

Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi)

DOI: <https://doi.org/10.35870/jtik.v9i2.3535>

Analisis Sentimen Berbasis Transformer: Persepsi Publik terhadap Nusantara pada Perayaan Kemerdekaan Indonesia yang Pertama

Triana Dewi Salma^{1*}, Muhammad Ferdi Kurniawan², Rizqi Darmawan³, Amat Basri⁴

^{1*,2,3} Program Studi Informatika, Fakultas Sains dan Bisnis, Universitas LLA, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta, Indonesia.

⁴ Program Studi Sistem Informasi, Fakultas Sains dan Bisnis, Universitas LLA, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta, Indonesia.

article info

Article history:

Received 10 December 2024

Received in revised form

20 December 2024

Accepted 10 January 2025

Available online April 2025.

Keywords:

Independence Day;

IndoBERT; Nusantara;

Sentiment Analysis;

TextBlob; Transformer.

Kata Kunci:

Analisis Sentimen; Hari

Kemerdekaan; IndoBERT;

Nusantara; TextBlob;

Transformer.

abstract

The inaugural Indonesian Independence Day celebration in the new capital, Nusantara, marked a historic milestone. This study analyzes public sentiment toward this event using the IndoBERT model. Data was collected from Twitter during the celebration period and classified into positive, negative, and neutral sentiments. Three main approaches were employed: IndoBERT as a baseline, IndoBERT fine-tuned with IndoNLU data, and IndoBERT applied to TextBlob-labeled data. Results indicate that the TextBlob-IndoBERT model outperforms the others, effectively processing informal Indonesian text with high accuracy. These findings provide strategic insights for the government in understanding public perception regarding the development of Nusantara and demonstrate the potential of Transformer-based sentiment analysis for the Indonesian language. The study recommends further exploration of factors influencing sentiment and analysis on other social media platforms.

abstrak

Perayaan Hari Kemerdekaan Indonesia yang pertama di ibu kota baru, Nusantara, menjadi momen yang bersejarah. Penelitian ini menganalisis sentimen publik terhadap perayaan tersebut dengan menggunakan model IndoBERT. Data diperoleh dari Twitter selama periode perayaan dan diklasifikasikan menjadi sentimen positif, negatif, dan netral. Tiga pendekatan utama digunakan: model IndoBERT sebagai baseline, IndoBERT yang di fine-tuning menggunakan data IndoNLU, serta penerapan IndoBERT pada data berlabel TextBlob. Hasil menunjukkan bahwa TextBlob-IndoBERT memiliki kinerja terbaik, mampu mengolah teks informal dengan akurasi tinggi. Temuan ini memberikan wawasan strategis bagi pemerintah dalam memahami persepsi publik terkait pengembangan Nusantara dan menunjukkan potensi analisis sentimen berbasis Transformer untuk bahasa Indonesia. Penelitian ini merekomendasikan eksplorasi lanjutan terhadap faktor-faktor sentimen dan analisis di platform media sosial lainnya.

Corresponding Author. Email: tdsalma@gmail.com ^{1}.

Copyright 2025 by the authors of this article. Published by Lembaga Otonom Lembaga Informasi dan Riset Indonesia (KITA INFO dan RISET). This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. 

1. Pendahuluan

Perkembangan teknologi kecerdasan buatan, khususnya di bidang pemrosesan bahasa alami atau *Natural Language Processing* (NLP), telah mengalami kemajuan pesat dalam beberapa tahun terakhir (Gunawan & Santoso, 2021; Kartika & Soca Baskara, 2024). Salah satu inovasi paling signifikan adalah hadirnya model transformer, yang memungkinkan analisis teks dengan pemahaman konteks yang lebih baik dan lebih efisien. Model transformer, seperti BERT (*Bidirectional Encoder Representations from Transformers*), telah mendominasi berbagai aplikasi NLP, termasuk analisis sentimen, dengan keunggulan dalam menangkap nuansa dan makna teks yang kompleks (Qiu *et al.*, 2020).

Analisis sentimen adalah proses pengumpulan, analisis, dan interpretasi data untuk menentukan pandangan, opini, dan emosi pengguna terhadap suatu isu, produk, atau peristiwa tertentu (Kaharudin *et al.*, 2023). Teknik ini telah menjadi metode yang sangat penting dalam memahami opini publik, terutama melalui media sosial, yang berfungsi sebagai sarana bagi masyarakat untuk mengekspresikan pendapat. Media sosial juga telah menjadi sumber data utama dalam menggali sentimen publik terhadap berbagai peristiwa (Fauzianto & Supatman, 2023). Dalam hal sosial Indonesia, peringatan Hari Kemerdekaan yang pertama kali diadakan di ibu kota baru, Nusantara, menjadi momen bersejarah yang menarik perhatian publik. Sebagai simbol kemajuan bangsa, peristiwa pemindahan kota memicu berbagai reaksi masyarakat yang disampaikan melalui media sosial seperti Twitter (Sutoyo & Almaarif, 2020).

Twitter, sebagai *platform* yang banyak digunakan masyarakat untuk menyuarakan pendapat, menyediakan data yang berharga untuk analisis sentimen. Namun, analisis ini menghadapi tantangan seperti volume data yang besar dan kompleksitas interpretasi sentimen dalam teks pendek. Untuk itu, model berbasis transformer seperti IndoBERT menunjukkan potensi besar dalam analisis ini, terutama melalui penerapan *fine-tuning* yang dapat meningkatkan akurasi. Penelitian sebelumnya menunjukkan efektivitas IndoBERT dalam berbagai domain, seperti analisis sentimen politik (Geni *et al.*, 2023), ulasan aplikasi layanan kesehatan (Imaduddin

et al., 2023), dan sentimen terhadap kendaraan listrik (Merdiansah & Ali Ridha, 2024), dengan hasil akurasi yang sangat baik. Di sisi lain, metode berbasis leksikon seperti TextBlob masih digunakan secara luas untuk analisis sentimen dasar. TextBlob telah berhasil diterapkan untuk analisis sentimen dalam berbagai konteks, seperti opini publik terhadap pandemi COVID-19 (Kaur *et al.*, 2020), layanan pelanggan PT Telkom Indonesia (Mas Diyasa *et al.*, 2021), dan analisis reputasi merek smartphone (K & P, 2022). Meskipun TextBlob tidak sekuat model transformer, alat ini dapat berfungsi sebagai pembanding atau pelengkap dalam pengembangan sistem analisis sentimen yang lebih kompleks.

Penelitian ini berfokus pada analisis sentimen publik terhadap perayaan Hari Kemerdekaan di ibu kota baru, Nusantara, dengan menggunakan data dari Twitter. Penelitian ini tidak hanya mengevaluasi efektivitas model IndoBERT melalui *fine-tuning* tetapi juga mengeksplorasi potensi kombinasi metode berbasis leksikon seperti TextBlob dengan transformer. Hasil dari penelitian ini dapat memberikan wawasan mendalam tentang opini publik terhadap Nusantara sebagai simbol kemajuan bangsa dan mendukung pengambilan keputusan strategis terkait pembangunan ibu kota baru.

2. Metodologi Penelitian

Tahapan penelitian secara keseluruhan dapat dilihat pada Gambar 1 berikut.



Gambar 1. Tahapan Penelitian dengan tiga skema

1) Pengumpulan Data

Langkah pertama dalam penelitian ini adalah melakukan pengumpulan data dari media sosial Twitter. Proses ini dilakukan menggunakan API Twitter yang diakses melalui *library* Tweepy pada Python. Pengumpulan data difokuskan pada percakapan yang relevan dengan topik Ibu Kota Nusantara (IKN) menggunakan kata kunci seperti #IKN, Nusantara, dan "ibukota baru", yang difilter berdasarkan bahasa Indonesia (*lang:id*) serta rentang waktu dari 22 Juli 2024 hingga 22 Agustus 2024. Sebanyak 2.010 data teks berhasil dikumpulkan dalam tahap ini. Data tersebut terdiri atas teks tweet, waktu posting, jumlah interaksi (seperti retweet, *likes*), dan atribut lain yang relevan untuk analisis sentimen. Gambar 2 memberikan contoh data yang berhasil dikumpulkan melalui skenario pengumpulan data yang telah dirancang.

```
full_text
hiks suka sdih liat hiburan humu2 ibukota soalnya gbs dateng jawah
Mantapkan ibukota Republik Indonesia...not IKN
Peringatan HUT Ke-79 Kemerdekaan RI Nusantara Baru Indonesia Maju di Bali Beach Hotel. Dirgahayu Indonesiaku Bangkitlah Bangsa
Kibarkan Sang Merah Putih di langit tinggi dan pancarkan rasa bangga ke penjuru nusantara. Nusantara Baru Indonesia Maju Merdeka!
ðŸŒŸ#9Yp0 https://t.co/zE3kdcC3r
Air bersih di IKN bisa diminum #ikn #ibukotanegara #airminum https://t.co/wOSY7jYyS
Apakah ini dalam rangka uji coba kalau sampeyan ntar ke ibukota baru?
Jelang upacara peringatan Hari Kemerdekaan RI ke-79 pada 17 Agustus di Ibu Kota Nusantara (IKN) Pertamina Gas Negara (PGN) memastikan
bahwa gas bumi telah mengalir dan siap digunakan. #PertaminaDukungIKN https://t.co/8otNhu4cWT
Pasokan listrik yang andal dan hijau PLN dalam mendukung pelaksanaan Upacara HUT ke 79 Kemerdekaan Republik Indonesia di Ibu Kota
Nusantara menuai apresiasi dari berbagai pihak. #PLNSukseskanUpacaraIKN https://t.co/84VB9ISS9M
Pawai is dalam memperingati HUT kemerdekaan RI yang ke 79 Nusantara baru Indonesia maju https://t.co/uAAMBC2jNC
@tulisnara harusnya yang di ikn bukan unesa bzir universitas negeri nusantara
@tvOneNews Seriusan nanya nih @ridwankamil... ibukota Indonesia dimana yaa?? Kaltim itu kan ibukota Nusantara bukan ibukota
Indonesia... apa sekarang udah diganti nama Indonesia jadi Nusantara... sampe2 aja juga Islam Nusantara... nanti adakah kristen nusantara
Kaltimk nusantara??
```

Gambar 2. Contoh data

2) Eksplorasi Data

Setelah pengumpulan data selesai, langkah berikutnya adalah melakukan eksplorasi data awal atau *Exploratory Data Analysis* (EDA). Tahap ini bertujuan untuk memahami distribusi data, mendeteksi pola, serta mengidentifikasi anomali seperti entri duplikat atau data yang tidak relevan (Gupta *et al.*, 2024). Hasil EDA membantu menentukan elemen-elemen yang memerlukan pembersihan lebih lanjut dalam tahap pemrosesan data. Gambar 3 menyajikan deskripsi mengenai data yang terkumpul, yang terdiri atas 2.010 baris dengan 14 kolom atau variabel, serta tipe data dari masing-masing variabel.

```
Info DataFrame:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2010 entries, 0 to 2009
Data columns (total 15 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                ---
0   conversation_id_str                  2010 non-null   int64
1   created_at                          2010 non-null   object
2   favorite_count                      2010 non-null   int64
3   full_text                           2010 non-null   object
4   id_str                              2010 non-null   int64
5   image_url                           1243 non-null   object
6   in_reply_to_screen_name             504 non-null   object
7   lang                                2010 non-null   object
8   location                            985 non-null   object
9   quote_count                         2010 non-null   int64
10  reply_count                         2010 non-null   int64
11  retweet_count                       2010 non-null   int64
12  tweet_url                           2010 non-null   object
13  user_id_str                         2010 non-null   int64
14  username                            2010 non-null   object
```

Gambar 2. Deskripsi dari data

3) Pra-Pemrosesan

Pra-pemrosesan data adalah langkah penting dalam penelitian ini untuk memastikan kualitas data Twitter yang tidak terstruktur menjadi lebih siap untuk analisis sentimen. Data Twitter, yang terdiri dari teks pendek dengan format beragam dan tidak konsisten, menimbulkan tantangan tersendiri. Oleh karena itu, tahapan ini dirancang untuk meningkatkan kualitas data, sebagaimana dinyatakan oleh (Salma *et al.*, 2021) bahwa data pelatihan berkualitas tinggi sangat penting untuk keberhasilan penerapan teknik NLP. Langkah-langkah yang dilakukan meliputi pembersihan data dengan menghapus elemen yang tidak relevan, seperti simbol, URL, emoji, dan karakter khusus lainnya. Selain itu, penghapusan entri duplikat dan stopwords dilakukan untuk menghilangkan data berulang serta kata-kata umum yang tidak memberikan kontribusi signifikan dalam analisis sentimen, seperti "dan", "di", atau "yang." Substitusi leksikal diterapkan untuk menyederhanakan variasi Bahasa (North *et al.*, 2024), sedangkan proses stemming digunakan untuk mengubah kata menjadi bentuk dasarnya agar analisis lebih konsisten (Rasappan *et al.*, 2024). Setelah melalui tahapan ini, data menjadi lebih bersih dan terstruktur, sebagaimana ditampilkan pada Gambar 4, yang menunjukkan contoh data setelah proses pra-pemrosesan dilakukan.

```
mantap ibukota republik indonesia not ikn
ingat hut ke merdeka ri nusantara indonesia maju bali beach hotel dirgahayu indonesia bangkit bangsa kibarkan sang merah putih langit pancar penjuru
nusantara nusantara indonesia maju merdeka
air bersih dan minum air ibukota nusantara
rangka uji coba sampeyan ntar ke ibukota
Jelang upacara ingat merdeka ri ke 79 agustus kota nusantara ikn pertamina gas negara pgn gas bumi air pertaminadukungin
pasok listrik andal hijau pln dukung laksana upacara hut merdeka republik indonesia kota nusantara tui apresiasi pnsukseskanupacaraan
pawai is ingat hut merdeka ri nusantara indonesia maju
ikn unesa bzir universitas negeri nusantara
serius nanya nih ibukota Indonesia mana yaa kaltim ibukota nusantara ibukota Indonesia udah ganti nama Indonesia nusantara sampe aja islam nusantara ada
kristen nusantara katolik nusantara
desain rumah ibukota agama damang kawasan kota nusantara bangun masjid negara karta nuarta dc lanjut basilika pasuk tahap tender source
harau akutu ibukota tp tetep ya posisi jakarta jd rebut pilkada
'membudi budi karya sumadi bandara kota nusantara ikn operasi september tepat rencana presiden joko widodo pindah ikn yanti beneran peluk aulia anies
menteri pagu pit kepala otorita kota nusantara oken basuki hadimujono aparatur sipil negara asin pindah ikn september yanti beneran peluk aulia anies ritsui
menteri hubung mentri budi karya sumadi bandara kota nusantara ikn operasi september tepat rencana presiden jokowi pindah ikn yanti beneran peluk aulia
```

Gambar 3. Data Teks setelah Pra-Pemrosesan

4) Pelabelan Data dan Pembagian Data

Setelah data melalui tahap pra-pemrosesan, langkah berikutnya adalah pelabelan sentimen. Proses pelabelan dilakukan menggunakan dua pendekatan, yaitu TextBlob dan model IndoBERT yang telah di-*pretrained*. Pelabelan dengan TextBlob mengklasifikasikan data ke dalam tiga kategori yaitu positif, negatif, dan netral, berdasarkan analisis berbasis aturan. Sementara itu, IndoBERT digunakan untuk memberikan pelabelan yang lebih mendalam melalui pendekatan berbasis transformer, yang dapat menangkap konteks sentimen dengan lebih akurat. Hasil dari proses pelabelan ini kemudian dibagi menjadi tiga bagian, yaitu data pelatihan sebesar 70%, data validasi sebesar 15%, dan data pengujian sebesar 15%. Dengan pembagian ini, model dapat dilatih secara optimal pada data pelatihan, dievaluasi secara objektif pada data validasi, dan diuji pada data pengujian untuk mengukur performanya.

5) Pemodelan IndoBERT

Pada tahap pemodelan, penelitian ini menggunakan tiga pendekatan untuk mengukur kinerja model dalam analisis sentimen. Pendekatan pertama adalah menggunakan model IndoBERT *pre-trained* tanpa pelatihan tambahan. Model ini langsung diterapkan pada data untuk menguji kemampuan dasar IndoBERT dalam memahami teks berbahasa Indonesia. Pendekatan kedua melibatkan *fine-tuning* model IndoBERT menggunakan dataset IndoNLU. Proses *fine-tuning* dilakukan untuk menyesuaikan parameter model agar lebih sesuai dengan karakteristik data dan konteks sentimen yang dianalisis. Pendekatan ketiga mengombinasikan hasil pelabelan sentimen berbasis TextBlob dengan model IndoBERT *pre-trained*. Dalam pendekatan ini, data yang telah dilabeli menggunakan TextBlob dimasukkan sebagai *input* untuk IndoBERT, sehingga memungkinkan evaluasi efektivitas integrasi metode berbasis aturan dengan transformer dalam meningkatkan akurasi analisis sentimen. Setiap pendekatan diuji menggunakan dataset validasi untuk menentukan parameter optimal sebelum model diterapkan pada dataset pengujian.

6) Evaluasi Model

Evaluasi kinerja model dalam analisis sentimen penting untuk mengukur kemampuan model

dalam mengklasifikasikan sentimen dengan akurasi yang tinggi. Metrik utama yang digunakan dalam evaluasi ini adalah akurasi, yang dihitung berdasarkan empat unsur: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). TP mengukur kemampuan model dalam mengidentifikasi sentimen positif yang benar, TN mengukur kemampuan model dalam mengidentifikasi sentimen negatif yang benar, sedangkan FP dan FN menunjukkan kesalahan klasifikasi model (Putri *et al.*, 2024). Akurasi dihitung dengan rumus seperti pada persamaan 1 yang mengukur proporsi prediksi yang benar terhadap total prediksi (M & M.N, 2015).

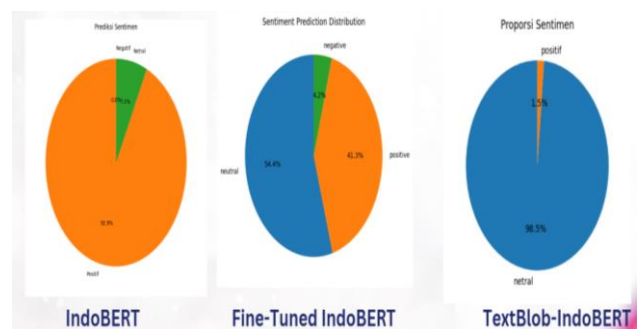
$$\text{Akurasi} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

Semakin tinggi nilai akurasi, semakin baik model dalam memprediksi kelas yang benar. Evaluasi ini memungkinkan peneliti untuk menilai efektivitas model dalam mengklasifikasikan sentimen dengan tepat.

3. Hasil dan Pembahasan

Hasil

Setelah melalui tahapan ekstraksi, pra-pemrosesan, dan pelabelan data, penelitian ini berhasil menganalisis 2.010 tweet yang dikumpulkan untuk menggambarkan sentimen publik terhadap Nusantara selama periode penelitian. Proses analisis dilakukan dengan menggunakan tiga model pendekatan yang berbeda, yaitu IndoBERT, Fine-Tuned IndoBERT, dan TextBlob-IndoBERT. Hasil analisis sentimen dari ketiga pendekatan ini dapat dilihat pada Gambar 4, yang menunjukkan visualisasi perbandingan sentimen yang diprediksi oleh masing-masing model.



Gambar 4. Hasil sentiment berdasarkan modelnya

Berdasarkan hasil analisis yang dilakukan, model IndoBERT menunjukkan hasil yang sangat dominan dengan sentimen positif, yakni mencapai 92,9%. Hal ini menunjukkan bahwa model IndoBERT cenderung mengklasifikasikan hampir seluruh tweet sebagai sentimen positif, sementara sentimen netral hanya ditemukan dalam proporsi kecil (7,1%), dan sentimen negatif hampir tidak terdeteksi (0%). Di sisi lain, model Fine-Tuned IndoBERT menghasilkan distribusi sentimen yang lebih seimbang. Model ini mengklasifikasikan sentimen netral sebagai yang paling dominan (54,4%), diikuti oleh sentimen positif (41,3%) dan sentimen negatif (4,2%). Pendekatan ini menunjukkan kemampuan yang lebih baik dalam membedakan berbagai jenis sentimen, terutama dalam mendeteksi sentimen negatif yang lebih sering terlewatkan pada model IndoBERT. Sementara itu, kombinasi antara TextBlob dan IndoBERT (TextBlob-IndoBERT) menghasilkan prediksi dengan 98,5% data diklasifikasikan sebagai sentimen netral. Sentimen positif hanya terdeteksi sebesar 1,5%, dan tidak ada sentimen negatif yang ditemukan.

Pembahasan

Hasil analisis sentimen dalam penelitian ini menunjukkan bahwa mayoritas sentimen yang teridentifikasi dalam dataset adalah sentimen netral. Temuan ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa tweet yang berisi informasi faktual sering kali tidak mengandung ekspresi emosional yang jelas, terutama dalam topik-topik teknis atau berita (Fauzianto & Supatman, 2023). Dalam kasus perayaan pembangunan Ibu Kota Negara (IKN) Nusantara, sebagian besar tweet lebih banyak berisi penyampaian informasi objektif mengenai perkembangan proyek, daripada menyuarakan opini atau perasaan yang jelas terkait isu tersebut. Sebagai contoh, banyak tweet yang memuat laporan tentang status pembangunan atau perubahan kebijakan, tanpa mencerminkan emosi yang kuat terhadap topik tersebut (Sutoyo & Almaarif, 2020). Pada eksperimen yang menggunakan kombinasi TextBlob dan IndoBERT (TextBlob-IndoBERT), akurasi tertinggi tercatat, yaitu 0,8, yang menunjukkan kecocokan model ini dengan distribusi data yang didominasi oleh sentimen netral. Pendekatan ini mengonfirmasi efektivitas TextBlob dalam menangkap pola sentimen yang lebih umum

pada data yang kurang ekspresif secara emosional. TextBlob yang berbasis aturan mampu mengidentifikasi pola-pola umum pada teks yang berfokus pada informasi dan bukan opini, yang selaras dengan temuan sebelumnya bahwa TextBlob efektif dalam analisis sentimen pada teks yang relatif tidak mengandung ekspresi emosional (Kaur *et al.*, 2020; K & P, 2022). Namun, eksperimen dengan Fine-Tuned IndoBERT menunjukkan hasil yang lebih kompleks. Meskipun akurasi model ini hanya mencapai 0,6, distribusi sentimen yang dihasilkan lebih seimbang, dengan kecenderungan sentimen netral-positif. Hal ini menunjukkan bahwa meskipun proses *fine-tuning* pada model IndoBERT meningkatkan kemampuan model untuk memahami bahasa Indonesia, ketidakseimbangan distribusi sentimen dalam data pelatihan tetap mempengaruhi performa model. *Fine-tuning* memang dapat meningkatkan kemampuan model dalam menangkap konteks yang lebih spesifik, namun jika data pelatihan tidak mencerminkan variasi sentimen yang seimbang, model tetap kesulitan untuk memprediksi dengan akurat (Geni, Yulianti, & Sensuse, 2023).

Sebaliknya, eksperimen dengan IndoBERT tanpa proses *fine-tuning* menunjukkan kecenderungan yang jelas terhadap klasifikasi sentimen positif, dengan sebagian besar tweet diklasifikasikan sebagai sentimen positif. Namun, analisis manual terhadap beberapa sampel menunjukkan bahwa banyak tweet tersebut sebenarnya bersifat netral atau bahkan negatif. Temuan ini menyoroti keterbatasan model *pre-trained* seperti IndoBERT dalam memahami konteks yang lebih kompleks dalam teks bahasa Indonesia tanpa *fine-tuning*. Tanpa penyesuaian pada data lokal atau dataset yang lebih spesifik, model *pre-trained* cenderung mengklasifikasikan teks berdasarkan pola yang sering dijumpai dalam data pelatihan awal, yang mungkin tidak sepenuhnya mencerminkan karakteristik tweet yang bersifat netral atau negatif (Qiu *et al.*, 2020). Penting juga untuk dicatat bahwa hasil akurasi yang diperoleh dalam setiap eksperimen ini hanya mencerminkan kemajuan model dalam memprediksi sentimen berdasarkan pola yang ada dalam dataset yang digunakan. Mengingat tidak adanya data manual yang dilabeli sebagai acuan pembandingan, akurasi ini mencerminkan kemampuan model untuk menghasilkan prediksi yang konsisten dengan data yang telah dilatih.

Evaluasi yang lebih komprehensif dan akurat akan memerlukan data referensi yang lebih lengkap dan representatif yang dilabeli secara manual, seperti yang disarankan oleh Fauzianto & Supatman (2023) dalam penelitian mereka yang membahas pentingnya akurasi dalam analisis sentimen pada teks sosial media. Temuan ini menunjukkan bahwa meskipun model berbasis *transformer* seperti *IndoBERT* menunjukkan potensi yang besar, penerapan *fine-tuning* dan penggunaan kombinasi metode berbasis leksikon seperti *TextBlob* sangat penting dalam meningkatkan akurasi analisis sentimen pada teks yang tidak selalu mengandung ekspresi emosional yang eksplisit (Gunawan & Santoso, 2021; Merdiansah & Ali Ridha, 2024). Penelitian ini membuka peluang untuk penerapan lebih lanjut pada data sosial media lainnya dan memberikan dasar untuk pengembangan teknik analisis sentimen berbasis model bahasa Indonesia yang lebih efisien.

4. Kesimpulan dan Saran

Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap Ibu Kota Negara (IKN) Nusantara pada Hari Kemerdekaan Indonesia menggunakan pendekatan berbasis *Transformer*, khususnya *IndoBERT*. Berdasarkan hasil yang diperoleh, dapat disimpulkan bahwa mayoritas sentimen yang teridentifikasi dalam dataset adalah sentimen netral, yang mencerminkan dominasi percakapan informasi mengenai pembangunan IKN tanpa ekspresi emosional yang jelas. Hasil eksperimen menunjukkan variasi dalam performa model yang digunakan. Model *TextBlob-IndoBERT* memberikan hasil terbaik dibandingkan model lainnya. Temuan ini menunjukkan bahwa meskipun model *pre-trained* seperti *IndoBERT* memiliki kemampuan dasar yang baik, pemahaman terhadap konteks bahasa Indonesia yang lebih kompleks memerlukan *fine-tuning* atau pelabelan manual. Penelitian ini memberikan gambaran yang jelas mengenai sentimen publik terhadap pembangunan IKN Nusantara, serta pentingnya pengembangan lebih lanjut untuk memperoleh analisis sentimen yang lebih akurat dan mendalam. Berdasarkan temuan penelitian ini, beberapa rekomendasi dapat diajukan untuk penelitian selanjutnya.

Pertama, disarankan untuk melakukan pelabelan data secara manual agar menghasilkan dataset yang lebih valid dan representatif, yang dapat digunakan sebagai acuan dalam mengevaluasi kinerja model dengan lebih akurat. Kedua, penting untuk menyeimbangkan distribusi data sentimen dalam dataset, sehingga model dapat belajar secara optimal dari berbagai jenis sentimen dan meningkatkan akurasi dalam klasifikasi sentimen. Penyeimbangan ini akan mengurangi potensi bias terhadap kategori sentimen tertentu, yang dapat mempengaruhi hasil analisis. Selanjutnya, penelitian lebih lanjut dapat difokuskan pada analisis penyebab sentimen untuk menggali faktor-faktor yang mendasari reaksi publik terhadap IKN Nusantara. Hal ini dapat memberikan wawasan yang lebih dalam mengenai dinamika opini publik yang terbentuk. Terakhir, disarankan untuk melakukan eksperimen yang menggabungkan pendekatan *TextBlob* dengan *fine-tuned IndoBERT*, yang diharapkan dapat meningkatkan akurasi analisis sentimen dan memungkinkan model untuk menangkap nuansa sentimen yang lebih kompleks dalam teks bahasa Indonesia. Dengan menerapkan rekomendasi-rekomendasi ini, diharapkan penelitian mendatang dapat menghasilkan analisis sentimen yang lebih akurat, *komprehensif*, dan representatif, serta memberikan kontribusi yang lebih besar dalam memahami opini publik terhadap kebijakan-kebijakan nasional strategis, seperti pembangunan IKN Nusantara.

5. Ucapan Terima Kasih

Ucapan terima kasih disampaikan kepada Universitas LIA atas segala dukungan, fasilitas, dan kesempatan yang diberikan selama pelaksanaan penelitian ini. Kami berharap hasil penelitian ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan serta memberikan kontribusi positif bagi kemajuan Universitas LIA.

6. Daftar Pustaka

- Fauzianto, R. A. (2022). Analisis Sentimen Opini Masyarakat Terhadap Tech Winter Pada Twitter Menggunakan Natural Language Processing. *Jurnal Syntax Admiration*, 3(9), 1577-1585.
<https://doi.org/10.46799/jsa.v3i9.909>.
- Geni, L., Yulianti, E., & Sensuse, D. I. (2023). Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using IndoBERT Language Models. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika (JITEKI)*, 9(3), 746-757.
- Gunawan, K. I., & Santoso, J. (2021). Multilabel text classification menggunakan svm dan doc2vec classification pada dokumen berita bahasa indonesia. *Journal of Information System, Graphics, Hospitality and Technology*, 3(01), 29-38.
<https://doi.org/10.37823/insight.v3i01.126>.
- Gupta, A. D., Singh, K., Singh, K. D., Kushwaha, P., Lohani, B. P., & Kumar, S. (2024). Unveiling Insights: Exploring Healthcare Data through Data Analysis. *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)*, 575–581.
<https://doi.org/10.1109/IC3SE62002.2024.10593333>.
- Imaduddin, H., A'la, F. Y., & Nugroho, Y. S. (2023). Sentiment analysis in Indonesian healthcare applications using IndoBERT approach. *International Journal of Advanced Computer Science and Applications*, 14(8).
- K, G., & P, S. (2022). Evaluating Popular Smartphone Brands Based On Twitter Sentiment Using Textblob. *ICTACT Journal on Image and Video Processing*, 12(3), 2655–2660.
<https://doi.org/10.21917/ijivp.2022.0377>.
- Kaharudin, A., Supriyadi, A. A., Baitika, H., & Derryanur, M. (2023). Analisis Sentimen pada Media Sosial dengan Teknik Kecerdasan Buatan Naïve Bayes: Kajian Literatur Review. *OKTAL: Jurnal Ilmu Komputer Dan Sains*, 2(06), 1642-1649.
- Kartika, I. D., & Baskara, D. S. (2024, October). Pengembangan Financial Documentation System Berbasis Artificial Intelligence Untuk Meningkatkan Self Service Dan Layanan Prima Bidang Keuangan. In *Prosiding Seminar Nasional Sains dan Teknologi Terapan* (No. 1).
- Kaur, C., & Sharma, A. (2020). Twitter sentiment analysis on coronavirus using textblob. *EasyChair* 2516-2314.
- M, H., & M.N, S. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2), 01–11.
<https://doi.org/10.5121/ijdkp.2015.5201>.
- Mas Diyasa, I. G. S., Marini Mandenni, N. M. I., Fachrurrozi, M. I., Pradika, S. I., Nur Manab, K. R., & Sasmita, N. R. (2021). Twitter Sentiment Analysis as an Evaluation and Service Base On Python Textblob. *IOP Conference Series: Materials Science and Engineering*, 1125(1), 012034.
<https://doi.org/10.1088/1757-899x/1125/1/012034>.
- Merdiansah, R., Siska, S., & Ridha, A. A. (2024). Analisis sentimen pengguna X Indonesia terkait kendaraan listrik menggunakan IndoBERT. *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, 7(1), 221-228.
<https://doi.org/10.55338/jikomsi.v7i1.2895>.
- North, K., Ranasinghe, T., Shardlow, M., & Zampieri, M. (2024). Deep learning approaches to lexical simplification: A survey. *Journal of Intelligent Information Systems*.
<https://doi.org/10.1007/s10844-024-00882-9>.
- Putri, A. D., Sholekhah, F., Dadynata, E., Efrizoni, L., Rahmadden, R., & Sapina, N. (2024). Penerapan Algoritma Decesion Tree C4.5 untuk Memprediksi Tingkat Kelangsungan Hidup Pasien Kanker Tiroid. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1485–1495.
<https://doi.org/10.57152/malcom.v4i4.1532>.

- Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2020). *Pre-trained Models for Natural Language Processing: A Survey*.
<https://doi.org/10.1007/s11431-020-1647-3>.
- Rasappan, P., Premkumar, M., Sinha, G., & Chandrasekaran, K. (2024). Transforming sentiment analysis for e-commerce product reviews: Hybrid deep learning model with an innovative term weighting and feature selection. *Information Processing & Management*, 61(3), 103654.
<https://doi.org/10.1016/j.ipm.2024.103654>.
- Salma, T. D., Saptawati, G. A. P., & Rusmawati, Y. (2021). Text Classification Using XLNet with Infomap Automatic Labeling Process. *2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, 1–6.
<https://doi.org/10.1109/ICAICTA53211.2021.9640255>.
- Sutoyo, E., & Almaarif, A. (2020). Twitter sentiment analysis of the relocation of Indonesia's capital city. *Bulletin of Electrical Engineering and Informatics*, 9(4), 1620–1630.
<https://doi.org/10.11591/eei.v9i4.2352>.