

MODEL ANALISIS SENTIMEN PADA KENDARAAN LISTRIK MENGGUNAKAN ALGORITMA INDOBERTWEET DAN INDOBERT

Diterima Redaksi: 25 Maret 2025; Revisi Akhir: 26 Agustus 2025; Diterbitkan Online: 7 September 2025

Belinda Eka Sarah Dewi

Teknik Informatika, Fakultas Teknologi Informasi dan Digital, Universitas Bani Saleh
Jalan Mayor M. Hasibuan No. 68, Kec. Bekasi Timur, Kota Bekasi, Jawa Barat, Indonesia, kode pos: 17113
e-mail: belinda@ubs.ac.id

Abstrak: Adopsi kendaraan listrik yang semakin meningkat di Indonesia telah memunculkan berbagai opini publik, sehingga diperlukan analisis sentimen untuk memahami perspektif Masyarakat. Penelitian ini bertujuan untuk membandingkan performa dua model berbasis transformer, yaitu IndoBERTweet dan IndoBERT dalam menganalisis sentimen terhadap kendaraan listrik di Indonesia. Dataset dikumpulkan dari tweet atau komentar online berbahasa Indonesia. Data ini kemudian melalui tahap prapemrosesan, pelabelan sentimen menjadi sentiment positif, netral, dan negative, serta fine-tuning pada kedua model. Model dievaluasi berdasarkan akurasi, presisi, recall, dan F1-score. Hasil eksperimen menunjukkan bahwa IndoBERTweet memiliki performa lebih baik dibandingkan IndoBERT dalam klasifikasi sentimen. Performa terbaik yang dicapai IndoBERTweet adalah akurasi sebesar 82,40% dengan F1-score sebesar 82,39%. Sedangkan, IndoBERT mencapai akurasi sebesar 75,98% dengan F1-score sebesar 75,46%. Penelitian ini menyoroti pentingnya penggunaan model domain spesifik untuk analisis sentimen dan berkontribusi pada kemajuan natural language processing (NLP).

Kata Kunci— analisis sentimen, IndoBERT, IndoBERTweet, kendaraan listrik

Abstract: The increasing adoption of electric vehicles in Indonesia has sparked various public opinions, necessitating sentiment analysis to understand societal perspectives. This study aims to compare the performance of two transformer-based models, IndoBERTweet and IndoBERT, in analyzing sentiments towards electric vehicles in Indonesia. Using a dataset collected from Indonesian language tweets and online comments, the data undergoes preprocessing, sentiment labelling into positive, negative, and neutral sentiments, and subsequent fine-tuning of both models. The models are evaluated based on accuracy, precision, recall, and F1-score. Experimental results demonstrate that IndoBERTweet achieves superior performance compared to IndoBERT in sentiment classification. The best performance recorded for IndoBERTweet was an accuracy of 82,40%, with an F1-score of 82,39%, while IndoBERT achieved an accuracy of 75,98% and an F1-score of 75,46%. These findings highlight the importance of using domain-specific models for sentiment analysis and contribute to advancements in Indonesia-language natural language processing (NLP).

Keywords— sentiment analysis, IndoBERT, IndoBERTweet, electric vehicles

I. PENDAHULUAN

ADOPSI kendaraan listrik yang pesat di Indonesia telah memunculkan beragam opini publik yang memengaruhi perspektif masyarakat terkait keberlanjutan, keterjangkauan, dan kesiapan infrastruktur. Memahami sentimen ini sangat penting bagi pembuat kebijakan, produsen, dan pemangku kepentingan untuk menanggapi kekhawatiran public serta menyusun strategi adopsi kendaraan listrik yang lebih efektif. Analisis sentimen yang memanfaatkan teknik *Natural Language Processing* (NLP) menyediakan pendekatan sistematis untuk menganalisis opini yang diungkapkan dalam diskusi daring, unggahan media sosial, dan artikel berita. Analisis sentimen menyediakan metode untuk mengekstrasi opini dari data teks.

Beberapa penelitian telah mengeksplorasi analisis sentimen terhadap kendaraan listrik menggunakan berbagai pendekatan *machine learning* dan *deep learning*. Salah satu metode konvensional yang banyak digunakan adalah metode Naïve Bayes. Metode Naïve Bayes dapat melakukan analisis sentimen kendaraan listrik di sosial media X dengan *accuracy* diatas 80 % [1]-[4]. Tetapi, ada

juga akurasi model yang menggunakan metode Naïve Bayes tidak mencapai 80% [5] Metode lain yang sering digunakan selain Naïve Bayes adalah metode *Support Vector Machine* (SVM). Dibandingkan dengan Naïve Bayes, SVM memberikan performa yang lebih baik. Akurasi yang diberikan SVM lebih tinggi dibandingkan dengan Naïve Bayes. Akan tetapi, waktu eksekusi SVM lebih lama dibandingkan dengan Naïve Bayes [2][6].

Studi lain menunjukkan penggunaan *Long Short-Term Memory* (LSTM) dalam klasifikasi sentimen kendaraan listrik di Indonesia. Penelitian tersebut menunjukkan peningkatan akurasi menggunakan jaringan LSTM dibandingkan dengan model konvensional. Meskipun LSTM mampu menangkap ketergantungan sekuensial, model tersebut masih mengalami kesulitan dalam memahami konteks jarak jauh [7][8]. Dalam beberapa tahun terakhir, model berbasis *transformer* telah diterapkan dalam tugas analisis sentimen. IndoBERT adalah salah satu model berbasis transformer yang sering digunakan untuk klasifikasi sentimen berbahasa Indonesia. Sudah banyak penelitian yang menggunakan model IndoBERT dan menunjukkan performa yang baik. IndoBERT memberikan akurasi yang lebih unggul dibandingkan dengan model-model konvensional, karena IndoBERT khusus dilatih untuk teks bahasa Indonesia [9]-[12]. Terdapat juga penelitian yang membandingkan antara performa IndoBERT dan mBERT. Dari penelitian tersebut, model IndoBERT lebih unggul dibandingkan mBERT[13].

IndoBERT memiliki performa lebih baik pada dataset terstruktur seperti artikel berita, tetapi kurang efektif dalam menangani teks di media sosial yang berbahasa Indonesia. IndoBERTweet yang dirancang khusus untuk teks media sosial berbahasa Indonesia telah diajukan sebagai alternatif potensial, tetapi efektivitasnya dalam domain sentimen kendaraan listrik di Indonesia belum banyak diteliti. Beberapa penelitian telah mendukung penggunaan IndoBERTweet untuk tugas NLP berbahasa Indonesia. Terdapat penelitian yang membandingkan antara performa LSTM dan IndoBERTweet. Dari penelitian tersebut, model IndoBERTweet memiliki performa yang lebih baik dari LSTM. Model IndoBERTweet tersebut digunakan untuk mengklasifikasi review Tiktok [14].

Berdasarkan temuan ini, IndoBERTweet diperkirakan memiliki keunggulan dalam analisis diskusi terkait kendaraan listrik di media sosial, dimana bahasa informal lebih sering digunakan di media sosial. Penelitian ini bertujuan untuk mengatasi kesenjangan tersebut dengan membandingkan IndoBERTweet dan IndoBERT dalam analisis sentimen terhadap diskusi mobil listrik di Indonesia. Dengan memanfaatkan model berbasis *transformer* tersebut, diharapkan dapat memberikan wawasan mengenai perbandingan performa IndoBERTweet dan IndoBERT dalam analisis sentimen kendaraan listrik di Indonesia dengan sumber data berasala dari komentar sosial media X (Twitter).

II. TINJAUAN PUSTAKA

A. Sentimen Analisis

Analisis sentimen melibatkan penilaian dan pengkategorian emosi secara otomatis dalam data berbasis teks, termasuk umpan balik tertulis dan pembaruan media sosial. Teknik ini berada di bawah lingkup *Natural Language Processing* (NLP) dan menggunakan teknik *machine learning* untuk mengklasifikasikan sentimen emosional yang terkandung dalam konten teks [12]. Teknik ini mencakup pengkategorian informasi menjadi sentimen positif, negatif, atau netral pada tingkat kalimat, dokumen, atau fitur. Secara keseluruhan, analisis sentimen berfungsi sebagai alat yang berharga untuk mengumpulkan informasi, membuat keputusan berdasarkan opini, serta mendapatkan wawasan tentang sentimen public. Proses analisis sentimen dimulai dengan pemilihan data. Selanjutnya, data diekstraksi dan sering diformat menjadi teks menggunakan teknik *scraping*. Setelah itu, data teks dianalisis untuk mengidentifikasi serta mengklasifikasikan emosi atau opini yang diekspresikan. Analisis sentimen dapat dilakukan melalui berbagai metode, termasuk strategi berbasis korpus, metode berbasis kamus, serta teknik *machine learning*.

B. IndoBERTweet

IndoBERTweet adalah model bahasa berbasis *Transformer* yang dikembangkan untuk menangani teks berbahasa Indonesia dari media sosial X (Twitter). Model ini merupakan turunan dari model IndoBERT yang telah disesuaikan dengan Bahasa informal dan karakteristik unik dari teks media sosial

X di Indonesia. Dalam berbagai penelitian, IndoBERTweet telah menunjukkan kinerja yang unggul dalam tugas analisis sentimen, deteksi hoaks, dan klasifikasi topik dibandingkan dengan model *Natural Language Processing* (NLP) lainnya yang tidak secara spesifik dilatih untuk teks informal. Penelitian ini menemukan bahwa IndoBERTweet memiliki akurasi yang lebih tinggi dalam analisis sentimen dibandingkan dengan model konvensional seperti LSTM dan CNN. Selain itu, IndoBERTweet mampu menangkap konteks dari Bahasa Indonesia sehari-hari yang sering digunakan dalam opini masyarakat di media sosial [15].

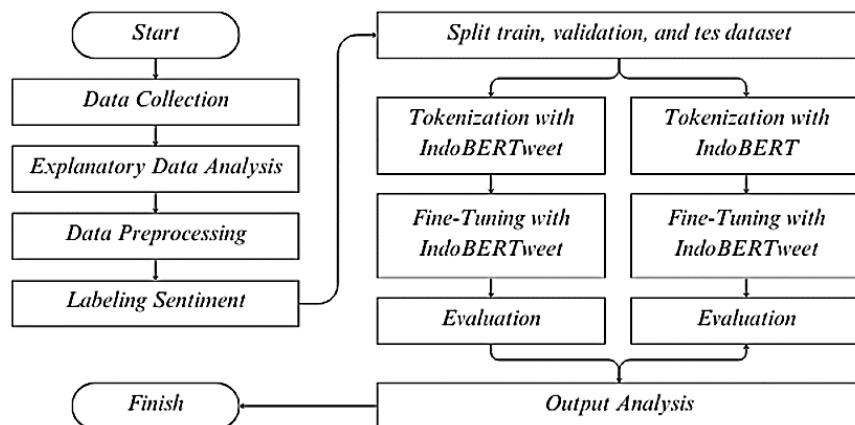
C. IndoBERT

IndoBERT adalah model bahasa yang telah dilatih sebelumnya dan dikembangkan secara khusus untuk bahasa Indonesia berdasarkan arsitektur Bidirectional Encoder Representations from Transformers (BERT) dari Google. Model ini memanfaatkan arsitektur transformer yang menggunakan mekanisme *self-attention* untuk memproses data masukan, memungkinkan model mempertimbangkan konteks dari kata-kata sebelum dan sesudahnya dalam sebuah kalimat.

D. Fine-Tuning

Fine-Tuning adalah teknik *transfer learning* yang digunakan untuk menyesuaikan model bahasa pra-latih dengan dataset spesifik tugas tertentu [16]. Dalam konteks IndoBERTweet, *fine-tuning* dilakukan dengan memberikan sampel data spesifik seperti *tweet* atau komentar mengenai kendaraan listrik, agar model dapat memahami pola sentimen yang lebih spesifik. Beberapa studi menunjukkan bahwa *fine-tuning* pada model Transformer dapat meningkatkan akurasi dalam tugas klasifikasi teks [17]. Hal ini karena model dapat menyesuaikan bobot parameter berdasarkan dataset yang digunakan, sehingga lebih optimal dalam memahami teks dalam domain tertentu, seperti opini publik tentang kendaraan listrik. Dengan menerapkan *fine-tuning* pada IndoBERTweet, penelitian ini bertujuan untuk meningkatkan akurasi model dalam memahami sentimen masyarakat terhadap kendaraan listrik di Indonesia.

III. METODE PENELITIAN



Gambar 1. Workflow menggunakan model IndoBERTweet dan IndoBERT

A. Data Collection

Penelitian ini memanfaatkan data yang diperoleh dari interaksi pengguna social media X (Twitter) dalam bentuk komentar pada akun pengguna X (Twitter). Data yang dikumpulkan dalam penelitian ini adalah sebanyak 3582 komentar.

B. Data Preprocessing

Preprocessing melibatkan serangkaian tindakan untuk mempersiapkan dan menyesuaikan data teks sebelum dilakukan analisis. Proses yang dilakukan pada tahap ini adalah *case folding*, *text cleansing*, dan *word normalization*. *Case folding* adalah tahap mengubah teks menjadi huruf kecil (*lowercase*). *Text*

cleansing dilakukan dengan beberapa tahap, yaitu menghilangkan *user hashtag*, URL, angka, *punctuation*, *emoticon*, *white space*, *single char*, symbol, kata duplikat.

C. Labelling Sentiment

Pelabelan dataset merupakan tahap penting dalam analisis sentiment. Dalam penelitian ini terdapat tiga kategori kelas sentiment, yaitu positif, negatif, dan netral. Proses pelabelan dilakukan secara manual dengan melihat kata pada komentar. Data yang sudah diberi label akan digunakan dalam pembentukan model analisis sentimen.

D. Split Train, Validation and Test Dataset

Pembagian dataset untuk proses *training* dan *testing* dilakukan sebelum tahap tokenisasi. Data dibagi menjadi 80% untuk data *training*, 10% untuk data validasi dan 10% untuk data *testing*. Data *training* digunakan untuk melatih model dengan memberikan contoh-contoh data beserta labelnya. Model belajar dari pola dalam data *testing* ini untuk menyesuaikan bobot pada arsitektur *transformer*. Data *validation* berfungsi untuk mengevaluasi performa model selama proses pelatihan. Data validation juga digunakan untuk *tuning hyperparameter* dan mencegah *overfitting*, sehingga model tidak hanya bekerja baik pada training data tetapi juga dapat menggeneralisasi ke data baru. Sedangkan Data *testing* digunakan untuk menguji performa akhir model setelah proses pelatihan selesai.

E. Tokenization

Tokenisasi adalah proses memecah teks menjadi unit-unit kecil yang disebut token. Pada konteks IndoBERTweet, tokenisasi dilakukan menggunakan *tokenizer* khusus yang telah disesuaikan dengan karakteristik Bahasa Indoneisa, terutama dari media social seperti X (Twitter). IndoBERTweet menggunakan *SentencePiece Unigram Tokenizer*, yang mirip dengan tokenisasi pada model BERT lainnya tetapi lebih optimal untuk bahasa Indonesia yang sering mengandung kata tidak baku, singkatan, atau bahasa gaul. Sedangkan, pada konteks IndoBERT, IndoBERT menggunakan WordPiece Tokenization, yaitu metode tokenisasi yang membagi kata menjadi sub-kata atau unit yang lebih kecil jika kata tersebut tidak ditemukan dalam kosakata model. WordPiece digunakan untuk menangani *out-of-vocabulary (OOV) words*, yaitu kata-kata yang tidak dikenal oleh model.

F. Fine-Tuning

Fine-Tuning adalah proses pelatihan ulang model yang sudah di-*pretrained* pada tugas spesifik dengan dataset yang lebih kecil tetapi lebih relevan. Teknik ini memungkinkan model yang sudah memiliki pemahaman dasar tentang bahasa untuk lebih disesuaikan dengan kebutuhan tertentu seperti analisis sentimen. Meskipun model IndoBERT dan IndoBERTweet telah dilatih pada data bahasa Indonesia, model ini masih bersifat generik.

G. Evaluation

Evaluasi dilakukan untuk menilai seberapa baik model dalam menyelesaikan tugas yang diberikan. Evaluasi dilakukan menggunakan metrics evaluasi yang terdiri dari *accuracy*, *precision*, *recall*, dan *F1-score*.

1) Accuracy

Accuracy digunakan untuk mengukur seberapa sering model membuat prediksi yang benar.

$$\text{Accuracy} = \frac{\text{Jumlah prediksi benar}}{\text{Total sampel}} \dots \quad (1)$$

2) Precision

Precision digunakan untuk mengukur ketepatan prediksi positif model.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \dots \quad (2)$$

3) Recall

Recall digunakan untuk mengukur seberapa banyak data positif yang benar-benar terdeteksi.

$$Recall = \frac{True\ Positive}{True\ Positive+False\ Negative} \quad \dots \dots \dots \quad (3)$$

4) F1-Score

F1-Score merupakan rata-rata harmonic antara *precision* dan *recall*. *F1-score* berguna jika ada ketidakseimbangan data antara kelas positif dan kelas negatif. Nilai *F1-score* yang tinggi menunjukkan model memiliki keseimbangan yang baik antara presisi dan sensitivitas.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad \dots \dots \dots \quad (4)$$

IV. HASIL DAN PEMBAHASAN

A. Hasil Data Collection

Berikut ini merupakan sampel hasil pengumpulan data dari sosial media X (Twitter). Data yang dikumpulkan dalam penelitian ini adalah sebanyak 3582 komentar. Sampel data ditunjukkan pada Gambar 2.

username	content
Andrysimp	@razkalog @zaxzone @sandiuno Yaelah...nanti mo...
Daniel13666	Bagaimana kalau kompensasi kenaikan BBM pemer...
indrash7	@gilbas Mungkin dia abis dapat sumbangan moto...
stanrobt	@gorofaiz @gilbas Nah atau motor listrik, pas...
gorofaiz	@stanrobt @gilbas Ga harus mobil listrik jd, ...
nevajifap	@gilbas Tapi jujur serem motor listrik kalau ...
GimoJohannes	@tribunnews Ayo bos percepat produksi motor/mo...
P Aryawidiastawa	Motor Listrik United T1800 Bisa Dicas di Rumah...
Tebobotee1	Atau tidak beat elektrik gitu, harga 12 juta b...
bersamasunu	Bbm naik, mendadak fyp tiktok + youtube isinya...
fajarfunbox	@agamasuk Itu kenapa sepeda bukan menjadi alte...

Gambar 2. Hasil data collection

B. Data Preprocessing

Proses yang dilakukan pada tahap ini adalah *case folding*, *text cleansing*, dan *word normalization*. *Case folding* adalah tahap mengubah teks menjadi huruf kecil (*lowercase*). *Text cleansing* dilakukan dengan beberapa tahap, yaitu menghilangkan *user hashtag*, URL, angka, *punctuation*, *emoticon*, *white space*, *single char*, symbol, kata duplikat. Pada kolom *content* merupakan komentar sebelum dilakukan data *preprocessing*, sedangkan pada kolom *clean_teks* merupakan komentar setelah dilakukan data *preprocessing* seperti yang ditunjukkan pada Gambar 3.

content	clean_teks
Fokus Solusi terhadap kenaikan BBM demi mengur...	fokus solusi terhadap kenaikan bbm demi mengur...
@bbyygum pake motor listrik aja	pake motor listrik aja
@itsmfe Beli motor listrik aja besok"	beli motor listrik aja besok
PLN juga memberikan dukungan kendaraan listrik...	pln juga memberikan dukungan kendaraan listrik...
Semoga harga motor listrik kaga naek seiring b...	semoga harga motor listrik kaga naek seiring b...
Mana motor listrik bentukannya ucul2 banget, h...	mana motor listrik bentukannya ucul banget har...
kalau beli motor listrik disini cocok bgt sih ...	kalau beli motor listrik disini cocok bgt sih ...
@Dennysiregar7 beli motor listrik nanti PLN yg...	beli motor listrik nanti pln yg dinaikin wkwk...
kepengen motor listrik seru kelihatannya	kepengen motor listrik seru kelihatannya
punya keinginan beli motor listrik aja kek nya...	punya keinginan beli motor listrik aja kek nya...
@kakabidaaa Mau pakai motor listrik aja lah	mau pakai motor listrik aja lah

Gambar 3. Hasil data preprocessing

C. Labelling Sentiment

Gambar 4 merupakan hasil *labeling sentiment*. Sentiment terdiri dari 3 klasifikasi yaitu, positif, negative, dan netral. Pada Tabel 2 menunjukkan tabel distribusi label yaitu dari 3.582 data terdapat 1.511 data berlabel positif (42,17%), 1.102 data netral (30,76%), dan 969 data negatif (27,07%). Secara keseluruhan, ketidakseimbangan distribusi antar kelas tidak terlalu ekstrem, sehingga masih memungkinkan untuk melakukan pelatihan model secara efektif tanpa perlu menerapkan teknik penyeimbangan data secara agresif (seperti oversampling atau undersampling).

content	label	clean_teks
Fokus Solusi terhadap kenaikan BBM demi mengur...	positif	fokus solusi terhadap kenaikan bbm demi mengur...
@bbyyygum pake motor listrik aja	positif	pake motor listrik aja
@itsmfefe Beli motor listrik aja besok"	positif	beli motor listrik aja besok
PLN juga memberikan dukungan kendaraan listrik...	positif	pln juga memberikan dukungan kendaraan listrik...
Semoga harga motor listrik kaga naek seiring b...	positif	semoga harga motor listrik kaga naek seiring b...
Mana motor listrik bentukannya ucul2 banget, h...	positif	mana motor listrik bentukannya ucul banget har...
kalau beli motor listrik disini cocok bgt sih ...	positif	kalau beli motor listrik disini cocok bgt sih ...
@Dennysiregar7 beli motor listrik nanti PLN yg...	negatif	beli motor listrik nanti pln yg dinaikin wkwk...
kepengen motor listrik seru kelihatannya	positif	kepengen motor listrik seru kelihatannya
punya keinginan beli motor listrik aja kek nya...	positif	punya keinginan beli motor listrik aja kek nya...
@kakabidaa Mau pakai motor listrik aja lah	positif	mau pakai motor listrik aja lah

Gambar 4. Hasil labelling sentiment

Tabel 1. Persebaran jumlah class labelling

Label	Jumlah data
Positif	1.511
Netral	1.102
Negatif	969

D. Split Train, Validation, and Test Dataset

Pembagian dataset untuk proses *training* dan *testing* dilakukan sebelum tahap tokenisasi. Data dibagi menjadi 80% untuk data *training*, 10% untuk data validasi dan 10% untuk data *testing*. Langkah pembagian ditunjukkan pada Gambar 5.

```
# Split data

# Pisahkan Training (80%) dan Sisa (20%)
train_texts, temp_texts, train_labels, temp_labels = train_test_split(df['clean_teks'], df['label'], test_size=0.2, random_state=42)

# Pisahkan Sisa menjadi Validation (10%) dan Test (10%)
val_texts, test_texts, val_labels, test_labels = train_test_split(temp_texts, temp_labels, test_size=0.5, random_state=42)

# Membuat dataset
from datasets import Dataset, DatasetDict

train_dataset = Dataset.from_pandas(pd.DataFrame({"clean_teks": train_texts, "label": train_labels}))
val_dataset = Dataset.from_pandas(pd.DataFrame({"clean_teks": val_texts, "label": val_labels}))
test_dataset = Dataset.from_pandas(pd.DataFrame({"clean_teks": test_texts, "label": test_labels}))
print(val_dataset)
```

Gambar 5. Proses split train, validation, and test Dataset

E. Tokenization

Setelah proses *split train, validation, and test dataset*, dilakukan tokenisasi untuk masing-masing dataset. Tokenisasi untuk IndoBERTweet menggunakan *autotokenizer pretrained "indolem/indobertweet-base-uncased"*. Sedangkan, Tokenisasi untuk IndoBERT menggunakan *autotokenizer pretrained "indolem/indobert-base-uncased"*.

```
# Tokenisasi Function IndoBERTweet
from transformers import AutoTokenizer, AutoModel

tokenizer_indobertweet = AutoTokenizer.from_pretrained("indolem/indobertweet-base-uncased", model_max_length=128)
def tokenize_indobertweet_function(examples):
| return tokenizer_indobertweet(examples["clean_teks"], padding="max_length", truncation=True, max_length=128)

tokenizer_indobert = AutoTokenizer.from_pretrained("indolem/indobert-base-uncased", model_max_length=128)
def tokenize_indobert_function(examples):
| return tokenizer_indobert(examples["clean_teks"], padding="max_length", truncation=True, max_length=128)

# Tokenisasi dataset IndoBERTweet
train_dataset_indobertweet = train_dataset.map(tokenize_indobertweet_function, batched=True, remove_columns=["clean_teks"])
val_dataset_indobertweet = val_dataset.map(tokenize_indobertweet_function, batched=True, remove_columns=["clean_teks"])
test_dataset_indobertweet = test_dataset.map(tokenize_indobertweet_function, batched=True, remove_columns=["clean_teks"])
datasets_indobertweet = DatasetDict({"train": train_dataset_indobertweet, "val": val_dataset_indobertweet, "test": test_dataset_indobertweet})
datasets_indobertweet
```

Map: 0%	0/2865 [00:00<?, ? examples/s]
Map: 100%	2865/2865 [00:00<00:00, 5637.17 examples/s]
Map: 100%	358/358 [00:00<00:00, 5538.49 examples/s]
Map: 100%	359/359 [00:00<00:00, 5521.47 examples/s]


```
# Tokenisasi dataset IndoBERT
train_dataset_indobert = train_dataset.map(tokenize_indobert_function, batched=True, remove_columns=["clean_teks"])
val_dataset_indobert = val_dataset.map(tokenize_indobert_function, batched=True, remove_columns=["clean_teks"])
test_dataset_indobert = test_dataset.map(tokenize_indobert_function, batched=True, remove_columns=["clean_teks"])
datasets_indobert = DatasetDict({"train": train_dataset_indobert, "val": val_dataset_indobert, "test": test_dataset_indobert})
datasets_indobert
```

Map: 0%	0/2865 [00:00<?, ? examples/s]
Map: 100%	2865/2865 [00:00<00:00, 5440.21 examples/s]
Map: 100%	358/358 [00:00<00:00, 5238.62 examples/s]
Map: 100%	359/359 [00:00<00:00, 5712.27 examples/s]

Gambar 6. Tokenisasi IndoBERTweet dan IndoBERT

F. Fine Tuning

Pada bagian ini, akan dijelaskan hasil dari proses *fine-tuning* model IndoBERTweet dan IndoBERT, serta evaluasi kinerja berdasarkan *recall*, *F1-score*, *accuracy*, dan *precision*. Berdasarkan *training arguments* yang digunakan pada Gambar 7, model IndoBERTweet dan IndoBERT dilatih dengan parameter yang sama untuk memastikan perbandingan yang adil. Salah satu parameter penting adalah banyaknya *training epoch* adalah 3. Kemudian, parameter *weight decay* sebesar 0,05. *Weight decay* adalah teknik regularisasi yang digunakan untuk mencegah *overfitting* dengan menambahkan penalti terhadap bobot model selama pelatihan. Nilai 0,05 menunjukkan bahwa setiap pembaruan bobot akan dikurangi sebagian kecil untuk mencegah bobot menjadi terlalu besar. Efeknya, model akan lebih stabil, memiliki generalisasi lebih baik, dan tidak terlalu bergantung pada pola spesifik dalam data pelatihan. Parameter *per_device_train_batch* = 16 dan *per_device_eval_size* = 16 digunakan karena adanya keterbatasan kapasitas GPU yang digunakan dalam penelitian.

```
# Training arguments
from transformers import AutoTokenizer, AutoModelForSequenceClassification, TrainingArguments, Trainer

training_args = TrainingArguments(
    output_dir=".results",
    evaluation_strategy="epoch",
    save_strategy="epoch",
    per_device_train_batch_size=16,
    per_device_eval_batch_size=16,
    num_train_epochs=3,
    weight_decay=0.05,
    logging_dir=".logs",
    save_total_limit=2, # Simpan hanya 2 model terbaru agar tidak boros penyimpanan
    load_best_model_at_end=True, # Ambil model terbaik setelah training selesai
    metric_for_best_model="eval_loss", # Pilih model terbaik berdasarkan loss terendah
    greater_is_better=False, # Loss harus semakin kecil
)
```

Gambar 7. Training Arguments untuk Fine Tuning IndoBERTweet dan IndoBERT

G. Hasil Fine-Tuning IndoBERTTweet

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
1	No log	0.493123	0.782123	0.778950	0.782123	0.776049
2	No log	0.579928	0.807263	0.821468	0.807263	0.810047
3	0.370000	0.699244	0.824022	0.824201	0.824022	0.823889

Gambar 8. Hasil Fine-Tuning IndoBERTTweet

Berdasarkan Gambar 8. hasil *Fine-Tuning* model IndoBERTTweet menunjukkan *validation loss* mengalami peningkatan untuk setiap *epoch* yaitu dari 0,4139 menjadi 0,6992. Akan tetapi, akurasi tetap mengalami peningkatan hingga *epoch* terakhir, yaitu dari 0,7821 menjadi 0,8240. Akurasi tertinggi diperoleh pada *epoch* ke-3 dengan 82,40%, yang menunjukkan bahwa model masih belajar dengan baik. *Precision*, *recall*, dan *F1-score* juga meningkat, menunjukkan keseimbangan performa model dalam klasifikasi sentimen positif, netral, dan negatif. Berdasarkan hasil *fine-tuning* tersebut, *epoch* ke-3 dipilih sebagai model final karena memiliki kombinasi terbaik antara akurasi, *precision*, dan *recall*.

H. Hasil Fine-Tuning IndoBERT

Berdasarkan gambar 9. hasil *Fine-Tuning* model IndoBERT menunjukkan *validation loss* menurun secara bertahap, yaitu dari 0,7102 menjadi 0,6694. Akurasi setiap *epoch* mengalami peningkatan, yaitu dari 0,7067 menjadi 0,7598. Akurasi tertinggi diperoleh pada *epoch* ke-3 dengan 75,98%. *Precision*, *recall*, dan *F1-score* juga meningkat, menunjukkan keseimbangan performa model dalam klasifikasi sentimen positif, netral, dan negatif. Berdasarkan hasil *fine-tuning* tersebut, *epoch* ke-3 dipilih sebagai model final untuk dibandingkan performanya dengan IndoBERTTweet karena memiliki kombinasi terbaik antara akurasi, *precision*, dan *recall*.

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
1	No log	0.710179	0.706704	0.721143	0.706704	0.672178
2	No log	0.666957	0.717877	0.755537	0.717877	0.725118
3	0.649500	0.669429	0.759777	0.755396	0.759777	0.754604

Gambar 9. Hasil Fine-Tuning IndoBERT

Akurasi setiap *epoch* mengalami peningkatan, yaitu dari 0,7067 menjadi 0,7598. Akurasi tertinggi diperoleh pada *epoch* ke-3 dengan 75,98%. *Precision*, *recall*, dan *F1-score* juga meningkat, menunjukkan keseimbangan performa model dalam klasifikasi sentimen positif, netral, dan negatif. Berdasarkan hasil *fine-tuning* tersebut, *epoch* ke-3 dipilih sebagai model final untuk dibandingkan performanya dengan IndoBERTTweet karena memiliki kombinasi terbaik antara akurasi, *precision*, dan *recall*.

I. Perbandingan Performa Fine-Tuning Model IndoBERTTweet dan IndoBERT

Tabel 2. Perbandingan evaluation matrix IndoBERT dan IndoBERTTweet

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
IndoBERT	75,98	75,54	75,98	75,46
IndoBERTTweet	82,40	82,42	82,40	82,38

Berdasarkan Tabel 2 hasil evaluasi menunjukkan bahwa IndoBERTTweet memiliki akurasi 82.40%, lebih tinggi 6.42% dibandingkan IndoBERT (75.98%). Peningkatan akurasi ini menunjukkan bahwa IndoBERTTweet lebih mampu menangkap pola dalam teks media sosial, yang sering menggunakan bahasa informal, slang, serta singkatan yang tidak umum dalam teks formal. IndoBERT memiliki keterbatasan dalam memahami teks informal, yang menyebabkan lebih banyak kesalahan prediksi dalam analisis sentimen.

Precision IndoBERTTweet (82.42%) lebih tinggi dibandingkan IndoBERT (75.54%), yang berarti model ini lebih sedikit menghasilkan *false positive* (prediksi salah sebagai positif). Recall IndoBERTTweet

(82.40%) lebih tinggi dari IndoBERT (75.98%), menunjukkan bahwa model ini lebih baik dalam menangkap sentimen positif, negatif, dan netral tanpa melewatkannya banyak data. *F1-Score* IndoBERTweet (82.38%) lebih tinggi dibandingkan IndoBERT (75.46%), menandakan keseimbangan yang lebih baik antara *precision* dan *recall*.

Hasil perbandingan performa antara IndoBERT dan IndoBERTweet menunjukkan bahwa IndoBERTweet secara konsisten mengungguli IndoBERT dalam semua metrik evaluasi, termasuk akurasi, presisi, recall, dan F1-score. Keunggulan IndoBERTweet dapat dijelaskan melalui beberapa faktor kunci yang berkaitan dengan arsitektur model, data pre-training, serta kesesuaian dengan domain data yang dianalisis dalam studi ini. Pertama, IndoBERTtweet merupakan model transformer yang secara khusus dilatih menggunakan data dari Twitter berbahasa Indonesia. Hal ini membuatnya lebih adaptif terhadap teks informal, slang, singkatan, emotikon, serta struktur kalimat khas media sosial yang tidak baku. Sebaliknya, IndoBERT dilatih menggunakan korpus teks formal seperti Wikipedia dan artikel berita, sehingga tidak optimal ketika diterapkan pada data sosial media yang bersifat noisy dan penuh variasi bahasa.

Kedua, tokenizer dan vocabulary yang digunakan oleh IndoBERTweet telah disesuaikan dengan karakteristik Twitter. Model ini mampu mengenali elemen-elemen khas seperti hashtags, mentions, dan kata-kata tidak baku yang sering muncul dalam percakapan online. Hal ini sangat penting dalam konteks analisis sentimen terhadap kendaraan listrik di Indonesia, di mana opini masyarakat cenderung disampaikan dalam bahasa yang santai, singkat, dan ekspresif. Sementara itu, IndoBERT menggunakan tokenizer standar untuk bahasa formal, yang cenderung gagal mengenali atau mengklasifikasikan kata-kata informal dengan tepat.

Ketiga, IndoBERTweet menunjukkan robustness yang lebih tinggi terhadap data berisik (noisy), seperti kesalahan ketik, struktur kalimat tidak lengkap, dan campuran bahasa. Hal ini membuat model lebih mampu menangkap konteks dan makna dalam opini yang disampaikan secara tidak terstruktur di media sosial. Model ini juga lebih efektif dalam memahami konteks kalimat pendek atau ringkas yang umum ditemukan dalam tweet, berkat pengalaman pre-training-nya yang memang berasal dari data serupa. Secara kuantitatif, peningkatan akurasi dari 75,98% pada IndoBERT menjadi 82,40% pada IndoBERTweet, serta peningkatan F1-score dari 75,46% menjadi 82,38%, menunjukkan bahwa IndoBERTweet tidak hanya akurat tetapi juga lebih seimbang dalam mengklasifikasikan opini positif, netral, dan negatif.

Secara keseluruhan, hasil ini menegaskan pentingnya pemilihan model yang sesuai dengan karakteristik data. Dalam kasus ini, penggunaan IndoBERTweet yang secara eksplisit dirancang untuk teks media sosial memberikan keunggulan signifikan dalam tugas analisis sentimen kendaraan listrik di Indonesia, dibandingkan model umum seperti IndoBERT yang lebih cocok untuk teks formal.

V. KESIMPULAN DAN SARAN

Penelitian ini membandingkan performa IndoBERTweet dan IndoBERT dalam analisis sentimen masyarakat terhadap kendaraan listrik di Indonesia. Berdasarkan hasil *fine-tuning* dan evaluasi model, IndoBERTweet menunjukkan performa yang lebih unggul dibandingkan IndoBERT dalam semua metrik evaluasi, dengan *accuracy* 82.40% dan *F1-score* 82.38%, sementara IndoBERT hanya mencapai *accuracy* 75.98% dan *F1-score* 75.46%. Hal ini menunjukkan bahwa IndoBERTweet lebih efektif dalam menangani bahasa informal, slang, dan karakteristik unik teks media sosial X (Twitter), yang menjadi sumber utama data penelitian ini. Hasil ini mengindikasikan bahwa pemilihan model yang sesuai dengan domain data sangat berpengaruh terhadap performa analisis sentimen.

Untuk meningkatkan keakuratan dan generalisasi model, penelitian selanjutnya dapat menggunakan dataset yang lebih luas dan beragam, mencakup berbagai platform media sosial dan berita online agar model dapat memahami opini dari berbagai sumber. Selain itu, eksplorasi lebih lanjut terhadap teknik *fine-tuning*, seperti *hyperparameter tuning* dan *data augmentation*, dapat dilakukan untuk meningkatkan performa model secara lebih optimal. Dari segi pengembangan model, penelitian lanjutan dapat mempertimbangkan penggabungan IndoBERTweet dengan teknik *ensemble learning*, seperti kombinasi dengan model LSTM atau CNN, dapat diteliti untuk meningkatkan performa analisis sentimen. Implementasi teknik self-supervised learning juga dapat menjadi alternatif dalam menangani keterbatasan

data berlabel, sehingga model dapat belajar dari teks dalam jumlah besar tanpa perlu banyak intervensi manusia. Dengan saran ini, diharapkan penelitian ke depan dapat semakin meningkatkan pemahaman terhadap opini masyarakat mengenai kendaraan listrik serta mendukung pengembangan teknologi *Natural Language Processing* (NLP) berbasis bahasa Indonesia yang lebih akurat dan aplikatif di berbagai bidang.

DAFTAR PUSTAKA

- [1] A. Farhani dan Sutisna, “Analisis Sentimen Terhadap Kendaraan Listrik di Indonesia Menggunakan Metode Klasifikasi Naïve Bayes,” *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 5, no. 3, hlm. 2680–2690, Sep 2024, doi: 10.35870/jimik.v5i3.983.
- [2] R. A. Ekatama, M. Rahardi, A. Aminuddin, dan F. F. Abdulloh, “Sentiment Analysis of Electric Vehicles in Indonesia Using Support Vector Machine and Naïve Bayes,” dalam 2023 3rd International Conference on Smart Cities, Automation & Intelligent Computing Systems (ICON-SONICS), IEEE, Des 2023, hlm. 120–125. doi: 10.1109/ICON-SONICS59898.2023.10435277.
- [3] A. Erfina dan R. A. Lestari, “Sentiment Analysis of Electric Vehicles using the Naïve Bayes Algorithm,” *SISTEMASI*, vol. 12, no. 1, hlm. 178, Jan 2023, doi: 10.32520/stmsi.v12i1.2417.
- [4] S. Alfarizi dan E. Fitriani, “Analisis Sentimen Kendaraan Listrik Menggunakan Algoritma Naive Bayes dengan Seleksi Fitur Information Gain dan Particle Swarm Optimization,” *Indonesian Journal on Software Engineering (IJSE)*, vol. 9, no. 1, hlm. 19–27, 2023, [Daring]. Tersedia pada: <http://ejournal.bsi.ac.id/ejurnal/index.php/ijse>
- [5] Nurul Afifah, Dony Permana, Dodi Vionanda, dan Dina Fitria, “Sentiment Analysis of Electric Cars Using Naive Bayes Classifier Method,” *UNP Journal of Statistics and Data Science*, vol. 1, no. 4, hlm. 289–296, Agu 2023, doi: 10.24036/ujstsds/vol1-iss4/68.
- [6] M. I. Alhari, O. N. Pratiwi, dan M. Lubis, “Sentiment Analysis of The Public Perspective Electric Cars in Indonesia Using Support Vector Machine Algorithm,” dalam 2022 International Conference of Science and Information Technology in Smart Administration (ICSINTESA), IEEE, Nov 2022, hlm. 155–160. doi: 10.1109/ICSINTESA56431.2022.10041604.
- [7] M. Daffa Attariq dan R. Jayadi, “Analysis Of Indonesian People’s Sentiment Towards Electric Cars On Social Media,” *J Theor Appl Inf Technol*, vol. 102, no. 8, 2024, [Daring]. Tersedia pada: www.jatit.org
- [8] A. Sri Widagdo, Ardiansyah, Krisna Nuresa Qodri, Fachruddin Edi Nugroho Saputro, dan Nisrina Akbar Rizky Putri, “Analisis Sentimen Mobil Listrik di Indonesia Menggunakan Long-Short Term Memory (LSTM),” *JURNAL FASILKOM*, vol. 13, no. 3, hlm. 416–423, Des 2023, doi: 10.37859/jf.v13i3.6303.
- [9] R. Merdiansah, S. Siska, dan A. Ali Ridha, “Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT,” *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 7, no. 1, hlm. 221–228, Mar 2024, doi: 10.55338/jikomsi.v7i1.2895.
- [10] M. N. Hidayat dan R. Pramudita, “Analisis Sentimen Terhadap Pembelajaran Secara Daring Pasca Pandemi Covid-19 Menggunakan Metode IndoBERT,” *Information Management For Educators And Professionals : Journal of Information Management*, vol. 8, no. 2, hlm. 161, Jan 2024, doi: 10.51211/imbi.v8i2.2719.
- [11] M. N. Hidayat dan R. Pramudita, “Analisis Sentimen Terhadap Pembelajaran Secara Daring Pasca Pandemi Covid-19 Menggunakan Metode IndoBERT,” *Information Management For Educators And Professionals : Journal of Information Management*, vol. 8, no. 2, hlm. 161, Jan 2024, doi: 10.51211/imbi.v8i2.2719.
- [12] P. Sayarizki dan H. Nurrahmi, “Implementation of IndoBERT for Sentiment Analysis of Indonesian Presidential Candidates,” *Journal on Computing*, vol. 9, no. 2, hlm. 61–72, 2024, doi: 10.34818/indojc.2024.9.2.934.
- [13] N. Nurhasiyah, R. Dwiyansaputra, S. Ika Murpratiwi, dan A. Aranta, “Analisis Sentimen Pengguna Platform Media Sosial X Pada Topik Pemilihan Presiden 2024 Menggunakan Perbandingan Model Monolingual Dan Multilingual BERT,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 1, hlm. 626–634, Des 2024, doi: 10.36040/jati.v9i1.12430.

- [14] J. C. Setiawan, K. M. Lhaksmana, dan B. Bunyamin, “Sentiment Analysis of Indonesian TikTok Review Using LSTM and IndoBERTweet Algorithm,” *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 3, hlm. 774–780, Agu 2023, doi: 10.29100/jipi.v8i3.3911.
- [15] F. Koto, A. Rahimi, J. H. Lau, dan T. Baldwin, “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” Nov 2020, [Daring]. Tersedia pada: <http://arxiv.org/abs/2011.00677>
- [16] J. Howard dan S. Ruder, “Universal Language Model Fine-tuning for Text Classification,” dalam *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2018, hlm. 328–339. doi: 10.18653/v1/P18-1031.
- [17] V. Chandradev, I. Made, A. Dwi Suarjaya, I. Putu, dan A. Bayupati, “Chandradev, Analisis Sentimen Review Hotel menggunakan Metode Deep Learning BERT 107 Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT.”