

# Quantifying Public Discourse Toward the 2024 Election: Approval/Disapproval Rates Utilizing the Fine Tune Indobert Pos Tagger Model

Dhafina Nadhira  
Data Science Technology  
Airlangga University  
Surabaya, Indonesia  
nadhiradhafina@gmail.com

Salsabila Dwi Septiani  
Data Science Technology  
Airlangga University  
Surabaya, Indonesia  
sseptiani1080@gmail.com

Ratu Clara Racheilla Afandi  
Data Science Technology  
Airlangga University  
Surabaya, Indonesia  
racheillaafandi@gmail.com

Freya Gracia  
Data Science Technology  
Airlangga University  
Surabaya, Indonesia  
freyagracia14@gmail.com

Ratu Noor Hasannah  
Data Science Technology  
Airlangga University  
Surabaya, Indonesia  
ratunoor003@gmail.com

**Abstract**—This project presents a novel information retrieval system aimed at understanding public discourse surrounding the 2024 Indonesian Presidential Election. Leveraging a dataset of 5,577 tweets, the system extracts and categorizes key opinions using fine-tuning techniques. By employing efficient tokenization, indexing, and a cosine similarity threshold of 0.7, the system retrieves relevant information that either aligns with or diverges from user-input opinions. This enables the quantification of public agreement and disagreement on specific viewpoints, offering valuable insights into real-time opinion dynamics leading up to the election.

**Keywords**—*presidential election 2024, natural language processing, information retrieval, indobert model, sentiment analysis.*

## I. PENDAHULUAN

Pemilihan umum (Pemilu) adalah proses untuk menduduki posisi politik dengan partisipasi warga negara yang memenuhi syarat. Dalam era modern, Pemilu memiliki peran sentral dalam menjaga demokrasi perwakilan dan menegaskan kedaulatan rakyat. Pemilu juga menjadi indikator utama negara demokratis, walaupun beberapa mungkin bersifat otoriter. Di Indonesia, Pemilu adalah implementasi nyata demokrasi, memberikan warga negara peran langsung dalam memilih pejabat publik. Keduanya saling terkait, di mana Pemilu dianggap sebagai prosedur untuk mencapai demokrasi dan mentransfer kedaulatan rakyat kepada kandidat politik. Pentingnya Pemilu tercermin dalam fakta bahwa setiap negara yang mengklaim demokratis harus melaksanakan Pemilu, bahkan jika bentuk pemerintahannya otoriter. Metode Pemilu dapat dijaga atau diubah sesuai dengan pandangan bahwa itu adalah jalur demokrasi yang benar. Oleh karena itu, penentuan suatu pilihan dalam Pemilu harus melibatkan pertimbangan dan refleksi mendalam sesuai prinsip konstitusi dan arah kebijakan negara.

Lembaga centre for strategic and international studies (CSIS) pada 2017 menyebut pengaruh kemajuan dan derasny arus informasi di media social semakin dipercaya bisa membentuk ruang generasi muda yang kuat dan peka terhadap politik [1]. Menurut Sukendar (2017), bahwa

pemilih muda memiliki ciri yang kritis, mandiri, visioner, independen, anti status quo dan berkemajuan, Pemuda identik dengan kemampuan adaptasi yang cepat dengan media sosial dan bisa dipandang sebagai kekuatan baru dalam dinamika politik di Indonesia.

Undang-undang Dasar Negara Republik Indonesia tahun 1945 Pasal 1 ayat (2) menegaskan bahwa kedaulatan berada di tangan rakyat dan dilaksanakan menurut Undang-Undang Dasar. Penyelenggaraan pemilu sebagai bentuk asas kedaulatan rakyat dan keterwakilan di dalam sistem pemerintahan.

Media sosial merupakan platform di mana individu dapat berbagi ide, gagasan, dan opini mereka, memberikan kesempatan bagi orang lain untuk mengetahui pandangan mereka. Twitter, sebagai salah satu media sosial yang populer, sering digunakan oleh masyarakat untuk berdiskusi dan menyampaikan opini mereka. Setelah pemilihan umum, banyak pembicaraan di dunia maya terkait hasil quick count yang muncul di media sosial, khususnya di laman Twitter. Opini-opini yang bermunculan di laman Twitter setelah pemilu menarik untuk diklasifikasikan ke dalam kategori sentimen positif atau negatif. Proses klasifikasi ini merupakan bagian dari data mining, khususnya dalam ranah text mining atau penambangan teks. pembelajaran induktif yang dianggap paling efektif dan efisien untuk keperluan machine learning dan data mining.

Dengan mengintegrasikan pemilihan umum 2024 dan teknologi, dapat menjadi sarana efektif untuk memvisualisasikan opini publik terkait Pemilu 2024. Teknik Information Retrieval adalah suatu metode untuk mengambil informasi yang relevan dari suatu kumpulan data besar, seperti data opini publik yang tersebar di media sosial, artikel berita, atau platform online lainnya. Dengan menganalisis kata-kata kunci, frasa, atau emoji yang digunakan oleh pengguna, kita dapat mengukur tingkat setuju atau tidak setuju terhadap kandidat atau isu tertentu. Pentingnya pemilih muda dalam dinamika politik juga dapat ditekankan dalam analisis opini publik. Dengan memfokuskan perhatian pada media sosial yang merupakan platform utama bagi pemilih muda, kita dapat melihat bagaimana generasi ini merespons kampanye politik, debat,

dan isu-isu terkini. Persentase setuju atau tidak setuju dari kalangan pemilih muda dapat memberikan gambaran yang jelas tentang arah dukungan publik. Penggunaan tagar atau hashtag yang populer dapat dijadikan indikator untuk mengukur sejauh mana suatu isu atau kandidat telah mendapatkan perhatian di media sosial. Analisis frekuensi penggunaan tagar tertentu dapat memberikan informasi tentang popularitas, polarisasi, atau tren opini publik.

Dengan menggunakan teknik IR, kita dapat menciptakan visualisasi yang dinamis dan informatif, memungkinkan para pemangku kepentingan untuk memahami dengan lebih baik dinamika opini publik menjelang Pemilu 2024. Dengan fokus pada aspek-aspek kunci seperti sentimen, visualisasi ini dapat menjadi alat strategis dalam membentuk strategi kampanye, menangkap tren, dan menginformasikan keputusan politik.

## II. TINJAUAN PUSTAKA

### A. *Natural Language Processing (Pemrosesan Bahasa Alami)*

Pada dasarnya, bahasa alami adalah suatu bentuk representasi dari suatu pesan yang ingin dikomunikasikan antar manusia yang bentuk utamanya berupa suara/ucapan, tetapi sering pula dinyatakan dalam bentuk tulisan [1]. Pemrosesan Bahasa Alami (Natural Language Processing atau NLP) adalah disiplin ilmu yang menyelidiki interaksi antara komputer dan bahasa manusia. NLP sering dianggap sebagai bagian dari kecerdasan buatan dan memiliki keterkaitan dengan linguistik komputasional. Dalam ruang lingkup kajiannya, NLP mencakup aspek-aspek seperti segmentasi tuturan (part-of-speech tagging) dan pengawataksaan makna (word sense disambiguation). Meskipun mampu menangani baik teks maupun tuturan, pemrosesan tuturan (speech processing) berkembang menjadi sub-bidang tersendiri.

Secara umum, aplikasi dalam NLP dapat dibagi menjadi dua jenis, yaitu aplikasi berbasis teks dan aplikasi berbasis dialog. Aplikasi berbasis teks menangani proses terhadap teks tertulis, seperti dokumen, surel, dan buku. Contoh aplikasi NLP berbasis teks meliputi program klasifikasi dan pengambilan isi dokumen, serta terjemahan mesin. Sedangkan aplikasi berbasis dialog, idealnya, melibatkan bahasa lisan atau pengenalan suara, meskipun juga dapat memperbolehkan interaksi melalui pengetikan teks. Beberapa contoh aplikasi NLP berbasis dialog mencakup asisten pribadi cerdas dan chatbot.

### B. *Pre-Train IndoBERT Model*

*Bidirectional Encoder Representations* dari *Transformers* atau biasa disebut *BERT* adalah model bahasa yang memanfaatkan mekanisme encoder pada *Transformers* dalam melakukan tugasnya pemodelan bahasa [5]. Sedangkan IndoBERT merupakan model bahasa pre-trained yang memberikan performa yang baik dalam berbagai tugas NLP bahasa Indonesia [6].

Terdapat dua penelitian yang menggunakan BERT dan melakukan analisis pada bahasa Indonesia. Pada penelitian pertama [5] bermaksud memanfaatkan m-BERT (BERT multibahasa) untuk meningkatkan performa ABSA pada kumpulan data ulasan Indonesia. Dengan menggabungkan m-BERT dengan metode transformasi tugas, penelitian ini berhasil mencapai peningkatan signifikan sebesar 8% pada skor F1 dibandingkan dengan studi sebelumnya. Hal ini

menunjukkan potensi besar m-BERT dalam meningkatkan akurasi ABSA untuk teks bahasa Indonesia, terutama dalam mengatasi tantangan kata OOV dan generalisasi ke data baru.

Penelitian kedua [6] mengeksplorasi penggunaan model bahasa pra-latih IndoBERT untuk PI dalam bahasa Indonesia. Penelitian ini berhasil meningkatkan F1 score dan akurasi IndoBERT untuk PI sebesar 2,2% dan 2,7% melalui penggunaan teknik augmentasi data, reformulasi tugas, dan penggunaan model DNN yang lebih kompleks. Temuan ini menunjukkan bahwa teknik-teknik tersebut efektif untuk meningkatkan kemampuan IndoBERT dalam memahami hubungan semantik antar kalimat, dan membuka jalan bagi pengembangan model Paraphrase Identification (PI) yang lebih baik untuk bahasa Indonesia.

Secara umum, kedua penelitian ini menunjukkan potensi besar model bahasa pra-latih untuk meningkatkan performa tugas NLP dalam bahasa Indonesia. Kedua penelitian ini juga menunjukkan bahwa teknik-teknik tertentu, seperti augmentasi data, reformulasi tugas, dan kombinasi model bahasa pra-latih dengan metode transformasi tugas, dapat meningkatkan performa model bahasa pra-latih untuk tugas NLP dalam bahasa Indonesia.

### C. *Fine-Tuning BERT*

Fine-tuning BERT, yaitu proses penyesuaian model pra-latih BERT untuk tugas spesifik, telah menjadi praktik yang penting dalam meningkatkan kinerja model NLP. Dua penelitian berikut menunjukkan bagaimana fine-tuning dapat diterapkan secara efektif untuk tugas yang berbeda.

Penelitian pertama [7] meneliti kemampuan model Masked Language Model (MLM) BERT dalam menentukan hubungan antara dua kalimat. Dengan menempatkan token [MASK] sebelum kalimat kedua, penelitian ini menguji apakah BERT dapat memprediksi konjungsi yang tepat untuk menghubungkan kedua kalimat tersebut. Hasil penelitian ini menunjukkan potensi MLM dalam memahami hubungan antar kalimat, yang dapat dimanfaatkan untuk berbagai tugas NLP seperti question answering dan natural language inference.

Penelitian kedua [8] berfokus pada fine-tuning BERT untuk tugas klasifikasi sentimen. Penelitian ini mengeksplorasi dua metode fine-tuning, yaitu static parameter fine-tuning dan dynamic layer frozen strategy. Hasil penelitian menunjukkan bahwa kedua metode ini berhasil meningkatkan akurasi prediksi model sentimen. Temuan ini menekankan pentingnya fine-tuning hyperparameter untuk mengoptimalkan kinerja BERT dalam berbagai tugas NLP.

Secara keseluruhan, kedua penelitian ini memberikan bukti bahwa fine-tuning BERT merupakan teknik yang efektif untuk meningkatkan kinerjanya dalam berbagai tugas NLP. Fine-tuning memungkinkan BERT untuk beradaptasi dengan konteks spesifik dari tugas yang diberikan, sehingga menghasilkan hasil yang lebih akurat. Penelitian-penelitian ini mendorong eksplorasi metode fine-tuning yang lebih canggih untuk memaksimalkan potensi BERT dalam berbagai domain NLP.

### D. *Sentiment Analysis (Analisis Sentimen)*

Sentiment analysis merupakan salah satu bidang dari Natural Language Processing (NLP) yang membangun sistem untuk mengenali dan mengekstraksi opini dalam

bentuk teks. Informasi berbentuk teks saat ini banyak terdapat di internet dalam format forum, blog, media sosial, serta situs berisi review. Dengan bantuan sentiment analysis, informasi yang tadinya tidak terstruktur dapat diubah menjadi data yang lebih terstruktur [4].

Data tersebut dapat menjelaskan opini masyarakat mengenai produk, merek, layanan, politik, atau topik lainnya. Perusahaan, pemerintah, maupun bidang lainnya kemudian memanfaatkan data-data tersebut untuk membuat analisis marketing, review produk, umpan-balik produk, dan layanan masyarakat. Guna menghasilkan opini yang dibutuhkan, sentiment analysis tidak hanya harus bisa mengenali opini dari teks. Proses yang juga disebut sebagai opini mining ini juga perlu bekerja dengan mengenali tiga aspek berikut [4]: Subjek: topik apa yang sedang dibicarakan. Polaritas: apakah opini yang diberikan bersifat positif atau negatif. Pemegang opini: seseorang yang mengeluarkan opini tersebut.

Sentiment analysis kemudian akan membedakan teks menjadi dua kategori, yakni fakta dan opini. Fakta merupakan ekspresi objektif mengenai sesuatu. Sementara opini adalah ekspresi subjektif yang menggambarkan sentimen, perasaan, maupun penghargaan terhadap suatu hal [4].

Analisis sentimen adalah teknik yang relatif baru, tetapi memiliki potensi untuk memberikan wawasan yang berharga tentang dunia di sekitar kita [9]. Analisis sentimen dapat dilakukan secara *unsupervised Learning*, yaitu algoritma yang dapat memberi sentimen pada data tanpa memerlukan data label untuk melatih model yang akan dibuat (Sinaga, 2022) [11]. terdapat penelitian sebelumnya dengan judul “Optimasi TextBlob Menggunakan Support Vector Machine untuk Analisis Sentimen (Studi Kasus Layanan Telkomsel)” yang menemukan polaritas sentimen dengan TextBlob, lalu optimasi model dengan SVM atau Support Vector Machine dan diperoleh akurasi sebesar 75% [10]

#### E. Information Retrieval (Pengambilan Informasi)

Information Retrieval (IR) merupakan suatu cara yang digunakan untuk menemukan kembali (retrieve) informasi-informasi yang relevan terhadap kebutuhan pengguna dari suatu kumpulan informasi secara otomatis. (Bunyamin, 2015) [2]. Proses IR umumnya berkaitan dengan pencarian informasi yang isinya tidak terstruktur. Demikian pula kata kunci pencarian pengguna yang disebut query, juga merupakan bentuk yang tidak struktur. Hal ini yang membedakan IR dengan sistem basis data. Dokumen adalah contoh informasi yang tidak terstruktur. Isi dari suatu dokumen umumnya berupa kumpulan teks yang sangat tergantung pada pembuat dokumen tersebut (Pardede, 2014) [2].

Bagian-bagian dari IR system menurut gambar 2 meliputi (Bunyamin, 2015) [2]:

1. Text Operations (operasi terhadap teks) yang meliputi pemilihan kata-kata dalam query maupun dokumen (term selection) dalam pentransformasian dokumen atau query menjadi term index (indeks dari kata-kata).
2. Query formulation (formulasi terhadap query) yaitu memberi bobot pada indeks katakata query.

3. Ranking (perangkingan), mencari dokumen-dokumen yang relevan terhadap query dan mengurutkan dokumen tersebut berdasarkan kesesuaiannya dengan query.
4. Indexing (pengindeksan), membangun basis data indeks dari koleksi dokumen. Dilakukan terlebih dahulu sebelum pencarian dokumen dilakukan.

Contoh aplikasi IR meliputi mesin pencari web seperti Google, CMS seperti WordPress, sistem pencarian dalam database seperti MySQL, dan platform e-commerce seperti Amazon. Pustaka digital seperti Google Books, sistem pencarian email dalam Gmail, dan sistem pencarian dokumen di perusahaan juga merupakan implementasi IR. Aplikasi ini mempermudah pengguna menemukan informasi dengan cepat, baik itu dalam konteks web, bisnis, atau media sosial.

Pencarian informasi (IR) memegang peran vital dalam membantu kita menemukan relevansi di tengah lautan data yang terus membesar. Tiga penelitian terbaru memberikan perspektif menarik tentang perkembangan dan tantangan di bidang ini:

1. Memahami Alur Pengambilan: Artikel pertama menawarkan tinjauan menyeluruh model dan teknik yang digunakan dalam dua tahap utama pemrosesan IR [14]. Penelitian ini menganalisis perkembangan historis, terobosan terkini, dan hambatan yang dihadapi para peneliti dan praktisi. Dengan pemahaman mendalam ini, artikel ini menjadi sumber pengetahuan berharga bagi siapa saja yang ingin berkecimpung dalam dunia IR, membuka jalan bagi inovasi dan pengembangan teknik-teknik baru.

2. Privasi dalam Pengambilan Informasi: Menyelim ke ranah privasi, artikel kedua meneliti varian Pencarian Informasi Pribadi Simetris (SPIR) dengan keacakan umum pihak pengguna [13]. Varian ini memungkinkan pengguna mengakses informasi rahasia dari beberapa basis data tanpa membeberkan pilihan mereka, sekaligus melindungi basis data dari pengungkapan data sensitif. Penelitian ini menentukan kapasitas optimal sistem tersebut, membuka pintu bagi pencarian informasi yang aman dan pribadi di masa depan.

3. Mengoptimalkan Kueri Penelitian Biomedis: Beralih ke kebutuhan informasi spesifik domain, artikel ketiga berfokus pada optimasi kueri dalam penelitian biomedis [12]. Penelitian ini mengusulkan teknik baru berbasis pembelajaran peringkat (LTR) untuk menggabungkan peningkatan dan perluasan kueri. Metode ini menyusun ulang kata-kata perluasan berdasarkan relevansinya dengan kueri asli, sehingga meningkatkan akurasi pengambilan informasi. Hasil penelitian menunjukkan peningkatan efisiensi sebesar 3,31% dibandingkan algoritma tradisional, menawarkan manfaat signifikan bagi peneliti yang menavigasi kumpulan data biomedis yang besar.

Ketiga penelitian ini menyoroti aspek-aspek beragam dalam IR. Dari pemahaman dasar model pengambilan hingga eksplorasi teknik privasi mutakhir dan optimasi domain-spesifik, bidang ini terus berkembang pesat. Dengan bekal wawasan baru ini, kita dapat bersama-sama membangun sistem IR yang lebih akurat, efisien, dan aman, membuka jalan bagi masa depan dengan akses informasi yang lebih demokratis dan inklusif di tengah banjir data yang terus mengepung kita.

#### F. Fine-tuning BERT for Information Retrieval (Pengambilan Informasi)

Dua penelitian [15] dan [16] memaparkan bagaimana fine-tuning model bahasa pra-latih BERT dapat meningkatkan efektivitas pencarian data dalam domain tertentu. Keduanya memanfaatkan kekuatan dan fleksibilitas BERT untuk mengatasi tantangan pencarian informasi yang kompleks.

Penelitian pertama [15] berfokus pada peningkatan sistem pengambilan trouble report (TR) duplikat dalam organisasi perangkat lunak berskala besar. Penelitian ini menguji berbagai strategi fine-tuning untuk mengintegrasikan pengetahuan domain telekomunikasi ke dalam model penelusuran TR berbasis BERT. Hasilnya menunjukkan bahwa menambahkan data spesifik domain meningkatkan performa model dalam mengidentifikasi TR duplikat, sekaligus meningkatkan generalisasi model terhadap data TR di luar domain.

Penelitian kedua [16] beralih ke ranah medis dengan mengembangkan pendekatan berbasis NLP untuk memprediksi jenis pemindaian dari laporan kedokteran nuklir. Penelitian ini menggunakan BERT yang telah difine-tune untuk mengklasifikasikan jenis pemindaian berdasarkan temuan yang dilaporkan. Model ini mencapai akurasi 98% dalam memprediksi 51 jenis pemindaian yang berbeda dari kumpulan data 36.000 laporan dari berbagai pusat kedokteran. Kemampuan ini berpotensi secara signifikan meningkatkan efisiensi pencarian data dalam database medis dengan mengotomatiskan proses ekstraksi nama pemindaian dari laporan radiologi.

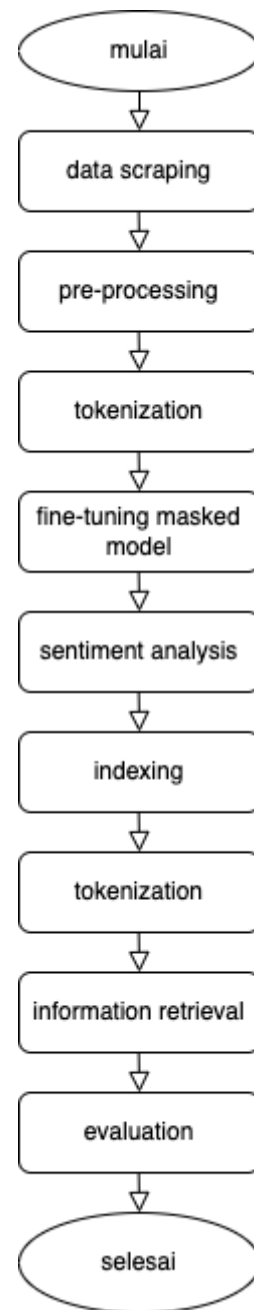
Secara keseluruhan, kedua penelitian ini memberikan contoh praktis pemanfaatan fine-tuning BERT untuk menyempurnakan pencarian data. Mereka menunjukkan bahwa mengintegrasikan pengetahuan domain dan data spesifik dapat meningkatkan akurasi dan generalisasi model, berujung pada peningkatan efisiensi dan efektivitas pencarian informasi dalam domain yang kompleks.

### III. SUMBER DATA DAN METODOLOGI

#### A. Sumber Data

Data yang digunakan dalam penelitian ini merupakan data tweet yang dikumpulkan dengan cara scrapping data dari X. Data yang dikumpulkan merupakan data tweet tentang Pemilu 2024 dalam rentang waktu 1 Januari 2023 hingga 18 Desember 2023, berbahasa Indonesia, dan memiliki likes minimal 25 dengan *keyword* “Pemilu 2024”, “Ganjar Mahfud”, “Prabowo Gibran”, “Anies Imin”, “Ganjar”, “Mahfud”, “Prabowo”, “Gibran”, “Anies”, dan “Imin”.

#### B. Metodologi



GAMBAR 1. FLOWCHART

- 1) *Pengumpulan Data*: Data tweet dapat dikumpulkan secara manual dengan cara membuka Twitter dan menyalin tweet yang diinginkan. Data tweet juga dapat dikumpulkan menggunakan tools scraping. Tools scraping adalah perangkat lunak yang dapat digunakan untuk mengunduh data dari internet secara otomatis. Pada penelitian ini, peneliti menggunakan metode scraping sesuai dengan query yang telah dipilih.
- 2) *Pre-Processing*: Proses pra-pemrosesan dilakukan untuk membersihkan data tweet dari berbagai hal yang dapat mengganggu proses analisis, seperti simbol, tag, dan tautan. Pada penelitian ini, preprocessing yang dilakukan antara lain:
  - Menghapus Tanda baca, seperti titik, koma, tanda seru, dan tanda tanya.

- Mengubah huruf besar atau kapital menjadi huruf kecil
  - Menghapus link
  - Menghapus Similar tweet, seperti tweet yang berisi sama persis dengan tweet lainnya
  - Menghapus stopwords
  - Menghapus word elongation
- 3) *Tokenisasi*: Tokenisasi dilakukan untuk memecah data tweet menjadi unit-unit terkecil yang disebut token. Token dapat berupa kata, frasa, atau kalimat. Tokenisasi dapat dilakukan dengan menggunakan berbagai algoritma, seperti algoritma tokenisasi berdasarkan spasi, algoritma tokenisasi berdasarkan karakter, dan algoritma tokenisasi berdasarkan sintaksis.
  - 4) *Fine-Tuning Masked Model*: Fine-Tuning pada model Indonesia roberta base dengan data tweet hasil scraping dilakukan untuk meningkatkan kinerja atau menyesuaikan model dengan gaya bahasa twitter sehingga model dapat memahami nuansa, ekspresi emosional yang diberikan bahasa slang atau singkatan yang umumnya muncul pada tweet agar hasil analisis sentimen serta information retrieval yang lebih baik lagi.
  - 5) *Analisis Sentimen*: Analisis sentimen adalah proses untuk menentukan sentimen yang terkandung dalam sebuah tweet, seperti positif, negatif, atau netral. Analisis sentimen dapat digunakan untuk berbagai keperluan, seperti untuk memantau opini publik, untuk melakukan pemasaran, dan untuk membuat keputusan bisnis. Pada Penelitian ini analisis sentimen dilakukan untuk menentukan tweet termasuk positif atau negatif.
  - 6) *Indeksasi*: Indeksasi dilakukan untuk membuat indeks dari data tweet yang telah ditokenisasi. Indeks ini digunakan untuk mempercepat proses pencarian informasi. Indeks dapat dibuat dengan menggunakan berbagai teknik, seperti teknik indeksasi berdasarkan frekuensi kemunculan kata, teknik indeksasi berdasarkan frekuensi kemunculan kata dalam konteks, dan teknik indeksasi berdasarkan semantik.
  - 7) *Information Retrieval*: Pencarian informasi dilakukan untuk mencari informasi yang diinginkan dari data tweet yang telah diindeks. Proses pencarian informasi dapat dilakukan dengan menggunakan berbagai metode, seperti metode pencarian berdasarkan pola, metode pencarian berdasarkan kata kunci, dan metode pencarian berdasarkan semantik. Pada penelitian ini pencarian informasi dilakukan dengan metode pencarian berdasarkan kata kunci atau query yang diinputkan.
  - 8) *Evaluasi*: Evaluasi dilakukan untuk mengevaluasi hasil analisis. Evaluasi dapat dilakukan dengan menggunakan berbagai metode, seperti menghitung akurasi, presisi, dan recall. Akurasi adalah persentase hasil analisis yang benar. Presisi adalah persentase hasil analisis yang benar-benar relevan dengan kebutuhan pengguna. Recall adalah persentase hasil analisis yang benar-benar relevan dengan kebutuhan pengguna, dibagi dengan jumlah total data yang relevan.

#### IV. HASIL DAN PEMBAHASAN

Dilakukan percobaan program “Tingkat Persetujuan Publik terhadap Suatu Pernyataan terkait Pemilihan Umum 2024” menggunakan kueri tentang pemilihan umum dan ketiga pasangan calon presiden-wakil presiden.

Kueri pertama yaitu “calon presiden tidak kompeten”. Analisis sentimen menunjukkan bahwa sentimen yang dihasilkan dari kueri tersebut adalah “negatif”. Selanjutnya, program mengambil dokumen yang relevan dengan kueri tersebut dengan menghitung nilai *cosine similarity* antara kueri dan dokumen pada dataset. Setelah didapatkan sentimen dari kueri yang dimasukkan, serta dokumen yang relevan dengan kueri, selanjutnya diperoleh persentase angka persetujuan publik terhadap kueri. Sebanyak 77.78% dokumen setuju atau memiliki sentimen yang sama dengan query dan 22.22% tidak setuju atau memiliki sentimen yang berbeda dengan kueri dengan dokumen. Diperoleh hasil lima teratas dokumen atau tweet yang relevan di antara lain:

TABEL I. calon presiden tidak kompeten

Dokumen	Similarity
semoga kecurangan pilpres pemilihan presiden pada pemilu ini tidak terjadi	0.7472
mahfud md mengaku tak terlibat perumusan visi misi capres calon presiden cawapres calon wakil presiden	0.7389
kayaknya timses mereka tau kalau gibran enggak bisa ngimbangin pengalaman dan ilmu cawapres calon wakil presiden lain maka nya acara debat cawapres calon wakil presiden dibikin enggak ada lah mengomong saja kayak begini	0.7279
ini bukan karena ada indikasi gibran berkesan dipaksakan jadi cawapres calon wakil presiden prabowo kalau pun maka meloloskan permohonan itu lalu gibran memutuskan untuk menjadi cawapres calon wakil presiden ganjar saya akan melawan dengan cara saya sendiri jika betul demikian meski saya bukan siapa siapa saya tidak akan	0.7188
fokus adu gagasan cawapres calon wakil presiden dengan sesama capres calon presiden cawapres calon wakil presiden lah sebagaimana publik inginkan jangan melebar ke debat lawan ini itu dulu takut adu gagasan bibit otoritarianisme gibran sudah tampak	0.7026

Kueri kedua yaitu “Anies Imin bagus”. Analisis sentimen menunjukkan bahwa sentimen yang dihasilkan dari kueri tersebut adalah positif. Selanjutnya, program mengambil dokumen yang relevan dengan kueri tersebut dengan menghitung nilai *cosine similarity* antara kueri dan dokumen pada dataset. Setelah didapatkan sentimen dari kueri yang dimasukkan, serta dokumen yang relevan dengan kueri, selanjutnya diperoleh persentase angka persetujuan publik terhadap kueri. Sebanyak 37.92% dokumen setuju atau memiliki sentimen yang sama dengan query dan 62.08% tidak setuju atau memiliki sentimen yang berbeda dengan kueri dengan dokumen.

Berikut adalah hasil lima teratas dokumen atau tweet yang relevan

TABEL II. Anies Imin bagus

Dokumen	Similarity
alhamdulillah pasangan anies cak imin telah menandatangani point pakta integritas ijtima ulama kepada seluruh umat mari kita dukung pasangan amin satu komando revolusi ahlak	0.8895
orasi berapi api pak anies di acara dekalarsi sahabat abi anies bersama imin minggu nov satu gerakan perubahan untuk kesetaraan dan keadilan semua lapisan masyarakat anies muhaimin amun aja dulu satukan indonesia	0.8875
kampanye perdana amin cak imin anies baswedan mojokerto kelahiran ibunda cak imin tak ingin masa depan anak buruk pergerakan perempuan nahdliyyin mojokerto dukung amin menang pilpres pemilihan presiden amin ajadulu anies muhaimin	0.8870
calon presiden nomor urut anies baswedan mengunjungi koperasi peternakan bandung selatan kpbs di pangalengan bandung jawa barat janji anies untuk reformasi tata niaga pangan anies janji permudah bantuan permodalan anies muhaimin amin aja dulu	0.8869
indonesia cerdas pilih amin anies muhaimin indonesia bermartabat pilih amin anies muhaimin indonesia berdaulat pilih amin anies muhaimin pilih amin nomor anies muhaimin amin aja dulu aminkan indonesia	0.8837

Kueri ketiga yaitu “dukung prabowo gibran”. Analisis sentimen menunjukkan bahwa sentimen yang dihasilkan dari kueri tersebut adalah ‘positif’. Selanjutnya, program mengambil dokumen yang relevan dengan kueri tersebut dengan menghitung nilai *cosine similarity* antara kueri dan dokumen pada dataset. Setelah didapatkan sentimen dari kueri yang dimasukkan, serta dokumen yang relevan dengan kueri, selanjutnya diperoleh persentase angka persetujuan publik terhadap kueri. Sebesar 39.24% dari dokumen yang relevan memiliki sentimen yang sama dengan kueri (sentimen ‘positif’) dan sebesar 60.76% memiliki sentimen berbeda dengan kueri (sentimen ‘negatif’) Dokumen yang memiliki sentimen sama dengan kueri, dapat diindikasikan bahwa dokumen tersebut setuju atau berada di pihak yang sama dengan pernyataan kueri.

Diperoleh hasil lima teratas dokumen atau tweet yang relevan di antara lain:

TABEL III. dukung prabowo gibran

Dokumen	Similarity
tebak tokoh prabowo gibran prabowo gibran istimewa prabowo gibran	0.9581
aku cinta kamu prabowo gibran tweet prabowo gibran nkri	0.9467
padati gbk anggota desa bersatu dukung prabowo gibran	0.9437
partaisocmed resmi deklarasi dukung prabowo gibran ajak relawan bersatu menangkan prabowo gibran	0.9405
semarang al prabowo gibran	0.9388

Kueri keempat yaitu “jangan pilih ganjar”. Analisis sentimen menunjukkan bahwa sentimen yang dihasilkan

dari kueri tersebut adalah ‘negatif’. Selanjutnya, program mengambil dokumen yang relevan dengan kueri tersebut dengan menghitung nilai *cosine similarity* antara kueri dan dokumen pada dataset. Setelah didapatkan sentimen dari kueri yang dimasukkan, serta dokumen yang relevan dengan kueri, selanjutnya diperoleh persentase angka persetujuan publik terhadap kueri. Sebesar 61,40% dari dokumen yang relevan memiliki sentimen yang sama dengan kueri (sentimen ‘negatif’) dan sebesar 38,60% memiliki sentimen berbeda dengan kueri (sentimen ‘positif’) Dokumen yang memiliki sentimen sama dengan kueri, dapat diindikasikan bahwa dokumen tersebut setuju atau berada di pihak yang sama dengan pernyataan kueri.

Berikut adalah hasil lima teratas dokumen atau tweet yang relevan

TABEL IV. jangan pilih ganjar

Dokumen	Similarity
ganjar paling tegas ganjar mahfud pilihan tepat orang cerdas pilih ganjar mahfud	0.8582
gibran pimpin kampanye di jateng jawa tengah yogyakarta suara ganjar mahfud bisa direbut	0.8580
nini mah pilih ganjar mahmud saja pilih gemoy takut ndasku di gowo	0.8574
lambungkan sampai ke kertanegara orang baik tidak pilih penculik orang baik pilih orang baik ganjar mahfud	0.8558
usai terima kunjungan gibran abuya muhtadi tetap serukan warga banten pilih ganjar mahfud	0.8532

## V. KESIMPULAN

Program “Tingkat Persetujuan Publik terhadap Suatu Pernyataan terkait Pemilihan Umum 2024” berhasil menganalisis dan mendapatkan informasi dari data tweet terkait Pemilu 2024. Hasil tersebut dapat digunakan untuk memahami pandangan dan opini masyarakat terkait pemilihan umum serta mengukur tingkat persetujuan dan ketidaksetujuan terhadap isu-isu tertentu. Beberapa kueri yang telah dicoba diantaranya: “calon presiden tidak kompeten” sebanyak 77.78% dokumen setuju atau memiliki sentimen yang sama dengan query dan 22.22% tidak setuju atau memiliki sentimen yang berbeda dengan kueri dengan dokumen; “Anies Imin bagus” sebanyak 37.92% dokumen setuju atau memiliki sentimen yang sama dengan query dan 62.08% tidak setuju atau memiliki sentimen yang berbeda dengan kueri dengan dokumen; “dukung prabowo gibran” Sebesar 39.24% dari dokumen yang relevan memiliki sentimen yang sama dengan kueri (sentimen ‘positif’) dan sebesar 60.76% memiliki sentimen berbeda dengan kueri (sentimen ‘negatif’); “jangan pilih ganjar” Sebesar 61,40% dari dokumen yang relevan memiliki sentimen yang sama dengan kueri (sentimen ‘negatif’) dan sebesar 38,60% memiliki sentimen berbeda dengan kueri (sentimen ‘positif’) Dokumen yang memiliki sentimen sama dengan kueri, dapat diindikasikan bahwa dokumen tersebut setuju atau berada di pihak yang sama dengan pernyataan kueri. Perlu diingat bahwa hasil analisis ini bersifat kontekstual dan dapat berubah seiring waktu. Oleh karena itu, evaluasi dan pemantauan terus-menerus diperlukan untuk

memastikan validitas dan relevansi informasi yang diperoleh dari analisis sentimen dan information retrieval. Ke depannya, penelitian ini dapat diperluas dengan melibatkan lebih banyak kategori isu dan melibatkan lebih banyak data untuk meningkatkan representasi hasil analisis. Selain itu, perlu dilakukan peningkatan pada model untuk memperoleh hasil yang lebih akurat dan relevan. Penelitian ini memberikan gambaran tentang bagaimana teknologi dapat digunakan untuk menganalisis opini publik dan mengambil informasi yang relevan dari data tweet di media sosial. Hal ini dapat menjadi dasar untuk pengembangan lebih lanjut dalam memahami dinamika politik dan opini masyarakat melalui analisis sentimen dan information retrieval.

#### REFERENCES

- [1] Astrianti Defretes, D., & Kleden, K. L. (2023). Efektivitas Pemilihan Umum Serentak Tahun 2024. *Jurnal Hasil Penelitian (JHP)* 17), 8(2), 49. <http://jurnal.untag-sby.ac.id/index.php/jhp17>
- [2] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
- [4] Prabowo, R., Sujaini, H., & Rismawan, T. (2023). Analisis Sentimen Pengguna Twitter Terhadap Kasus COVID-19 di Indonesia Menggunakan Metode Regresi Logistik Multinomial. *Jurnal Teknologi dan Sistem Komputer*, 11(1), 1- [halaman terakhir artikel]. DOI: 10.26418/justin.v11i1.57450
- [5] A. Azhar and M. Khodra, "Fine-tuning Pretrained Multilingual BERT Model for Indonesian Aspect-based Sentiment Analysis." Accessed: Dec. 24, 2023. [Online]. Available: <https://arxiv.org/ftp/arxiv/papers/2103/2103.03732.pdf>
- [6] K. Benedictus Visto, A. Martin Jason, and K. Gede Putra. (2023). "Fine-Tuned IndoBERT Based Model and Data Augmentation for Indonesian Language Paraphrase Identification | IIETA," *iieta.org*, 2022. <https://iieta.org/journals/ria/paper/10.18280/ria.370322> (accessed Dec. 24, 2023).
- [7] Zhao, Y., Cao, R., Bai, J., Ma, W., & Shinnou, H. (2020). Determining the Logical Relation between Two Sentences by Using the Masked Language Model of BERT. 2020 International Conference on Technologies and Applications of Artificial Intelligence (TAAI).
- [8] Li, X., Wang, X., & Liu, H. (2021). Research on fine-tuning strategy of sentiment analysis model based on BERT. 2021 International Conference on Communications, Information System and Computer Engineering (CISCE). doi:10.1109/cisce52179.2021.94458.
- [9] Bobby Siswanto, Ford Lumban Gaol, Benefano Soewito, and L. Hendric, "Sentiment Analysis of Big Cities on The Island of Java in Indonesia from Twitter Data as A Recommender System," Oct. 2021, doi: <https://doi.org/10.1109/icimcis53775.2021.9699147>.
- [10] Perdana, K., Pricillia, T., & Zulfachmi, Z. (2021). Optimasi TextBlob Menggunakan Support Vector Machine untuk Analisis Sentimen (Studi Kasus Layanan Telkomsel). *Jurnal Bangkit Indonesia*, 10(1), 13-15.
- [11] Sinaga, A. F. (2022). Analisis Sentimen pada Aplikasi Halodoc Berbasis Clustering dengan Word2vec (Doctoral dissertation, Universitas Sumatera Utara).
- [12] Y. Peng and Y. Yang, "Learning to Rank for Biomedical Information Retrieval," Mar. 2021, doi: <https://doi.org/10.1109/icitbs53129.2021.00143>.
- [13] Z. Wang and Şennur Ulukuş, "Symmetric Private Information Retrieval at the Private Information Retrieval Rate," *IEEE journal on selected areas in information theory*, vol. 3, no. 2, pp. 350–361, Jun. 2022, doi: <https://doi.org/10.1109/jsait.2022.3188610>.
- [14] K. A. Hambarde and H. Proença, "Information Retrieval: Recent Advances and Beyond," *IEEE Access*, vol. 11, pp. 76581–76604, Jan. 2023, doi: <https://doi.org/10.1109/access.2023.3295776>.
- [15] A. Saberi Manesh et al., "Fine-tuned BERT Language Model for Efficient Nuclear Medicine Data Retrieval," Nov. 2023, doi: <https://doi.org/10.1109/nssmicrtsd49126.2023.10338102>.
- [16] N. S. Bosch, Serveh Shalmashi, F. Yaghoubi, H. Holm, Fitsum Gaim, and A. H. Payberah, "Fine-Tuning BERT-based Language Models for Duplicate Trouble Report Retrieval," 2022 IEEE International Conference on Big Data (Big Data), Dec. 2022, doi: <https://doi.org/10.1109/bigdata55660.2022.10020825>.