

Analisis Sentimen Masyarakat terhadap Kereta Cepat Whoosh pada Platform X menggunakan IndoBERT

Gibran Hakim¹, Tirana Noor Fatyanosa², Agus Wahyu Widodo³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹gibranhakim@student.ub.ac.id, ²fatyanosa@ub.ac.id, ³a_wahyu_w@ub.ac.id

Abstrak

Infrastruktur kereta api cepat di Indonesia relatif tertinggal dibandingkan negara lain. Perkembangannya baru dimulai dengan proyek kereta cepat Jakarta-Bandung (Whoosh) yang baru dimulai pada 2015 dan selesai pada 2023. Proyek kerja sama Indonesia-China ini sangat menarik perhatian publik, terutama di media sosial seperti X. Penelitian ini bertujuan mengklasifikasikan sentimen publik sebagai positif, negatif, atau netral menggunakan pemrosesan bahasa alami (NLP). IndoBERT sebagai model NLP yang dilatih khusus untuk bahasa Indonesia, digunakan untuk menganalisis sentimen secara efektif. Penelitian ini melakukan pengujian pada dataset Whoosh hasil scraping pada platform X dengan konfigurasi hyperparameter seperti jumlah epoch, learning rate, dan batch size. Performa terbaik dicapai dengan konfigurasi 3 epoch, learning rate 2e-5, dan batch size 32 yang menghasilkan metrik evaluasi accuracy, recall, precision, dan f1-score sebesar 0,78. Hasil ini menunjukkan learning rate yang lebih kecil memberikan pembelajaran yang stabil, sementara batch size yang lebih besar memberikan estimasi gradien yang konsisten. Namun, model kesulitan mengklasifikasikan sentimen netral sehingga sering salah mengklasifikasikannya sebagai positif atau negatif. Hasil ini juga menunjukkan adanya overfitting, di mana model tampil baik pada data pelatihan tetapi menurun pada data pengujian yang menunjukkan fokus berlebihan pada detail dan noise dari data pelatihan. Penelitian ini menyoroti pentingnya pemilihan model dan konfigurasi hyperparameter dalam analisis sentimen.

Kata kunci: Whoosh, X, analisis sentimen, IndoBERT

Abstract

Indonesia's high-speed rail infrastructure is relatively lagging behind other countries. Its development only started with the Jakarta-Bandung high-speed rail project (Whoosh), which started in 2015 and will be completed in 2023. This Indonesia-China cooperation project has attracted a lot of public attention, especially on social media such as X. This research aims to classify public sentiment as positive, negative or neutral using natural language processing (NLP). IndoBERT as an NLP model specially trained for the Indonesian language, is used to effectively analyze sentiment. This research conducted tests on Whoosh dataset scraped on X platform with hyperparameter configurations such as number of epochs, learning rate, and batch size. The best performance was achieved with a configuration of 3 epochs, learning rate 2e-5, and batch size 32 which resulted in an accuracy, recall, precision, and f1-score evaluation metric of 0.78. These results show that a smaller learning rate provides stable learning, while a larger batch size provides consistent gradient estimation. However, the model struggles to classify neutral sentiments, often misclassifying them as positive or negative. The results also indicate overfitting, where the model performed well on the training data but degraded on the test data indicating an excessive focus on the details and noise of the training data. This research highlights the importance of model selection and hyperparameter configuration in sentiment analysis.

Keywords: Whoosh, X, sentiment analysis, IndoBERT

1. PENDAHULUAN

Perkembangan infrastruktur kereta api di Indonesia, khususnya kereta cepat relatif tertinggal dibandingkan negara lain. Jepang

meluncurkan Shinkansen pada 1964, sementara Indonesia baru memulai pada 2015, yang selesai pada 2023. Kereta ini menghubungkan Jakarta dan Bandung yang kini disebut sebagai "Whoosh" (Waktu Hemat, Operasi Optimal,

Sistem Hebat). Proyek ini merupakan kerja sama Indonesia-China melalui konsorsium PT Kereta Cepat Indonesia China (Wijaya, 2020) dan kini menjadi proyek infrastruktur strategis yang menarik perhatian publik.

Sebagai kereta cepat pertama di Asia Tenggara, Whoosh menghubungkan Jakarta dan Bandung dalam sekitar 40 menit (PT KCIC, 2023) jauh lebih cepat dibandingkan kereta Argo Parahyangan yang memakan waktu 3 jam. Proyek ini merupakan lompatan besar dalam transportasi Indonesia. Whoosh tidak hanya mempercepat perjalanan, tetapi juga mendorong pertumbuhan ekonomi, pariwisata, dan pertukaran sosial-budaya antara kedua kota. Selain itu, proyek ini diharapkan meningkatkan kualitas infrastruktur dan mendorong inovasi di sektor transportasi.

Sebagai proyek dengan investasi besar dan kerja sama internasional, Whoosh memicu berbagai opini di masyarakat. Beberapa pihak melihat potensi peningkatan efisiensi dan pertumbuhan ekonomi, sementara yang lain mengkhawatirkan biaya, penggusuran, dan keterjangkauan tiket. Dinamika opini ini berkembang sejak operasi komersial dimulai pada Oktober 2023. Penelitian ini menggunakan analisis sentimen untuk mengidentifikasi dan mengklasifikasikan opini masyarakat sebagai positif, negatif, atau netral (Hoang et al., 2019).

Di era digital, media sosial menjadi saluran utama bagi masyarakat untuk menyampaikan pandangan, termasuk tentang kereta cepat Whoosh. Platform X (sebelumnya Twitter) sangat populer di Indonesia, dengan lebih dari 24 juta pengguna aktif (DataReportal, 2023), menjadikannya barometer efektif untuk opini publik. X memfasilitasi diskusi real-time dan penyebaran informasi cepat, mempengaruhi wacana publik dan opini masyarakat tentang isu-isu besar seperti Whoosh. Volume percakapan yang besar di X membuka peluang untuk analisis sentimen komprehensif, memberikan wawasan tentang persepsi publik, tren opini, dan isu-isu penting, yang bermanfaat untuk evaluasi kebijakan di masa depan.

Kompleksitas dan volume data dari percakapan media sosial membuat pendekatan konvensional tidak praktis dan kurang akurat. Di sinilah kecerdasan buatan, khususnya pemrosesan bahasa alami, menjadi penting dalam analisis sentimen. Kemampuan menganalisis data skala besar dengan cepat dan konsisten membuka peluang memahami sentimen publik secara lebih komprehensif.

IndoBERT, model yang dilatih untuk bahasa Indonesia, menawarkan potensi besar dalam analisis sentimen lokal. Dikembangkan oleh IndoNLP (Wilie et al., 2020), IndoBERT telah menunjukkan performa unggul dalam berbagai tugas bahasa Indonesia, termasuk analisis sentimen, menjadikannya alat potensial untuk memahami opini publik terhadap isu-isu besar.

Penelitian ini bertujuan untuk membandingkan performa model *pre-trained* IndoBERT dalam analisis sentimen opini masyarakat terhadap kereta cepat Whoosh berdasarkan data *tweet* pada X. Selain itu, penelitian ini juga bertujuan untuk memahami pengaruh konfigurasi *hyperparameter* terhadap performa model tersebut. Dengan mengeksplorasi kedua aspek ini, penelitian diharapkan dapat memberikan wawasan yang lebih mendalam tentang efektivitas IndoBERT dalam menangani analisis sentimen di media sosial, khususnya dalam konteks isu-isu besar seperti kereta cepat Whoosh.

2. LANDASAN KEPUSTAKAAN

2.1. Whoosh

Kereta Cepat Jakarta-Bandung, dikenal sebagai "Whoosh", adalah proyek kereta api berkecepatan tinggi yang menghubungkan Jakarta dan Bandung. Proyek ini merupakan kolaborasi Indonesia dan China melalui konsorsium PT Kereta Cepat Indonesia China (Wijaya, 2020). Sebagai kereta cepat pertama di Asia Tenggara, Whoosh dapat mengurangi waktu tempuh dari 3 jam menjadi 40 menit. Kereta ini dirancang untuk mengatasi kepadatan transportasi yang selama ini bergantung pada jalan tol (PT KCIC, 2023), serta diharapkan dapat mendorong pertumbuhan ekonomi, pariwisata, dan sosial-budaya antara kedua kota.

2.2. X

X, sebelumnya dikenal sebagai Twitter, adalah platform mikroblog dengan lebih dari 370 juta pengguna global. Pengguna dapat menulis *tweet* hingga 280 karakter dan mengikuti pengguna lain. Di Indonesia, terdapat 24 juta pengguna X, menjadikannya negara pengguna terbesar keempat setelah AS, Jepang, dan India (DataReportal, 2023). Platform ini berperan penting dalam diskusi publik dan mempengaruhi pandangan masyarakat tentang berbagai topik, termasuk kebijakan pemerintah dan proyek besar seperti kereta cepat Whoosh.

2.3. Analisis Sentimen

Analisis sentimen adalah metode untuk mengekstraksi opini atau sikap dari teks yang menggunakan teknik pemrosesan bahasa alami (NLP) untuk mengidentifikasi polaritas sebagai positif, negatif, atau netral (Alsaedi & Khan, 2019). Metode ini membantu mengidentifikasi emosi, opini, sikap, atau penilaian tentang produk, individu, atau isu-isu dari berbagai sumber seperti ulasan, komentar, atau unggahan pada media sosial.

2.4. Preprocessing

Preprocessing adalah langkah penting dalam analisis teks untuk mengubah teks mentah menjadi format yang dapat dipahami serta siap untuk diolah lebih lanjut sesuai kebutuhan. Proses ini mencakup beberapa teknik seperti *cleaning*, *case folding*, dan normalisasi. Memilih kombinasi tahapan *preprocessing* yang tepat dapat meningkatkan akurasi klasifikasi teks secara signifikan tergantung pada domain & bahasa yang dipelajari (Uysal & Gunal, 2014).

2.5. Pemrosesan Bahasa Alami

Pemrosesan Bahasa Alami (NLP) adalah cabang ilmu komputer dan linguistik yang mengkaji interaksi antara komputer dan bahasa manusia. NLP telah berkembang dari representasi berbasis aturan dan statistik menjadi representasi yang peka terhadap konteks, sehingga meningkatkan pemahaman sintaksis, semantik, dan sentimen bahasa (Patil et al., 2023). Kemajuan ini telah menghasilkan aplikasi nyata seperti sistem dialog, penerjemahan ucapan, dan analisis sentimen di berbagai industri.

2.6. IndoBERT

IndoBERT adalah model *pre-trained* khusus untuk bahasa Indonesia yang berbasis arsitektur BERT (Devlin et al., 2019). Model ini adalah salah satu model BERT monolingual pertama untuk bahasa Indonesia (Koto et al., 2020) yang dirancang untuk meningkatkan performa tugas NLP dalam konteks bahasa Indonesia.

IndoBERT menggunakan *dataset* Indo4B, yang mencakup lebih dari 23 GB dan terdiri dari 4 miliar kata data teks bahasa Indonesia dari berbagai sumber, termasuk media sosial, blog, dan berita (Wilie et al., 2020). Keberagaman sumber data ini memungkinkan IndoBERT

menangkap nuansa bahasa dan konteks yang lebih kaya, menjadikannya alat efektif untuk tugas NLP seperti analisis sentimen dalam bahasa Indonesia.

2.7. Fine-tuning

Fine-tuning model yang sudah dilatih sebelumnya terbukti efektif untuk tugas spesifik, terutama dalam NLP (Fu et al., 2023). Proses ini menyesuaikan model *pre-trained* agar optimal pada *dataset* tertentu dengan melatih kembali beberapa lapisan akhir, sementara lapisan awal tetap untuk mempertahankan fitur umum bahasa. Hal ini memungkinkan model mempelajari pola khusus tanpa kehilangan pengetahuan umum selama proses *pre-training* (Fu et al., 2023).

2.8. Confusion Matrix

Confusion matrix adalah metode evaluasi performa model klasifikasi dalam *machine learning*. Metode ini menyajikan informasi tentang klasifikasi aktual dan prediksi yang dilakukan oleh sistem. Kinerja dievaluasi menggunakan data dalam matriks (Ting, 2017), yang memvisualisasikan perbandingan antara label sebenarnya dan prediksi model, seperti yang ditunjukkan pada Gambar 1.

		Actual Values	
		Positive	Negative
Predicted Values	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Gambar 1. *Confusion Matrix*

Struktur dari *confusion matrix* umumnya sebagai berikut.

- *True Positive* (TP): Klasifikasi dengan hasil positif pada data asli yang positif.
- *True Negative* (TN): Klasifikasi hasil negatif pada data asli yang negatif.
- *False Positive* (FP): Klasifikasi dengan hasil positif pada data asli yang negatif.
- *False Negative* (FN): Klasifikasi dengan hasil negatif pada data asli yang positif.

Hasil perbandingan dari kelas prediksi yang dihasilkan oleh algoritma dibandingkan dengan data aktual tersebut kemudian digunakan untuk menghitung nilai metrik evaluasi yaitu *accuracy*, *precision*, *recall*, dan *f1-score* untuk mengukur kinerja model klasifikasi.

Accuracy pada Persamaan (2.1) adalah metrik yang mengukur proporsi dari total jumlah prediksi yang benar terhadap keseluruhan data yang diprediksi.

$$\text{accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (2.1)$$

Precision pada Persamaan (2.2) adalah metrik yang menunjukkan proporsi hasil positif yang diprediksi oleh model yang benar-benar positif.

$$\text{precision} = \frac{TP}{TP+FP} \quad (2.2)$$

Recall pada Persamaan (2.3) adalah metrik yang mengukur proporsi dari semua kasus positif aktual yang berhasil diidentifikasi dengan benar oleh model sebagai positif.

$$\text{recall} = \frac{TP}{TP+FN} \quad (2.3)$$

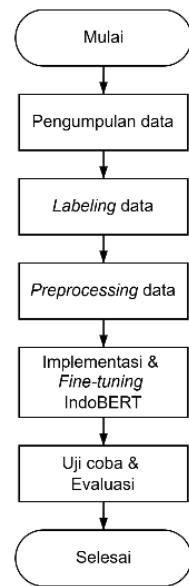
F1-score pada Persamaan (2.4) adalah *harmonic mean* dari *precision* dan *recall*, yang memberikan keseimbangan antara kedua metrik tersebut. Metrik ini akan memberikan gambaran yang lebih baik ketika kelas tidak seimbang.

$$F1 - score = 2 \times \frac{(precision \times recall)}{(precision + recall)} \quad (2.4)$$

3. METODOLOGI

3.1. Strategi Penelitian

Penelitian ini menggunakan model *pre-trained* IndoBERT yang *di-fine-tuning* untuk menganalisis sentimen terhadap kereta cepat Whoosh di platform X. Tahapan penelitian dimulai dengan pengumpulan data dari media sosial X menggunakan teknik *scraping*. Data yang terkumpul kemudian dilabeli secara manual menjadi tiga kategori yaitu positif, negatif, dan netral. Selanjutnya dilakukan *preprocessing* yang meliputi *cleaning*, *case folding*, dan normalisasi. Setelah itu, model IndoBERT *di-fine-tune* secara khusus untuk melatih model menganalisis sentimen terhadap *dataset* Whoosh. Kemudian dilakukan evaluasi untuk membantu memahami sentimen publik serta dibandingkan dengan hipotesis yang telah dirumuskan, sebagaimana yang ditunjukkan pada diagram alir penelitian pada Gambar 2.



Gambar 2. Diagram Alir Penelitian

3.2. Tools dan Perangkat Keras

Penelitian ini menggunakan bahasa pemrograman Python sebagai alat utama karena kelebihannya dalam mendukung implementasi pemrosesan bahasa alami (NLP) dan ketersediaan pustaka yang luas. Proses penelitian dilakukan dengan memanfaatkan layanan komputasi awan gratis dari Google Colab, yang menyediakan GPU Tesla T4 dengan 16GB VRAM GDDR6 dan 2.560 CUDA cores.

3.3. Dataset Penelitian

Penelitian ini menggunakan *dataset* primer yang berasal dari *tweet* pada media sosial X mengenai kereta cepat Whoosh, dengan kata kunci "Whoosh", "KCIC", atau "kereta cepat". Data dikumpulkan dengan rentang waktu satu tahun pengoperasian dari 1 Oktober 2023 hingga 31 Oktober 2024 menggunakan *tools* untuk *scraping* yaitu "TweetHarvest", yang menyimpan hasil dalam format CSV dengan atribut "date", "username", dan "tweet". Dalam melakukan analisis, hanya kolom "tweet" yang digunakan, ditambah kolom "sentiment" dari labelisasi manual. Data yang berhasil dikumpulkan adalah sebanyak 7500 tweet dengan komposisi masing-masing 2500 data untuk setiap kelasnya. Dari 7500 tweet, 70% (5250 tweet) digunakan untuk pelatihan, 20% (1500 tweet) untuk validasi, dan 10% (750 tweet) untuk pengujian. Data pelatihan digunakan untuk melatih model, data validasi untuk mengoptimalkan model, dan data pengujian digunakan untuk mengevaluasi performa model.

3.4. Konfigurasi Hyperparameter

Menurut penelitian pada model BERT, *hyperparameter* yang optimal untuk memperoleh performa model terbaik meliputi *batch size* 16 dan 32, nilai *learning rate* 5e-5, 3e-5, dan 2e-5, serta penerapan *optimizer Adam* dengan jumlah *epoch* 2, 3, dan 4 (Devlin et al., 2019). Rancangan pengujian untuk konfigurasi *hyperparameter* model IndoBERT ini ditunjukkan pada Tabel 1.

Tabel 1. Konfigurasi *Hyperparameter*

Hyperparameter	Ukuran
<i>Batch Size</i>	16 dan 32
<i>Learning Rate</i>	5e-5, 3e-5, dan 2e-5
<i>Epoch</i>	2, 3, dan 4

4. PENGUJIAN DAN ANALISIS

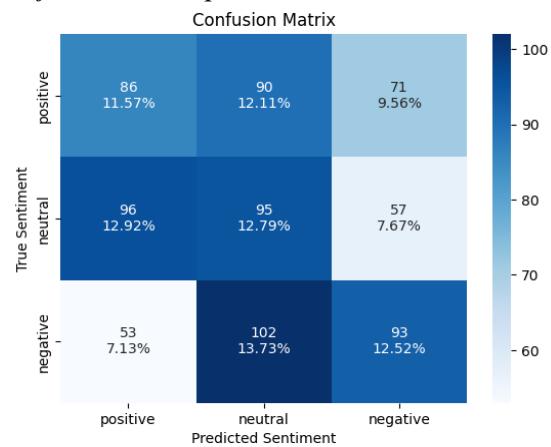
4.1. Pengujian Model *Pre-trained* IndoBERT

Pada penelitian ini, dilakukan pengujian terhadap model *pre-trained* dari IndoBERT yang terdapat pada situs Hugging Face, sebelum nantinya dipilih satu model paling baik yang akan dilakukan pengujian dengan melakukan *fine-tuning*. Pengujian ini dilakukan berdasarkan perancangan yang telah dibuat sebelumnya dengan menggunakan dataset Whoosh yang memiliki 7500 baris data. Model yang diujikan antara lain *indobert-base-p1*, *indobert-large-p1*, *indobert-base-p2*, *indobert-large-p2*, *indobert-base-uncased*, *indobertweet-base-uncased*, dan *multilingualbert*. Perbandingan hasil pengujian model *pre-trained* IndoBERT ini ditunjukkan pada Tabel 2.

Tabel 2. Perbandingan Pengujian Model *Pre-trained* IndoBERT

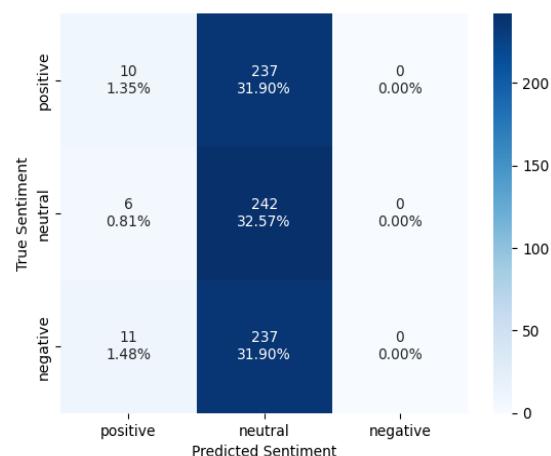
No	Model	Metrik Evaluasi			
		Acc	Prec	Rec	F1
1	<i>indobert-base-p1</i>	0,39	0,39	0,39	0,37
2	<i>indobert-large-p1</i>	0,33	0,33	0,33	0,25
3	<i>indobert-base-p2</i>	0,40	0,40	0,40	0,38
4	<i>indobert-large-p2</i>	0,32	0,26	0,32	0,21
5	<i>indobert-base-uncased</i>	0,37	0,37	0,37	0,37
6	<i>indobertweetbaseuncased</i>	0,34	0,36	0,34	0,28
7	<i>multilingualbert</i>	0,34	0,24	0,34	0,19

Berdasarkan hasil pengujian tersebut, didapatkan model *pre-trained* IndoBERT yaitu *indobert-base-p2* dari IndoBenchmark, menghasilkan performa dengan metrik evaluasi yang paling tinggi dibandingkan dengan model *pre-trained* IndoBERT lainnya. Sementara itu, model yang menghasilkan performa yang paling rendah adalah model *multilingualbert* (mBERT). Hal ini juga dibuktikan dengan *confusion matrix* pada Gambar 3 dan Gambar 4.



Gambar 3. *Confusion Matrix* *indobert-base-p2*

Pada Gambar 3, diagram tersebut menunjukkan bahwa model *indobert-base-p2* menghasilkan distribusi prediksi yang cukup merata, walaupun masih terdapat cukup banyak kesalahan. Prediksi benar untuk sentimen positif adalah 86 kasus (11,57%), netral 95 kasus (12,79%), dan negatif 93 kasus (12,52%). Namun, terdapat 90 kasus positif (12,11%), 96 kasus netral (12,92%), serta 102 kasus negatif (13,73%) yang masih salah diprediksi.



Gambar 4. *Confusion Matrix* *bert-base-multilingual-cased*

Sementara itu pada Gambar 4, diagram menunjukkan bahwa model *multilingualbert* (mBERT) memiliki bias ekstrem terhadap

prediksi netral, dengan 242 kasus (32,57%) benar diprediksi sebagai netral. Namun, hampir semua sentimen positif dan negatif salah diprediksi sebagai netral, masing-masing 237 kasus (31,90%). Prediksi benar untuk sentimen positif dan negatif juga sangat rendah, hanya 10 kasus (1,35%) dan 0 kasus (0,00%). Ini menunjukkan model kesulitan dalam membedakan sentimen positif dan negatif.

4.2. Pengujian Hyperparameter

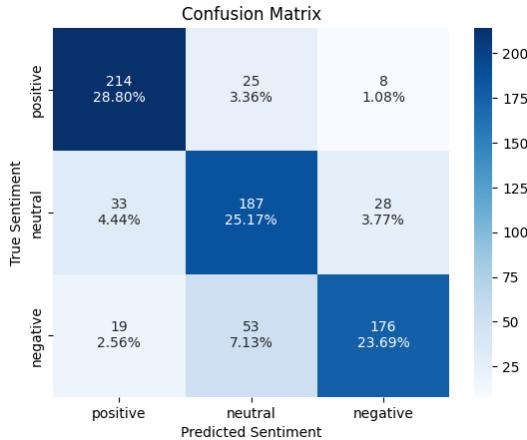
Pengujian model IndoBERT dilakukan menggunakan data pengujian dengan konfigurasi *hyperparameter* yang telah dirancang sebelumnya, yaitu *Epoch* sebanyak 2, 3, dan 4; *Learning Rate* sebesar 5e-5, 3e-5, dan 2e-5; serta *Batch Size* sebesar 16 dan 32. Hasil pengujian *hyperparameter* dari IndoBERT ini ditunjukkan pada Tabel 3.

Tabel 3. Hasil Pengujian Hyperparameter

Epoch	Lrate	Batch	Acc	Prec	Rec	F1
2	5e-5	16	0,74	0,74	0,74	0,74
		32	0,76	0,75	0,76	0,75
	3e-5	16	0,74	0,74	0,74	0,73
		32	0,76	0,76	0,76	0,76
	2e-5	16	0,75	0,75	0,75	0,75
		32	0,76	0,76	0,76	0,76
3	5e-5	16	0,74	0,74	0,74	0,73
		32	0,75	0,77	0,75	0,75
	3e-5	16	0,76	0,77	0,76	0,76
		32	0,76	0,76	0,76	0,75
	2e-5	16	0,75	0,76	0,75	0,75
		32	0,78	0,78	0,78	0,78
4	5e-5	16	<u>0,71</u>	<u>0,72</u>	<u>0,71</u>	<u>0,70</u>
		32	0,75	0,75	0,75	0,75
	3e-5	16	0,76	0,75	0,76	0,75
		32	0,75	0,75	0,75	0,75
	2e-5	16	0,76	0,76	0,76	0,76
		32	0,76	0,76	0,76	0,76

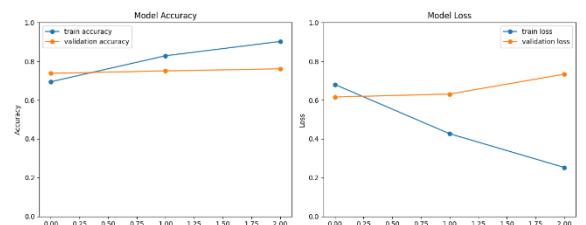
Berdasarkan pengujian tersebut, IndoBERT memperoleh akurasi rata-rata 0,75 dengan 0,78 untuk akurasi tertinggi dan 0,71 untuk akurasi terendah. Akurasi tertinggi yang ditunjukkan dengan teks bercetak tebal diperoleh dengan konfigurasi *hyperparameter Epoch* 3, *Learning Rate* 2e-5, dan *Batch Size* 32. Sementara itu akurasi terendah yang ditunjukkan dengan teks

yang digarisbawahi diperoleh dengan *Epoch* 4, *Learning Rate* 5e-5, dan *Batch Size* 16.



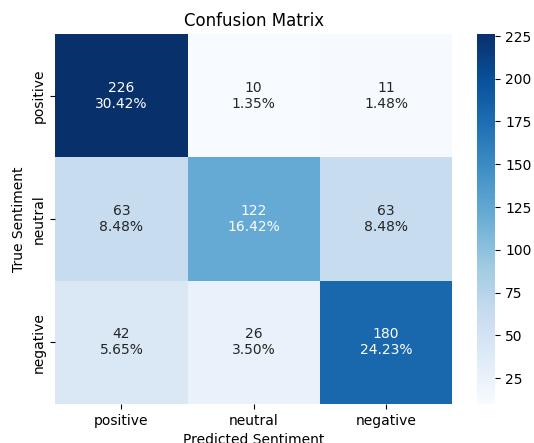
Gambar 5. Confusion Matrix Performa Tertinggi

Confusion matrix dari hasil pengujian tertinggi pada Gambar 5 menunjukkan performa model yang relatif seimbang, dengan prediksi benar untuk sentimen positif sebanyak 214 kasus (28,80%), netral 187 kasus (25,17%), dan negatif 176 kasus (23,69%). Namun, terdapat masih terdapat beberapa kesalahan prediksi, terutama pada sentimen netral yang salah diprediksi sebagai positif sebanyak 33 kasus (4,44%) dan negatif 28 kasus (3,77%). Sentimen negatif juga memiliki beberapa kesalahan dengan 53 kasus (7,13%) diprediksi sebagai netral. Ini menunjukkan bahwa meskipun model memiliki akurasi yang cukup baik, masih diperlukan peningkatan untuk mengurangi kesalahan prediksi antar kategori sentimen.



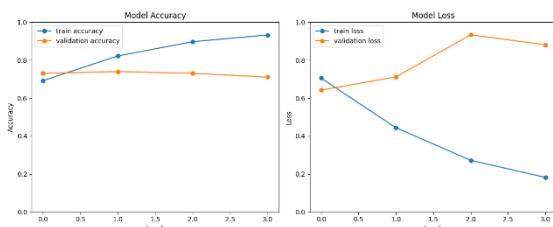
Gambar 6. Learning Curve Performa Tertinggi

Grafik *learning curve* pada Gambar 6 menunjukkan di grafik model *accuracy* (kiri), akurasi pada data pelatihan terus meningkat secara konsisten dengan bertambahnya *epoch*, namun akurasi pada data validasi cenderung stagnan dan tidak menunjukkan peningkatan. Sementara itu pada grafik model *loss* (kanan), *train loss* menurun secara konsisten, namun *validation loss* justru meningkat.



Gambar 7. Confusion Matrix Performa Terendah

Sementara itu, *confusion matrix* performa terendah yang ditunjukkan pada Gambar 7, menunjukkan performa model yang cukup baik dalam memprediksi sentimen positif dengan 226 kasus (30,42%) benar dan sentimen negatif dengan 180 kasus (24,23%). Namun, terdapat kesalahan signifikan pada sentimen netral yang salah diprediksi sebagai positif dan negatif, masing-masing sebanyak 63 kasus (8,48%). Sentimen negatif juga salah diprediksi sebagai positif sebanyak 42 kasus (5,65%). Ini menunjukkan bahwa meskipun akurasi untuk kategori positif dan negatif cukup baik, diperlukan peningkatan untuk mengurangi kesalahan pada kategori netral. Hal ini menunjukkan bahwa meskipun akurasi untuk kategori positif dan negatif cukup baik, namun diperlukan peningkatan untuk mengurangi kesalahan, khususnya pada kategori netral.



Gambar 8. Learning Rate Performa Terendah

Grafik *learning curve* pada Gambar 8 menunjukkan hasil yang kurang lebih serupa dengan hasil pengujian dengan performa tertinggi. Pada grafik model *accuracy* (kiri), akurasi pada data pelatihan terus meningkat secara konsisten, sedangkan akurasi pada data validasi hanya sedikit meningkat di awal dan kemudian stagnan, bahkan sedikit menurun di akhir. Sementara itu pada grafik model *loss* (kanan) *Train loss* menurun secara konsisten, sedangkan *validation loss* mengalami peningkatan setelah *epoch* pertama.

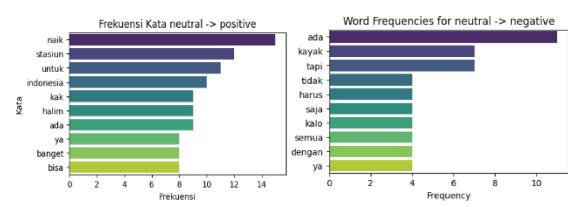
4.3. Analisis Pengujian Model *Pre-trained* IndoBERT

Berdasarkan pengujian, model *pre-trained* indobert-base-p2 dari IndoBenchmark menunjukkan performa terbaik dibandingkan model IndoBERT lainnya, menunjukkan efektivitasnya dalam klasifikasi data kereta cepat Whoosh. Sebaliknya, model bert-base-multilingual-cased memiliki performa terendah. Hal ini mendukung temuan sebelumnya bahwa model monolingual seperti IndoBERT, yang dilatih dengan fokus pada satu bahasa, cenderung mengungguli model multilingual seperti mBERT dalam tugas bahasa spesifik. Model monolingual dapat lebih efektif karena kapasitasnya difokuskan pada satu bahasa, memungkinkan pemahaman yang lebih mendalam dan efisien dalam menangkap struktur serta istilah unik bahasa tersebut (Koto et al., 2020; Wu & Dredze, 2020).

4.4. Analisis Pengujian Hyperparameter

Hasil pengujian dengan *confusion matrix* pada Gambar 5 dan Gambar 7 menunjukkan bahwa model memiliki performa yang cukup baik dalam mengklasifikasikan kategori positif dan negatif. Namun, terdapat kekurangan dalam klasifikasi sampel netral, yang sering salah dikategorikan sebagai positif dan negatif, ataupun sebaliknya. Kesulitan ini dapat disebabkan oleh ambiguitas data netral, yang sering tumpang tindih dengan karakteristik data positif dan negatif. Untuk meningkatkan akurasi, diperlukan penyesuaian lebih lanjut seperti peningkatan kualitas *dataset*, penambahan data pelatihan, atau penggunaan teknik pemrosesan fitur yang lebih canggih.

Lebih lanjut, dapat dilakukan analisis terhadap kata yang sering muncul dan contoh sampel kalimat pada data netral yang salah diklasifikasikan (misklasifikasi) sebagai data positif atau negatif, seperti yang ditunjukkan pada Gambar 9 dan Tabel 4.



Gambar 9. Frekuensi Kata pada Misklasifikasi Sentimen Netral sebagai Positif atau Negatif

Pada Gambar 9 menunjukkan data netral yang diklasifikasi sebagai positif mengandung

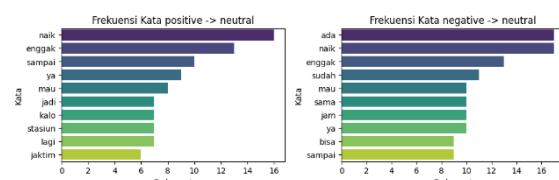
kata-kata seperti "naik," "ya," "banget," dan "bisa." Sedangkan, di data netral yang diklasifikasi sebagai negatif mengandung kata-kata seperti "kayak," "tapi," "tidak," dan "harus." Contoh kalimat dari data netral yang salah klasifikasi ini dapat dilihat pada Tabel 4.

Tabel 4. Sampel Kalimat pada Misklasifikasi Sentimen Netral sebagai Positif atau Negatif

Kalimat	Kelas Aktual	Kelas Prediksi
dengan skema ini tiket whoosh bisa lebih murah dan lebih mahal pada waktu tertentu.	Netral	Positif
bejir aku sudah sedih karena nanti juli pindah ke nangor tapi ternyata tiap weekend boleh whoosh balik saja ke jakarta	Netral	Negatif
hunting foto kcic itu ilmu prakiraannya kepake banget	Netral	Positif
kayak nya kalo sekarang bengong di lrt mrt adalah bagian dari healing beberapa tahun lagi gue butuh kcic jakarta surabaya biar bengongnya lama	Netral	Negatif
dari mrt ke lrt ke whoosh lumayan banget jalannya bisa dapat macam steps	Netral	Positif

Kata-kata seperti "naik," "ya," "banget," "bisa," "kayak," "tapi," "tidak," dan "harus" dapat mempengaruhi prediksi sentimen karena sering digunakan dalam konteks ambigu dalam kalimat. Misalnya, "naik" bisa berarti peningkatan (positif) atau sekadar aktivitas (netral). "Banget" dapat berarti menekankan sesuatu, yang bisa dianggap positif tanpa konteks yang jelas. "Bisa" dapat menunjukkan kemungkinan netral, tetapi bisa dianggap positif jika terkait hasil yang diinginkan. "Tapi" dan "tidak" sering muncul dalam negasi, sehingga dapat membingungkan model jika konteksnya tidak dianalisis dengan benar. "Kayak" dan "harus" bisa menunjukkan ketidakpastian atau keharusan, yang dapat disalahartikan sebagai sentimen negatif atau positif.

Dapat dilakukan juga analisis kata yang sering muncul dan contoh sampel kalimat pada data positif ataupun negatif yang sering diklasifikasikan sebagai data netral, seperti yang ditunjukkan pada Gambar 10 dan Tabel 6.4.



Gambar 10 Frekuensi Kata pada Misklasifikasi Sentimen Positif atau Negatif sebagai Netral

Pada Gambar 10 menunjukkan data positif yang diklasifikasi sebagai netral mengandung kata-kata seperti "naik," "enggak," "sampai," dan "mau." Demikian pula data negatif yang diklasifikasi sebagai netral juga mengandung kata-kata yang serupa. Contoh kalimat dari data positif atau negatif yang salah klasifikasi sebagai netral dapat dilihat pada Tabel 5.

Tabel 5. Sampel Kalimat pada Misklasifikasi Sentimen Positif atau Negatif sebagai Netral

Kalimat	Kelas Aktual	Kelas Prediksi
mau testimoni naik whoosh. kemaren lupa mau lanjutin twit ini. Ini aku cek di aplikasi enggak bisa detect gps nya. Saking cepatnya	Positif	Netral
mau naik sih whoosh ini enggak sempat ih naiknya jauh benar kudu ke halim	Negatif	Netral
dengan harga segini siapa coba yang enggak mau naik kcic. wkwkwk	Positif	Netral
kayaknya enggak sempat deh mencoba whoosh harga promo	Negatif	Netral

Kata-kata seperti "naik," "enggak," "sampai," dan "mau" dapat mempengaruhi prediksi sentimen karena juga sering digunakan dalam konteks ambigu dalam kalimat. Misalnya, "naik" bisa berarti sesuatu yang positif, seperti kenaikan gaji, atau netral seperti naik kendaraan. "Enggak" sebagai negasi dapat membalikkan sentimen, tetapi jika konteksnya tidak ditangkap dengan baik, dapat menyebabkan terjadinya misklasifikasi. Hal ini menunjukkan bahwa kata-kata tertentu dapat memengaruhi keputusan klasifikasi model, meskipun konteks keseluruhan kalimat seharusnya mengarah pada klasifikasi yang berbeda.

Sementara itu pada laporan klasifikasi di Tabel 3, model IndoBERT menunjukkan performa yang bervariasi berdasarkan konfigurasi *hyperparameter* seperti *epoch*, *learning rate*, dan *batch size*. Model mencapai akurasi tertinggi 0,78 pada *epoch* 3 dengan *learning rate* 2e-5 dan *batch size* 32, menunjukkan kemampuan klasifikasi yang benar lebih dari 78% dari total data uji. *Precision*, *recall*, dan *f1-score* juga mencapai nilai tertinggi pada konfigurasi ini, menunjukkan keseimbangan baik antara *true positive* dan *false positive*. Sebaliknya, akurasi terendah 0,71 dicapai pada *epoch* 4 dengan *learning rate* 5e-5 dan *batch size* 16, di mana model masih cukup baik dalam mengidentifikasi *instance* positif, meskipun tidak sebaik konfigurasi sebelumnya.

Pengujian ini menunjukkan bahwa *learning rate* berpengaruh signifikan, dengan *learning rate* $2e-5$ memberikan performa terbaik di sebagian besar konfigurasi. *Learning rate* yang lebih kecil ini menghasilkan pembelajaran yang lebih stabil dan efektif, membantu model mencapai konvergensi yang lebih baik. Sebaliknya, *learning rate* $5e-5$ cenderung menyebabkan *overfitting* atau kesulitan konvergensi. Pengaruh *batch size* juga signifikan; *batch size* 32 umumnya menghasilkan kinerja lebih baik dibandingkan *batch size* 16, karena memungkinkan pembelajaran yang lebih efisien dan estimasi gradien yang stabil. Namun, penurunan performa pada *epoch* 4 dengan *batch size* 32 menunjukkan potensi *overfitting*. Tren performa meningkat dari *epoch* 2 ke 3, tetapi menurun pada *epoch* 4, mengindikasikan titik jenuh atau *overfitting*, terutama dengan *learning rate* yang lebih tinggi. Meskipun beberapa metrik stabil pada *epoch* 4, penurunan performa dengan *learning rate* $5e-5$ menunjukkan bahwa model tidak lagi belajar dengan baik.

Grafik *learning curve* pada Gambar 6 dan Gambar 8 juga menunjukkan indikasi terjadinya *overfitting*. Akurasi pelatihan yang meningkat menandakan model belajar baik dari data pelatihan, tetapi akurasi validasi yang stabil menunjukkan batas performa pada data validasi. *Train loss* yang menurun konsisten menunjukkan pembelajaran yang baik, namun peningkatan *validation loss* mengindikasikan *overfitting*, di mana performa pada data validasi memburuk. Untuk mengatasi ini, dapat diterapkan beberapa teknik seperti *regularization* (misalnya, Dropout atau L2), *early stopping*, atau *data augmentation* untuk meningkatkan generalisasi model. Eksplorasi konfigurasi *hyperparameter* juga dapat membantu menemukan pengaturan optimal.

Overfitting bisa terjadi ketika model terlalu menyesuaikan dengan data pelatihan hingga menangkap *noise* yang tidak relevan, sehingga performanya baik pada data pelatihan tetapi buruk pada data pengujian. Penyebabnya termasuk model yang terlalu kompleks, jumlah data yang minim, atau kualitas data yang buruk. Penggunaan teknik seperti regularisasi atau pengurangan *learning rate* secara bertahap dapat membantu menghindari *overfitting*. Eksperimen lebih lanjut dengan *learning rate* yang lebih kecil, teknik augmentasi data, dan perbaikan data juga dapat meningkatkan performa model.

Secara keseluruhan, model IndoBERT menunjukkan potensi yang baik, tetapi membutuhkan penyesuaian lebih lanjut, seperti peningkatan kualitas *dataset*, pengaturan *hyperparameter* yang lebih hati-hati, dan teknik pelatihan lanjutan. Dengan optimalisasi yang tepat, performa model diharapkan dapat meningkat secara signifikan.

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil dan analisis penelitian mengenai sentimen masyarakat terhadap kereta cepat Whoosh pada platform X menggunakan IndoBERT yang telah dilaksanakan sebelumnya, dapat disimpulkan beberapa hal yaitu, dari hasil pengujian model *pre-trained* IndoBERT pada *dataset* Whoosh sebelum *fine-tuning*, menunjukkan bahwa model *indobert-base-p2* memberikan performa terbaik dibandingkan model *pre-trained* IndoBERT lainnya, dengan metrik evaluasi berupa *accuracy*, *precision*, dan *recall* sebesar 0,40, serta *f1-score* sebesar 0,38.

Selain itu, pengujian juga mengindikasikan bahwa model monolingual seperti IndoBERT lebih unggul dalam menangani tugas bahasa spesifik dibandingkan model multilingual seperti MultilingualBERT, karena model monolingual akan lebih fokus dalam memahami nuansa dan struktur satu bahasa secara mendalam.

Sementara itu, hasil pengujian *hyperparameter* menunjukkan bahwa performa model IndoBERT bervariasi berdasarkan konfigurasi seperti *epoch*, *learning rate*, dan *batch size*, dengan konfigurasi terbaik dicapai pada *epoch* 3, *learning rate* $2e-5$, dan *batch size* 32, dengan menghasilkan metrik evaluasi tertinggi berupa *accuracy*, *precision*, *recall*, dan *f1-score* sebesar 0,78. *Learning rate* $2e-5$ memberikan pembelajaran yang lebih stabil dibandingkan nilai yang lebih tinggi seperti $5e-5$, karena mendukung proses pembelajaran yang bertahap dan mencegah perubahan parameter yang terlalu drastis, sementara *batch size* 32 memberikan estimasi gradien yang lebih stabil dibandingkan *batch size* 16. Namun, model masih menghadapi tantangan, seperti kesulitan mengklasifikasikan sentimen netral akibat pengaruh kata-kata berkonotasi positif atau negatif meskipun konteks kalimat tidak mendukung sentimen tersebut, serta model masih mengalami *overfitting* yang ditunjukkan

oleh performa tinggi pada data pelatihan tetapi menurun pada data pengujian, yang mengindikasikan kurangnya kemampuan generalisasi model terhadap data baru.

5.2. Saran

Berdasarkan analisis dari penelitian yang telah dilakukan, berikut adalah beberapa saran dan rekomendasi untuk pengembangan penelitian di masa mendatang.

1. Meningkatkan jumlah dan kualitas *dataset* untuk membantu model membedakan kategori sentimen, terutama netral. Oleh karena itu disarankan untuk menggunakan *annotator* berpengalaman untuk pelabelan data agar kualitas dan konsistensi terjaga, yang akan meningkatkan performa model.
2. Melakukan lebih banyak pengujian dengan konfigurasi *hyperparameter* sehingga dapat menemukan kombinasi yang lebih optimal untuk meningkatkan performa model.
3. Menerapkan teknik-teknik yang dapat menghindari atau mengurangi kemungkinan terjadinya *overfitting* seperti *regularization* dan *early stopping*.
4. Melakukan perbandingan pengujian dengan model lain selain IndoBERT untuk mengevaluasi keunggulan dan kelemahan masing-masing model, serta memberikan wawasan lebih komprehensif tentang performa relatif dari berbagai pendekatan dalam pemrosesan bahasa alami (NLP) untuk bahasa Indonesia.

DAFTAR PUSTAKA

- Alsaedi, A. & Khan, M.Z. 2019. A Study on Sentiment Analysis Techniques of Twitter Data. *International Journal of Advanced Computer Science and Applications (IJACSA)*. <http://dx.doi.org/10.14569/IJACSA.2019.0100248>.
- DataReportal. 2023. Digital 2023: Indonesia — DataReportal – Global Digital Insights. <https://datareportal.com/reports/digital-2023-indonesia> [Diakses 23 Agustus 2024]
- Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*. <https://aclanthology.org/N19-1423>.
- Fu, Z., Yang, H., So, A.M.-C., Lam, W., Bing, L. & Collier, N. 2023. On the Effectiveness of Parameter-Efficient Fine-Tuning. *Proceedings of the AAAI Conference on Artificial Intelligence*. 12799–12807. <http://arxiv.org/abs/2211.15583>.
- Hoang, M., Bihorac, O.A. & Rouces, J. 2019. Aspect-Based Sentiment Analysis Using BERT. *Proceedings of the 22nd Nordic Conference on Computational Linguistics*.
- Koto, F., Rahimi, A., Lau, J.H. & Baldwin, T. 2020. IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. *Proceedings of the 28th International Conference on Computational Linguistics*, International Committee on Computational Linguistics. <https://aclanthology.org/2020.coling-main.66>.
- Patil, R., Boit, S., Gudivada, V. & Nandigam, J. 2023. A Survey of Text Representation and Embedding Techniques in NLP. *IEEE Access*, 11: 36120–36146.
- PT KCIC. 2023. Tentang KCIC. <https://kcic.co.id/tentang-kami/profil/> [Diakses 23 Agustus 2024].
- Ting, K.M. 2017. Confusion Matrix. *Encyclopedia of Machine Learning and Data Mining Second Edition*.
- Uysal, A.K. & Gunal, S. 2014. The Impact of Preprocessing on Text Classification. *Information Processing and Management*. <https://doi.org/10.1016/j.ipm.2013.08.006>.
- Wijaya, H. 2020. Aktualisasi Kebijakan China One Belt and One Road Di Indonesia Melalui Pembangunan Kereta Cepat Jakarta-Bandung. *Jurnal Dinamika Global*, 5(1).
- Wilie, B., Vincentio, K., Indra Winata, G., Cahyawijaya, S., Li, X., Lim, Z.Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S. & Purwarianti, A. 2020. IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics: 843–857. <https://doi.org/10.48550/arXiv.2009.05387>