

PERBANDINGAN KINERJA PRE-TRAINED INDOBERT-BASE DAN INDOBERT-LITE PADA KLASIFIKASI SENTIMEN ULASAN TIKTOK TOKOPEDIA SELLER CENTER DENGAN MODEL INDOBERT

Wildan Amru Hidayat¹, Vinna Rahmayanti Setyaning Nastiti²

^{1,2} Jurusan Informatika Fakultas Teknik Universitas Muhammadiyah Malang
Jln. Raya Tlogomas – Lowokwaru Kota Malang

¹wah110803gmailcom@webmail.um.ac.id

²vinastiti@umm.ac.id

Abstract

Era digital telah membawa revolusi dalam dunia *e-commerce* dengan mengintegrasikan *platform* media sosial dan *platform e-commerce*, yang menghasilkan inovasi seperti aplikasi *TikTok Tokopedia Seller Center*. Aplikasi ini menggabungkan *platform e-commerce* dengan fitur media sosial, memungkinkan pengguna untuk mengelola penjualan sekaligus memperluas jangkauan pasar dan mempromosikan produk melalui video pendek yang interaktif pada *platform* media sosial *TikTok*. Dengan adanya inovasi fitur baru dalam aplikasi ini, penelitian ini melakukan analisis sentimen untuk memahami persepsi dan ulasan berbahasa Indonesia dari para pengguna aplikasi *TikTok Tokopedia Seller Center* menggunakan model *deep learning IndoBERT*. Data ulasan dikumpulkan menggunakan teknik *scraping* pada *Google Play Store* sebanyak 3.145 ulasan yang dilabeli secara manual menjadi 1.755 klasifikasi sentimen negatif dan 1390 klasifikasi sentimen positif. Tahapan *preprocessing* seperti teks *cleaning*, *case folding*, normalisasi teks, dan *stopword removal* dilakukan untuk memberihkan data teks sebelum digunakan untuk pelatihan model. Data yang sudah dibersihkan terbagi menjadi 64% data *training* sebesar 2.012 data, 16% data *validation* sebesar 504 data, dan 20% data *testing* sebesar 629 data. Dua varian *pre-trained* model *IndoBERT*, yaitu *Indobert-base-p2* versi besar dan *Indobert-lite-base-p2* versi lebih ringan digunakan dalam penelitian ini untuk pemrosesan bahasa alami khusus bahasa Indonesia. Hasil penelitian menunjukkan bahwa komparasi model *IndoBERT* dengan kedua *pre-trained* menunjukkan bahwa *pre-trained Indobert-base-p2* mendapatkan hasil akurasi yang lebih unggul dibandingkan *Indobert-lite-base-p2*, dengan akurasi sebesar 97%, presisi sebesar 97%, *recall* sebesar 97%, dan *f1-score* sebesar 97%, sedangkan *pre-trained Indobert-lite-base-p2* dengan akurasi sebesar 94%, presisi sebesar 94%, *recall* sebesar 94%, dan *f1-score* sebesar 94%.

Keywords: Analisis Sentimen, IndoBERT, Pre-trained, E-commerce

I. PENDAHULUAN

Pada era digital saat ini, kolaborasi antara platform media sosial dan e-commerce mencerminkan adaptasi terhadap tren global dalam perdagangan digital. Aplikasi *TikTok Tokopedia Seller Center* adalah contoh terbaru dari integrasi ini, yang menggabungkan *TikTok* sebagai platform media sosial video pendek dengan infrastruktur *e-commerce* dari *Tokopedia*. Pengguna dapat menggunakan fitur ini untuk mempromosikan produk mereka melalui konten video menarik, yang bertujuan untuk meningkatkan interaksi dengan konsumen dan potensi penjualan melalui media sosial *TikTok* [1]. Namun, dengan adanya fitur-fitur baru dalam aplikasi ini, terdapat isu dan tantangan yang dihadapi oleh para penjual, seperti keluhan mengenai fitur, pelayanan, dan pengalaman pengguna secara keseluruhan. Isu-isu ini mendorong penelitian ini untuk

melakukan analisis sentimen, dengan tujuan memahami persepsi dan umpan balik pengguna dengan lebih baik.

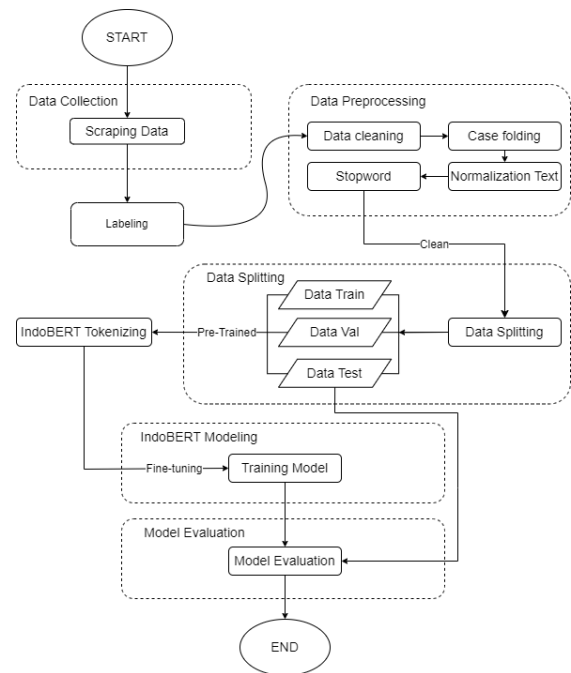
Penelitian terdahulu mengenai analisis sentimen telah banyak dilakukan pada berbagai *platform* media sosial dan *e-commerce*, yang menunjukkan keunggulan berbagai metode machine learning dan deep learning. Pada penelitian oleh M. Isnan et al. [2] menggunakan model *machine learning Support Vector Machine* (SVM) mencapai *F1-score* terbaik dengan akurasi 80%. Penelitian oleh M. E. Purbaya et al. [3] menunjukkan bahwa model *Support Vector Machine* (SVM) mencapai akurasi terbaik sebesar 89.60%. Penelitian lainnya oleh Z. A. Diekson et al. [4], T. Wilianto et al. [5], dan M. J. Hossain et al. [6] menggunakan dataset ulasan *e-commerce* dan membandingkan beberapa model *machine learning*, menemukan bahwa model SVM memberikan akurasi terbaik masing-masing sebesar 84.5%, 78%, dan 94%.

Penelitian mengenai aplikasi *TikTok Shop* telah dilakukan oleh beberapa peneliti dengan hasil yang bervariasi. Penelitian oleh S. J. Sidiq dan A. N. Rachman [7] menggunakan 1000 datasets ulasan *TikTok Shop* yang dilabeli secara manual dan melakukan perbandingan dua model *machine learning*, *Naïve Bayes* dan *Decision Tree*. Hasilnya menunjukkan bahwa model *Decision Tree* mampu menghasilkan akurasi terbaik sebesar 89.6%. Penelitian oleh Lestari et al. [8] menggunakan 875 datasets yang dilabeli secara manual dan hasilnya menunjukkan bahwa model *K-Nearest Neighbor* mencapai akurasi sebesar 91%. Adapun penelitian yang dilakukan C. M. T. Yunanda et al. [9] menggunakan 1229 dataset ulasan *TikTok Shop* yang dilabeli berdasarkan score rating, hasil metode *deep learning Long Short-Term Memory Method (LSTM)* mendapat akurasi sebesar 74%. Penelitian oleh N. Z. AlHabesyah et al. [10] menggunakan 3000 datasets *TikTok Shop* yang dilabeli secara manual dan membandingkan model *machine learning*, *Support Vector Machine (SVM)*, *Decision Tree*, *Random Forest*, dan *deep learning H20*. Hasilnya metode *deep learning H20* mampu memberikan hasil terbaik dengan akurasi terbesar 85%.

Penelitian sebelumnya juga telah menunjukkan bahwa model *deep learning IndoBERT* mampu memberikan kinerja yang baik dalam berbagai tugas pemrosesan bahasa alami (NLP) untuk bahasa Indonesia. Penelitian oleh M. A. Hadiwijaya et al. [11] menemukan bahwa model *IndoBERT* mengungguli model *Logistic Regression* dan *Support Vector Classification (SVC)* dengan akurasi sebesar 91%. Penelitian oleh [12] juga menemukan bahwa model *IndoBERT* yang dikombinasikan dengan *RCNN* menghasilkan akurasi tertinggi sebesar 95.16%. Adapun penelitian oleh W. M. Baihaqi dan A. Munandar [13] menemukan bahwa model *IndoBERT* mengungguli model *Naïve Bayes* dengan akurasi sebesar 85%.

Berdasarkan literatur review, terdapat peluang untuk memberikan kontribusi lebih lanjut dalam analisis sentiment terhadap *TikTok Tokopedia Seller Center* dengan menggunakan model *deep learning IndoBERT*. Model *IndoBERT* ini dirancang untuk memahami konteks dalam teks berbahasa Indonesia. Oleh karena itu, penggunaan *IndoBERT* diharapkan dapat meningkatkan akurasi analisis sentiment terhadap ulasan *TikTok Tokopedia Seller Center* dengan bahasa Indonesia.

II. METODOLOGI PENELITIAN



Gambar 1. Contoh gambar dengan resolusi kurang

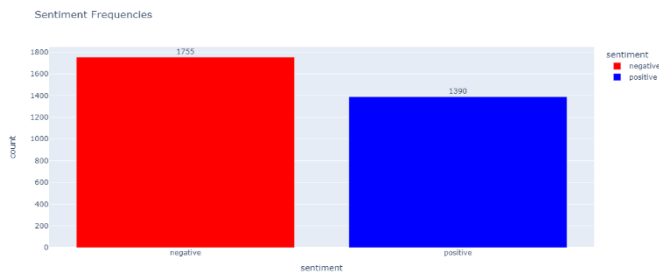
Penelitian ini dimulai dari tahap pengumpulan data. Data yang telah dikumpulkan kemudian melalui proses *preprocessing*, yang mencakup pembersihan untuk menghilangkan *noise* data. Setelah melalui *preprocessing*, dataset akan di *split* menjadi tiga bagian yaitu data *training*, data *validation*, dan data *test*. Pada tahap berikutnya, data yang telah di *split* akan di *pre-trained* dan proses tokenisasi dari *IndoBERT*. Selanjutnya, dataset dimodelkan menggunakan *IndoBERT* melalui proses *fine-tuning*, yang diharapkan menghasilkan evaluasi model yang lebih baik. Berikut dibawah ini merupakan penjelasan tahapan penelitian pada Gambar 1.

A. Data Collection

Dataset yang akan digunakan dalam penelitian ini yaitu data ulasan *TikTok Tokopedia Seller Shop* yang dikumpulkan melalui *platform Google Play Store*. Pengumpulan data dilakukan menggunakan teknik *web scraping* dengan *package* dari *google-play-scraper* yang menyediakan API (*Application Programming Interface*) untuk *python* [14]. *Web scraping* ini menggunakan token '*com.tiktokshop.seller*' dan ulasan disimpan dalam format CSV (*Comma Separated Values*). Dari proses ini, didapatkan dataset sebanyak 3.145 ulasan.

B. Labeling

Proses *labeling* adalah proses memberikan label atau klasifikasi pada data yang akan dianalisis untuk memudahkan model dalam mengklasifikasikan sentimen. Dalam proses pelabelan data penelitian ini memberikan *labeling* secara manual untuk menentukan label setiap ulasan dalam dataset dengan sentimen positif dan negatif.



Gambar 2. Hasil Label

Gambar 2 menunjukkan hasil *Labeling* manual dilakukan oleh tiga orang penilai yang menentukan 1.755 ulasan positif dan 1.390 ulasan negatif.

C. Data Preprocessing

Sebelum pengolahan dataset, dilakukan data preprocessing untuk dapat diklasifikasikan dengan lebih mudah. *Preprocessing* merupakan Langkah penting untuk membersihkan data mentah, yang bertujuan untuk meningkatkan akurasi dan efisiensi model [15]. Berikut ini adalah tahapan dalam preprocessing.

1) Data Cleaning

Data cleaning merupakan proses mengurangi *noise* dengan dilakukan pembersihan dari elemen-elemen teks yang tidak diperlukan [16].

Tabel 1. Data Cleaning

Original Teks	Data Cleaning
Parah, cape bgt sama aplikasinya Ⓟ tiap masukin produk selalu kena pelanggaran, padahal hasil foto sendiri, hasil deskripsi sendiri, udah cape foto, ngetik dan upload, eh pelanggaran. nyesel deh	Parah cape bgt sama aplikasinya tiap masukin produk selalu kena pelanggaran padahal hasil foto sendiri hasil deskripsi sendiri udah cape foto ngetik dan upload eh pelanggaran nyesel deh

2) Case Folding

Case folding merupakan proses yang dilakukan untuk merubah setiap kata yang ada didalam dataset menjadi huruf kecil [17].

Tabel 2. Case Folding

Data Cleaning	Case Folding
Parah cape bgt sama aplikasinya tiap masukin produk selalu kena pelanggaran padahal hasil foto sendiri hasil deskripsi sendiri udah cape foto ngetik dan upload eh pelanggaran nyesel deh	parah cape bgt sama aplikasinya tiap masukin produk selalu kena pelanggaran padahal hasil foto sendiri hasil deskripsi sendiri udah cape foto ngetik dan upload eh pelanggaran nyesel deh

Proses *case folding* pada Tabel 2 menunjukkan bahwa setiap huruf kapital dalam teks diubah menjadi huruf kecil.

3) Normalization Text

Normalization text merupakan proses yang dilakukan untuk mengubah kata singkatan atau kata slang menjadi bentuk normal bahasa Indonesia [18].

Tabel 3. Normalization Text

Case Folding	Normalization Text
parah cape bgt sama aplikasinya tiap masukin produk selalu kena pelanggaran padahal hasil foto sendiri hasil deskripsi sendiri udah cape foto ngetik dan upload eh pelanggaran nyesel deh	parah capek banget sama aplikasinya tiap masukin produk selalu kena pelanggaran padahal hasil foto sendiri hasil deskripsi sendiri udah capek foto ngetik dan upload eh pelanggaran nyesel deh

Proses *normalization text* pada Tabel 3 mengubah kata singkatan seperti 'cape' dan 'bgt' menjadi kata 'capek' dan 'banget' sesuai bentuk normal bahasa Indonesia.

4) Stopword Removal

Stopword Removal merupakan proses yang dilakukan untuk menghilangkan kata-kata yang sering muncul tetapi tidak memberikan kontribusi penting terhadap makna [19].

Tabel 4. Stopword Removal

Normalization Text	Stopword Removal
parah capek banget sama aplikasinya tiap masukin produk selalu kena pelanggaran padahal hasil foto sendiri hasil deskripsi sendiri udah capek foto ngetik dan upload eh pelanggaran nyesel deh	parah capek banget aplikasinya masukin produk kena pelanggaran hasil foto hasil deskripsi udah capek foto ngetik upload eh pelanggaran nyesel deh

Proses *stopword removal* pada Tabel 4 menghilangkan kata-kata seperti 'sama', 'tiap', dan 'padahal', yang termasuk dalam kategori *stopword*.

D. Data Splitting

Data splitting atau pembagian dataset adalah proses penting untuk kinerja yang optimal pada sebuah model mencegah *overfitting* dan memvalidasi kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya [20].

Tabel 5. Hasil Pembagian Data

Split Data	Total
Data Training	2012
Data Validation	504
Data Test	629

Pada Tabel 5, penelitian ini membagi data menjadi tiga proporsi, yaitu data *training* untuk pelatihan dan *fine-tuning* model, data *validation* digunakan untuk mengevaluasi kinerja model selama proses pelatihan dan mencegah *overfitting*, dan data *testing* untuk mengevaluasi kinerja akhir model setelah proses pelatihan selesai. Rasio data dibagi dengan 64% sebagai

data *training*, 16% sebagai data *validation*, dan 20% sebagai data *testing*.

E. IndoBERT Tokenization

Untuk membangun model dengan menggunakan *pre-trained IndoBERT* yang dirancang khusus untuk pemrosesan bahasa Indonesia, diperlukan teknik tokenisasi khusus [21].

Tabel 6. Representasi *IndoBERT Tokenization*

Input	Tokenized	Encode
parah capek banget sama aplikasinya tiap masukin produk selalu kena pelanggaran padahal hasil foto sendiri hasil deskripsi sendiri udah capek foto ngetik dan upload eh pelanggaran nyesel deh	[CLS], parah, capek, banget, aplikasi, ##nya, masuk, ##in, produk, kena, pelanggaran, hasil, foto, hasil, deskripsi, udah, capek, foto, ngetik, upload, eh, pelanggaran, nyesel, deh, [SEP], [PAD], [PAD], [PAD], ... (max length)	[[2, 4735, 11863, 4771, 3376, 2653, 4987, 2604, 5815, 5107, 2294, 3761, 2294, 4386, 2363, 11863, 3761, 7611, 2641, 6385, 2268, 5107, 7593, 2910, 3, 0, 0, 0, ..., (max length)

Representasi *IndoBERT tokenization* pada Tabel 6 dalam penelitian ini menentukan panjang teks maksimum 128 token. *IndoBERT* sendiri memiliki token khusus yaitu [CLS] untuk menentukan kalimat yang menjadi masukan sebagai kategori klasifikasi, [SEP] untuk menandai akhir dari sebuah kalimat, dan [PAD] untuk menambahkan *padding* pada teks yang lebih pendek sehingga semua *input* memiliki panjang yang sama [22]. Tokenisasi *IndoBERT* tidak hanya memecah teks menjadi token-token yang lebih kecil, tetapi juga mampu menangkap makna yang kompleks dari teks. Setelah tahap tokenisasi, dilakukan encoding untuk mengonversi setiap token menjadi representasi numerik yang dapat dimengerti oleh model [23].

F. IndoBERT Modeling

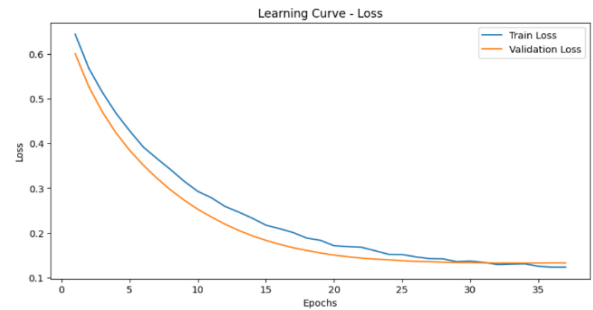
Dalam tahap pemodelan menggunakan *IndoBERT*, penelitian ini menggunakan *Indobert-base-p2* dan *Indobert-lite-base-p2*, dua varian *pre-trained* model *IndoBERT* khusus bahasa Indonesia yang dikembangkan dengan menggunakan arsitektur BERT (*Bidirectional Encoder Representations from Transformers*). Pada proses pelatihan model menggunakan pendekatan *trial and error* atau mencoba berbagai konfigurasi secara berulang untuk menemukan kombinasi parameter yang paling optimal [24]. Setelah menemukan kombinasi parameter, dilakukan *fine-tuning* pada model *pre-trained IndoBERT*. *Fine-tuning* merupakan proses penyesuaian lebih lanjut dari model *pre-trained* dengan tujuan meningkatkan performa model sehingga model mampu memberikan hasil prediksi yang lebih akurat dalam analisis sentimen [25].

Tabel 7. Training Model Parameter

Parameter	Indobert-base-p2	Indobert-lite-base-p2
Optimizer	Adam	Adam

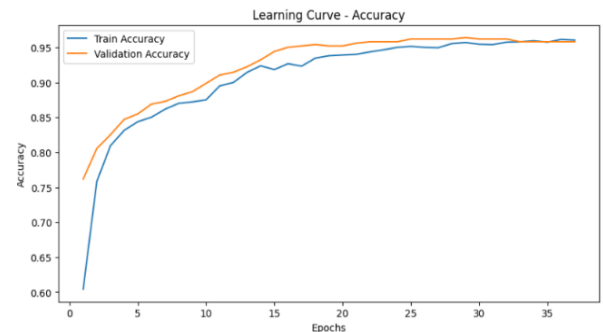
Learning rate	1e-7 (0.0000001)	1e-7 (0.0000001)
Batch size	16	16
Max length	128	128
epoch	37	37

Pada Tabel 7, terdapat hasil kombinasi parameter yang digunakan dalam penelitian ini. Terdapat dua varian *pre-trained* yaitu *Indobert-base-p2* dan *Indobert-lite-base-p2*, dengan kombinasi parameter yang setara di antara keduanya.



G. Model Evaluation

Tahap evaluasi model ini bertujuan untuk menganalisis hasil dari prediksi sentimen yang dilakukan oleh model terhadap ulasan-ulasan dalam dataset. Evaluasi mencakup beberapa metrik seperti akurasi, presisi, *recall* dan *F1-score*, untuk menilai seberapa baik model dalam mengklasifikasikan sentimen [26].



Accuracy suatu algoritma dapat diukur dengan membandingkan jumlah data yang diklasifikasikan dengan benar (TP + TN) dengan total jumlah data yang ada (TP + TN + FP + FN) [27]. Berikut representasi matematis dari metrik *accuracy* pada Rumus 1.

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

Precision suatu algoritma dapat diukur dengan membandingkan jumlah data yang terklasifikasi dengan benar (TP) dengan total data yang telah diprediksi dengan benar (TP

+ FP) [27]. Berikut representasi matematis dari metrik *precision* pada Rumus 2.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall merupakan perbandingan antara jumlah data yang terklasifikasi dengan benar (TP) dengan total jumlah data yang sebenarnya positif (TP + FN) [27]. Berikut representasi matematis dari metrik *recall* pada Rumus 3.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

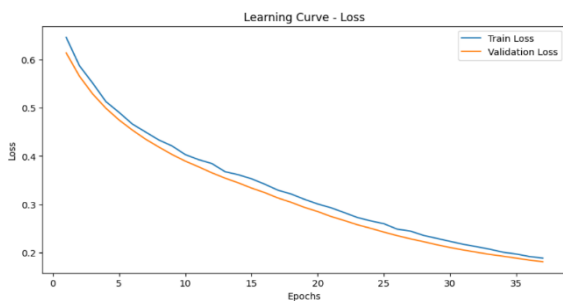
F1-Score adalah pengukuran yang menggabungkan *precision* dan *recall* untuk mengindikasikan keseimbangan antara keduanya [27]. Berikut representasi matematis dari metrik *F1-score* pada Rumus 4.

$$F1 - Score = \frac{2 * precision * recall}{TP + FNprecision + recall} \quad (4)$$

III. HASIL DAN PEMBAHASAN

A. Performa Model

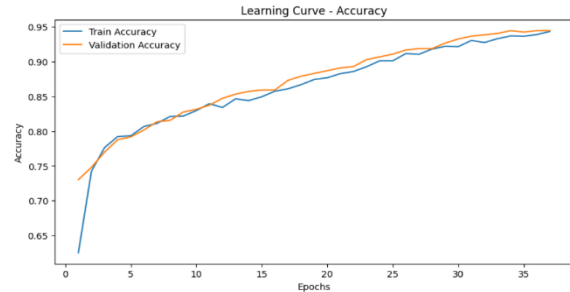
Evaluasi performa model dilakukan dengan menggunakan kurva pembelajaran (*learning curve*) yang mencakup metrik *loss* dan *accuracy*, dengan tujuan memantau dan memahami performa model selama proses pelatihan [28]. Jumlah epoch diperoleh pada setiap pre-trained model dengan menggunakan teknik early stopping. Pada gambar dibawah ini, ditampilkan kurva pembelajaran untuk *loss* dan *accuracy* dari model *IndoBERT* dengan *pre-trained indobert-base-p2* dan *indobert-lite-base-p2* yang sudah disesuaikan dengan best training model parameternya pada Tabel 7.



Gambar 3. Learning Curve *Indobert-base-p2*

Gambar 3 menunjukkan kurva pembelajaran (*learning curve*) untuk *loss* dan akurasi selama 37 epoch. Penurunan *loss* yang stabil pada *train loss* dan *validation loss*, yang mencapai nilai 0.12, mengindikasikan bahwa *pre-trained Indobert-base-p2* memiliki performa yang baik dalam mengurangi *loss* selama

pelatihan. Selain itu, kurva akurasi juga menunjukkan peningkatan yang stabil, di mana *train accuracy* dan *validation accuracy* mencapai nilai 0.97 pada akhir pelatihan. Hasil ini menunjukkan bahwa *pre-trained Indobert-base-p2* mampu memberikan performa yang baik dalam peningkatan akurasi model secara keseluruhan.

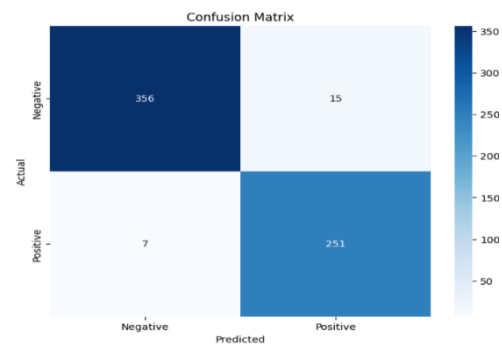


Gambar 4. Learning Curve *Indobert-lite-base-p2*

Gambar 4 menunjukkan kurva pembelajaran (*learning curve*) untuk *loss* dan akurasi pada *pre-trained Indobert-lite-base-p2*. Pada akhir pelatihan, *train loss* dan *validation loss* mencapai nilai 0.18, yang menunjukkan bahwa *loss* pada model ini sedikit lebih tinggi dibandingkan dengan *pre-trained Indobert-base-p2*. Selain itu, kurva akurasi memperlihatkan peningkatan, di mana *train accuracy* dan *validation accuracy* mencapai 94% pada akhir pelatihan, mengindikasikan performa yang cukup baik dari *pre-trained Indobert-lite-base-p2* dalam meningkatkan akurasi model.

B. Evaluasi Model

Hasil klasifikasi yang dilakukan dengan model *IndoBERT* menggunakan *pre-trained Indobert-base-p2* dan *Indobert-lite-base-p2* serta *fine-tuning*, penelitian ini mengevaluasi performa model menggunakan berbagai metrik evaluasi. Metrik ini meliputi *confusion matrix*, akurasi, presisi, *recall*, dan *F1-score*.



Gambar 5. Confusion Matrix *Indobert-base-p2*

Pada gambar 5 merupakan *confusion matrix* dari performa model dengan *pre-trained Indobert-base-p2*. Model berhasil memprediksi 356 sentimen negatif dengan benar (TN) dan 7 salah dari total 363, serta berhasil memprediksi 251 sentimen positif dengan benar (TP) dan 15 salah dari total 266.

Tabel 8. Identifikasi FP dan FN Indobert-base-p2

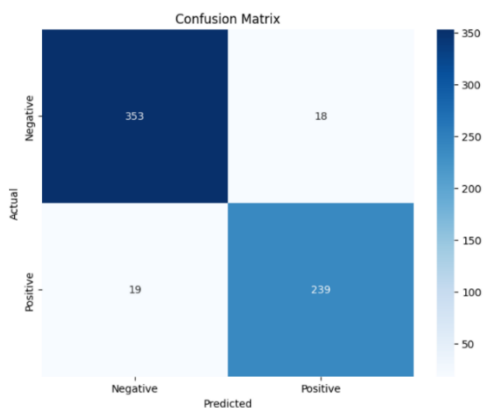
NO	False Positive (FP)	False Negatif (FN)
1	Belibet kategori tolong berguna terima kasih	Management tiktok tolong dipertimbangkan kebijakan berat sebelah kebijakan seller
2	Semoga sistem familiar ekspedisi bertambah pilihannya	dana caricair parah
3	Tolong permudah live kasih trafik	Pelanggaran ampun
4	Membantu afiliator pemula sayajd gampang deskripsiharga promo	Aplikasi bagus bingung penjual promosi

Pada Tabel 8 menunjukkan contoh kesalahan prediksi oleh model *IndoBERT* dengan *pre-trained Indobert-base-p2*, dimana prediksi yang seharusnya positif diprediksi menjadi negatif (FP) dan seharusnya negatif diprediksi menjadi positif (FN).

Tabel 9. Laporan Klasifikasi Indobert-base-p2

	Precision	Recall	F1-Score	Accuracy
Negative	98%	96%	97%	
Positive	94%	97%	96%	
Macro Avg	96%	97%	97%	97%
Weighted Avg	97%	97%	97%	

Pada Tabel 9 menunjukkan laporan klasifikasi model *IndoBERT* dengan *pre-trained Indobert-base-p2* menggunakan metrik *precision*, *recall*, *F1-score*, dan *accuracy*. Untuk masing masing kelas positif dan negatif keseluruhan mencapai akurasi 97%.



Gambar 6. Confusion Matrix Indobert-lite-base-p2

Pada gambar 6 merupakan *confusion matrix* dari performa model dengan *pre-trained Indobert-lite-base-p2*. Model berhasil memprediksi 353 sentimen negatif dengan benar dan 19 salah dari total 372, serta berhasil memprediksi 239 sentimen positif dengan benar dan 18 salah dari total 257.

Tabel 10. Identifikasi FP dan FN Indobert-lite-base-p2

NO	False Positive (FP)	False Negatif (FN)
1	Tolong aplikasi diinstal	Management tiktok tolong dipertimbangkan kebijakan berat sebelah kebijakan seller
2	Aplikasi ok	Pelanggaran ampun
3	Ya mudah mudahan akun bantu penjualan	Pelanggaran ampun
4	Daftar bisa	Bahasanya tolong diganti pakai bahasa indonesia sih pilihannya ktp pakai paspor asli indonesia

Pada Tabel 10 menunjukkan contoh kesalahan prediksi oleh model *IndoBERT* dengan *pre-trained Indobert-lite-base-p2*, dimana prediksi yang seharusnya positif diprediksi menjadi negatif (FP) dan seharusnya negatif diprediksi menjadi positif (FN).

Tabel 11. Laporan Klasifikasi Indobert-lite-base-p2

	Precision	Recall	F1-Score	Accuracy
Negative	95%	95%	95%	
Positive	93%	93%	93%	
Macro Avg	94%	94%	94%	94%
Weighted Avg	94%	94%	94%	

Pada Tabel 11 menunjukkan laporan klasifikasi model *IndoBERT* dengan *pre-trained Indobert-lite-base-p2* menggunakan metrik *precision*, *recall*, *F1-score*, dan *accuracy*. Untuk masing masing kelas positif dan negatif keseluruhan mencapai akurasi 94%.

IV. KESIMPULAN

Penelitian ini membandingkan kinerja dua varian model *pre-trained IndoBERT*, yaitu *IndoBERT-base-p2* dan *IndoBERT-lite-base-p2*, dalam klasifikasi sentimen ulasan berbahasa Indonesia pada aplikasi *TikTok Tokopedia Seller Center*. Penelitian ini menggunakan *optimizer Adam* dengan *learning rate* sebesar $1e-7$ (0.0000001), *batch size* 16, dan *max length* 128, yang dilatih selama 37 *epoch*. Data dibagi dengan rasio 64% untuk data *training*, 16% untuk data *validation*, dan 20% untuk data *testing*. Hasil analisis menunjukkan bahwa model *IndoBERT-base-p2* memiliki kinerja lebih unggul dengan akurasi 97%, sementara model *IndoBERT-lite-base-p2* mencapai akurasi 94%. Selain itu, metrik evaluasi lainnya seperti *precision*, *recall*, dan *F1-score* juga menunjukkan performa yang lebih baik pada model *IndoBERT-base-p2* dibandingkan dengan *IndoBERT-lite-base-p2*. Oleh karena itu, penggunaan model *IndoBERT* dengan *pre-trained* yang lebih besar seperti *IndoBERT-base-p2* dinilai lebih efektif untuk mencapai akurasi yang terbaik dalam analisis sentimen.

V. SARAN

Berdasarkan penelitian yang telah dilakukan terdapat saran untuk penelitian selanjutnya, untuk mengembangkan model dengan menggunakan *pre-trained* berbahasa Indonesia yang lebih besar, yaitu *Indobert-large-p2*. Diharapkan dengan penggunaan model yang lebih besar, akurasi dan performa dalam klasifikasi sentimen dapat lebih ditingkatkan lagi.

REFERENSI

- [1] A. Andini, D. Ramadani, F. H. Jafar, and R. E. Mayasari, "Legal Review of Tik Tok Shop Re-Operation on The Tik Tok Social Media Application," vol. 1, no. 1, pp. 1–9, 2024.
- [2] M. Isnain, G. N. Elwirehardja, and B. Pardamean, "Sentiment Analysis for TikTok Review Using VADER Sentiment and SVM Model," *Procedia Comput. Sci.*, vol. 227, pp. 168–175, 2023, doi: 10.1016/j.procs.2023.10.514.
- [3] M. E. Purbaya, D. Putra Rakhmadani, M. Puspa Arum, and L. Zian Nasifah, "Comparison of Kernel Support Vector Machines in Conducting Sentiment Analysis Review of Buying Chips on the Shopee E- Marketplace in Indonesian," in *2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, IEEE, Nov. 2022, pp. 435–440. doi: 10.1109/ICIMCIS56303.2022.10017546.
- [4] Z. A. Diekson, M. R. B. Prakoso, M. S. Q. Putra, M. S. A. F. Syaputra, S. Achmad, and R. Sutoyo, "Sentiment analysis for customer review: Case study of Traveloka," *Procedia Comput. Sci.*, vol. 216, pp. 682–690, 2023, doi: 10.1016/j.procs.2022.12.184.
- [5] T. Willianto, Supryadi, and A. Wibowo, "Sentiment Analysis on E-commerce Product using Machine Learning and Combination of TF-IDF and Backward Elimination," *Int. J. Recent Technol. Eng.*, vol. 8, no. 6, pp. 2862–2867, Mar. 2020, doi: 10.35940/ijrte.F7889.038620.
- [6] M. J. Hossain, D. Das Joy, S. Das, and R. Mustafa, "Sentiment Analysis on Reviews of E-commerce Sites Using Machine Learning Algorithms," in *2022 International Conference on Innovations in Science, Engineering and Technology (ICISSET)*, IEEE, Feb. 2022, pp. 522–527. doi: 10.1109/ICISSET54810.2022.9775846.
- [7] S. Jafar Sidiq and A. Nur Rachman, "Analysis Of Twitter User Sentiment To Tiktok Shop Using Naïve Bayes And Decision Tree Algorithms," *Int. J. Appl. Inf. Syst. Informatics*, vol. 1, no. 1, Nov. 2023, doi: 10.37058/jaisi.v1i1.8990.
- [8] J. Mantik *et al.*, "Application Of N-Gram On K-Nearest Neighbor Algorithm To Sentiment Analysis Of TikTok Shop Shopping Features," *J. Mantik*, vol. 6, no. 3, pp. 2685–4236, 2022.
- [9] C. M. T. Y. M. H. W. M. P. Dhuhita, "Sentiment Analysis on TikTok Shop Reviews Using Long Short-Term Memory Method to Find Business Opportunity," *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, no. Vol. 9 No. 1 (2024), pp. 1–7, 2024, [Online]. Available: <https://ejournal.unitomo.ac.id/index.php/inform/article/view/6524/3258>
- [10] N. Z. Al Habesyah, R. Herteno, F. Indriani, I. Budiman, and D. Kartini, "Sentiment Analysis of TikTok Shop Closure in Indonesia on Twitter Using Supervised Machine Learning," *J. Electron. Electromed. Eng. Med. Informatics*, vol. 6, no. 2, pp. 148–156, Apr. 2024, doi: 10.35882/jeeemi.v6i2.381.
- [11] M. A. Hadiwijaya, F. P. Pirdaus, D. Andrews, S. Achmad, and R. Sutoyo, "Sentiment Analysis on Tokopedia Product Reviews using Natural Language Processing," in *2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS)*, IEEE, Nov. 2023, pp. 380–386. doi: 10.1109/ICIMCIS60089.2023.10348996.
- [12] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, "Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN," *Ilk. J. Ilm.*, vol. 14, no. 3, pp. 348–354, Dec. 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.
- [13] W. M. Baihaqi and A. Munandar, "Sentiment Analysis of Student Comment on the College Performance Evaluation Questionnaire Using Naïve Bayes and IndoBERT," *JUITA J. Inform.*, vol. 11, no. 2, p. 213, Nov. 2023, doi: 10.30595/juita.v11i2.17336.
- [14] P. Kaur, "Sentiment analysis using web scraping for live news data with machine learning algorithms," *Mater. Today Proc.*, vol. 65, pp. 3333–3341, 2022, doi: 10.1016/j.matpr.2022.05.409.
- [15] S. G. C. G and B. S. -, "Grid Search Tuning of Hyperparameters in Random Forest Classifier for Customer Feedback Sentiment Prediction," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 9, 2020, doi: 10.14569/IJACSA.2020.0110920.
- [16] M. Khader, A. Awajan, and G. Al-Naymat, "The Effects of Natural Language Processing on Big Data Analysis: Sentiment Analysis Case Study," in *2018 International Arab Conference on Information Technology (ACIT)*, IEEE, Nov. 2018, pp. 1–7. doi: 10.1109/ACIT.2018.8672697.
- [17] C. Slamet, A. R. Atmadja, D. S. Maylawati, R. S. Lestari, W. Darmalaksana, and M. A. Ramdhani, "Automated Text Summarization for Indonesian Article Using Vector Space Model," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 288, p. 012037, Jan. 2018, doi: 10.1088/1757-899X/288/1/012037.
- [18] F. Hemmatian and M. K. Sohrabi, "A survey on classification techniques for opinion mining and sentiment analysis," *Artif. Intell. Rev.*, vol. 52, no. 3, pp. 1495–1545, Oct. 2019, doi: 10.1007/s10462-017-

- 9599-6.
- [19] J. Singh and P. Tripathi, "Sentiment analysis of Twitter data by making use of SVM, Random Forest and Decision Tree algorithm," in *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*, IEEE, Jun. 2021, pp. 193–198. doi: 10.1109/CSNT51715.2021.9509679.
- [20] Y. Xu and R. Goodacre, "On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning," *J. Anal. Test.*, vol. 2, no. 3, pp. 249–262, Jul. 2018, doi: 10.1007/s41664-018-0068-2.
- [21] J. H. Computer, S. M. Honova, V. P. Computer, C. A. Setiawan, I. H. Parmonangan, and Diana, "Sentiment Analysis of Skincare Product Reviews in Indonesian Language using IndoBERT and LSTM," in *2023 IEEE 9th Information Technology International Seminar (ITIS)*, IEEE, Oct. 2023, pp. 1–6. doi: 10.1109/ITIS59651.2023.10420222.
- [22] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. Mlm, pp. 4171–4186, 2019.
- [23] A. Nayak, H. Timmapathini, K. Ponnalagu, and V. Gopalan Venkoparao, "Domain adaptation challenges of BERT in tokenization and sub-word representations of Out-of-Vocabulary words," in *Proceedings of the First Workshop on Insights from Negative Results in NLP*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 1–5. doi: 10.18653/v1/2020.insights-1.1.
- [24] H. D. Sharma and P. Goyal, "An Analysis of Sentiment: Methods, Applications, and Challenges," in *RAiSE-2023*, Basel Switzerland: MDPI, Dec. 2023, p. 68. doi: 10.3390/engproc2023059068.
- [25] K. S. Nugroho, A. Y. Sukmadewa, H. W. DW, F. A. Bachtiar, and N. Yudistira, "BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews," Jul. 2021, doi: 10.1145/3479645.3479679.
- [26] R. Qasim, W. H. Bangyal, M. A. Alqarni, and A. Ali Almazroi, "A Fine-Tuned BERT-Based Transfer Learning Approach for Text Classification," *J. Healthc. Eng.*, vol. 2022, pp. 1–17, Jan. 2022, doi: 10.1155/2022/3498123.
- [27] K. Bhowmick and V. Sarvaiya, "A Comparative Study Of The Different Classification Algorithms On Football Analytics," *Int. J. Adv. Res.*, vol. 9, no. 08, pp. 392–407, Aug. 2021, doi: 10.21474/IJAR01/13280.
- [28] M. Totox and H. F. Pardede, "Exploring the Effectiveness of Deep Learning in Analyzing Review Sentiment," *JIKO (Jurnal Inform. dan Komputer)*, vol. 6, no. 2, Aug. 2023, doi: 10.33387/jiko.v6i2.6372.