

Analisis Sentimen Ulasan Aplikasi DLU Ferry Pada Google Play Store Menggunakan *Bidirectional Encoder Representations from Transformers*

(Sentiment Analysis of DLU Ferry Application Reviews on the Google Play Store Using Bidirectional Encoder Representations from Transformers)

Ekka Pujo Ariesanto Akhmad

**Program Studi Manajemen Pelabuhan, Program Diploma Pelayaran,
Universitas Hang Tuah**

Abstrak: *DLU Ferry* merupakan aplikasi yang dikeluarkan oleh PT. Dharma Lautan Utama untuk memudahkan pelanggan dalam melakukan pemesanan tiket. Aplikasi *DLU Ferry* masih memiliki kelemahan, oleh karena itu perlu perbaikan untuk meningkatkan kualitas aplikasi ini. Untuk mengetahui kelemahan aplikasi ini dapat diperoleh dari ulasan yang ditulis konsumen pada *Google Play Store*. Penelitian ini berisi tentang analisis data ulasan konsumen terhadap aplikasi *DLU Ferry*. Penelitian ini menggunakan data sebanyak 1575 ulasan yang terdiri dari sentimen positif, netral, dan negatif. Penelitian ini akan mengukur performa metode *Bidirectional Encoder Representations from Transformers (BERT)* dalam melakukan klasifikasi sentimen menggunakan *Pretrained* model *IndoBERT-base* dengan teknik *fine-tuning*. Hasil pengujian pada penelitian memperoleh akurasi sebesar 86% dengan pemilihan *hyperparameter*, yaitu *batch size* 32, *learning rate* $3e-6$, dan *epoch* 5.

Kata kunci: Analisis Sentimen, *DLU Ferry*, *Google Play Store*, *BERT*, *hyperparameter*

Abstract: *DLU Ferry* is an application issued by PT. Dharma Lautan Utama to make it easier for customers to order tickets. The *DLU Ferry* application still has weaknesses, therefore it needs improvement to improve the quality of this application. To find out the weaknesses of this application can be obtained from reviews written by consumers on the *Google Play Store*. This research contains analysis of consumer review data on the *DLU Ferry* application. This study uses data from 1575 reviews consisting of positive, neutral and negative sentiments. This research will measure the performance of the *Bidirectional Encoder Representations from Transformers (BERT)* method in classifying sentiments using the *IndoBERT-base Pretrained* model with *fine-tuning* techniques. The test results in this study obtained an accuracy of 86% with the selection of *hyperparameters*, namely *batch size* 32, *learning rate* $3e-6$, and *epoch* 5.

Keywords: *Sentiment Analysis*, *DLU Ferry*, *Google Play Store*, *BERT*, *hyperparameter*

Alamat Korespondensi:

Ekka Pujo Ariesanto Akhmad, Program Studi Manajemen Pelabuhan dan Logistik Maritim, Fakultas Vokasi Pelayaran, Universitas Hang Tuah, Jalan A. R. Hakim 150, Surabaya. e-mail: jurnal.pdp@hangtuah.ac.id

PENDAHULUAN

PT. Dharma Lautan Utama (DLU) adalah sebuah perusahaan yang melayani transportasi laut dan penyeberangan feri di Indonesia (Jawa, Kalimantan, Sulawesi, dan Nusa Tenggara).

PT. DLU mengeluarkan aplikasi berbasis android, yaitu *DLU Ferry*. Aplikasi ini dibuat untuk memudahkan pelanggan dalam melakukan pemesanan tiket secara *online*.

Layanan tiket secara *online* kini sudah menjadi keharusan bagi perusahaan jasa. Pengadaan pembelian tiket elektronik ini untuk mengendalikan penjualan, sehingga perusahaan mengetahui jumlah penumpang yang naik secara *online*. Selain itu, penerapan pembelian tiket elektronik ini juga untuk menghilangkan penumpang gelap dalam kapal.

Aplikasi *DLU Ferry* telah diunduh dan mendapat ulasan oleh konsumen atau calon penumpang kapal

DLU. Ulasan tersebut dapat digunakan sebagai indikator kualitas layanan dan sumber informasi bagi perusahaan (Anand dan Naorem, 2016).

Namun, banyaknya ulasan membuat analisis menjadi sulit bagi perusahaan. Oleh karena itu, analisis sentimen diperlukan untuk mengolah data dan menganalisis ulasan yang ada untuk mengklasifikasikannya menjadi ulasan positif, netral, atau negatif secara otomatis.

Analisis sentimen adalah teknik yang mengkaji pemikiran pengguna yang diungkapkan melalui berbagai media, termasuk media cetak, media sosial, blog, dan situs web, untuk menentukan kepuasan pengguna dan persepsi mereka terhadap suatu isu (Pang dan Lee, 2008). Analisis sentimen ulasan digunakan dalam berbagai konteks, seperti ulasan produk (Rahul dan Monika, 2019).

Tujuan dalam penelitian ini adalah

Mengetahui tingkat akurasi dan kinerja *IndoBERT* melakukan analisis sentimen ulasan aplikasi *DLU Ferry*.

Penelitian Terdahulu

Warga negara Indonesia menjadi ekspresif terhadap vaksin *Covid-19*, sehingga mereka menyampaikan pendapat dengan mengunggah teks di Twitter. Ekspresi tersebut dipelajari dengan menggunakan kerangka pembelajaran mendalam, yaitu *BERT*, *CNN-LSTM*, dan *IndoBERTweet* untuk memperoleh pengetahuan tentang kategori ucapan negatif seperti emosi, kecemasan, panik, atau ucapan positif seperti vaksin apakah bekerja dengan baik. Lalu, ketiga metode ini berhasil melakukan prediksi sentimen tentang vaksinasi menggunakan dataset tweets di Twitter bulan Januari 2021 hingga Maret 2022. Penelitian menggunakan

IndoBERT berhasil mengklasifikasikan sentimen menjadi sentimen positif sekitar 80%, kemudian *IndoBERTweet* pada 68%, lalu menggunakan *CNN-LSTM* mencapai 53% dengan penggunaan sebanyak 2020 dataset dari Twitter (Saadah, dkk., 2022).

Atmaja dan Yustanti (2021) melakukan analisis sentimen aplikasi Ruang Guru di *Google Play Store*. Data ulasan diambil dari fitur komentar yang ada di *Google Play Store* menggunakan teknik *scraping*. Data memakai 5437 baris. Luaran analisis sentimen dimanfaatkan untuk melihat reaksi pengguna Ruang Guru kemudian pengembang dapat meningkatkan fitur yang dianggap minim oleh pengguna. *Wordcloud* pada ulasan Ruang Guru memuat keluhan aplikasi, video yang berbayar, dan aplikasi yang crash. Metode yang digunakan adalah model *pre-trained BERT*. Model meraih nilai *F1-Score* 98,9% dengan proporsi data latih dan data uji 70:30. Lalu, evaluasi dilakukan terhadap model menghasilkan akurasi sebesar 99%, presisi sebesar 64,13%, *recall* sebesar 60,51%. Nilai kriteria persentase sentimen memiliki bobot 99% dan bisa dinyatakan valid menurut 10 kali *epoch* di rata-rata dengan peningkatan yang konsisten.

Prananda dan Thalib (2020) melakukan analisis sentimen untuk mengidentifikasi analisis bisnis intelijen di GO-JEK. Penelitian menggunakan posting Twitter yang dikumpulkan dari pustaka Twint terdiri dari 3111 tweet. Karena kumpulan data tidak memberikan kebenaran dasar, penelitian menggunakan *Microsoft Text Analytic* untuk menentukan sentimen positif, netral, dan negatif. Sebelum menerapkan *Microsoft Text Analytic*, penelitian melakukan langkah pra-pemrosesan untuk menghapus data yang tidak

diinginkan, seperti duplikat tweet, gambar, alamat situs web, dan lain-lain. Menurut *Microsoft Text Analytic*, hasilnya adalah 666 sentimen positif, 2055 sentimen netral, dan 127 sentimen negatif. Berdasarkan hasil tersebut, penelitian menyimpulkan sebagian besar pelanggan GO-JEK adalah puas dengan layanan GO-JEK. Penelitian ini juga mengembangkan model klasifikasi untuk memprediksi analisis sentimen dari data baru. Penelitian menggunakan beberapa algoritma *classifier* seperti *Decision Tree*, *Naïve Bayes*, *Support Vector Machine*, dan *Neural Network*. Hasilnya, menunjukkan *Decision Tree* memberikan kinerja terbaik.

Penelitian ini menggunakan pijakan algoritma *BERT* oleh Rauf, dkk. (2019). Penelitian mempunyai perbedaan data dan penerapan terhadap subjek yang berbeda. Penulis menggunakan data ulasan *DLU Ferry* yang discraping dari *Google Play Store*, sedangkan penelitian milik Rauf, dkk. (2019) menggunakan data *IMDB Movie Reviews*. Pada penelitian ini penulis menggunakan algoritma *BERT*, karena menurut penelitian sebelumnya dinilai mempunyai tingkat akurasi lebih baik daripada metode lainnya. Penelitian ini memiliki persamaan dengan penelitian Rauf, dkk. (2019) pada penggunaan *library deep learning*. Penelitian ini menggunakan *Pytorch* dan *Tensorflow*. *Wordcloud* digunakan penulis untuk menggambarkan *corpus* luaran *text preprocessing*.

Bidirectional Encoder Representations from Transformers (BERT)

Bidirectional Encoder Representations from Transformers (BERT) merupakan model representasi bahasa terlatih yang pertama kali dipopulerkan oleh Google pada 11

Oktober 2018. Tidak seperti model bahasa lainnya, *BERT* dibuat sebagai *pre-trained model* atau model yang sudah dilatih secara *bidirectional* dari data teks yang tidak berlabel dengan cara menggabungkan konteks dari sisi kiri dan kanan layer. Dengan cara ini model *BERT* dapat *finetuning* dengan tambahan satu layer saja (Devlin, dkk., 2019).

IndoBERT

IndoBERT merupakan pre-trained model berdasarkan *BERT* untuk bahasa Indonesia. IndoBERT menggunakan dataset Indo4B yang berisi lebih dari 23 GB data teks bahasa Indonesia, terdiri dari 4 Milliar kata baik formal maupun *colloquial* (bahasa sehari-hari) dari berbagai sumber diantaranya media sosial, blog, berita, dan *website* (Wilie, dkk., 2020). IndoBERT menggunakan *SentencePiece* dengan *Byte Pair Encoding (BPE)* tokenizer sebagai metode pembentukan *vocabulary*. *SentencePiece* merupakan *language-independent* subword tokenizer dan detokenizer yang didesain untuk neural-based textprocessing.

Terdapat empat model pada IndoBERT yaitu IndoBERT_{BASE}, IndoBERT_{LARGE}, IndoBERT-lite_{BASE}, IndoBERT-lite_{LARGE}. Semua model IndoBERT dilatih menggunakan TPU v3-8 dalam dua fase. Pada fase pertama, model dilatih dengan *maximum sequence length* 128, dan pada fase kedua 512. Dari segi akurasi, IndoBERT_{LARGE} memiliki akurasi yang lebih tinggi, namun *memory* yang dibutuhkan sangat besar.

Confusion Matrix

Kinerja model klasifikasi dievaluasi menggunakan confusion matrix seperti akurasi, precision, recall, specificity, dan F1 score (Hemmatian dan Sohrabi, 2019). Tabel 1 menunjukkan struktur confusion matrix.

Tabel 1. Confusion Matrix

Predicted	Actual	
	Positive	Negative
Positive	TP	FP
Negative	FN	TN

TP, True Positive; FN, False Negative;
FP, False Positive; TN, True Negative

Penelitian ini menggunakan akurasi untuk mengukur seberapa akurat model dapat mengklasifikasikan data dengan benar.

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

Precision menggambarkan tingkat keakuratan antara data prediksi benar positif yang diminta dengan hasil prediksi yang diberikan oleh model.

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

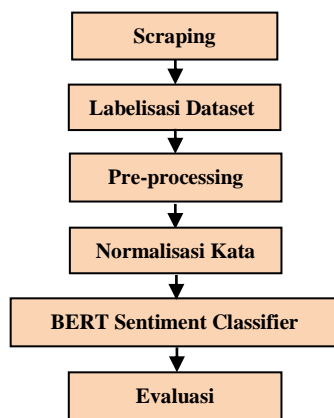
Recall menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi.

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

F1 score merupakan perbandingan rata-rata precision dan recall yang dibobotkan.

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

METODE PENELITIAN

**Gambar 1. Alur Penelitian**

Gambar 1 menjelaskan analisis sentimen diawali dengan mengerjakan *scraping* atau ekstraksi data dari halaman web Google Play Store. Luaran dari *scraping* lalu disatukan jadi sebuah dataset. Kemudian, dataset diberi label *positive*, *neutral*, dan *negative*. Berikutnya dataset masuk ke dalam tahap *pre-processing*. *Pre-processing* dataset adalah tahap untuk mempersiapkan data yang awalnya tidak terstruktur menjadi data yang lebih terstruktur agar dapat diproses untuk langkah berikutnya. Beberapa tahapan pada *pre-processing*, yaitu *case folding*, *data cleaning*, tokenisasi, dan normalisasi kata tidak baku. Setelah melewati proses tersebut, kemudian dataset dilatih menggunakan IndoBERT untuk diklasifikasi menjadi tiga bagian, yaitu positif, negatif, dan netral. Data yang sudah diklasifikasi lalu dievaluasi untuk melihat akurasi prediksi (Putra, 2021).

HASIL DAN PEMBAHASAN

Scraping

Scraping dikerjakan dengan *library google-play-scraper* yang diinstall lewat *pip python*. Kemudian, penyeleksian fitur dikerjakan pada dataset. Seleksi fitur dikerjakan dengan menghapus kolom *reviewid*, *username*, *userimage*, *thumbsUpCount*, *reviewCreatedVersion*, *tanggal (at)*, *replyContent*, *repliedAt*, dan *appVersion* karena dianggap tidak dibutuhkan untuk analisis sentimen. Lalu, dataset disimpan dalam format *tab separated value* (.tsv). Luaran *scraping* dapat dilihat pada Gambar 2.

A	B
1. Tiket kapal jadwalnya gk nongol. Menghubungi contact person tapi gk direpson	1
2. Saya lupa password gimana cara masuk lagi	4
3. Aplikasi apaan nih msi bolehin pake akun random menugikan banget ada orang ga di kenal pake	1
4. Kenapa kok mau masuk salah email nya	1
5. Layanan aplikasi sangat baik dan memuaskan... menuding aplikasinya di tutup aja	5
6. 1 keluhan yang utama, coba di kasih fitur cancel (refund) agar memudahkan ketika terjadi klu	3
7. Klaimmu gllleeeeeeeeeee kau mau jual tiket atau jual kltan. Burukkkk benar	1
8. Ijin cerita min... kmren mudik saya beli tiket untuk kelas VIP... saya kira emang ada private untuk k	3
9. Jadwal jam 12 berangkat jam 2	3
10. Ipp	5
11. Aplikasi nya sangat menyenangkan	5
12. Saya kecewa menggunakan kapal Darma Lautan Utama KDM. Dharmia Kartika 9 kelas III tdk ada privi	1
13. Lebih baik apl ini dibubarkan aja... tiket kendaraan sudah dibeli sudah dibayar giriran tiket sopirny	3
14. Hati2 pembelian tiket melalui aplikasi ini, uang sudah diklaim tapi tiket tidak terbit dan itu artiny	3
15. Cari tiket nya sulit mending cari emas bawa uang langsung dapet emas fa ini cari tiket aja sulit bar	2
16. Trims aplikasi	5
17. Ada yg tau cara dapetin tiket kelas 7? Karena tiap buka aplikasi kok ga ada... apalagi tebaran... Gila	1
18. Semua pembayaran online gabis dilakukan, makudnya apa yah? Bim juga sepuluh menit siese i	3
19. Terakhir pake km kirana parepare balikpapan, parah delaynya sampai 7 jam, loading kendaraan g	3
20. Petugas informasinya ga ramah, mau upgrade ke kelas 1 gabis dg muka jutek, padahal kelas2 kele	3
21. Trk apakah aplikasi error di klu pesan kok tidak bisa??	1
22. Kurang siso, msh sering error, nesi data cacat2 error hmmm	1

Gambar 2. Luaran scraping Labelisasi Dataset

Karena, analisis sentimen menggunakan metode *supervised learning* dibutuhkan dataset yang telah mempunyai label. Labelisasi dikerjakan penulis dengan membubuhkan label *negative* pada data rating 1 dan 2, label *neutral* pada data rating 3, dan label *positive* pada data rating 4 dan 5. Sesudah itu dijalankan pemeriksaan kesesuaian label dengan teks ulasan secara langsung, karena ada ulasan yang tidak signifikan dengan rating yang dituliskan, dan cukup banyak data yang bersifat mendua. Labelisasi data luaran scraping dapat dilihat pada Gambar 3.

1. Tiket kapal jadwalnya gk nongol. Menghubungi contact person tapi gk direpson	negative
2. Saya lupa password gimana cara masuk lagi	positive
3. Aplikasi apaan nih msi bolehin pake akun random menugikan banget ada orang ga di kenal pake	negative
4. Kenapa kok mau masuk salah email nya	negative
5. Layanan aplikasi sangat baik dan memuaskan... menuding aplikasinya di tutup aja	negative
6. 1 keluhan yang utama, coba di kasih fitur cancel (refund) agar memudahkan ketika terjadi klu	neutral
7. Klaimmu gllleeeeeeeeeee kau mau jual tiket atau jual kltan. Burukkkk benar	negative
8. Ijin cerita min... kmren mudik saya beli tiket untuk kelas VIP... saya kira emang ada private untuk k	neutral
9. Jadwal jam 12 berangkat jam 2	neutral
10. Ipp	positive
11. Aplikasi nya sangat menyenangkan	positive
12. Saya kecewa menggunakan kapal Darma Lautan Utama KDM. Dharmia Kartika 9 kelas III tdk ada privi	negative
13. Lebih baik apl ini dibubarkan aja... tiket kendaraan sudah dibeli sudah dibayar giriran tiket sopirny	negative
14. Hati2 pembelian tiket melalui aplikasi ini, uang sudah diklaim tapi tiket tidak terbit dan itu artiny	negative
15. Cari tiket nya sulit mending cari emas bawa uang langsung dapet emas fa ini cari tiket aja sulit bar	negative
16. Trims aplikasi	positive
17. Ada yg tau cara dapetin tiket kelas 7? Karena tiap buka aplikasi kok ga ada... apalagi tebaran... Gila	negative
18. Semua pembayaran online gabis dilakukan, makudnya apa yah? Bim juga sepuluh menit siese i	negative
19. Terakhir pake km kirana parepare balikpapan, parah delaynya sampai 7 jam, loading kendaraan g	negative
20. Petugas informasinya ga ramah, mau upgrade ke kelas 1 gabis dg muka jutek, padahal kelas2 kele	negative
21. Trk apakah aplikasi error di klu pesan kok tidak bisa??	negative
22. Kurang siso, msh sering error, nesi data cacat2 error hmmm	negative

Gambar 3. Labelisasi Data Pre-processing

Pre-processing dilakukan untuk mengubah dataset yang tidak terstruktur menjadi terstruktur, sehingga memudahkan data untuk diproses dengan melakukan beberapa tahapan, yaitu case folding, data cleaning, tokenisasi, dan normalisasi kata tidak baku.

Case Folding

Case folding digunakan untuk mengubah semua karakter menjadi huruf kecil. Tahapan ini dilakukan karena data yang diperoleh tidak selalu terstruktur dan konsisten dalam penggunaan huruf kapital, maka case folding dilakukan untuk menyamaratakan penggunaan huruf kapital. *Case folding* dilakukan dengan menggunakan fungsi `lower()` yang telah tersedia pada *library Python*.

Data Cleaning

Data Cleaning dipakai untuk menghilangkan angka, beberapa simbol, url, *username* (@username), *hashtag* (#), spasi berlebih, tanda baca, emoji, dan pengulangan karakter yang ada pada kalimat. *Data cleaning* menggunakan *regular expression* untuk menemukan karakter yang akan dihapus.

Tokenisasi

Tokenisasi memakai *regular expression* untuk menemukan karakter yang akan dihapus. Tokenisasi digunakan untuk memecah kalimat menjadi list kata. Proses ini menggunakan fungsi `word_tokenize` yang disediakan oleh *library NLTK*.

Normalisasi Kata Tidak Baku

Normalisasi diterapkan pada dataset yang memiliki kata-kata tidak baku diubah menjadi kata yang baku atau sesuai dengan ejaan. Hal ini dilakukan karena cukup banyak kalimat yang menggunakan kata gaul, misal gk, ga, gak, tdk. Jika kata tersebut tidak melalui tahap normalisasi, maka sistem akan menganggap kata-kata itu adalah kata yang berbeda, padahal kata tersebut mempunyai makna yang seharusnya sama, yakni tidak.

Dataset Splitting

Dataset *splitting* meliputi

training, *validation*, dan *testing*.

Dataset *training* digunakan untuk melatih model, dataset *validation* digunakan untuk meminimalisir *overfitting* yang sering terjadi pada jaringan syaraf tiruan, sedangkan dataset *testing* sendiri digunakan sebagai tes akhir untuk mengetahui keakuratan jaringan yang sudah dilatih dengan dataset *training*. Proporsi split dataset pada penelitian ini adalah 70% *train set*, 20% *validation set*, dan 10% *test set*.

BERT Sentiment Classifier

Pada penelitian ini, penulis menggunakan teknik *fine-tuning* dengan model *IndoBERT-base-pl*, salah satu model yang menggunakan arsitektur *BERT-base*. Teknik ini menggunakan model yang telah dilatih sebelumnya dan hanya belajar sedikit lagi untuk mencapai titik optimal pada task yang baru. Model ini telah dilatih menggunakan 4 miliar kata dengan sekitar 250 juta kalimat formal dan sehari-hari dalam Bahasa Indonesia (Wilie, dkk., 2020). Pada penelitian ini penulis menggunakan *library Transformers* yang disediakan oleh *HuggingFace*. Library ini menyediakan ribuan *pre-trained model* yang dapat digunakan untuk melakukan tugas-tugas klasifikasi, ekstraksi informasi, tanya jawab, *summarization*, translasi, *text generation*, dan lain-lain dalam 100 bahasa. *Transformers* didukung oleh dua *library deep learning* yang terkenal, yaitu *PyTorch* dan *TensorFlow*.

Data Preparation

Sebelum dilakukan training pada *BERT*, dataset harus disesuaikan dengan input yang dapat diterima oleh *BERT*. Oleh karena itu dibutuhkan *BertTokenizer*, sebuah *tokenizer* yang bertujuan untuk melakukan tokenisasi pada kalimat-kalimat dan menghasilkan input yang sesuai. Hal ini dilakukan

karena *BERT* menggunakan *vocabulary* yang spesifik yang tergantung dengan model apa yang dipakai. Proses mempersiapkan kalimat menjadi representasi input pada *BERT* dilakukan oleh *tokenizer*.

Tahap training dan evaluasi

Training tidak dilakukan dari awal, melainkan dengan menggunakan model yang telah dilatih sebelumnya dan hanya belajar sedikit lagi untuk mencapai titik optimal pada task yang baru. Teknik training ini disebut sebagai *fine-tuning*. Dengan cara ini, tidak perlu melakukan *training* dari awal, tetapi kita dapat mendownload model yang telah dilatih sebelumnya (*pre-trained model*).

BERT akan menerima urutan kata-kata atau sebuah kalimat sebagai *input* yang akan terus melalui tumpukan *encoder*. Tiap *encoder* mengaplikasikan *self-attention* dan memberikan *output* melalui *feed-forward network* yang kemudian dilanjutkan oleh *encoder* selanjutnya. Karena penelitian ini menggunakan model *IndoBERT-base*, proses ini berlanjut sebanyak 12 kali.

Setelah melewati semua *encoder*, tiap *token* per posisi memberikan *output* berupa vektor dengan ukuran *hidden size* yaitu 768. Untuk proses analisis sentimen, *output* yang diperhatikan adalah *output* dari posisi pertama yaitu *token* [CLS], karena *token* tersebut dianggap melakukan pengumpulan rata-rata atas *token* kata untuk mendapatkan vektor dari kalimat. Vektor tersebut digunakan sebagai *input* untuk *classifier*. *Layer* terakhir pada *classifier layer* menghasilkan *logits*. *Logits* adalah *output* yang berupa prediksi probabilitas kasar dari kalimat yang akan diklasifikasikan. *Softmax* akan mengubah *logits* tersebut menjadi probabilitas dengan mengambil eksponen dari tiap nilai *logit*, sehingga total probabilitasnya adalah tepat 1. Nilai

probabilitas akan berada di antara 0 dan 1.

BERT dapat di-*fine-tuning* dengan menyesuaikan *hyperparameters*-nya (Devlin, dkk., 2019). Kebanyakan *hyperparameters* ketika *fine-tuning BERT* tetap sama seperti *training BERT* biasanya. *Hyperparameters* yang digunakan dalam *training BERT* antara lain

1. *Batch size* adalah jumlah sampel yang dimasukkan ke dalam *network* sebelum *weight* disesuaikan. Semakin besar *batch size*, maka semakin lama waktu yang dibutuhkan untuk menyelesaikan satu *batch*.

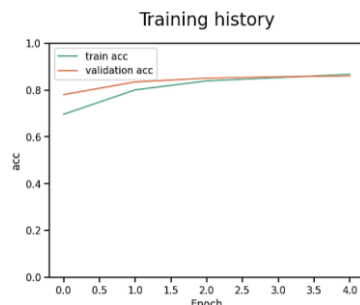
2. *Epoch* adalah jumlah berapa kali jaringan melihat seluruh dataset. Satu *epoch* terjadi ketika semua contoh sudah melewati jaringan baik *forward pass* dan *backward pass*.

3. *Learning rate* menentukan seberapa banyak *weight* pada *neural network* yang akan diubah. Semakin tinggi *rate*-nya, semakin cepat *gradient* bergerak menuju *landscapes*.

Hyperparameters untuk *fine-tuning IndoBERT* yang digunakan dalam penelitian ini, yaitu *batch size* 32, *learning rate* (Adam) $3e-6$, dan *epoch* 5.

Evaluasi

Setelah melalui proses perulangan per *epoch* pada model, hasil pelatihan disimpan. Gambar 4 menunjukkan hasil perbandingan antara akurasi yang didapat saat *training* dan validasi.

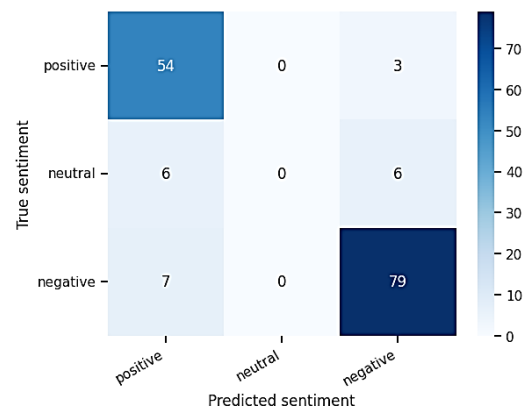


Gambar 4. Luaran Perbandingan Training dan Validasi

Gambar 4 menjelaskan akurasi saat *training* lebih baik daripada hasil saat validasi. Kurva menunjukkan adanya peningkatan hasil akurasi yang diperoleh saat *training*. Tetapi untuk validasi, hasilnya lebih rendah.

Selanjutnya model diuji dengan menggunakan dataset *testing*. Hasil uji coba model dengan data *testing* akan ditunjukkan dengan *confusion matrix* untuk mengukur seberapa banyak model berhasil memprediksi dengan benar semua sentimen pada data *test*.

Gambar 5 dan 6 menunjukkan hasil dari *confusion matrix* dan *classification report* untuk prediksi pada data *testing*.



Gambar 5. Confusion Matrix

	precision	recall	f1-score	support
positive	0.81	0.95	0.87	57
neutral	0.00	0.00	0.00	12
negative	0.90	0.92	0.91	86
accuracy				155
macro avg	0.57	0.62	0.59	155
weighted avg	0.79	0.86	0.82	155

Gambar 6. Classification Report

Gambar 5 menunjukkan model memiliki kesulitan dalam memprediksi data dengan sentimen netral, namun untuk sentimen positif dan negatif sudah cukup baik. Gambar 6 menjelaskan model mendapat skor cukup tinggi pada sentimen negatif, dan skor relatif lebih rendah pada sentimen positif dan netral.

Kata-kata yang kerap muncul dari data ulasan dicitrakan dalam bentuk *WordCloud*. Gambar 7 menunjukkan luaran penggambaran ulasan dari *corpus*.



Gambar 7. *WordCloud* dari *corpus*

Gambar 7 menjelaskan kata-kata yang biasa timbul dari *corpus*, yaitu “aplikasi”, “enggak”, “bisa”, “tiket”, “tidak”, “jadwal”, “ada”, “sangat”, “yang”, “saya”, “untuk”, “membantu”.

KESIMPULAN

Kesimpulan berdasarkan hasil penelitian dan pembahasan adalah sebagai berikut.

Analisis sentimen ulasan aplikasi DLU Ferry sukses dilaksanakan memakai *pre-trained* model *IndoBERT* base dengan teknik *fine tuning*. Model *BERT* menghasilkan akurasi sebesar 86% dengan pemilihan *hyperparameter*, yaitu *batch_size* 32, *learning rate* 3e-6, dan *epoch* 5. Karena jumlah dataset yang tidak besar untuk pemrosesan kata, model menghadapi kesukaran untuk klasifikasi sentimen netral. Ada ketidakteraturan dalam proses labelisasi pada dataset yang bersifat mendua. Analisis sentimen dilakukan menggunakan model *IndoBert-base* yang menerapkan arsitektur *BERT-base*, yakni jenis sederhana dari *BERT*.

DAFTAR PUSTAKA

- Atmaja, Raden Mas Rizqi Wahyu Panca Kusuma; Yustanti, Wiyli. (2021). Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers). *Journal of Emerging Information Systems and Business Intelligence*, Volume 02 Number 03, 2021.
- Deepa Anand and Deepan Naorem. (2016). "Semi-supervised Aspect Based Sentiment Analysis for Movies Using Review Filtering," in *Procedia Computer Science*, vol. 84, pp. 86–93.
- Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT 2019*, pages 4171–4186.
- Fatemeh Hemmatian and Mohammad Karim Sohrabi. (2019). A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review* 52, 1495–1545.
- Pang, B. and Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval*, 2, 1-135.
- Prananda, Alifia Revan & Thalib, Irfandy. (2020). Sentiment Analysis for Customer Review: Case Study of GO-JEK Expansion. *Journal of Information Systems Engineering and Business Intelligence*. Vol. 6, No. 1.
- Putra, Eza Ananda. (2021). Sentiment Analysis Using Bidirectional Encoder Representations from

Transformers (BERT). Diakses dari
<https://medium.com/@eza.a.putra/implementasi-bert-untuk-analisis-sentimen-terhadap-ulasan-aplikasi-flip-berbahasa-indonesia-557d691e0440>

Linguistics and the 10th International Joint Conference on Natural Language Processing, pages 843–857.

Rahul, V. Raj and Monika. (2019). "Sentiment Analysis on Product Reviews," *2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, Greater Noida, India, pp. 5-9.

Rauf, Saad Abdul; Qiang, Yan; Ali, Syed Basit; Ahmad, Waqas. (2019). Using BERT for Checking the Polarity of Movie Reviews. *International Journal of Computer Applications* (0975 – 8887) Volume 177 – No. 21.

Saadah, Siti; Auditama, Kaenova Mahendra; Fattahila, Ananda Affan; Amorokhman Fendi Irfan; Aditsania, Annisa; Rohmawati, Aniq Atiqi. (2022). Implementation of BERT, IndoBERT, and CNN-LSTM in Classifying Public Opinion about COVID-19 Vaccine in Indonesia. *Jurnal Rekayasa Sistem dan Teknologi Informasi (RESTI)*. Vol. 6 No. 4 (2022) 648 – 655.

Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., and 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*