset ini diperoleh dengan memproses daftar tautan artikel yang sebelumnya dikumpulkan dan disusun dalam dataset awal. Tujuan utama dari dataset ini adalah menyediakan isi teks artikel secara lengkap beserta metadata penting untuk keperluan analisis teks lanjutan seperti analisis sentimen, frekuensi kata, POS tagging, dan Named Entity Recognition (NER).

Tabel 5.2 Deskripsi Dataset Konten Artikel Berita

Kolom	Deskripsi
title	Judul artikel berita yang diperoleh.

authors	Nama penulis artikel atau informasi penulis jika tersedia
source	Nama domain atau portal berita asal artikel
published_date	Tanggal dan waktu publikasi artikel.
summary	Ringkasan otomatis artikel (hasil fitur summarization dari newspaper3k).
content	Isi lengkap teks utama artikel berita.
url	Tautan sumber artikel yang diambil dari dataset awal.
tag	Label setiap tautan artikel (1) *Economy*, (2) *Local News*, (3) *Foreign News*, (4) *Opinion*, dan (5) *Education*.

Dataset ini digunakan sebagai dasar utama dalam seluruh tahap analisis Natural Language Processing (NLP), di mana kolom *content* berisi isi teks utama artikel yang menjadi objek utama dalam berbagai proses seperti praproses teks (cleaning, tokenisasi, stopword removal, dan stemming), analisis sentimen (polarity dan subjectivity), ekstraksi entitas (Named Entity Recognition/NER), serta analisis linguistik lainnya seperti POS tagging dan visualisasi frekuensi kata. Dengan memanfaatkan library newspaper3k, proses ekstraksi konten dilakukan secara otomatis dan konsisten dari berbagai portal berita daring, sehingga mempercepat akuisisi data dalam skala besar sekaligus menjaga validitas dan kebersihan struktur teks yang diperoleh.

5.3 Dataset Hasil Preprocessing dan Analisis Sentimen

Dataset ini merupakan hasil dari proses scraping dan praproses teks terhadap artikel-artikel berita yang membahas kebijakan Danantara. Selain itu, dataset ini juga memuat hasil analisis sentimen menggunakan library TextBlob, yang direpresentasikan dalam dua kolom yaitu polarity dan subjectivity.

Tabel 5.3 Deskripsi Dataset Hasil Preprocessing dan Analisis Sentimen

Kolom	Deskripsi
title	Judul artikel berita yang diperoleh.
authors	Nama penulis artikel atau informasi penulis jika tersedia
source	Nama domain atau portal berita asal artikel
published_date	Tanggal dan waktu publikasi artikel.

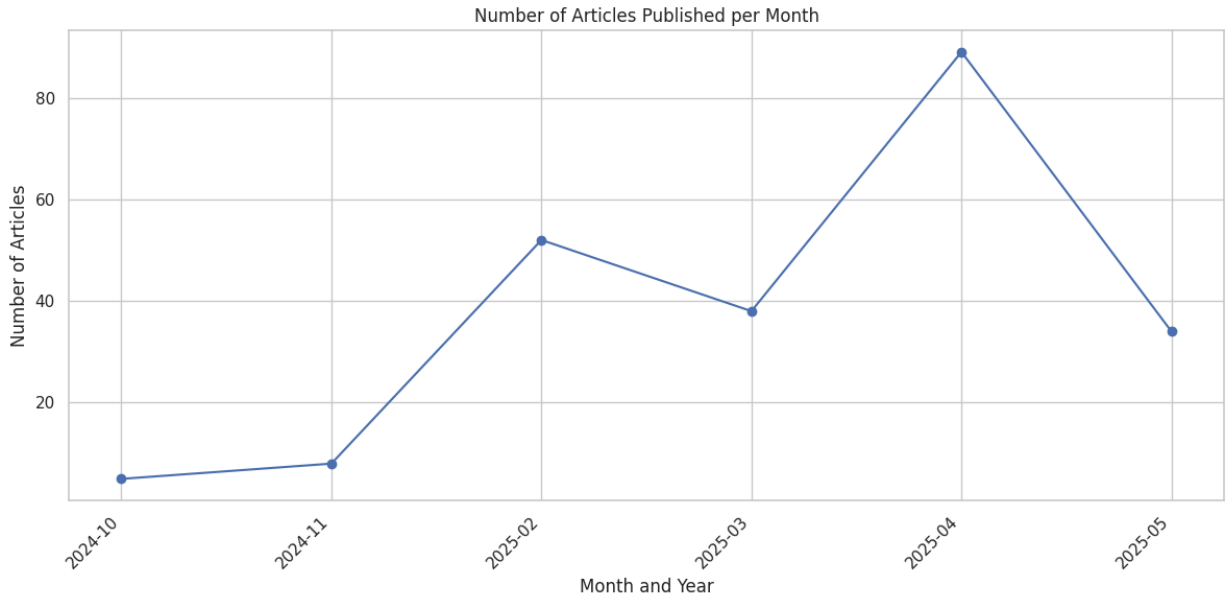
summary	Ringkasan otomatis artikel (hasil fitur summarization dari newspaper3k).
content	Isi lengkap teks utama artikel berita.
url	Tautan sumber artikel yang diambil dari dataset awal.
polarity	Skor sentimen (-1 hingga 1) untuk menunjukkan apakah teks bernada negatif/positif
subjectivity	Skor subjektivitas (0 hingga 1) untuk menunjukkan apakah teks bersifat opini atau fakta
original	Teks artikel yang telah dibersihkan dan dikonversi ke huruf kecil
tokens_awal	Tokenisasi awal sebelum penghapusan stopword
no_stopwords	Tokenisasi setelah penghapusan kata umum yang tidak informatif (stopword)
stemmed	Token hasil stemming ke bentuk kata dasar menggunakan Sastrawi

Nilai polarity menunjukkan kecenderungan opini dalam teks (dari -1 untuk sentimen negatif hingga +1 untuk sentimen positif), sedangkan subjectivity mengukur tingkat subjektivitas teks (dari 0 sebagai objektif hingga 1 sebagai sangat subjektif). Dataset disimpan dalam file berformat CSV dengan nama Danantara_Preprocessed.csv, yang berisi total 13 kolom yang mencakup informasi metadata, isi artikel, hasil analisis sentimen, dan hasil praproses.

6. HASIL

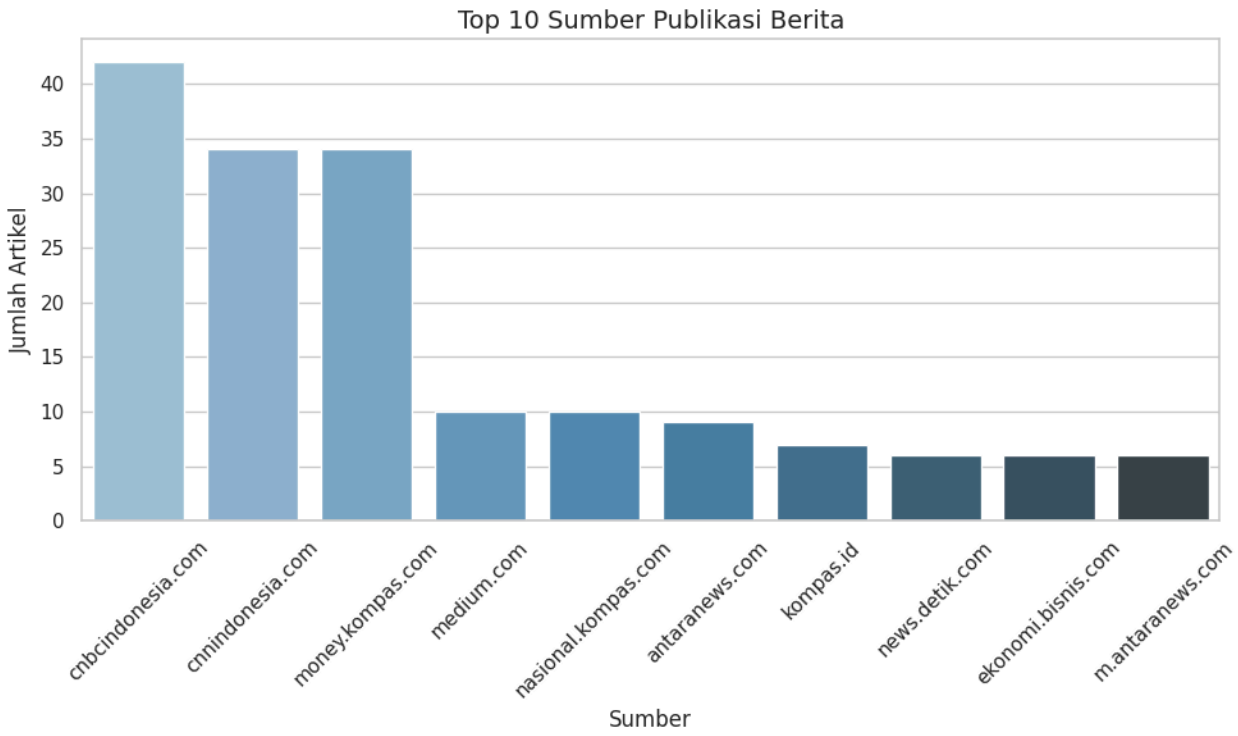
6.1 Analisis Tren Artikel

Berdasarkan hasil proses scrapping isi konten dari link artikel yang telah dikumpulkan menggunakan library newspaper3k, diperoleh sebanyak 226 artikel yang berhasil diambil isi konten dan publish date dari total 369 link artikel yang telah dikumpulkan. Dengan grafik distribusi jumlah artikel per bulan seperti pada gambar 6.1



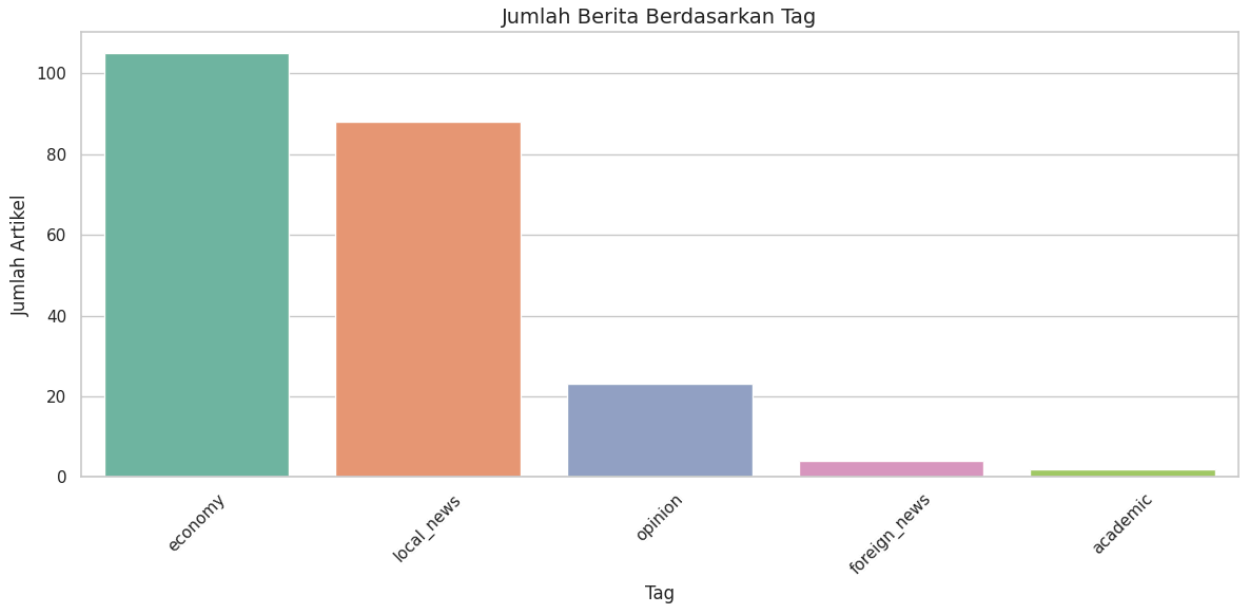
Gambar 6.1 Jumlah Artikel yang Diterbitkan tiap Bulan

Gambar 6.1 menunjukkan tren jumlah artikel yang diterbitkan tiap bulan dari Oktober 2024 hingga Mei 2025. Dari grafik tersebut terlihat peningkatan signifikan pada bulan februari dengan total 52 artikel, hal ini mungkin terjadi karena bertepatan dengan pengumuman resmi dari Presiden Prabowo Subianto tentang peluncuran danantara. Dilanjutkan dengan penurunan dan kenaikan kembali sampai di titik tertinggi pada bulan April, dari sebelumnya sebanyak 38 artikel pada bulan maret menjadi 89 artikel pada bulan april. Peningkatan ini kemungkinan terjadi karena bertepatan dengan pengumuman kerjasama antara Danantara dan Qatar Investment Authority yang diumumkan pada tanggal 15 April 2025.



Gambar 6.2 Top 10 Sumber Publikasi Berita

Gambar 6.2 menunjukkan 10 sumber berita dengan jumlah artikel paling banyak pada dataset. Dapat dilihat bahwa artikel paling banyak berasal dari cnbcindonesia.com (), diikuti dengan cnnindonesia.com (), money.kompas.com (), dan sisanya tersebar di beberapa media lain, seperti medium.com, nasional.kompas.com, dan antaranews.com.



Gambar 6.3 Persebaran Artikel berdasarkan Tag

Berdasarkan gambar 6.3 terlihat bahwa pemberitaan tentang danantara sangat didominasi oleh aspek ekonomi dan lokal. Hal ini dapat disebabkan oleh Danantara sendiri yang merupakan badan pengelolaan kekayaan negara dan dampaknya terhadap berbagai sektor strategis di indonesia. Jumlah berita yang rendah pada kategori “foreign_news” dan “academic” menunjukkan potensi pengembangan lebih lanjut dalam kajian akademik dan minat media internasional terhadap isu ini.

6.2 Analisis Frekuensi Kata

Data yang telah melalui proses preprocessing selanjutnya di analisis untuk frekuensi kata yang muncul. Berikut merupakan tabel yang menampilkan 10 kata dengan frekuensi kemunculan terbanyak pada seluruh artikel terkait danantara.

Tabel 6.1 Top 15 Frekuensi data Muncul

No	Kata	Frekuensi
1	danantara	2430
2	indonesia	985
3	bumn	967
4	investasi	881
5	negara	689
6	prabowo	562
7	aset	522
8	presiden	422
9	badan	403
10	bpi	386
11	rp	385
12	triliun	364
13	ekonomi	352
14	perusahaan	333

15	pemerintah	319

Pada Tabel 6.1 dapat dilihat bahwa kata terbanyak yang ada pada artikel adalah danantara dengan frekuensi kemunculan sebanyak 2430 kata, hal ini sesuai dengan topik utama dalam analisis. Kata lain yang juga terlihat cukup menonjol adalah indonesia, bumn, investasi, dan prabowo, yang dapat menandakan isu-isu terkait kebijakan, aktor politik, serta entitas ekonomi yang terlibat dalam pemberitaan terkait danantara.

Analisis dilanjutkan secara visual dengan menggunakan wordcloud terpisah berdasarkan sentimen positif dan negatif. Tujuannya adalah untuk mengetahui perbedaan narasi atau fokus kata dalam artikel dengan sentimen positif dan negatif.

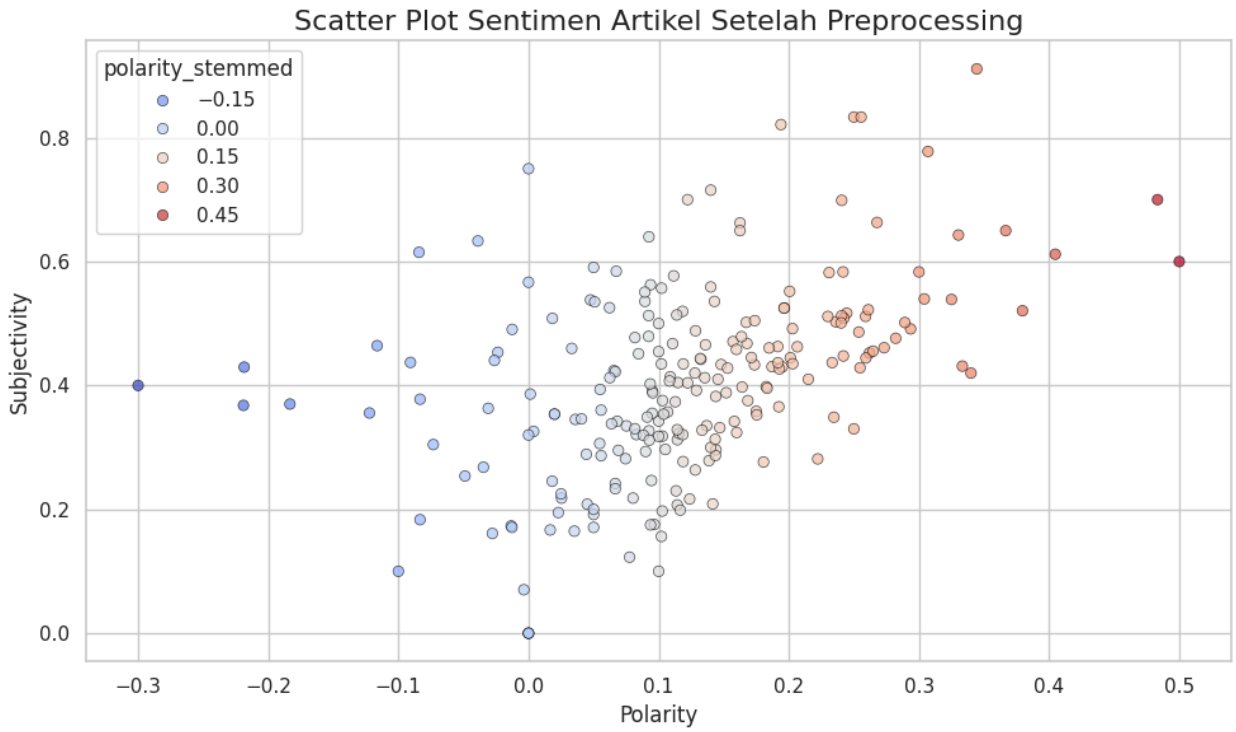
Gambar 6.4 Wordcloud Sentimen Negatif

Gambar 6.5 Wordcloud Sentimen Positif

Dari gambar 6.4 dan gambar 6.5 terlihat kesamaan beberapa kata umum, seperti danantara, indonesia, dan bumn yang menandakan bahwa ketiganya adalah kata kunci yang konsisten dibahas pada pemberitaan danantara baik dalam narasi positif maupun negatif. Selain kata umum tersebut dapat terlihat dari gambar 6.4 bahwa kata yang cukup spesifik dibahas pada artikel dengan sentimen negatif adalah menteri, dana, proyek, aset, dan prabowo. Kumpulan kata ini menunjukkan bahwa narasi negatif lebih banyak berfokus pada aktor politik, alokasi dana, serta pengelolaan proyek dan aset. Sehingga dapat disimpulkan bahwa kritik media atau publik terkait danantara cenderung negatif pada pengelolaan, keuangan, transparansi, dan kepemimpinan. Sedangkan pada gambar 6.5 kata-kata yang menonjol pada artikel dengan sentimen positif adalah investasi, kelola, ekonomi, dan perintah. Kumpulan kata ini menunjukkan bahwa narasi positif lebih banyak berfokus pada optimisme potensi investasi, pengelolaan aset, kontribusi terhadap ekonomi, dan arahan kebijakan pemerintah. Sentimen positif ini umumnya muncul ketika membahas manfaat dan peluang dari kebijakan Danantara.

6.3 Analisis Sentimen

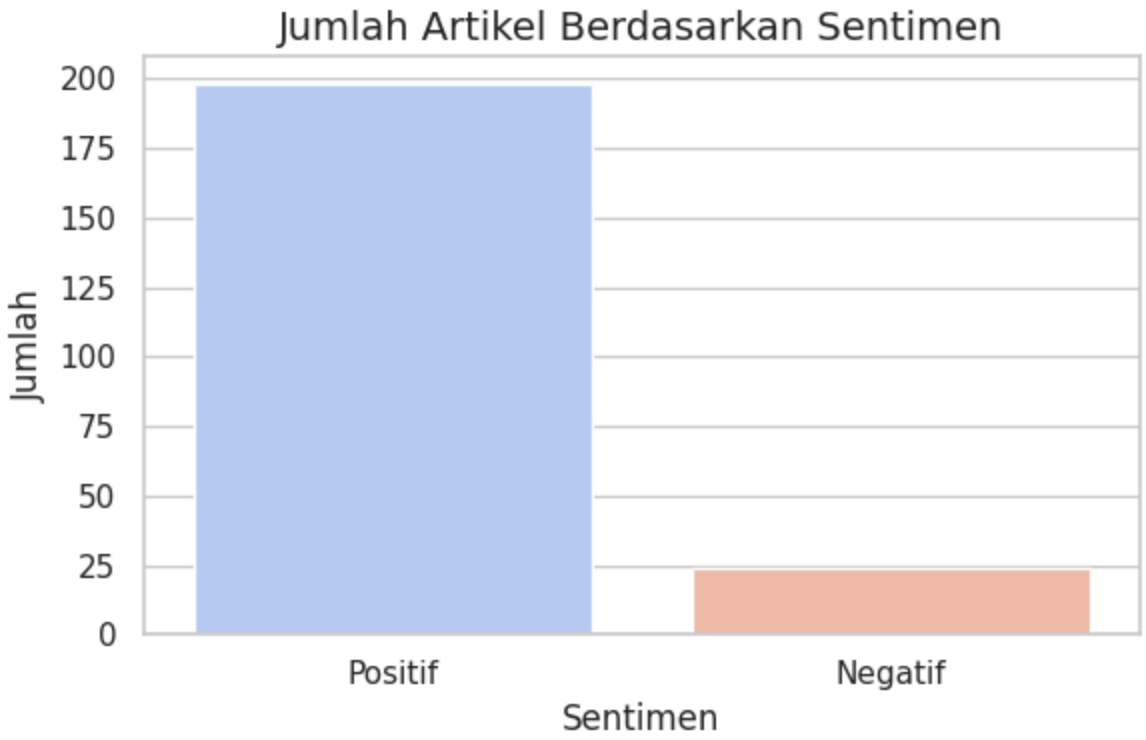
Analisis sentimen dilakukan menggunakan library TextBlob yang melalui proses translasi terlebih dahulu. Translasi diperlukan karena TextBlob dibangun di atas lexicon dan model analisis seperti *Pattern Analyzer* yang lebih banyak mengenali kata-kata dalam bahasa Inggris. Oleh karena itu diperlukan translasi menggunakan *library* Google Translate. Setelah menggunakan TextBlob, skor *Polarity* dan *Subjectivity* dihasilkan dan dapat digunakan untuk pengambilan insight.



Gambar 6.6 Scatter Plot Sentimen Artikel Setelah Preprocessing

Scatter plot ini menunjukkan hubungan antara nilai polarity dan subjectivity dari artikel-artikel berita yang telah melalui proses preprocessing teks. Sumbu X menunjukkan Polaritas yaitu seberapa positif atau negatif sentimen yang dimiliki suatu artikel berkisar -1 (sangat negatif) hingga +1 (sangat positif). Dalam plot ini titik paling ekstrim dari polaritas nya adalah -0.3 dan 0.5. Sumbu Y menunjukkan subjektivitas yang mengukur berapa subjektif atau objektif teks bersibut. Nilai ini berkisar antara 0 (sangat objektif) hingga 1 (sangat subjektif). Sebagian besar data terkonsentrasi di sekitar nilai polarity netral hingga positif ringan (sekitar 0 hingga 0.2), dengan tingkat subjectivity yang bervariasi dari rendah hingga sedang. Hal ini menunjukkan bahwa mayoritas artikel bersifat netral atau memiliki sentimen positif ringan, yang sesuai dengan karakteristik umum berita yang cenderung menjaga objektivitas. Namun, terdapat pula beberapa artikel dengan polaritas negatif ringan, meskipun jumlahnya relatif lebih sedikit. Warna titik dalam grafik mewakili tingkat polaritas yang telah distem, dengan gradasi dari biru (negatif), putih (netral), hingga merah (positif). Secara umum, plot ini menunjukkan bahwa meskipun sebagian besar berita cenderung objektif, beberapa di antaranya mengandung opini atau nada emosional yang mencerminkan sentimen positif maupun negatif.

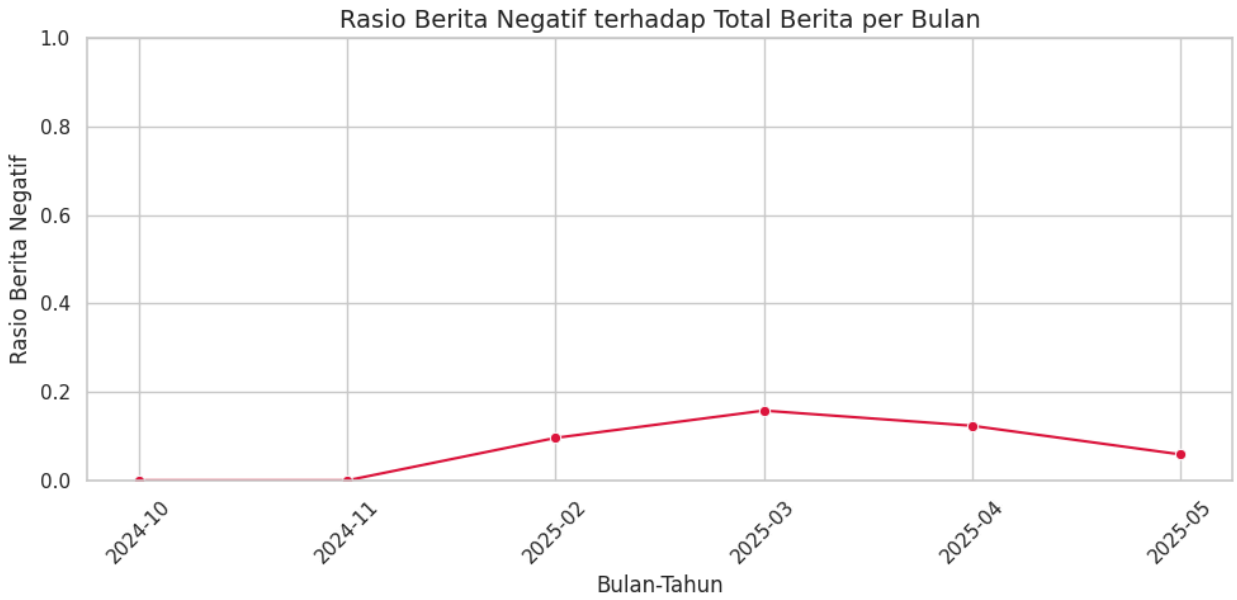
6.3.1 Jumlah Berita berdasarkan Sentimen Polaritas



Gambar 6.7 Bar Plot Jumlah Berita berdasarkan Sentimen

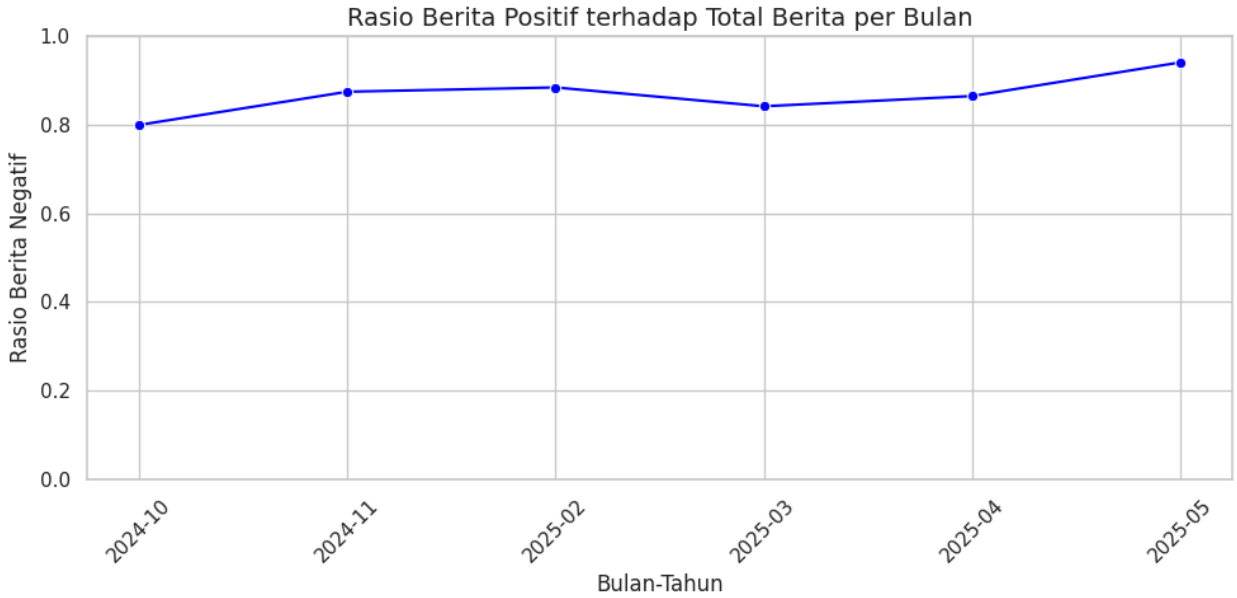
Berdasarkan grafik di atas, dapat dilihat bahwa jumlah artikel dengan sentimen positif jauh lebih banyak dibandingkan dengan artikel bersentimen negatif. Tercatat sekitar 198 artikel dikategorikan sebagai positif, sementara hanya sekitar 22 artikel yang termasuk dalam kategori negatif. Hasil ini menunjukkan bahwa sebagian besar artikel berita yang dianalisis memiliki nada yang positif atau netral-positif setelah melalui proses analisis sentimen.

6.3.2 Analisis Sentimen Berdasarkan Waktu



Gambar 6.8 Rasio Berita Negatif terhadap Total Berita per Bulan

Grafik di atas menunjukkan rasio berita negatif terhadap total berita yang dipublikasikan setiap bulan. Terlihat bahwa pada bulan Oktober dan November 2024, tidak ada berita negatif yang terdeteksi. Rasio mulai meningkat pada Februari 2025, mencapai puncaknya di bulan Maret 2025 dengan sekitar 16% dari total berita bersentimen negatif. Setelah itu, rasio mengalami penurunan bertahap pada April dan Mei 2025. Hal ini mengindikasikan bahwa meskipun sentimen negatif sempat meningkat, secara umum proporsinya tetap rendah dibandingkan jumlah total berita setiap bulan.

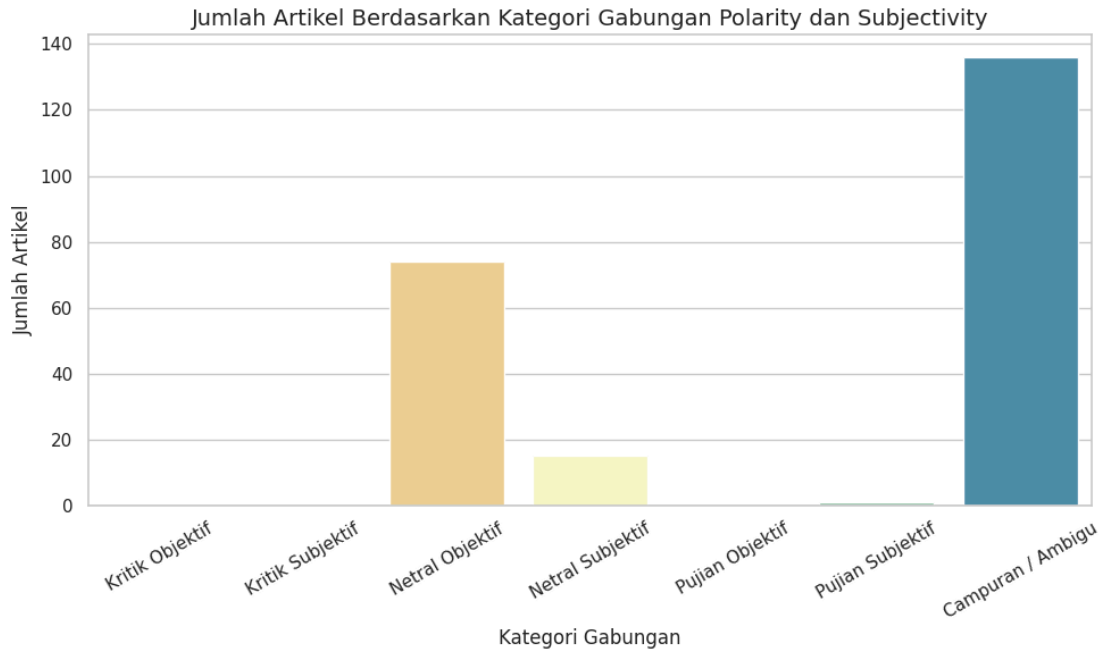


Gambar 6.9 Rasio Berita Positif terhadap Total Berita per Bulan

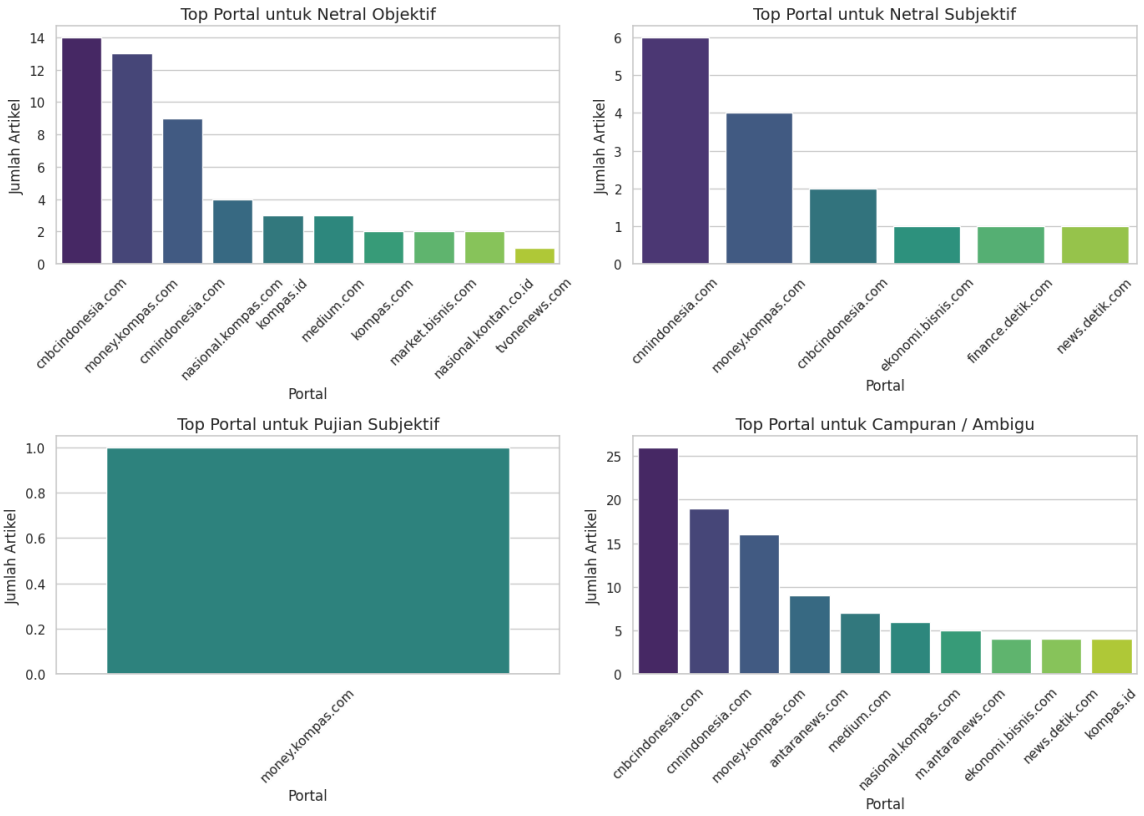
Grafik di atas memperlihatkan rasio berita positif terhadap total berita yang dipublikasikan setiap bulan. Secara umum, rasio berita positif tetap tinggi di setiap periode, berada di atas 80%. Rasio ini mengalami peningkatan dari bulan Oktober 2024 hingga Mei 2025, dengan sedikit penurunan pada Maret 2025 sebelum kembali naik dan mencapai puncaknya pada Mei 2025 di angka mendekati 95%. Temuan ini menunjukkan bahwa mayoritas berita yang dipublikasikan setiap bulan cenderung bersentimen positif, dengan tren yang relatif stabil dan meningkat seiring waktu.

6.3.3 Analisis Sentimen Berdasarkan Portal

Klasifikasi gabungan polaritas dan subjektivitas artikel dilakukan dengan membagi ke dalam tujuh kategori berdasarkan nilai polarity dan subjectivity. Artikel dengan polaritas sangat negatif (≤ -0.5) dikategorikan sebagai Kritik, dan yang sangat positif (≥ 0.5) sebagai Pujian, masing-masing dibedakan lagi menjadi Objektif (subjektivitas ≤ 0.5) dan Subjektif (subjektivitas > 0.5). Artikel dengan polaritas netral (antara -0.1 hingga 0.1) dibagi menjadi Netral Objektif dan Netral Subjektif berdasarkan subjektivitasnya. Sementara itu, artikel yang tidak memenuhi kriteria tersebut dimasukkan ke dalam kategori Campuran / Ambigu. Klasifikasi ini bertujuan untuk menangkap tidak hanya sikap positif atau negatif, tetapi juga sejauh mana narasi bersifat fakta atau opini. Pendekatan ini penting untuk mengidentifikasi peran bahasa dalam membentuk opini publik, serta untuk mengevaluasi kecenderungan editorial dari masing-masing media dalam menyampaikan informasi, khususnya yang berkaitan dengan isu-isu ekonomi.



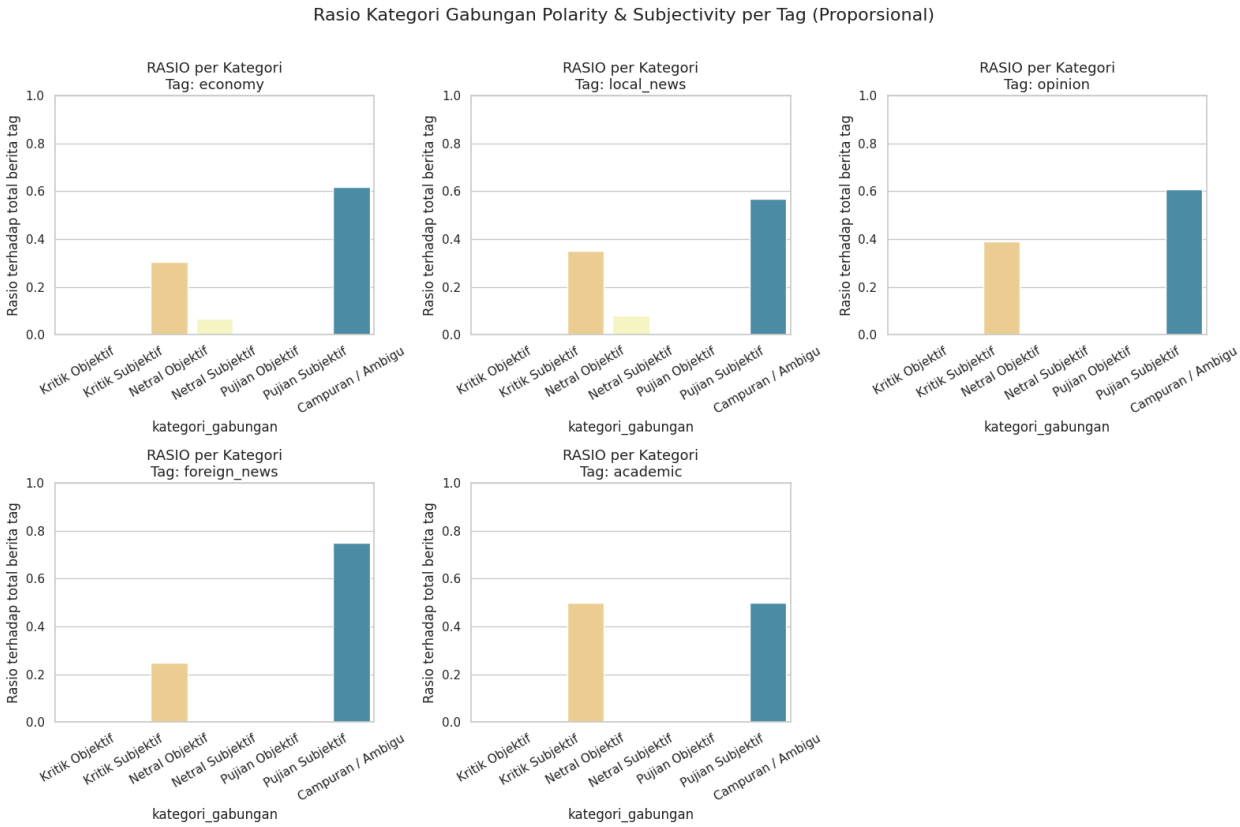
Gambar 6.10 Bar Plot Jumlah Artikel berdasarkan Kategori Gabungan Polarity dan Subjectivity



Gambar 6.11 Bar Plot Jumlah Publikasi Berdasarkan Sumber

Hasil visualisasi menunjukkan bahwa portal cnbcindonesia.com dan money.kompas.com mendominasi kategori Netral Objektif, yang berarti mereka lebih banyak mempublikasikan berita ekonomi yang bersifat faktual dan netral. Sementara itu, cnnindonesia.com lebih banyak muncul dalam kategori Netral Subjektif, yang mengindikasikan adanya penyampaian informasi netral namun dibalut dengan opini atau gaya bahasa yang subjektif. Kategori Pujian Subjektif hanya diisi oleh satu artikel dari money.kompas.com, menunjukkan bahwa gaya pemberitaan yang sangat positif dan sarat opini jarang digunakan oleh portal-portal yang dianalisis. Sebaliknya, kategori Campuran/Ambigu cukup mendominasi, dengan cnbcindonesia.com sebagai penyumbang terbanyak, diikuti oleh cnnindonesia.com dan money.kompas.com. Ini menunjukkan bahwa banyak artikel yang memiliki karakteristik bahasa yang tidak bisa digolongkan secara pasti sebagai positif, negatif, netral, ataupun faktual sepenuhnya, sehingga dinilai bersifat campuran atau ambigu.

6.3.4 Analisis Sentimen berdasarkan Tag Artikel



Gambar 6.12 Rasio Klasifikasi Sentimen berdasarkan Tag Artikel

Visualisasi pada Gambar menunjukkan distribusi proporsional dari kategori gabungan polaritas dan subjektivitas berdasarkan lima kategori tag utama: economy, local_news, opinion, foreign_news, dan academic. Secara umum, kategori Campuran / Ambigu mendominasi hampir semua tag, yang menandakan bahwa banyak artikel memuat elemen bahasa yang tidak secara konsisten objektif maupun subjektif, serta tidak menunjukkan arah sentimen yang jelas. Pada tag economy, terlihat bahwa proporsi artikel Netral Objektif relatif tinggi dibandingkan tag lainnya, menunjukkan adanya kecenderungan untuk menyampaikan berita ekonomi secara faktual. Sementara pada tag opinion, proporsi Netral Objektif dan Campuran / Ambigu hampir seimbang, mencerminkan adanya opini yang tetap menjaga keseimbangan tanpa kecenderungan emosional yang kuat.

Untuk tag *foreign_news*, *Campuran / Ambigu* mencapai lebih dari 70%, mengindikasikan bahwa liputan luar negeri sering kali ditulis dengan gaya bercampur atau tidak eksplisit dalam mengekspresikan opini atau emosi. Tag *academic* memperlihatkan dua kutub utama yaitu *Netral Objektif* dan *Campuran / Ambigu*, memperkuat asumsi bahwa berita akademik berusaha menyajikan data dan fakta, namun tidak sepenuhnya bebas dari interpretasi atau penyusunan narasi yang mengandung ambiguitas. Tag *local_news* memiliki komposisi serupa dengan

economy, meskipun porsi *Netral Subjektif* juga mulai terlihat, yang dapat dikaitkan dengan peran narasi lokal dan kedekatan emosional terhadap isu di sekitar masyarakat.

6.4 TF-IDF Konten Artikel

Analisis Term Frequency-Inverse Document Frequency (TF-IDF) dilakukan untuk mengidentifikasi kata-kata yang memiliki bobot penting dalam artikel. Hasil dari TF-IDF dapat menunjukkan kata-kata yang sering muncul tapi tidak umum, atau bisa dibilang spesifik kepada topik tertentu. Pada analisis ini, TF-IDF digunakan untuk mencari top words dari seluruh artikel untuk melihat topik secara keseluruhan dan dengan mengelompokkan berdasarkan periode bulan untuk mengetahui dinamika topik pembicaraan setiap bulannya.

Tabel 6.2 Top 5 TF-IDF untuk seluruh Artikel

No	Kata	TF-IDF Score
1	bumn	0.0729
2	aset	0.0545
3	prabowo	0.0477
4	negara	0.0448
5	ros	0.0390

Berdasarkan hasil pada tabel 6.2 dapat dilihat bahwa mayoritas artikel danantara berhubungan dengan pembahasan terkait bumn, aset, prabowo, dan negara. Hal ini wajar mengingat danantara adalah proyek besar negara dimasa kepemimpinan Presiden Prabowo dan merupakan badan yang mengelola aset dan bumn.

Tabel 6.3 Top 5 TF-IDF Bulan Oktober

No	Kata	TF-IDF Score
1	kelola	0.2284
2	investasi	0.1822
3	bumn	0.1513
4	badan	0.1344
5	aset	0.1223

Berdasarkan analisis pada Tabel 6.2 kemungkinan topik utama pada bulan oktober tentang pengelolaan investasi BUMN, hal ini menandakan artikel pada bulan ini berfokus tentang strategi atau aktivitas pengelolaan aset dan investasi dana pemerintah.

Tabel 6.4 Top 5 TF-IDF Bulan November

No	Kata	TF-IDF Score
1	bumn	0.1516
2	presiden	0.1404
3	bentuk	0.1294
4	luncur	0.0760
5	bp	0.0732

Tabel 6.3 menunujukkan bahwa topik utama pada bulan November adalah seputar presiden, bumn, bentuk, serta luncur. Hal ini mengindikasikan artikel mengenai peluncuran kebijakan oleh presiden atau pembentukan badan baru yang melibatkan BUMN dan pemerintah pusat.

Tabel 6.5 Top 5 TF-IDF Bulan Februari

No	Kata	TF-IDF Score
1	negara	0.0608
2	bank	0.0565
3	dana	0.0540
4	ekonomi	0.0524
5	awas	0.0471

Berdasarkan Tabel 6.4 kemungkinan topik pada bulan Februari didominasi oleh isu keuangan yang berhubungan dengan danantara seperti pengelolaan dana, bank, dan ekonomi. Hasil ini kemungkinan membahas tentang dampak danantara terhadap kondisi negara, bank, dana, dan ekonomi yang sedang berbahaya.

Tabel 6.6 Top 5 TF-IDF Bulan Maret

No	Kata	TF-IDF Score
1	bumn	0.0853
2	kelola	0.0697
3	investasi	0.0693
4	proyek	0.0624
5	indonesia	0.0613

Tabel 6.5 menunujukkan bahwa pada bulan maret isu yang membahas bumn, pengelolaan, dan investasi kembali menjadi sorotan. Diiktui dengan kata proyek dan indonesia, mengindikasikan kemungkinan pengelolaan proyek investasi bumn oleh danantara

Tabel 6.7 Top 5 TF-IDF Bulan April

No	Kata	TF-IDF Score
1	qatar	0.0695
2	bumn	0.0695
3	prabowo	0.0663
4	aset	0.0632
5	ros	0.0531

Tabel 6.6 menunjukkan adanya topik baru pada bulan April yang membahas tentang kerjasama bilateral antara indonesia dan qatar yang melibatkan Presiden Prabowo. Kerjasama ini juga sepertinya berdampak pada bumn serta aset.

Tabel 6.8 Top 5 TF-IDF Bulan Mei

No	Kata	TF-IDF Score
1	aset	0.0859
2	gates	0.0748
3	ros	0.0615

4	bumn	0.0600
5	trilliun	0.0584

Tabel 6.7 menunjukkan topik pembahasan baru pada bulan Mei, yaitu Bill Gates yang kemungkinannya akan berinvestasi dengan nilai yang cukup besar (trilliunan). Selain itu bumn dan aset masih menjadi pembahasan berulang.

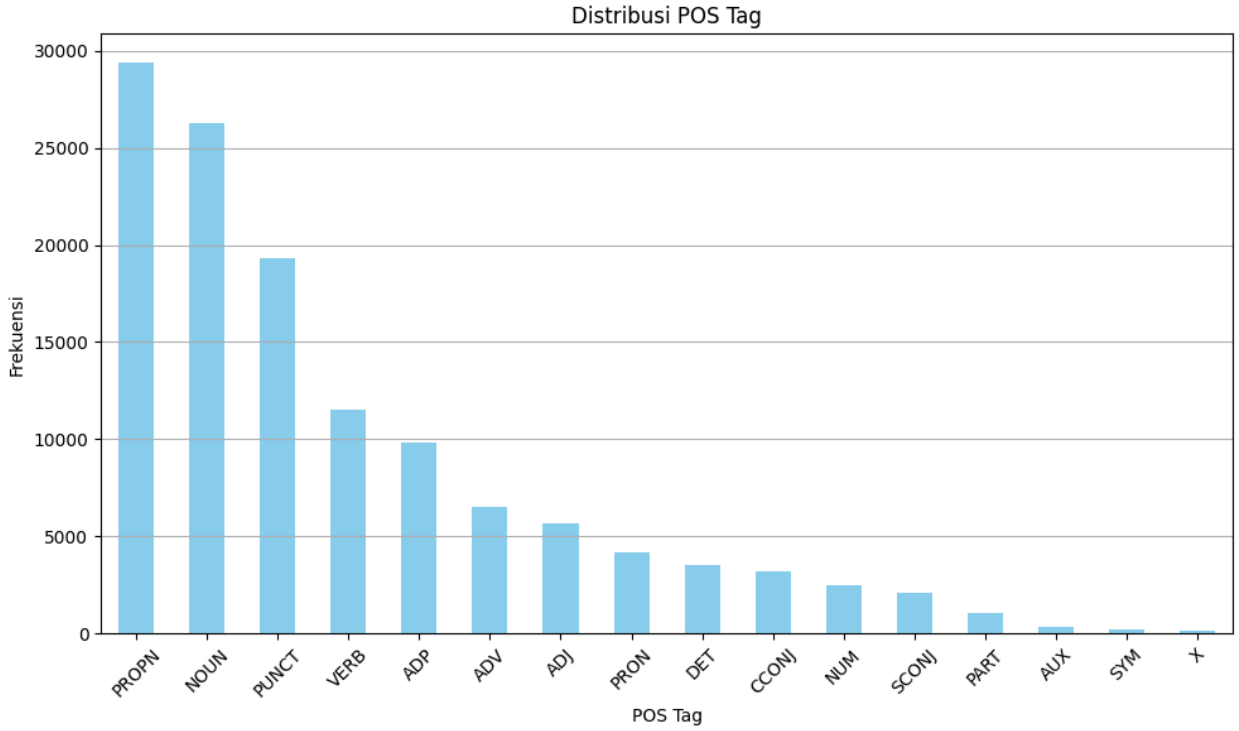
6.4 POS dan NER Konten Artikel Berita

Tabel 6.9 Data POS Universal

Tag	Nama Lengkap	Penjelasan Singkat
NOUN	Noun	Kata benda umum (buku, meja, presiden)
PROPN	Proper Noun	Nama khusus (Indonesia, Jokowi, Danantara)
VERB	Verb	Kata kerja utama (makan, pergi, membaca)
AUX	Auxiliary Verb	Kata kerja bantu (telah, akan, sedang)
ADJ	Adjective	Kata sifat (besar, cepat, hijau)
ADV	Adverb	Kata keterangan (cepat, sangat, dulu)
PRON	Pronoun	Kata ganti (saya, kamu, dia, mereka)
DET	Determiner	Penentu (ini, itu, setiap, semua)
ADP	Adposition	Preposisi (di, ke, dari, pada)
SCONJ	Subordinating Conj.	Konjungsi subordinatif (karena, jika, agar)
CCONJ	Coordinating Conj.	Konjungsi koordinatif (dan, atau, tetapi)
PART	Particle	Partikel (lah, pun, kah)
INTJ	Interjection	Seruan/emosi (hai, wah, aduh)
NUM	Numeral	Angka/bilangan (dua, 10, ketiga)
PUNCT	Punctuation	Tanda baca (., !, :)

SYM	Symbol	Simbol (%, $, #)
X	Other/Unknown	Kategori tidak dikenali (biasanya error tagging)

Tabel 6.9 menyajikan daftar lengkap tag Part-of-Speech (POS) berdasarkan skema Universal POS Tagging yang digunakan dalam analisis sintaksis Bahasa Indonesia. Setiap tag merupakan representasi kelas kata seperti kata benda (NOUN), kata kerja (VERB), kata sifat (ADJ), dan lain-lain. Penjelasan singkat dan contoh penggunaan disertakan untuk mempermudah pemahaman terhadap fungsi masing-masing tag dalam struktur kalimat.



Gambar 6.12 Grafik Distribusi POS Tag

Dari grafik distribusi POS Tag diperoleh informasi PROPN menjadi tag paling dominan dengan hampir 30.000 kemunculan. Hal ini mengindikasikan teks yang dianalisis **sangat kaya dengan nama entitas seperti orang, tempat, atau institusi**. Ini umum dijumpai dalam teks seperti berita, laporan, atau artikel formal yang menyebut banyak nama diri. Kemudian NOUN menempati posisi kedua, menunjukkan banyak kata benda umum digunakan yang memperkuat indikasi bahwa teks banyak menyampaikan **objek, konsep,** atau **peristiwa**. Di peringkat ketiga ditemukan banyak PUNCT (tanda baca) yang diasumsikan karena diperoleh dari artikel yang umumnya memiliki **gaya penulisan** dengan struktur kalimat yang **lengkap** dan **tertulis formal**.

Hal itu juga menyebabkan kemunculan SYM (simbol), X (unspecified), AUX (kata kerja bantu), dan PART (partikel) **muncul sangat jarang** mengindikasikan bahwa teks tidak mengandung banyak simbol atau kata yang tidak dikenali, bentuk kalimat pasif atau progresif kurang dominan, dan teks mungkin tidak bersifat sangat percakapan/informal

Gambar 6.13 *Display POS Tag Result* pada Dataset Danantara

Gambar 6.13 menampilkan hasil visualisasi Part-of-Speech (POS) tagging terhadap salah satu konten artikel berbahasa Indonesia. Setiap kata diberi label kelas katanya sesuai dengan skema Universal POS Tagging, seperti NOUN (kata benda), VERB (kata kerja), ADJ (kata sifat), PROPN (nama diri), dan sebagainya. Warna latar yang berbeda digunakan untuk membedakan jenis tag, sehingga memudahkan dalam mengidentifikasi struktur gramatikal kalimat secara visual. Visualisasi ini membantu memahami distribusi dan fungsi sintaktik kata-kata dalam sebuah teks, serta mendukung analisis linguistik maupun praproses data dalam tugas-tugas NLP.

Tabel 6.11 Top 10 POS Tag *counts*

No	Word	POS	Count
1	Danantara	PROPN	2157
2	BUMN	PROPN	911
3	Indonesia	PROPN	901
4	investasi	NOUN	574
5	Prabowo	PROPN	553
6	negara	NOUN	473
7	aset	NOUN	403
8	BPI	PROPN	383
9	Presiden	PROPN	375
10	ekonomi	NOUN	318

Tabel 6.11 memberikan gambaran tentang sepuluh kata yang paling sering muncul dalam korpus teks yang dianalisis beserta kategori kelas katanya (POS). Hasil ini secara langsung **menjawab rumusan masalah terkait identifikasi kata kunci dominan dan pola linguistik yang muncul dalam pembahasan isu Danantara Indonesia.** Terlihat bahwa Danantara muncul secara signifikan lebih tinggi dibandingkan kata lainnya, dengan jumlah kemunculan sebanyak 2.157 kali dan dikategorikan sebagai PROPN (proper noun). Hal ini menunjukkan bahwa topik Danantara merupakan fokus utama dalam narasi dokumen yang diteliti.

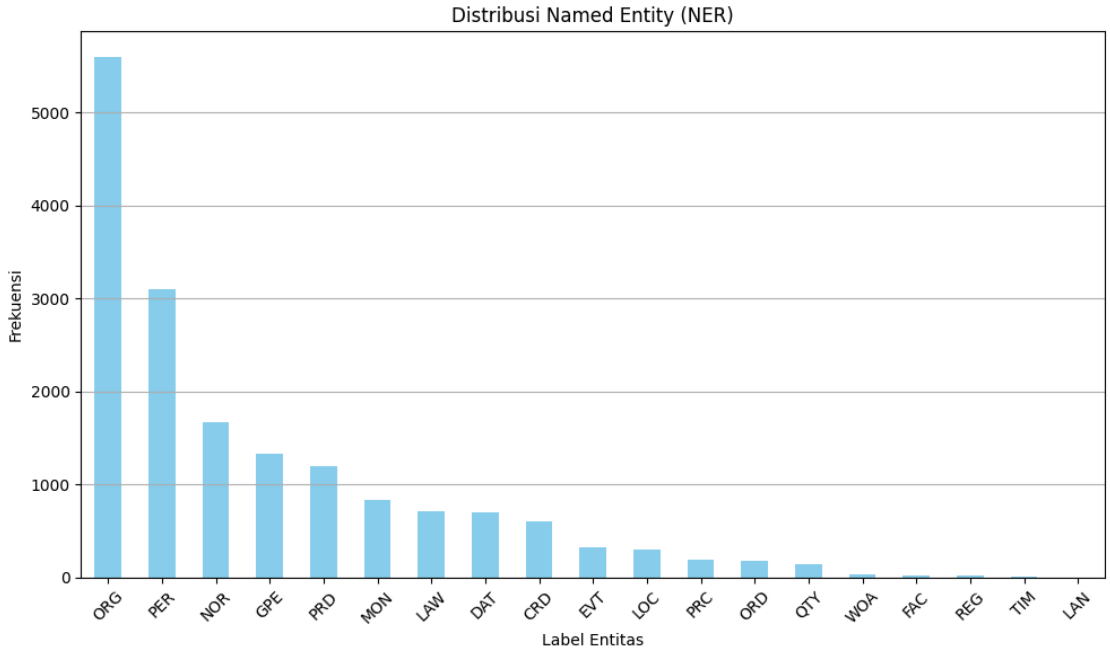
Selain Danantara, kata-kata seperti **BUMN, Indonesia, Prabowo, Presiden, dan BPI** juga muncul dengan frekuensi tinggi dalam kategori **PROPN**, mengindikasikan bahwa pembahasan banyak berkaitan dengan **tokoh publik** dan **institusi formal**. Sementara itu, kata-kata seperti **investasi, negara, aset,** dan **ekonomi** yang masuk dalam kategori **NOUN**, memperkuat bahwa topik yang diangkat mencakup isu-isu ekonomi dan tata kelola. Temuan ini selaras dengan fokus penelitian terhadap representasi kebijakan publik dan wacana ekonomi dalam teks media atau dokumen resmi.

Tabel 6.12 Data NER Universal

Label	Nama Lengkap	Contoh
PERSON	Nama Orang	Jokowi, Prabowo
ORG	Organisasi	BRI, Danantara, PLN
GPE	Lokasi/Negara	Indonesia, Kalimantan

LOC	Lokasi (non-negara)	Gunung Merapi, Jalan Sudirman
DATE	Tanggal/Waktu	2023, 27 Februari
TIME	Waktu spesifik	08:00, pagi
MISC	Lain-lain	UMKM, kategori khusus
NUMBER	Angka	10 juta, 45%

Tabel 6.10 menunjukkan daftar label entitas yang digunakan dalam proses Named Entity Recognition (NER) berdasarkan skema Universal. Label-label ini mengkategorikan kata atau frasa penting dalam teks ke dalam tipe-tipe entitas seperti nama orang (PERSON), organisasi (ORG), lokasi (GPE atau LOC), waktu (DATE atau TIME), angka (NUMBER), serta entitas lain yang bersifat umum atau tidak terklasifikasi (MISC).



Gambar 6.14 Grafik Distribusi NER

Gambar 6.14 menampilkan distribusi untuk label Named Entity Recognition (NER) yang dihasilkan dari proses ektraksi entitas pada datase artikel. Grafik ini menunjukkan bahwa ORG (Organization) menjadi label yang paling dominan dengan lebih dari 5.000 kali kemunculan. Hal ini menunjukkan bahwa dataset artikel mengandung banyak sekali pembahasan terkait organisasi, lembaga, atau perusahaan. Dominasi ini sesuai dengan topik utama yang dibawa, yaitu organisasi yang baru didirikan pemerintah dengan nama danantara yang merupakan badan investasi negara.

Label selanjutnya yang paling banyak muncul adalah PER (Person) dengan lebih dari 3.000 kemunculan yang mengindikasikan seringnya topik danantara dihubungkan dengan individu tertentu, seperti pejabat negara, pimpinan organisasi, atau tokoh publik lain.

Label dengan urutan terbanyak selanjutnya secara berurutan adalah NOR (Nationality, religious, or political group), GPE (Geo-Political Entity) dan PRD (Product). Distribusi yang menonjol pada label ORG dan PER menadakan bahwa teks yang dianalisis berfokus pada pemberitaan mengenai hubungan antar organisasi, peristiwa bisinis, maupun politik yang melibatkan individu atau institusi tertentu. Dominasi ini relevan dengan konteks berita ekonomi, pemerintahan, dan korporasi.

Gambar 6.15 *Display NER Tag Result* pada Dataset Danantara

Gambar 6.15 menampilkan hasil visualisasi tagging untuk Named Entity Recognition (NER) secara langsung pada salah satu artikel. Dapat dilihat bahwa setiap kata atau frasa penting yang teindikasi sebagai entitas diberi label sesuai dengan kategorinya, seperti ORG, PER, GPE, LOC, dan lainnya. Visualisasi ini memudahkan pembaca untuk mengidentifikasi dan memahami konteks berita melalui highlight entitas yang penting dalam artikel. Selain itu, hasil ini juga menunjukkan keberhasilan sistem NER dalam mendeteksi berbagai jenis entitas, meskipun masih ada kemungkinan kesalahan atau

keterbatasan dalam mendeteksi entitas yang ambigu. Kemunculan banyak entitas ORG, PER, dan NOR sangat mewakili hasil pada gambar 6.14, dimana ketiga entitas tersebut memang adalah entitas dengan jumlah kemunculan terbanyak. Kemunculan ini juga menunjukkan fokus utama teks berada pada perusahaan, pejabat, dan lokasi yang relevan dengan pemberitaan ekonomi, politik, atau bisnis.

Tabel 6.13 Top 10 NER *Tag counts*

No	Entitas	Label	Count
1	Danantara	ORG	964
2	Indonesia	GPE	518
3	Prabowo	PER	356
4	Jakarta	GPE	249
5	Pemerintah	NOR	245
6	Bpi Danantara	ORG	193
7	Prabowo Subianto	PER	178
8	Rosan	PER	171
9	Bumn	ORG	129
10	Kpk	NOR	114

Tabel 6.13 menyajikan sepuluh entitas bernama (named entities) yang paling sering muncul dalam korpus teks berdasarkan hasil analisis Named Entity Recognition (NER). Entitas Danantara menduduki peringkat tertinggi dengan jumlah kemunculan sebanyak 964 kali dan diklasifikasikan sebagai ORG (organisasi), yang menunjukkan bahwa topik utama dalam dokumen sangat terfokus pada entitas ini. Selain Danantara, entitas seperti Bpi Danantara dan Bumn juga termasuk dalam kategori organisasi, memperkuat temuan bahwa korpus banyak membahas institusi yang berkaitan dengan sektor publik dan kebijakan ekonomi. Entitas bertipe GPE (Geo-Political Entity) seperti Indonesia dan Jakarta menunjukkan bahwa aspek geografis dan nasionalitas juga menjadi elemen penting dalam pembahasan. Sementara itu, tokoh-tokoh seperti Prabowo, Prabowo Subianto, dan Rosan muncul sebagai PER (person), menandakan bahwa teks juga menyoroti aktor-aktor politik atau tokoh penting yang relevan dalam konteks pembahasan. Kehadiran entitas seperti Pemerintah dan KPK, yang diklasifikasikan sebagai NOR, mengindikasikan bahwa terdapat muatan kebijakan dan penegakan hukum yang cukup dominan dalam isi teks. Secara keseluruhan, hasil ini menunjukkan bahwa wacana dalam korpus teks sangat terkait dengan organisasi besar, tokoh publik, dan institusi negara. Distribusi entitas ini membantu

menjawab rumusan masalah yang berkaitan dengan fokus aktor, lembaga, dan wilayah geografis yang dominan dalam narasi kebijakan publik dan ekonomi di Indonesia.

Jika dibandingkan secara langsung, Tabel 6.11 dan Tabel 6.13 menunjukkan keterkaitan yang kuat antara kata-kata dengan frekuensi tinggi (berdasarkan POS tagging) dan entitas-entitas yang berhasil diidentifikasi secara spesifik melalui proses Named Entity Recognition (NER). Kata Danantara, misalnya, tidak hanya muncul sebagai kata dengan frekuensi tertinggi dalam POS tagging dan diklasifikasikan sebagai PROPN, tetapi juga menjadi entitas organisasi (ORG) yang paling dominan dalam hasil NER. Ini menandakan bahwa Danantara merupakan topik utama dalam narasi wacana yang dianalisis, baik dari sisi struktur sintaktik maupun dari segi entitas konseptual. Korelasi yang sama dapat dilihat pada kata-kata seperti Prabowo, BUMN, dan Indonesia, yang tidak hanya menduduki posisi tinggi dalam POS tagging tetapi juga diidentifikasi secara jelas sebagai entitas bernama dalam hasil NER, masing-masing dikategorikan sebagai PER, ORG, dan GPE. Hal ini memperlihatkan bahwa kata-kata yang sering muncul sebagai proper noun (PROPN) dalam POS tagging umumnya berkaitan erat dengan entitas penting dalam konteks kebijakan dan wacana publik. Selain itu, keberadaan entitas seperti Pemerintah, KPK, dan Jakarta menunjukkan bahwa konteks pembahasan dalam teks tidak hanya terfokus pada individu atau organisasi, tetapi juga menyentuh aspek struktural negara dan lokasi geografis yang relevan.

7. KESIMPULAN

Penelitian ini berhasil menerapkan teknik NLP untuk menganalisis persepsi media terhadap kebijakan Danantara. Dari 226 artikel yang berhasil dianalisis, ditemukan bahwa mayoritas berita memiliki sentimen positif atau netral, dengan peningkatan jumlah publikasi pada momen penting seperti pengumuman resmi dan kerja sama internasional. Analisis POS dan NER memperlihatkan dominasi entitas seperti nama tokoh, organisasi, dan lokasi, yang mencerminkan tingginya intensitas pemberitaan politis dan ekonomi. Penggunaan TF-IDF juga membantu mengidentifikasi topik yang sedang hangat setiap bulan. Secara keseluruhan, pendekatan ini menunjukkan bahwa NLP dapat menjadi alat yang kuat untuk memahami dinamika wacana publik dalam media online, serta memberikan insight penting bagi pemerintah, media, dan akademisi dalam mengevaluasi respon publik terhadap kebijakan.

DAFTAR PUSTAKA

Abdurrohim, I., & Rahman, A. P. (2024). Penerapan natural language processing untuk analisis sentimen terhadap kebijakan pemerintah. Jurnal Kebanggaan RI, 1, 55–60.

Chiche, A., & Yitagesu, B. (2022). Part of speech tagging: A systematic review of deep learning and machine learning approaches. Journal of Big Data, 9(1), 10. https://doi.org/10.1186/s40537-022-00561-y

Hatwar, S., Partridge, V., Bhargava, R., & Bermejo, F. (2024). Author unknown: Evaluating performance of author extraction libraries on global online news articles [Preprint]. arXiv. https://arxiv.org/abs/2410.19771

Montoyo, A., Martínez-Barco, P., & Balahur, A. (2012). Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. Decision Support Systems, 53(4), 675–679. https://doi.org/10.1016/j.dss.2012.05.022

Naseer, S., Ghafoor, M. M., Alvi, S. B. K., Kiran, A., Rahman, S. U., Murtazae, G., & Murtaza, G. (2021). Named entity recognition (NER) in NLP techniques, tools accuracy and performance. Pakistan Journal of Multidisciplinary Research, 2(2), 293–308.

Thota, P., & Ramez, E. (2021, June). Web scraping of COVID-19 news stories to create datasets for sentiment and emotion analysis. In Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference (pp. 306–314). ACM. https://doi.org/10.1145/3453892.3461333