

Klasifikasi Sentimen Untuk Mengetahui Kecenderungan Politik Pengguna X Pada Calon Presiden Indonesia 2024 Menggunakan Metode IndoBert

Indro Abri Oktariansyah*, Fajri Rakhmat Umbara, Fatan Kasyidi

Fakultas Sains dan Informatika, Teknik Informatika, Universitas Jenderal Achmad Yani, Cimahi, Indonesia

Email: indroabrio20@if.unjani.ac.id, fajri.rakhmat@lecture.unjani.ac.id, fatan.kasyidi@lecture.unjani.ac.id

Email Penulis Korespondensi: indroabrio20@if.unjani.ac.id

Submitted: 27/06/2024; Accepted: 04/07/2024; Published: 07/09/2024

Abstrak—X telah berkembang menjadi salah satu *platform* media sosial yang paling populer di dunia. Di Indonesia, penggunaan X cukup luas, terutama dalam diskusi pemilihan presiden yang sedang ramai diperbincangkan. Setiap orang memiliki pandangan yang berbeda terhadap para calon, baik positif maupun negatif. Dengan banyaknya data *tweet* dari pengguna, informasi tersebut dapat dijadikan sebagai sumber data untuk diolah dan dianalisis. Terdapat berbagai metode yang dapat digunakan untuk menganalisis dan mengklasifikasikan sentimen dari data tersebut, salah satunya yaitu menggunakan *BERT*. Pada penelitian ini dilakukan klasifikasi sentimen menggunakan *BERT* dengan model *IndoBert* dengan tujuan penelitian yaitu, melakukan klasifikasi sentimen terhadap *tweet* terkait pemilihan presiden Indonesia 2024 guna memahami kecenderungan politik pengguna X, mengevaluasi kinerja model *IndoBert* dalam klasifikasi sentimen, dan menilai sejauh mana teknik augmentasi *back translation* dan augmentasi sinonim dapat meningkatkan performa model. Data dikumpulkan dengan teknik *crawling* selama 7 hari menjelang pemilu dan dilabeli secara manual oleh *annotator*. Teknik augmentasi sinonim dan *back translation* digunakan untuk menyeimbangkan data pada kelas minoritas. Data dibagi menjadi 80% data latih, 10% data uji, dan 10% data validasi. Proses klasifikasi dilakukan dengan model *IndoBert* yang telah di *fine-tuning*. Hasil penelitian menunjukkan bahwa *IndoBert* dengan augmentasi sinonim mencapai akurasi tertinggi, yaitu 82% pada eksperimen pertama dan 81% pada eksperimen kedua. Di sisi lain, *back translation* hanya mencapai akurasi 78% pada eksperimen pertama dan 74% pada eksperimen kedua. Hal ini menunjukkan bahwa augmentasi sinonim terbukti lebih efektif dalam meningkatkan variasi data dan kinerja model pada *dataset* yang digunakan dalam penelitian ini.

Kata Kunci: Pemilu; Klasifikasi Sentimen; Augmentasi; IndoBert

Abstract—X has evolved into one of the most popular social media platforms in the world. In Indonesia, the use of X is quite widespread, especially in discussions about the presidential election, which is currently a hot topic. Everyone has different views on the candidates, both positive and negative. With a large amount of tweet data from users, this information can serve as a data source for processing and analysis. Various methods can be used to analyze and classify sentiment from this data, one of which is using BERT. This research conducts sentiment classification using BERT with the IndoBert model. The research aims to classify sentiments towards tweets related to the 2024 Indonesian presidential election to understand the political inclinations of X users, evaluate the performance of the IndoBert model in sentiment classification, and assess the extent to which back translation augmentation and synonym augmentation techniques can enhance the model's performance. Data was collected using crawling techniques for seven days leading up to the election and manually labeled by annotators. Synonym augmentation and back translation techniques were used to balance data in minority classes. The data was divided into 80% training data, 10% test data, and 10% validation data. The classification process was conducted using the IndoBert model that had been fine-tuned. The research results show that IndoBert with synonym augmentation achieved the highest accuracy, which was 82% in the first experiment and 81% in the second experiment. On the other hand, back translation only reached an accuracy of 78% in the first experiment and 74% in the second experiment. This indicates that synonym augmentation proved to be more effective in increasing data variation and model performance on the dataset used in this research.

Keywords: Election; Sentiment Classification; Augmentation; IndoBert

1. PENDAHULUAN

X telah berkembang menjadi salah satu *platform* media sosial yang paling populer di dunia. Setiap harinya, jutaan pengguna aktif menggunakan media sosial X untuk berbagi pendapat, informasi, dan berinteraksi dengan orang lain. Menurut data dari We Are Social, pada Juli 2023 tercatat ada 564,1 juta pengguna media sosial X di seluruh dunia. Dari angka tersebut, Indonesia menempati peringkat keempat dengan jumlah pengguna sebanyak 25,25 juta. Dengan pertumbuhan jumlah pengguna media sosial X yang terus meningkat, terdapat peluang besar untuk mengklasifikasi sentimen yang terkait dengan topik populer di *platform* ini.

Salah satu topik yang sedang hangat diperbincangkan di media sosial X adalah pemilihan presiden di Indonesia. Pengguna *platform* tersebut memiliki kebebasan untuk menyampaikan pendapat mereka mengenai calon presiden melalui unggahan singkat yang dikenal sebagai *tweet*. Pengguna juga dapat mengekspresikan perasaan atau pandangan terhadap kandidat presiden dengan menggunakan tanda pagar (*hashtag*). Dengan banyaknya data *tweet* dari pengguna, informasi tersebut dapat dijadikan sebagai sumber untuk diolah dan dianalisis. Terdapat berbagai metode yang dapat digunakan untuk mengklasifikasi sentimen dari data tersebut [1].

Pada penelitian sebelumnya, dilakukan klasifikasi sentimen terhadap *tweet-tweet* pengguna X terkait pemilihan presiden Indonesia 2019. Penelitian ini menggunakan berbagai metode *machine learning* dan *deep neural network* untuk proses klasifikasi sentimen. Data yang digunakan dikumpulkan melalui proses *crawling* menggunakan *tweetpy* dan teknik *pseudo-labelling* diterapkan untuk proses pelabelan data. Hasil penelitian menunjukkan bahwa algoritma

deep neural network mencapai akurasi yang lebih tinggi dibandingkan dengan tiga algoritma *machine learning* tradisional lainnya (*SVM*, *Logistic Regression*, dan *Multinomial Naïve Bayes*). Penggunaan *Bidirectional LSTM* berhasil mengungguli metode *deep learning* lainnya dengan akurasi mencapai 84,60% [2].

Pada penelitian terdahulu, sebuah penelitian telah dilakukan terhadap *tweet-tweet* terkait pemilihan presiden Indonesia 2019 dengan menggunakan analisis sentimen. Metode *Naïve Bayes* dan *SVM* digunakan dalam penelitian tersebut untuk mengklasifikasikan *tweet* terkait pemilu presiden Indonesia 2019. Data *tweet* dikumpulkan dengan proses *crawling* dan dilabel secara otomatis menggunakan teknik *lexicon*. Hasil penelitian menunjukkan bahwa nilai akurasi tertinggi diperoleh saat menggunakan kombinasi algoritma *machine learning SVM* dan tokenisasi alfabetik, yang menghasilkan nilai akurasi 79,02%. Nilai akurasi terendah dalam penelitian ini diperoleh untuk algoritma *machine learning NBC* dengan tokenisasi *N-gram*, yang memiliki nilai akurasi 44,94% [3].

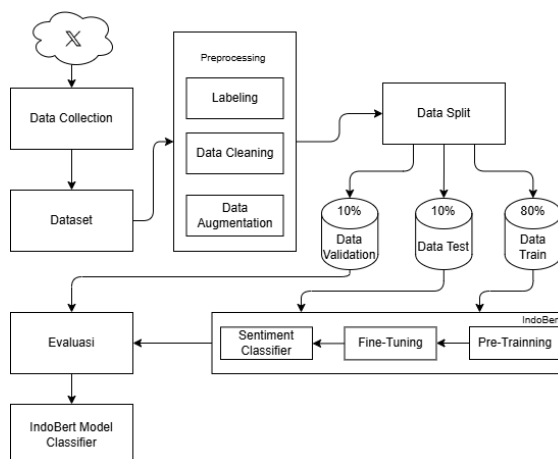
Pembicaraan publik mengenai terpilihnya tiga pasangan calon Presiden Republik Indonesia Tahun 2024 di media sosial X menjadi topik yang sangat menarik bagi para pengguna dengan beragam sentimen masyarakat, baik positif maupun negatif. Mengingat besarnya pengaruh opini publik terhadap dinamika politik dan hasil pemilihan, diperlukan suatu metode yang membantu melihat opini masyarakat secara efektif dan akurat. Pada penelitian ini, dilakukan klasifikasi sentimen *tweet* pengguna di media sosial X terhadap calon presiden Indonesia 2024 menggunakan metode *BERT* dengan model *IndoBert*. Penelitian ini bertujuan untuk mengevaluasi kinerja model *IndoBert* dalam klasifikasi sentimen, dan menilai sejauh mana teknik augmentasi *back translation* dan augmentasi sinonim dapat meningkatkan akurasi model. Augmentasi dilakukan karena *dataset* yang digunakan dalam penelitian ini menunjukkan ketidakseimbangan kelas. Data yang tidak seimbang dapat menyebabkan masalah seperti akurasi prediksi yang rendah untuk kelas-kelas yang jarang muncul [4]. Berdasarkan penelitian sebelumnya, salah satu cara untuk menangani kelas yang tidak seimbang dan meningkatkan variasi data pada kelas minoritas adalah dengan menerapkan teknik augmentasi sinonim dan *back translation* [5], [6], [7].

Penelitian ini memiliki empat tahapan utama yaitu, pengumpulan data, praproses data, pemodelan *IndoBert*, dan evaluasi. Penelitian ini dilakukan dengan merujuk pada berbagai penelitian yang telah dilakukan sebelumnya. Meskipun sudah ada penelitian sebelumnya yang menggunakan metode *machine learning* dan *deep learning* untuk klasifikasi sentimen terkait pemilihan presiden, penggunaan *BERT* dengan model *IndoBert* serta teknik augmentasi sinonim dan *back translation* masih tergolong baru dalam konteks pemilihan presiden Indonesia 2024. Penelitian sebelumnya cenderung menggunakan metode tradisional atau model-model yang belum sepenuhnya dioptimalkan untuk bahasa Indonesia. Selain itu, penelitian ini turut berperan dalam menciptakan *dataset* yang berisi *tweet-tweet* dari pengguna media sosial X terkait dengan pemilihan presiden. *Dataset* ini dapat menjadi sumber data bagi penelitian selanjutnya dalam klasifikasi sentimen, pengembangan model *NLP* (*Natural Language Processing*), dan penelitian-penelitian lainnya. Pada penelitian ini metode *BERT* digunakan karena *BERT* memiliki kinerja yang lebih baik dari metode yang lain dalam klasifikasi sentimen [8], [9], [10], [11]. Selanjutnya, *IndoBert* dipilih sebagai model yang digunakan karena dianggap sebagai model yang paling canggih dan inovatif untuk pengolahan bahasa Indonesia [12]. Augmentasi sinonim dan *back translation* dipilih untuk menangani ketidakseimbangan kelas karna dapat meningkatkan varian data dan kinerja model pada kelas minoritas[5], [6], [7].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Dalam tahapan penelitian, dilakukan serangkaian langkah yang terstruktur untuk memastikan kelancaran proses analisis data dan pencapaian tujuan penelitian. Tahapan ini mengikuti sebuah diagram alur yang menggambarkan setiap tahap yang dilakukan. Diagram alur tersebut ditunjukkan dalam Gambar 1 berikut:



Gambar 1. Tahapan Penelitian

a. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini dikumpulkan melalui proses *crawling* atau pengambilan data menggunakan *Tweet Harvest* versi 2.2.8. *Tweet Harvest* adalah sebuah alat atau perangkat lunak yang digunakan untuk mengumpulkan data dari media sosial X melalui proses yang dikenal sebagai web *crawling* atau *scraping*. *Tweet Harvest* memanfaatkan API (*Application Programming Interface*) X untuk mengambil data berupa *tweet* berdasarkan kata kunci, *hashtag*, pengguna, atau parameter lainnya. Dalam penelitian ini, *dataset* yang dikumpulkan terdiri dari *tweet* pengguna media sosial X selama 7 hari menjelang pemilihan Presiden Indonesia 2024.

b. Praproses

Pada tahap praproses, dilakukan pengolahan *dataset* untuk mempersiapkan data tersebut agar dapat dengan mudah diolah oleh algoritma *machine learning*. Proses ini dapat mengoptimalkan data sehingga data dapat memberikan hasil maksimum selama proses klasifikasi [12]. Beberapa langkah yang dilakukan dalam tahap ini adalah sebagai berikut:

1. Pelabelan Data

Pada tahapan ini, proses pelabelan data pada setiap *tweet* dilakukan. Label data diterapkan ke *dataset tweet* mentah dan membuat data *tweet* yang dibagi berdasarkan labelnya [12]. Pada tahapan ini dilakukan pelabelan secara manual merujuk pada penelitian sebelumnya [13]. *Tweet* dikategorikan berdasarkan sentimen yang diekspresikan terhadap masing-masing calon presiden. Pelabelan dilakukan oleh tiga *annotator* yang memiliki latar belakang pendidikan sebagai sarjana psikologi, sehingga mereka memiliki pemahaman yang mendalam tentang analisis sentimen. *Tweet-tweet* tersebut dikategorikan ke dalam tujuh kategori sentimen utama:

1. Positif Anies: *Tweet* yang dengan jelas menunjukkan dukungan atau apresiasi terhadap Anies Baswedan.
2. Positif Prabowo: *Tweet* yang dengan jelas menunjukkan dukungan atau apresiasi terhadap Prabowo Subianto
3. Positif Ganjar: *Tweet* yang dengan jelas menunjukkan dukungan atau apresiasi terhadap Ganjar Pranowo.
4. Negatif Anies: *Tweet* yang berisi kritik, celaan, atau sentimen negatif terhadap Anies Baswedan.
5. Negatif Prabowo: *Tweet* yang berisi kritik, celaan, atau sentimen negatif terhadap Prabowo Subianto.
6. Negatif Ganjar: *Tweet* yang berisi kritik, celaan, atau sentimen negatif terhadap Ganjar Pranowo.
7. Netral: *Tweet* yang tidak menunjukkan sentimen mendukung ataupun menolak terhadap calon presiden manapun.

2. Pembersihan Data

Pembersihan data merupakan tahap penting untuk mengurangi *noise* yang mungkin terdapat dalam kumpulan data *tweet*. Proses pembersihan data mencakup serangkaian langkah terinci yang bertujuan untuk meningkatkan kualitas dan kebersihan data yang digunakan dalam klasifikasi sentimen. Langkah-langkah tersebut meliputi :

1. penghapusan entitas ganda atau duplikat, tautan, nama pengguna yang diawali dengan simbol "@" (*@username*), dan karakter numerik. Selain itu, simbol, spasi ekstra, dan tanda baca yang tidak diperlukan juga dihapus, sehingga menghasilkan kumpulan data yang lebih bersih dan siap untuk digunakan dalam klasifikasi sentimen [14], [15].
2. Menghapus terlalu banyak aspek dari *dataset* akan mengurangi keragaman kata. oleh karena itu kata-kata *slang* yang terdapat dalam data tidak dihapus. model dapat memproses dan belajar dari berbagai jenis masukan dalam proses pelatihan [4].

3. Augmentasi Data

Setelah melakukan pembersihan pada data, dilakukan proses augmentasi untuk menambah jumlah dan variasi data yang ada dalam kelas minoritas. Dua teknik augmentasi digunakan pada tahapan ini, yaitu augmentasi sinonim dan augmentasi *back translation*. Teknik augmentasi sinonim meningkatkan variasi dalam *dataset* dengan mengganti kata-kata dalam kalimat menggunakan sinonim yang memiliki makna yang serupa [5]. Sementara itu, teknik *back translation* dilakukan dengan menerjemahkan teks awal ke bahasa lain dan kemudian menerjemahkannya kembali ke bahasa asalnya, menciptakan variasi dalam data dengan mempertahankan makna asli teks. proses menerjemahkan teks bisa menggunakan layanan terjemahan seperti Google Translate dan Amazon [5], [16]. Dengan *dataset* yang besar, proses penerjemahan membutuhkan banyak sumber daya [5]. Untuk menangani permasalahan ini dapat dilakukan dengan menggunakan model penerjemahan [5], [16].

4. Pembagian Data

Setelah melalui tahap augmentasi, langkah berikutnya adalah membagi data. Untuk mengevaluasi kinerja dari model, *dataset* dibagi menjadi data latih, validasi, dan uji. *Dataset* dibagi dengan perbandingan 80% untuk data latih dan 20% untuk data uji dan validasi. Pembagian ini berdasarkan penelitian sebelumnya [4].

c. Klasifikasi Sentimen

Setelah melalui tahap pembagian data, langkah selanjutnya adalah membuat model klasifikasi menggunakan *IndoBert*. *IndoBert* adalah varian dari *BERT* yang merupakan model berbasis *transformer* [12]. Model dibangun menggunakan data *training*, diperiksa kinerja model selama pelatihan untuk melihat model tersebut *overfitting* atau tidak menggunakan data validasi dan dievaluasi dengan menggunakan data *testing*.

d. Evaluasi

Dua metrik evaluasi utama yang sering digunakan adalah akurasi dan skor F1. Kedua metrik ini memberikan gambaran yang berbeda tetapi saling melengkapi tentang kinerja model.

Akurasi adalah metrik evaluasi yang mengukur sejauh mana model klasifikasi dapat memprediksi semua kasus dalam suatu *dataset* dengan benar. Secara formal, akurasi didefinisikan sebagai rasio antara jumlah prediksi yang benar (*True Positive* dan *True Negative*) dengan total jumlah prediksi yang dibuat oleh model. Akurasi memberikan gambaran umum tentang seberapa baik model dapat membuat prediksi yang benar secara keseluruhan.

$$\text{Akurasi} = \frac{TP + TN}{TP + FN + FP + FN} \quad (1)$$

Skor F1 adalah metrik evaluasi yang menggabungkan presisi dan *recall*, memberikan gambaran yang lebih menyeluruh tentang kinerja model klasifikasi, terutama dalam konteks ketidakseimbangan kelas. Skor F1 adalah rata-rata harmonis dari presisi dan *recall*, sehingga memberikan penilaian yang lebih seimbang ketika keduanya penting.

$$\text{Skor F1} = 2 \times \left(\frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \right) \quad (2)$$

Presisi mengukur sejauh mana model memberikan hasil yang relevan atau *true positive*, dan dihitung sebagai rasio antara jumlah *true positive* dan total jumlah prediksi positif (*true positive* dan *false positive*)

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (3)$$

Recall mengukur sejauh mana model dapat menemukan atau mengambil semua kasus positif yang ada, dan dihitung sebagai rasio antara *true positive* dan total jumlah kasus positif dalam dataset. TN (*True Negative*): Jumlah data yang benar diklasifikasikan sebagai negatif.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

Presisi dan *recall* adalah dua metrik evaluasi yang sangat penting terutama dalam konteks ketidakseimbangan kelas, dimana jumlah kasus positif dan negatif tidak seimbang. Presisi penting ketika nilai kesalahan *false positive* tinggi, sementara *recall* penting ketika nilai kesalahan *false negative* tinggi. Oleh karena itu, skor F1 menjadi metrik yang sangat berguna karena mempertimbangkan kedua aspek ini secara bersamaan, memberikan penilaian yang lebih holistik tentang kinerja model [13].

2.2 IndoBert

IndoBert merupakan model berbasis transformasi yang dilatih menggunakan kerangka kerja *Huggingface* dengan konfigurasi *default BERT-Base (uncased)* [17]. Lebih dari 220 juta kata digunakan dalam pelatihan *IndoBert*, yang berasal dari tiga sumber utama: Wikipedia (74 juta kata), berita dari Kompas, Tempo, dan Liputan 6 (55 juta kata), serta situs web korpus Indonesia (90 juta kata) [17]. *IndoBert* memiliki 12 lapisan tersembunyi dengan ukuran 768d per lapisan, 12 *attention head*, dan satu lapisan tersembunyi *feed-forward* berukuran 3.072d [17].

2.3 Augmentasi Sinonim

Augmentasi sinonim adalah proses meningkatkan variasi dan volume data teks dengan mengganti kata-kata atau frasa dengan sinonimnya. Tujuan teknik ini yaitu untuk memperluas *dataset* tanpa mengubah makna atau konteks aslinya. Metode ini menggunakan *WordNet* atau *database leksikal* lain untuk mengelompokkan kata-kata ke dalam sinonim kognitif, memungkinkan penggantian yang sesuai dengan konteks kalimat. Teknik ini membantu meningkatkan keragaman data pelatihan dan dapat digunakan dalam berbagai aplikasi seperti pemodelan bahasa dan pemahaman teks untuk meningkatkan kinerja model[5].

2.4 Augmentasi Back Translation

Back translation adalah sebuah teknik augmentasi data yang melibatkan proses menerjemahkan teks dari bahasa sumber ke bahasa target menggunakan sistem terjemahan mesin, kemudian menerjemahkan kembali teks tersebut dari bahasa target ke bahasa sumber. Dalam konteks teks tersebut, *back translation* dilakukan dengan menggunakan API Google Translate. *Back translation* dapat digunakan sebagai teknik augmentasi untuk meningkatkan data pelatihan[7].

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data pada penelitian melalui proses *crawling* menggunakan *Tweet Harvest* versi 2.2.8. teknik ini melibatkan penggunaan kata kunci spesifik, yaitu nama-nama dari calon Presiden, Secara rinci, berikut adalah data yang berhasil dikumpulkan:

1. **Keyword "Anies":** Menghasilkan 4379 *tweet*. Data ini mencakup berbagai macam *tweet* yang menyebutkan nama Anies, termasuk *tweet* dari pendukung, kritik, berita terkait, dan percakapan umum yang melibatkan nama calon Presiden Anies Baswedan.
2. **Keyword "Prabowo":** Menghasilkan 3368 *tweet*. Data ini mencakup semua *tweet* yang menyebutkan nama Prabowo Subianto, yang mencakup dukungan, kritik, liputan media, dan diskusi terkait pencalonannya sebagai Presiden.
3. **Keyword "Ganjar":** Menghasilkan 2298 *tweet*. Data ini mencakup berbagai *tweet* yang menyebutkan nama Ganjar Pranowo, termasuk *tweet* dari pendukung, kritik, serta berita dan diskusi terkait pencalonannya.

Secara keseluruhan, jumlah data yang berhasil dikumpulkan adalah sebanyak 10045 *tweet* dengan atribut data seperti yang ditunjukkan pada Tabel 1:

Tabel 1. Atribut Data Hasil *Crawling*

No.	Atribut	Deskripsi
1	<i>Created_at</i>	Menunjukkan tanggal dan waktu kapan <i>tweet</i> tersebut diunggah.
2	<i>Id_Str</i>	Merupakan <i>ID</i> unik yang diberikan oleh X untuk setiap <i>tweet</i> .
3	<i>Full_Text</i>	Isi lengkap dari <i>tweet</i> yang diunggah.
4	<i>Quote_Count</i>	Jumlah <i>tweet di-quote</i> oleh pengguna lain.
5	<i>Reply_Count</i>	Jumlah balasan atau <i>reply</i> yang diterima.
6	<i>Retweet_Count</i>	Jumlah <i>tweet</i> ulang yang diterima.
7	<i>Favorite_Count</i>	Jumlah ' <i>likes</i> ' atau ' <i>favorit</i> ' yang diterima.
8	<i>Lang</i>	Bahasa yang digunakan dalam <i>tweet</i> .
9	<i>User_Id_Str</i>	<i>ID</i> unik dari pengguna X yang mengunggah <i>tweet</i>
10	<i>Conversation_id_Str</i>	<i>ID</i> unik yang mengidentifikasi percakapan atau <i>thread</i>
11	<i>Username</i>	Nama pengguna X yang memposting <i>tweet</i>
12	<i>Tweet_URL</i>	<i>URL</i> lengkap yang dapat digunakan untuk mengakses <i>tweet</i>

Tabel 1 menyajikan atribut-atribut data yang berhasil dikumpulkan dari proses *crawling tweet*, dengan total jumlah 10,045 *tweet*. Atribut yang dicatat meliputi waktu pembuatan *tweet* (*Created_at*), *ID* unik *tweet* (*Id_Str*), isi lengkap *tweet* (*Full_Text*), jumlah kutipan (*Quote_Count*), jumlah balasan (*Reply_Count*), jumlah *retweet* (*Retweet_Count*), jumlah *favorit* (*Favorite_Count*), bahasa *tweet* (*Lang*), *ID* unik pengguna X (*User_Id_Str*), *ID* unik percakapan atau *thread* (*Conversation_id_Str*), nama pengguna X (*Username*), dan *URL* lengkap *tweet* (*Tweet_URL*).

Tabel 2. Atribut Data Yang Digunakan

No.	Atribut	Deskripsi
1	<i>Full_Text</i>	Isi lengkap dari <i>tweet</i> yang diunggah
2	<i>Label</i>	Kategori atau klasifikasi sentimen dari <i>tweet</i>

Tabel 2 merupakan daftar atribut data yang digunakan dalam penelitian ini. Atribut pertama adalah *Full_Text*, yang berisi teks lengkap dari setiap *tweet* yang diunggah. Atribut kedua adalah *Label*, yang merupakan kategori atau klasifikasi sentimen yang diberikan oleh *annotator* untuk setiap *tweet* berdasarkan analisisnya.

3.2 Praproses Data

Setelah melalui proses pengumpulan data, *dataset* memasuki tahap *preprocessing*, dimana data dioptimalkan sehingga dapat diproses dalam model.

3.2.1 Pelabelan Data

Pada tahap ini, dilakukan proses penentuan label untuk setiap *tweet* berdasarkan sentimen yang dinyatakan terhadap masing-masing calon. Distribusi data hasil pelabelan tersebut dapat dilihat dalam Tabel 3.

Tabel 3 Distribusi Data Hasil Pelabelan

	Positif Anies	Positif Prabowo	Positif Ganjar	Negatif Anies	Negatif Prabowo	Negatif Ganjar	Netral
<i>Dataset</i>	1901	888	1525	264	971	93	708

Tabel 3 menampilkan distribusi hasil pelabelan sentimen terhadap calon dalam *dataset tweet*. Jumlah *tweet* yang diklasifikasikan sebagai positif terhadap Anies sebanyak 1901, positif terhadap Prabowo sebanyak 888, dan positif terhadap Ganjar sebanyak 1525. Di sisi lain, terdapat 264 *tweet* yang diklasifikasikan sebagai negatif terhadap Anies, 971 negatif terhadap Prabowo, dan 93 negatif terhadap Ganjar. Selain itu, terdapat 708 *tweet* yang dikategorikan sebagai netral.

3.2.2 Pembersihan Data

Setelah proses pelabelan, dilakukan pembersihan data. Langkah-langkah tersebut mencakup: penghapusan entitas ganda atau duplikat, tautan, nama pengguna yang diawali dengan simbol "@" (*@username*), dan karakter numerik. Selain itu, simbol, spasi ekstra, dan tanda baca yang tidak diperlukan juga dihapus, sehingga menghasilkan kumpulan

data yang lebih bersih dan siap untuk digunakan dalam klasifikasi sentimen. Pembersihan data dapat dilihat dalam Tabel 4 dan 5.

Tabel 4. Penghapusan Data Duplikat

	Sebelum	Sesudah
Jumlah Data Duplikat	207	0

Tabel 4 menunjukkan hasil penghapusan data duplikat dalam *dataset*. Sebelum proses penghapusan, terdapat 207 data duplikat yang terdeteksi. Setelah proses penghapusan, jumlah data duplikat menjadi 0.

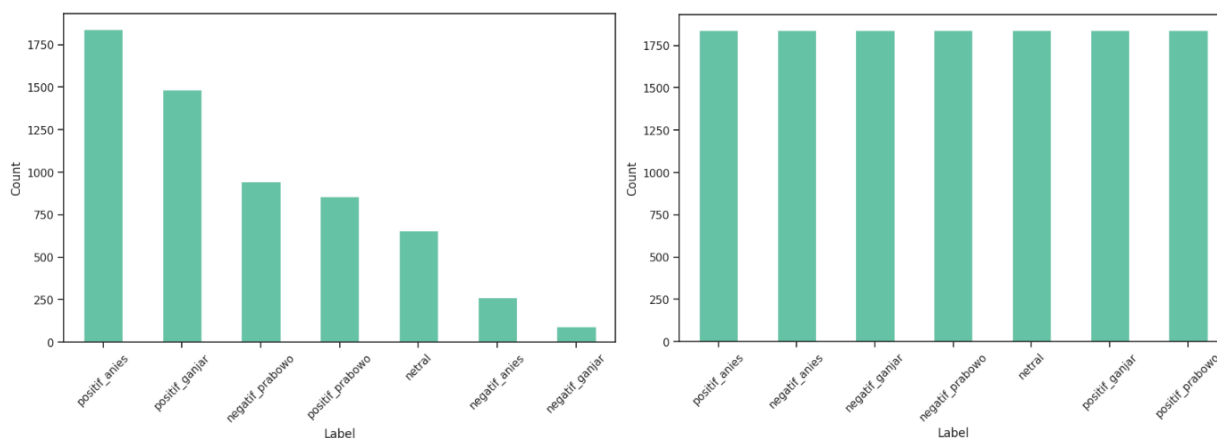
Tabel 5. Pembersihan Data

Sebelum	Sesudah
@aniesbaswedan @SHIXUNINLUV @_teww Ya Allah aku mau pa anies jadi presiden Indonesia 2024 ya Allah 🙄👊	ya allah aku mau pa anies jadi presiden indonesia 2024 ya allah

Tabel 5 menunjukkan proses pembersihan data pada teks sebelum dan sesudah diproses. Sebelum pembersihan, teks mengandung banyak elemen yang tidak relevan untuk analisis seperti *username* (@aniesbaswedan, @SHIXUNINLUV, @_teww), *emoticon*, dan variasi kapitalisasi. Setelah pembersihan, teks menjadi lebih sederhana dan konsisten dengan semua karakter dalam huruf kecil dan penghapusan elemen-elemen yang tidak diperlukan, seperti *username* dan *emoticon*. Proses pembersihan ini penting untuk meningkatkan kualitas data sehingga model dapat lebih fokus pada isi teks yang sebenarnya, yang dapat membantu meningkatkan akurasi dalam analisis selanjutnya.

3.2.3 Augmentasi Data

Dataset yang digunakan dalam penelitian ini mengalami ketidakseimbangan kelas, yang dapat mengakibatkan masalah seperti akurasi prediksi yang rendah untuk kelas-kelas yang jarang muncul [4]. Untuk mengatasi hal ini, dilakukan proses augmentasi data menggunakan augmentasi sinonim dan *back translation* untuk meningkatkan variasi data pada kelas minoritas. Dua teknik augmentasi tersebut kemudian dibandingkan dalam penelitian ini.



Gambar 2 Distribusi Data Sebelum Dan Sesudah Augmentasi

Gambar 2 menampilkan distribusi data sebelum dan sesudah augmentasi. Pada grafik sebelah kiri, distribusi data asli terlihat tidak seimbang, dengan label positif Anies memiliki jumlah tertinggi di atas 1750, diikuti oleh positif Ganjar dan negatif Prabowo yang masing-masing berjumlah sekitar 1250 dan 1000, sementara beberapa label lain memiliki jumlah yang jauh lebih sedikit. Grafik sebelah kanan menunjukkan distribusi data setelah augmentasi, dimana semua label memiliki jumlah data yang seimbang, masing-masing sekitar 1750. Augmentasi ini berhasil mengatasi ketidakseimbangan kelas dalam *dataset*, yang diharapkan dapat meningkatkan kinerja model dalam mengklasifikasikan berbagai kelas dengan lebih akurat.

3.2.4 Pembagian Data

Setelah melalui tahapan augmentasi, langkah berikutnya adalah membagi data. Untuk mengevaluasi kinerja dari model, *dataset* dibagi menjadi data latih, validasi, dan uji dengan proporsi data 80%:10%:10%. Distribusi data setelah proses pembagian ditunjukkan dalam tabel

Tabel 6. Distribusi Data Setelah Pembagian

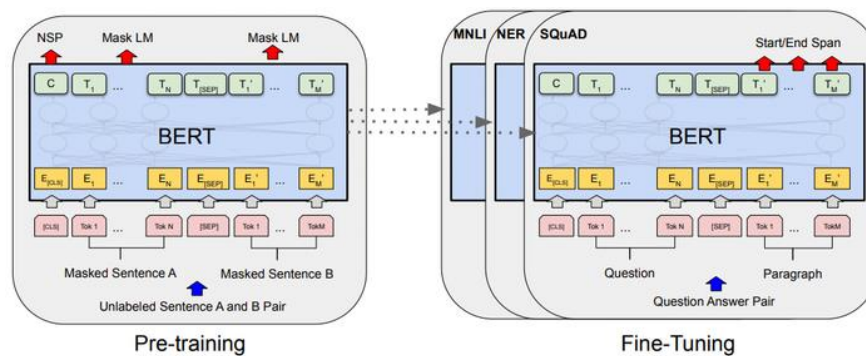
	Positif Anies	Positif Prabowo	Positif Ganjar	Negatif Anies	Negatif Prabowo	Negatif Ganjar	Netral
Data Latih	1471	1472	1471	1471	1471	1471	1471

Data Val.	184	184	184	184	184	184	184
Data Uji	184	183	184	184	184	184	184

Tabel 6 menunjukkan distribusi data setelah proses pembagian menjadi data latih, validasi, dan uji untuk evaluasi model klasifikasi sentimen terhadap calon. Data latih, yang digunakan untuk melatih model, memiliki proporsi yang seimbang antara kategori sentimen terhadap setiap calon (Anies, Prabowo, Ganjar), dengan masing-masing kategori memiliki 1471 sampel. Data validasi dan data uji masing-masing memiliki 184 sampel untuk setiap kategori sentimen, memastikan bahwa evaluasi model dilakukan secara seimbang dan akurat di setiap tahap pengembangan dan penilaian model.

3.3 Proses Klasifikasi

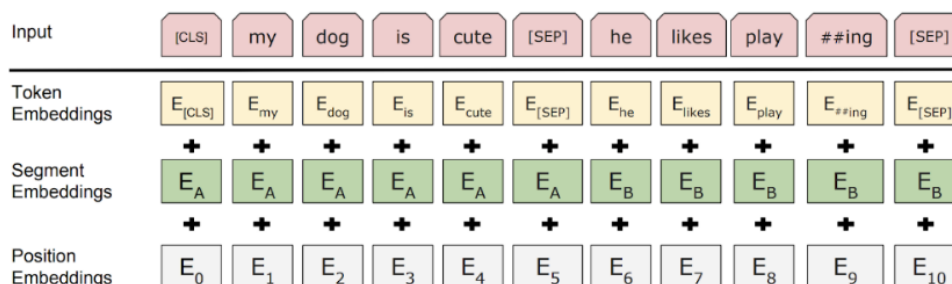
Setelah data dibagi, langkah berikutnya adalah menerapkan algoritma *IndoBERT* untuk proses pelatihan dan klasifikasi. *BERT* (*Bidirectional Encoder Representations from Transformer*) adalah model *NLP* yang melatih representasi dua arah yang mendalam dari teks tak berlabel dengan mempertimbangkan konteks kiri dan kanan di semua lapisan secara bersamaan [18]. *IndoBERT* merupakan model berbasis transformasi yang dilatih menggunakan kerangka kerja *Huggingface* dengan konfigurasi *default BERT-Base (uncased)* [17]. Lebih dari 220 juta kata digunakan dalam pelatihan *IndoBERT*, yang berasal dari tiga sumber utama: Wikipedia (74 juta kata), berita dari Kompas, Tempo, dan Liputan 6 (55 juta kata), serta situs web korpus Indonesia (90 juta kata) [17]. *IndoBERT* memiliki 12 lapisan tersembunyi dengan ukuran 768d per lapisan, 12 *attention head*, dan satu lapisan tersembunyi *feed-forward* berukuran 3.072d [17].



Gambar 3. Pre-training dan Fine-Tuning model BERT

Model *BERT* harus dilatih terlebih dahulu agar dapat digunakan dengan baik, seperti yang ditunjukkan pada Gambar 3. Dalam proses *pre-trained*, model *BERT* menggunakan dua tugas tanpa supervisi. Tugas pertama disebut *Masked LM (MLM)*, di mana model *BERT* memanfaatkan kata-kata lain dalam konteks sekitarnya untuk memprediksi kata [MASK] [18], [19]. Model *BERT* dilatih dengan menyisipkan [MASK] pada persentase *token input* secara acak [18], [19].

Setelah proses *Masked LM* selesai, dilanjutkan dengan *Next Sentence Prediction (NSP)*, di mana model mempelajari dan memprediksi kalimat berikutnya dalam dokumen sebenarnya [19]. Selama pelatihan, 50% dari *input* terdiri dari pasangan di mana kalimat kedua adalah kalimat berikutnya dari dokumen asli, sementara 50% sisanya terdiri dari kalimat acak yang dipilih sebagai kalimat kedua dari korpus [18], [19] Representasi dari *input* proses yang diberikan ke model *BERT* dan hasil dari data yang telah distandarisasi setelah proses *NSP* ditunjukkan pada Gambar 4.



Gambar 4. Representasi masukan model BERT

Model ini memanfaatkan arsitektur *pre-trained IndoBERT* untuk klasifikasi biner urutan. Model menerima teks sebagai *input*, mengolahnnya, dan kemudian memprediksi probabilitas masuk ke salah satu dari dua kelas. Model ini telah disetel dengan baik untuk mencapai akurasi klasifikasi yang tinggi selama proses pelatihan.

3.4 Pengujian Dan Evaluasi

Dalam eksperimen yang dilakukan dengan *dataset* yang telah dibuat, dilakukan pengamatan performa model dalam melakukan klasifikasi sentimen. Pada proses eksperimen, model dilatih menggunakan *hyperparameter* yang telah di *fine-tuning* dengan *epoch* 50, *learning rate* 3×10^{-7} , dan *batch size* 8. *Fine-tuning* model dilakukan karena secara signifikan meningkatkan kualitas klasifikasi model [20]. Eksperimen pertama akan melibatkan tiga kasus. Pada kasus pertama, seluruh *dataset* digunakan untuk pelatihan, sedangkan pada kasus kedua, dilakukan klasifikasi dengan augmentasi data menggunakan teknik augmentasi sinonim, kemudian pada kasus ketiga dilakukan klasifikasi dengan augmentasi data menggunakan teknik augmentasi *back translation*. augmentasi data dilakukan karena terdapat perbedaan jumlah sampel pada setiap kelas. Eksperimen pertama ini dilakukan augmentasi tanpa mempertimbangkan keunikan data. Berikut adalah distribusi data pada setiap kelas sebelum dan sesudah dilakukan augmentasi data menggunakan augmentasi sinonim dan *back translation* dapat dilihat dalam Tabel 7.

Tabel 7. Distribusi Data Sebelum Dan Sesudah Augmentasi.

	Positif Anies	Positif Prabowo	Positif Ganjar	Negatif Anies	Negatif Prabowo	Negatif Ganjar	Netral
Dataset Awal	1839	860	1486	264	944	92	658
Dataset Akhir	1839	1839	1839	1839	1839	1839	1839

Tabel 7 menunjukkan perbandingan distribusi data sebelum dan setelah proses augmentasi. *Dataset* awal memiliki jumlah *tweet* yang bervariasi antara kategori sentimen terhadap calon Anies, Prabowo, dan Ganjar, dengan jumlah yang lebih sedikit untuk beberapa kategori dibandingkan dengan kategori lainnya. Setelah augmentasi, jumlah *tweet* untuk setiap kategori sentimen menjadi 1839 untuk semua kategori, menciptakan keseragaman dalam representasi data untuk analisis dan pelatihan model.

Pada eksperimen kedua, augmentasi dilakukan dengan mempertimbangkan keunikan data. Eksperimen ini melibatkan dua kasus. Pada kasus pertama, dilakukan klasifikasi dengan augmentasi sinonim pada *dataset*. Pada kasus kedua, dilakukan klasifikasi dengan augmentasi *back translation*. Distribusi data pada setiap kelas dapat dilihat dalam Tabel 8.

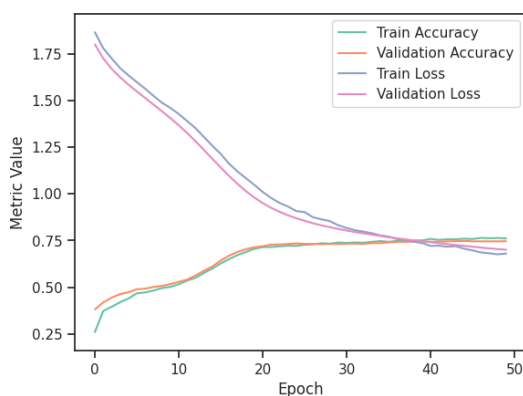
Tabel 8. Distribusi Data Sebelum Dan Sesudah Augmentasi Dengan Mempertimbangkan Keunikan Data

	Positif Anies	Positif Prabowo	Positif Ganjar	Negatif Anies	Negatif Prabowo	Negatif Ganjar	Netral
Dataset Awal	1839	860	1486	264	944	92	658
Sinonim Augmentation	1839	1795	1835	1618	1811	1412	1734
Back Translation Aug.	1839	1434	1801	526	1511	184	1202

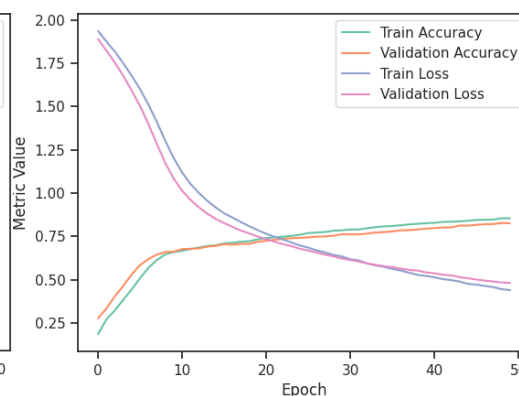
Tabel 8 memberikan gambaran distribusi data sebelum dan setelah augmentasi dengan mempertimbangkan keunikan data dalam konteks penelitian ini. *Dataset* awal menunjukkan jumlah *tweet* untuk masing-masing kategori sentimen terhadap calon Anies, Prabowo, dan Ganjar. Setelah dilakukan augmentasi dengan dua metode yang berbeda, yaitu Sinonim Augmentation dan Back Translation Augmentation, terlihat peningkatan jumlah data dalam beberapa kategori sentimen.

3.4.1 Kurva ROC-AUC

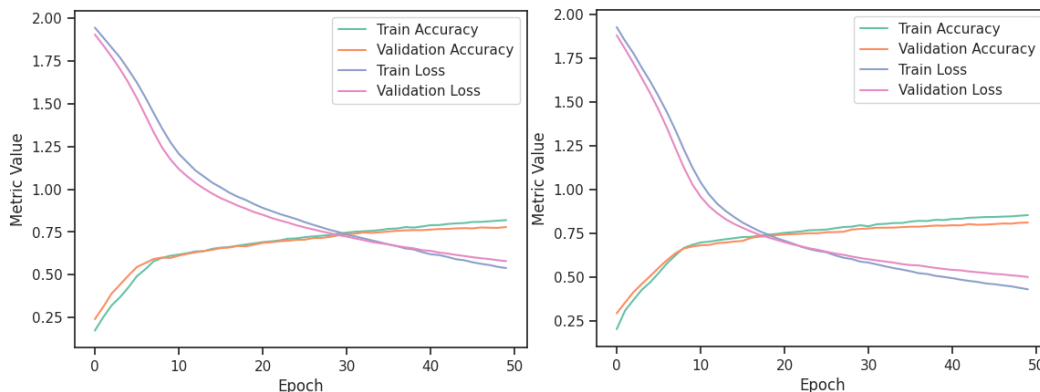
Pada tahap ini, evaluasi dilakukan menggunakan kurva *ROC-AUC* untuk mengukur kinerja model dalam memprediksi probabilitas kelas pada data uji. Kurva *ROC-AUC* menunjukkan seberapa efektif model dalam membedakan antara kelas positif dan negatif, dengan area di bawah kurva (*AUC*) mencerminkan tingkat akurasi prediksi model tersebut.



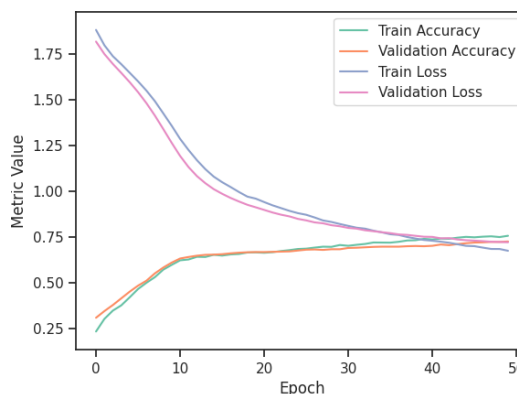
Gambar 5. IndoBert Tanpa Augmentasi



Gambar 6. IndoBert Dengan Augmentasi Sinonim



Gambar 7. *IndoBert* Dengan Augmentasi *Back Translation* **Gambar 8.** Augmentasi Sinonim Dengan *Hash* (Data Unik)

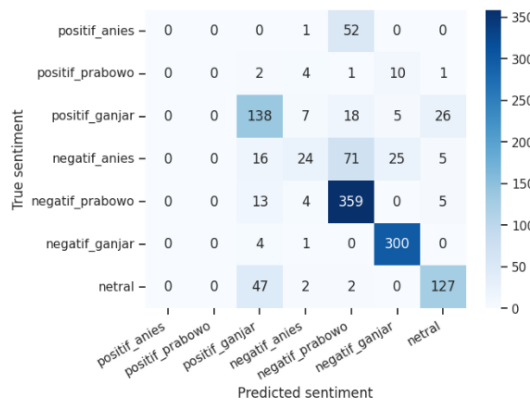


Gambar 9. Augmentasi *Back Translation* Dengan *Hash* (Data Unik)

Kurva dalam Gambar 5,6,7,8, dan 9 menunjukkan performa model *IndoBert* pada berbagai skenario augmentasi data. Pada Gambar 5, Akurasi pelatihan dan validasi meningkat, sementara *loss* menurun secara konsisten, menunjukkan model belajar dengan baik. Tidak ada tanda *overfitting* yang signifikan, karena kurva akurasi dan *loss* pelatihan serta validasi mendekati satu sama lain. Gambar 6 memperlihatkan klasifikasi *IndoBert* dengan augmentasi sinonim. Akurasi pelatihan dan validasi dengan teknik ini meningkat, tetapi mulai ada perbedaan setelah sekitar 10 *epoch*, menunjukkan model mungkin mulai mengalami *overfitting* ringan. *Loss* validasi menurun lebih lambat dibandingkan pelatihan, dan ada sedikit perbedaan antara akurasi pelatihan dan validasi, menandakan *overfitting* ringan. Gambar 7 dengan augmentasi *back translation* menunjukkan peningkatan yang signifikan dalam akurasi validasi dan penurunan *loss* secara konsisten, menunjukkan bahwa model belajar dengan baik tanpa tanda *overfitting* yang signifikan. Gambar 8, menggunakan kombinasi sinonim dan *hash* (data unik). Dengan teknik ini, akurasi pelatihan dan validasi meningkat, tetapi mulai *divergen* setelah sekitar 20 *epoch*, dan *loss* validasi menurun lebih lambat dibandingkan pelatihan. Ini menunjukkan awal dari *overfitting*, Gambar 9 dengan kombinasi *back translation* dan *hash* menunjukkan akurasi pelatihan dan validasi meningkat, dan *loss* menurun secara konsisten, serupa dengan Gambar 7. Namun, perbedaan antara akurasi pelatihan dan validasi sedikit lebih besar, menunjukkan sedikit *overfitting*.

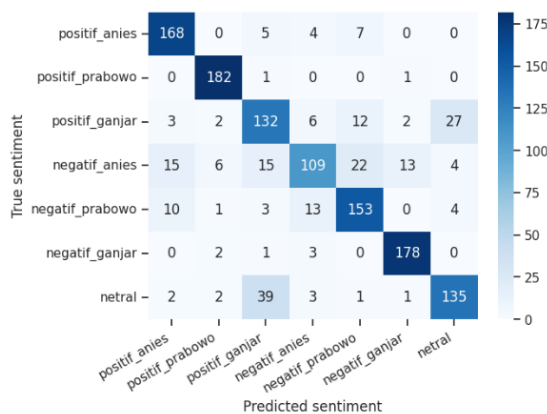
3.4.2 Confusion Matrix

Pada tahap ini, evaluasi kinerja model klasifikasi dilakukan menggunakan *confusion matrix*. *Confusion matrix* merupakan tabel yang berisi empat kategori hasil klasifikasi: *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. Tabel ini memberikan gambaran mengenai berbagai kemungkinan hasil dari proses klasifikasi.



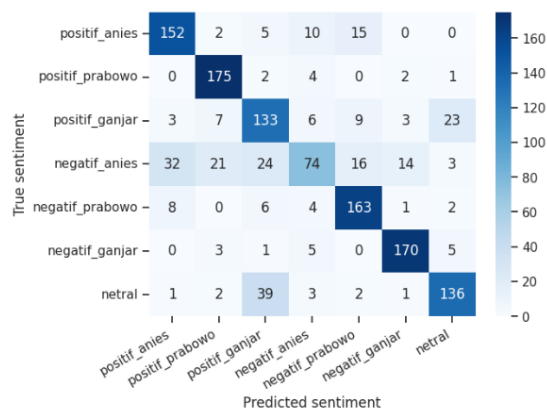
Gambar 10. IndoBert Tanpa Augmentasi

Pada Gambar 10, ditampilkan jumlah hasil prediksi model terhadap sentimen setiap calon. Warna biru yang lebih gelap menunjukkan jumlah yang lebih tinggi dari prediksi di kategori tersebut. Dapat dilihat bahwa model cenderung lebih akurat dalam mengklasifikasikan sentimen negatif Prabowo dengan 359 prediksi yang benar dan sentimen negatif Ganjar dengan 300 prediksi yang benar. Namun, model mengalami kesulitan dalam mengklasifikasikan kategori lainnya, terutama untuk sentimen negatif Anies, positif Prabowo dan Positif Anies, di mana terjadi banyak kesalahan prediksi. Hal ini terlihat dari banyaknya angka di luar diagonal utama, menunjukkan bahwa model sering salah memprediksi kategori-kategori tersebut.



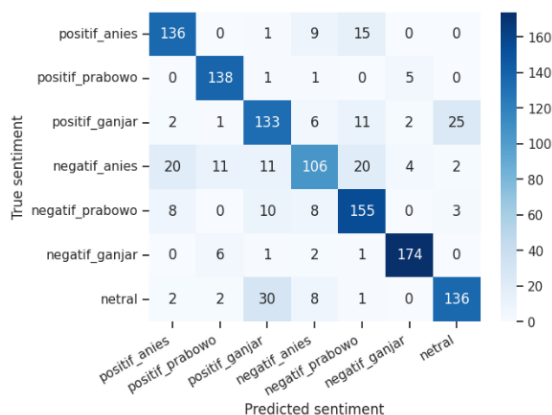
Gambar 11. IndoBert Dengan Augmentasi Sinonim

Pada Gambar 11 terlihat bahwa model mampu melakukan klasifikasi dengan baik pada setiap kelas, hal ini dapat dilihat dari warna biru yang gelap secara diagonal. Data yang berada di luar diagonal merupakan data prediksi yang salah. Akurasi tertinggi terlihat pada kategori positif Prabowo dan negatif Ganjar dengan masing-masing 182 dan 178 instance yang diklasifikasikan dengan benar, menunjukkan kinerja yang kuat dalam kategori sentimen ini. Positif Anies dan netral juga menunjukkan jumlah klasifikasi yang benar yang tinggi, masing-masing 168 dan 135. Namun, model mengalami kesulitan dengan klasifikasi negatif Anies, positif Ganjar dan negatif Prabowo, menunjukkan banyak salah klasifikasi di berbagai kategori lainnya. Sentimen positif Ganjar menunjukkan akurasi moderat dengan salah klasifikasi yang signifikan ke dalam kategori netral dan negatif Ganjar.



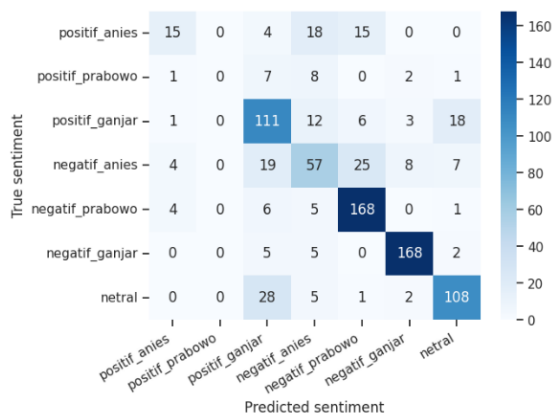
Gambar 12. IndoBert Dengan Augmentasi Back Translation

Confusion matrix pada Gambar 12 menunjukkan bahwa model *IndoBert* dengan augmentasi *back translation* memiliki kinerja yang cukup baik di beberapa kategori. Kategori positif Prabowo dan negatif Ganjar memiliki jumlah klasifikasi benar tertinggi dengan masing-masing 175 dan 170 *instance* yang diklasifikasikan dengan benar. Positif Anies dan netral juga menunjukkan performa yang baik dengan masing-masing 152 dan 136 *instance* yang diklasifikasikan dengan benar. Namun, terdapat kesulitan dalam mengklasifikasikan negatif Anies dan positif Ganjar, yang terlihat dari tingginya jumlah salah klasifikasi pada kategori-kategori lain. Salah klasifikasi yang signifikan juga terlihat antara positif Ganjar dan kategori netral serta negatif Ganjar.



Gambar 13. Augmentasi Sinonim Dengan Hash (Data Unik)

Confusion matrix pada Gambar 13 menunjukkan bahwa model memiliki kinerja yang cukup baik di beberapa kategori. Kategori negatif Ganjar dan negatif Prabowo memiliki jumlah klasifikasi benar tertinggi dengan masing-masing 174 dan 155 *instance* yang diklasifikasikan dengan benar. Namun model mengalami kesulitan dalam mengklasifikasikan negatif Anies dan positif Ganjar, yang terlihat dari tingginya jumlah salah klasifikasi pada kategori-kategori lain.



Gambar 14. Augmentasi Back Translation Dengan Hash (Data Unik)

Confusion matrix pada Gambar 14 menunjukkan ada beberapa kesalahan prediksi yang cukup signifikan, seperti negatif Anies yang sering diprediksi sebagai negatif Prabowo (25 kali) dan positif Anies yang diprediksi sebagai negatif Anies (18 kali), negatif Prabowo (15 kali), dan pada positif Prabowo tidak ada yang diprediksi dengan benar. Kesalahan ini menunjukkan adanya kebingungan model dalam membedakan beberapa kategori sentimen, terutama antara sentimen positif dan negatif untuk kandidat tertentu. Secara keseluruhan, meskipun model memiliki banyak prediksi yang benar, terdapat banyak kesalahan yang perlu diperbaiki untuk meningkatkan akurasi.

3.4.3 Classification Report

Selanjutnya evaluasi menggunakan *classification report*. *Classification report* memberikan gambaran komprehensif mengenai kinerja model dalam menangani berbagai kelas dalam *dataset*.

Tabel 9. Classification Report

	Precision	recall	F1-score	Support	Accuracy
Tanpa Augmentasi	0.69	0.75	0.70	1270	0.75
Augmentasi Sinonim	0.82	0.82	0.82	1287	0.82
Augmentasi Back Translation	0.78	0.78	0.77	1288	0.78
Augmentasi Sinonim Dengan Data Unik	0.81	0.81	0.81	1207	0.81

Augmentasi <i>Back Translation</i> Dengan Data Unik	0.72	0.74	0.72	850	0.74
---	------	------	------	-----	------

Hasil *classification report* pada Tabel 9 menunjukkan hasil evaluasi metode klasifikasi *IndoBert* dengan beberapa teknik augmentasi data pada model klasifikasi. Tanpa augmentasi, model mencapai presisi 0.69, *recall* 0.75, dan *F1-score* 0.70, dengan akurasi keseluruhan sebesar 0.75, menunjukkan performa dasar dari model tanpa bantuan teknik tambahan. Penggunaan augmentasi sinonim secara signifikan meningkatkan kinerja model, dengan presisi, *recall*, dan *F1-score* masing-masing mencapai 0.82, serta akurasi keseluruhan juga meningkat menjadi 0.82. Hal ini menunjukkan bahwa augmentasi sinonim efektif dalam menambah variasi data tanpa mengorbankan kualitas klasifikasi. Augmentasi dengan *back translation* juga meningkatkan kinerja model, meskipun tidak sebesar augmentasi sinonim, dengan presisi dan *recall* 0.78 serta *F1-score* 0.77, dan akurasi 0.78. Kombinasi augmentasi sinonim dengan data unik menunjukkan hasil yang sangat baik, dengan semua metrik mencapai 0.81, dan akurasi yang sama, memperlihatkan bahwa penambahan data unik dapat mempertahankan atau sedikit meningkatkan kinerja model. Di sisi lain, kombinasi augmentasi *back translation* dengan data unik menghasilkan presisi 0.72, *recall* 0.74, *F1-score* 0.72, dan akurasi 0.74, menunjukkan penurunan kinerja dibandingkan dengan augmentasi sinonim, namun tetap lebih baik dibandingkan tanpa augmentasi.

4. KESIMPULAN

Hasil penelitian menunjukkan bahwa penggunaan augmentasi data dapat meningkatkan performa model secara signifikan. Secara khusus, penggunaan augmentasi sinonim pada *dataset* ini terbukti lebih efektif dibandingkan dengan augmentasi *back translation*. Hal ini dapat dilihat dari hasil eksperimen yang dilakukan. Pada eksperimen pertama, augmentasi sinonim menghasilkan akurasi, presisi, *recall*, dan *F1-score* sebesar 0.82. Sedangkan pada eksperimen kedua, nilai akurasi yaitu 0.81, dengan presisi, *recall*, dan *F1-score* masing-masing sebesar 0.81. Di sisi lain, *back translation* hanya mencapai akurasi 0.78 pada eksperimen pertama, dengan nilai presisi, *recall*, dan *F1-score* berturut-turut sebesar 0.78 dan 0.77. Pada eksperimen kedua, tercatat penurunan akurasi menjadi 0.74, dengan nilai presisi, *recall*, dan *F1-score* sebesar 0.72. Selain itu, dalam penelitian ini augmentasi sinonim menghasilkan lebih banyak varian teks dibandingkan dengan *back translation* pada *dataset* yang digunakan. Augmentasi sinonim menghasilkan 5901 data baru, sedangkan augmentasi *back translation* hanya menghasilkan 2354 data baru. Hal ini menunjukkan bahwa augmentasi sinonim tidak hanya meningkatkan akurasi, tetapi juga memperkaya variasi data, yang dapat membantu model untuk melakukan *train* lebih baik dari data yang lebih beragam. Dengan demikian, augmentasi sinonim direkomendasikan sebagai teknik augmentasi yang lebih unggul dengan *dataset* yang digunakan dalam penelitian ini. Perbedaan hasil antara penelitian ini dan penelitian sebelumnya yang menunjukkan bahwa augmentasi *back translation* lebih baik dari augmentasi sinonim dapat disebabkan oleh perbedaan proporsi antara kelas mayoritas dan minoritas dalam *dataset* yang digunakan. Dalam kasus ini, augmentasi sinonim lebih efektif pada *dataset* dengan perbandingan kelas mayoritas dan minoritas yang terpaut jauh.

REFERENCES

- [1] M. Mujahid *et al.*, "Sentiment analysis and topic modeling on tweets about online education during covid-19," *Applied Sciences (Switzerland)*, vol. 11, no. 18, Sep. 2021, doi: 10.3390/app11188438.
- [2] A. F. Hidayatullah, S. Cahyaningtyas, and A. M. Hakim, "Sentiment Analysis on Twitter using Neural Network: Indonesian Presidential Election 2019 Dataset," *IOP Conf Ser Mater Sci Eng*, vol. 1077, no. 1, p. 012001, Feb. 2021, doi: 10.1088/1757-899x/1077/1/012001.
- [3] G. A. BUNTORO, R. ARIFIN, G. N. SYAIFUDDIIN, A. SELAMAT, O. KREJCAR, and H. FUJITA, "Implementation of a Machine Learning Algorithm for Sentiment Analysis of Indonesia's 2019 Presidential Election," *IJUM Engineering Journal*, vol. 22, no. 1, pp. 78–92, 2021, doi: 10.31436/IJUM.EJ.V22I1.1532.
- [4] S. M. Isa, G. Nico, and M. Permana, "INDOBERT FOR INDONESIAN FAKE NEWS DETECTION," *ICIC Express Letters*, vol. 16, no. 3, pp. 289–297, Mar. 2022, doi: 10.24507/icicel.16.03.289.
- [5] A. Holzinger, P. Kieseberg, A. M. Tjoa, and E. Weippl, Eds., *Machine Learning and Knowledge Extraction*, vol. 12279. in Lecture Notes in Computer Science, vol. 12279. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-57321-8.
- [6] A. N. Azizah, M. Falach Asy'ari, I. Wisma, D. Prastya, and D. Purwitasari, "EASY DATA AUGMENTATION UNTUK DATA YANG IMBALANCE PADA KONSULTASI KESEHATAN DARING," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 10, no. 5, pp. 1095–1104, 2023, doi: 10.25126/jtiik.2023107082.
- [7] I. Athiyyah Rahma and L. Hulliyyatus Suadaa, "PENERAPAN TEXT AUGMENTATION UNTUK MENGATASI DATA YANG TIDAK SEIMBANG PADA KLASIFIKASI TEKS BERBAHASA INDONESIA STUDI KASUS: DETEKSI JUDUL CLICKBAIT DAN KOMENTAR HATE SPEECH PADA BERITA ONLINE," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 2023, doi: 10.25126/jtiik.2023107325.
- [8] A. Bello, S. C. Ng, and M. F. Leung, "A BERT Framework to Sentiment Analysis of Tweets," *Sensors*, vol. 23, no. 1, Jan. 2023, doi: 10.3390/s23010506.
- [9] A. Roy and M. Ojha, "Twitter sentiment analysis using deep learning models," in *2020 IEEE 17th India Council International Conference, INDICON 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020. doi: 10.1109/INDICON49873.2020.9342279.



- [10] A. Singh, A. Kumar, N. Dua, V. K. Mishra, D. Singh, and A. Agrawal, “Predicting Elections Results using Social Media Activity A Case Study: USA Presidential Election 2020,” in *2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021*, Institute of Electrical and Electronics Engineers Inc., Mar. 2021, pp. 314–319. doi: 10.1109/ICACCS51430.2021.9441835.
- [11] A. J. Nair, G. Veena, and A. Vinayak, “Comparative study of Twitter Sentiment on COVID - 19 Tweets,” in *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, Institute of Electrical and Electronics Engineers Inc., Apr. 2021, pp. 1773–1778. doi: 10.1109/ICCMC51019.2021.9418320.
- [12] A. B. Y. A. Putra, Y. Sibaroni, and A. F. Ihsan, “Disinformation Detection on 2024 Indonesia Presidential Election using IndoBERT,” in *2023 International Conference on Data Science and Its Applications, ICoDSA 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 350–355. doi: 10.1109/ICoDSA58501.2023.10277572.
- [13] L. Geni, E. Yulianti, and D. I. Sensuse, “Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using IndoBERT Language Models,” *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 3, pp. 746–757, 2023, doi: 10.26555/jiteki.v9i3.26490.
- [14] T. M. Fagbola and S. C. Thakur, “Lexicon-based Bot-aware Public Emotion Mining and Sentiment Analysis of the Nigerian 2019 Presidential Election on Twitter,” 2019. [Online]. Available: www.ijacsa.thesai.org
- [15] D. Fimoza, A. Amalia, and T. Henny Febriana Harumy, “Sentiment Analysis for Movie Review in Bahasa Indonesia Using BERT,” in *2021 International Conference on Data Science, Artificial Intelligence, and Business Analytics, DATABIA 2021 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 27–34. doi: 10.1109/DATABIA53375.2021.9650096.
- [16] M. Bucos and B. Drăgulescu, “Enhancing Fake News Detection in Romanian Using Transformer-Based Back Translation Augmentation,” *Applied Sciences (Switzerland)*, vol. 13, no. 24, Dec. 2023, doi: 10.3390/app132413207.
- [17] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” Nov. 2020, [Online]. Available: <http://arxiv.org/abs/2011.00677>
- [18] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” May 2019, [Online]. Available: <http://arxiv.org/abs/1810.04805>
- [19] I. R. Hidayat and W. Maharani, “General Depression Detection Analysis Using IndoBERT Method,” *International Journal on Information and Communication Technology (IJoICT)*, vol. 8, no. 1, pp. 41–51, Aug. 2022, doi: 10.21108/ijoiict.v8i1.634.
- [20] G. A. Pradnyana, W. Anggraeni, E. M. Yuniarno, and M. H. Purnomo, “Fine-Tuning IndoBERT Model for Big Five Personality Prediction from Indonesian Social Media,” in *2023 International Seminar on Intelligent Technology and Its Applications: Leveraging Intelligent Systems to Achieve Sustainable Development Goals, ISITIA 2023 - Proceeding*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 93–98. doi: 10.1109/ISITIA59021.2023.10221074.