ANALISIS SENTIMEN PADA *GAME* EFOOTBALL DI *GOOGLE PLAY STORE* MENGGUNAKAN ALGORITMA INDOBERT

Muhammad Haris, Aries Suharso, E. Haodudin Nurkifli
Program Studi Informatika, Universitas Singaperbangsa Karawang
Jl. HS. Ronggo Waluyo, Telukjambe Timur, Karawang, Indonesia
2010631170098@student.unsika.ac.id

ABSTRAK

Industri *game* di Indonesia tumbuh pesat, dengan 53,4 juta pemain PC dan 133,8 juta pemain mobile, serta pendapatan *game* mobile diproyeksikan mencapai US$291,10 juta pada 2023. Salah satu *game* populer, eFootball, bersaing dengan FC Mobile dan *game* sepak bola lainnya. Namun, eFootball menghadapi ketidakpuasan pemain, hanya menempati peringkat ke-12 dalam unduhan kategori olahraga pada 2023. Analisis sentimen ulasan pengguna penting untuk mengidentifikasi fitur yang perlu ditingkatkan guna memperbaiki pengalaman bermain. Penelitian ini bertujuan untuk menganalisis sentimen ulasan *game* eFootball di *Google play store* menggunakan algoritma IndoBERT. Tujuan utama penelitian ini adalah untuk membuktikan bahwa penggunaan validasi *thumbsUp* dapat meningkatkan performa model secara signifikan dibandingkan tanpa validasi *thumbsUp*. Validasi *thumbsUp* diambil dari jumlah dukungan pengguna lain terhadap suatu ulasan, yang menandakan relevansi ulasan tersebut. Penelitian ini menggunakan metode *random oversampling* untuk menangani ketidakseimbangan data dan membagi dataset menjadi data *training* (80%), data *validation* (10%), dan data *testing* (10%). Pada skenario pertama, dengan validasi *thumbsUp*, model mencapai akurasi 92%, *precision* 92%, *recall* 92%, dan *f1-score* 92%. Pada skenario kedua tanpa validasi *thumbsUp*, akurasi yang dicapai adalah 84%, dengan *precision*, *recall*, dan *f1-score* yang sama. Hasil ini menunjukkan bahwa validasi *thumbsUp* mampu meningkatkan performa model secara signifikan.

Kata kunci : *analisis sentimen, IndoBERT, thumbsUp, eFootball*

1. PENDAHULUAN

Menurut laporan We Are Social, Indonesia berada di peringkat ketiga dunia dalam jumlah pemain video *game*. Sebanyak 94,5% pengguna internet berusia 16-64 tahun di Indonesia bermain video *game* pada Januari 2022 [1].

Industri *game* di Indonesia terus berkembang, dengan 53,4 juta pemain *game* PC dan 133,8 juta pemain *game* mobile [2].

Industri *game* mobile di Indonesia diproyeksikan akan mencapai pendapatan sebesar US$291,10 juta pada tahun 2023. Bahkan akan terus mengalami pertumbuhan pendapatan dalam 5 tahun ke depan yaitu periode 2023 sampai 2028 [3].

Salah satu *game mobile* yang populer di kalangan masyarakat Indonesia adalah *eFootball*. Terdapat dua raksasa dalam industri *game* sepak bola, yaitu *eFootball* dan *FC Mobile*. Meskipun banyak pesaing seperti *Dream League Soccer* dan *Football Manager Mobile*, kedua *game* ini tetap yang paling populer di kalangan penggemar sepak bola [4].

FC Mobile 24 memiliki keunggulan dalam pembentukan skuad dengan *Ultimate Team* yang mencakup pemain wanita dan pilihan formasi serta taktik yang lebih beragam dibandingkan *eFootball* 2024 yang lebih terbatas. Visual FC Mobile 24 juga mengesankan dengan model pemain, tekstur, dan cuaca pertandingan yang realistis, sementara *eFootball* 2024 masih dikritik meskipun ada beberapa perbaikan. *FC Mobile* 24 memperkenalkan fitur-fitur baru seperti Agen untuk jalur karier dan berbagai mode permainan, sedangkan *eFootball* 2024 hanya menambahkan fitur minor. Namun, *eFootball* 2024 unggul dalam *game*play dengan gerakan dan keahlian yang realistis serta pertahanan yang adaptif, sementara FC Mobile 24 memiliki masalah dengan pemain bertahan yang sering keluar posisi [5].

Berdasarkan data dari data.ai (2023), *eFootball* hanya berada di peringkat 12 dalam jumlah unduhan *game* kategori olahraga sepanjang 2023, tertinggal dari para kompetitornya yang masuk 10 besar, menunjukkan ada ketidakpuasan pemain terhadap *eFootball* [6].

Urgensi penelitian terkait analisis sentimen dalam industri game menjadi semakin jelas karena beberapa alasan. Dengan semakin meningkatnya ketergantungan konsumen pada ulasan dan opini pengguna, analisis sentimen menjadi alat yang vital untuk memahami persepsi pelanggan secara real-time terhadap game tertentu. Misalnya, ulasan game yang positif atau negatif dapat langsung mempengaruhi kesuksesan sebuah game di pasar. Game developer menggunakan analisis sentimen untuk mengidentifikasi area yang perlu diperbaiki, seperti grafik, gameplay, atau fitur spesifik, yang sering kali menjadi keluhan atau pujian utama dalam ulasan pengguna [7][8].

Selain itu, analisis sentimen juga membantu dalam benchmarking dengan pesaing, memungkinkan perusahaan untuk melihat bagaimana produk mereka dibandingkan dengan kompetitor. Hal ini tidak hanya membantu dalam pengembangan produk tetapi juga dalam meningkatkan pengalaman pengguna dan

loyalitas merek dengan menanggapi preferensi serta kritik pengguna secara proaktif [9].

Penting untuk mengidentifikasi dan memperbaiki fitur-fitur yang kurang disukai oleh pemain serta meningkatkan fitur-fitur yang disukai. Ulasan pengguna sangat penting karena biasanya calon pengguna melihat ulasan, rating, dan peringkat *game* sebelum mengunduhnya [10].

Pemanfaatan analisis sentimen menjadi sangat berguna ketika *developer game* berkeinginan untuk memahami pandangan pelanggan terkait pengalaman bermain *game*. Membaca ulasan dari berbagai platform menjadi metode yang sangat sesuai untuk menentukan arah pengembangan dan perbaikan untuk *game* selanjutnya. Selain itu, analisis sentimen juga berperan dalam mengonfirmasi tingkat popularitas produk [11].

Pada penelitian sebelumnya yang dilakukan oleh Kang, dkk. (2021) menggunakan enam model untuk analisis sentimen terhadap Malaysian Airlines, termasuk *Linear Support Vector Classifier, BERT Model, Ensemble Method, Multinomial Naïve Bayes, Bi-LSTM, dan Random Forest*. Hasilnya menunjukkan bahwa pendekatan *deep learning* berkinerja lebih baik dibandingkan metode pembelajaran mesin tradisional. Secara khusus, *Bi-LSTM* mencapai akurasi 77%, sementara model BERT unggul dengan akurasi 86% [12].

Pada penelitian lain yang dilakukan oleh Septian, dkk. (2023) menggunakan pemodelan *pre-trained* BERT dari IndoBERT menghasilkan kinerja dari IndoBERT menunjukkan akurasi 98% pada data latih serta 93% pada data validasinya. *Hyperparameter* yang digunakan termasuk *batch size* 32 dan *epoch* pelatihan 10 dengan proporsi data latih dan data uji 70:30. Selain itu, hasil pengujian model dengan menggunakan *confusion matrix* menunjukkan akurasi sebesar 91% serta presisi sebesar 92%, *recall* 88%, dan *f1-score* 90% untuk kategori negatif. Sedangkan untuk kategori positif, diperoleh presisi sebesar 89%, *recall* 93%, dan *f1-score* 91% [13].

Pada penelitian lain yang dilakukan oleh Braja dan Kodar (2023) penggunaan dua model *pre-trained*, yaitu BERTbase Multilingual dan IndoBERTBase, dalam pemodelan, dengan dua pendekatan pelabelan berbeda berdasarkan skor dan *TextBlob*, menghasilkan model *fine-tuned* IndoBERT yang menunjukkan kinerja yang lebih baik dalam pelabelan data berdasarkan *TextBlob*. Model ini mencapai tingkat akurasi tertinggi sebesar 94% dengan menggunakan *learning rate* 2^{-5}, *batch size* 32, 5 *epoch*, dan waktu pelatihan selama 12 menit [14].

Dengan demikian, penelitian ini mengonfirmasi bahwa model IndoBERT, mampu memberikan performa yang sangat baik dalam analisis sentimen, sejalan dengan temuan penelitian sebelumnya yang menunjukkan superioritas model BERT dibandingkan dengan metode tradisional dan beberapa model *deep learning* lainnya. Peneliti melakukan penyesuaian *hyperparameters* untuk mencapai hasil optimal. *Hyperparameters* yang digunakan dalam penetian ini yaitu *learning rate* 0,000003, *dropout* 0,1 *batch size* 32 dan *epoch* 10. Dataset ulasan *game eFootball* dikumpulkan dari *Google play store*berdasarkan ulasan terbaru. Karena dataset tidak memiliki label, peneliti menerapkan pelabelan berdasarkan *score* atau rating, lalu mengklasifikasikannya menjadi tiga kategori: positif, netral, dan negatif. Model IndoBERT dipilih karena penelitian sebelumnya menunjukkan bahwa IndoBERT memiliki tingkat akurasi yang tinggi. Selain itu, model IndoBERT dipilih juga karena mampu *meng-handle* data ulasan berbahasa Indonesia dengan baik, karena model ini dimodifikasi khusus untuk mengatasi data ulasan yang berbahasa Indonesia.

2. TINJAUAN PUSTAKA

2.1. *Text Mining*

Text mining adalah proses ekstraksi pengetahuan implisit dari data tekstual yang tidak tersimpan secara eksplisit [15].

Berbeda dengan informasi yang diperoleh dari penyimpanan data, pengetahuan ini ditemukan melalui tugas-tugas seperti klasifikasi teks, pengelompokan, dan asosiasi. Text mining merupakan varian dari data mining, dengan fokus pada teks sebagai data tak terstruktur yang terdiri dari string atau kata-kata [16].

Dalam konteks ini, teks merujuk pada artikel yang terdiri dari paragraf dan ditulis dalam bahasa alami, tidak termasuk kode sumber atau persamaan matematika.

Teknik *text mining* menganalisis jumlah besar teks dalam bahasa alami untuk mendeteksi pola leksikal dan mengekstrak informasi yang berguna. Ada beberapa tahap dalam text mining, yang meliputi pengumpulan dokumen, pra-pemrosesan, transformasi teks, seleksi fitur, seleksi pola, dan evaluasi [17].

2.2. Analisis Sentimen

Analisis sentimen atau opinion mining adalah salah satu cabang *Natural Language Processing* (NLP) yang memfokuskan diri pada evaluasi pendapat, perilaku, penilaian, dan emosi individu terkait dengan berbagai entitas seperti produk, topik, layanan, organisasi, individu, atau kegiatan lainnya yang diekspresikan dalam bentuk teks [18].

Dalam proses analisis sentimen, teks subjektif dapat dikelompokkan ke dalam berbagai klasifikasi, termasuk positif, negatif, netral, atau dijelaskan dalam bentuk emosi seperti kegembiraan, kemarahan, atau kesedihan [19].

Berbagai tugas dapat dilakukan dalam analisis sentimen, antara lain klasifikasi subjektivitas, klasifikasi sentimen, deteksi spam opini, deteksi bahasa implisit, dan ekstraksi aspek. Terdapat tiga pendekatan utama dalam analisis sentimen, yaitu *machine learning* (*unsupervised dan semi-supervised*), berbasis lexicon (berbasis korpus dan berbasis kamus), dan pendekatan *hybrid* yang menggabungkan keduanya [20].

2.3. *Bidirectional Encoder Representation from Tranformers* (BERT)

Bidirectional Encoder Representation from Transformers (BERT) merupakan sebuah algoritma baru dalam bidang *Natural Language Processing* (NLP) yang dikembangkan oleh Google. Pertama kali diperkenalkan oleh para peneliti di Google AI pada tahun 2018, BERT telah terbukti efektif dalam menangani berbagai tugas NLP seperti menjawab pertanyaan, menyimpulkan bahasa alami, klasifikasi, dan evaluasi pemahaman bahasa secara umum. BERT memiliki dua versi utama, yakni BERT-base dan BERT-large. BERT-base memiliki 12 *layers* (L=12), 768 ukuran *embedding* (H=768), 12 *attention head* (A=12), dengan total parameter sebanyak 110 juta. Sedangkan BERT-large memiliki 24 lapisan (L=24), 1024 ukuran *embedding* (H=1024), 16 *attention head* (A=16), dengan total parameter sebanyak 340 juta [21].

2.4. IndoBERT

IndoBERT merupakan sebuah model BERT yang dikhususkan untuk Bahasa Indonesia. Model ini terdiri dari empat versi yang berbeda, yaitu IndoBERTBase, IndoBERTLarge, serta dua versi berbasis model ALBERT, yaitu IndoBERT-Litebase dan IndoBERT-Litelarge [22].

Proses pelatihan IndoBERT menggunakan dataset Indo4B dan TPUv3-8 dalam dua tahap. IndoBERT menggunakan dataset Indo4B yang berisi lebih dari 23 GB data teks bahasa Indonesia, mencakup 3,6 miliar kata yang mencakup bahasa formal dan sehari-hari, yang diambil dari media sosial, situs web, blog, dan berita [23].

Untuk membentuk kosakata atau daftar subword token, IndoBERT menerapkan SentencePiece dengan tokenizer *Byte Pair Encoding* (BPE). *Sentence Piece* berfungsi untuk memecah teks menjadi subword token yang lebih kecil, dan BPE adalah salah satu algoritma yang efektif untuk membentuk subword token.

2.5. *Random Oversampling* (ROS)

Oversampling adalah metode untuk menyeimbangkan jumlah data antara kelas minoritas dan mayoritas sehingga data menjadi jumlah data mayoritas [24].

Dalam penelitian ini, metode yang digunakan adalah *random oversampling* (ROS). *Random oversampling* (ROS) adalah teknik di mana data dari kelas minoritas ditambahkan ke dalam data pelatihan secara acak. Proses ini diulang hingga jumlah data dari kelas minoritas menjadi setara dengan kelas mayoritas. Langkah pertama adalah menghitung perbedaan jumlah antara kelas mayoritas dan minoritas. Kemudian, dilakukan pengulangan sebanyak perbedaan tersebut dengan menambahkan data dari kelas minoritas secara acak ke dalam data pelatihan [25].

2.6. *Hyperparameter*

Hyperparameter adalah parameter yang tidak teratur oleh mesin dan tidak memerlukan pengujian. *Hyperparameter* diatur secara manual sesuai dengan kebutuhan peneliti untuk membantu memperkirakan parameter model. *Hyperparameter* yang digunakan dalam penelitian ini meliputi:

a. *Batch Size*
   Batch size adalah *hyperparameter* yang menentukan jumlah sampel yang diproses sebelum bobot model diperbaharui. Proses satu batch akan memakan waktu lebih lama jika ukuran batch lebih besar [26].
b. *Epoch*
   Epoch adalah *hyperparameter* yang menentukan berapa kali jaringan neural memproses data pelatihan. Satu *epoch* menunjukan bahwa algoritma *deep learning* telah mempelajari seluruh dataset pelatihan [27].
c. *Learning Rate*
   Learning rate adalah *hyperparameter* yang digunakan untuk mengatur nilai pembaruan bobot selama proses pelatihan. Semakin besar nilai learning rate, semakin cepat proses pelatihan berlangsung.

2.7. *Loss Function*

Loss function adalah komponen penting dalam proses training model *machine learning*, terutama dalam jaringan saraf tiruan (*neural networks*). Fungsi ini berperan sebagai indikator seberapa baik atau buruk model memprediksi *output* yang diinginkan dibandingkan dengan nilai sebenarnya. Ketika proses training selesai, jaringan saraf tiruan akan menunjukkan nilai *loss*. Nilai *loss* ini seharusnya terus menurun untuk setiap iterasi, jika tidak, berarti jaringan tidak belajar dari proses yang dilalui. Untuk menghindari hal tersebut, diperlukan sebuah fungsi *loss* yang berperan dalam mengatur *output* dari jaringan saraf agar sesuai dengan tujuan penelitian (optimisasi). Hasil dari fungsi *loss* ini digunakan sebagai sinyal umpan balik untuk menyesuaikan bobot-bobot jaringan guna mengurangi nilai *loss*. Salah satu fungsi *loss* yang umum digunakan adalah *cross-entropy loss*, yang sering dikombinasikan dengan *softmax* [28]. *Cross-entropy loss* ini mengukur kinerja model klasifikasi *multiclass* yang *output*nya berupa nilai probabilitas antara 0 dan 1.

2.8. *Confusion Matrix*

Confusion matrix adalah matriks dengan ukuran M x N yang digunakan untuk mengevaluasi performa model klasifikasi, di mana N merepresentasikan jumlah kelas target [29].

Matriks ini membandingkan nilai target aktual dengan nilai yang diprediksi oleh model machine learning [30].

Hal ini memberikan penjelasan holistik tentang seberapa baik kinerja model klasifikasi yang dibangun dan jenis-jenis kesalahan.

Tabel 1 Kriteria Confusion Matrix

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

Keterangan:

a. *True Positive* (TP) mengacu pada data yang memiliki kondisi aktual positif dan diprediksi dengan benar.
b. *True Negative* (TN) merujuk pada data negatif yang diprediksi secara akurat.
c. *False Positive* (FP) terjadi ketika data yang seharusnya negatif diprediksi sebagai positif.
d. *False Negative* (FN) terjadi ketika data yang seharusnya positif diprediksi sebagai negatif.

Terdapat beberapa metrik evaluasi pada *confusion matrix* yaitu:

a. Accuracy

$$Accuracy = \frac{TN + TP}{TN + FN + TP + FP} \times 100\% \quad (1)$$

b. Precision

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

c. Recall

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

d. F1-score

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (4)$$

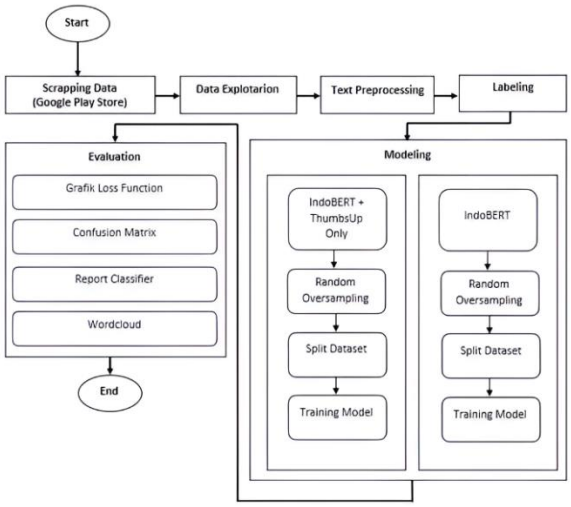
2.9. *Wordcloud*

Pendekatan visualisasi data yang dikenal sebagai *wordcloud* sangat populer dan sudah umum digunakan. Dalam visualisasi *wordcloud*, ukuran *font* yang digunakan untuk kata akan berbanding lurus dengan berapa kali kata tersebut muncul dalam sekelompok dokumen. Artinya, semakin sering suatu subjek disebutkan, semakin besar ukuran yang akan muncul dalam visualisasi [31].

Wordcloud merupakan bagian penting dari data mining, yang belakangan ini semakin mendapatkan perhatian dan menyediakan lebih banyak opsi aplikasi. Saat ini, terdapat berbagai generator *wordcloud online* yang dapat diakses oleh pengguna dengan kebutuhan dasar, seperti mengulang frasa tertentu atau mengumpulkan data teks dari situs web. Pengguna dapat memanfaatkan generator *wordcloud* ini [32].

3. METODE PENELITIAN

Adapun langkah-langkah dalam penelitian ini mengikuti diagram alur yang menjelaskan setiap tahap yang dilakukan selama penelitian, seperti yang ditunjukan pada Gambar 1 berikut.



Gambar 1. Tahapan Penelitian

3.1. Pengumpulan Data

Dalam penelitian ini, data yang akan diambil berasal dari *platform Google play store*pada *game* eFootball dalam bahasa Indonesia. Dataset diperoleh dengan menggunakan package dari *google-play-scraper* yang menyediakan API pada bahasa pemrograman Python, dan data akan disimpan dalam format CSV.

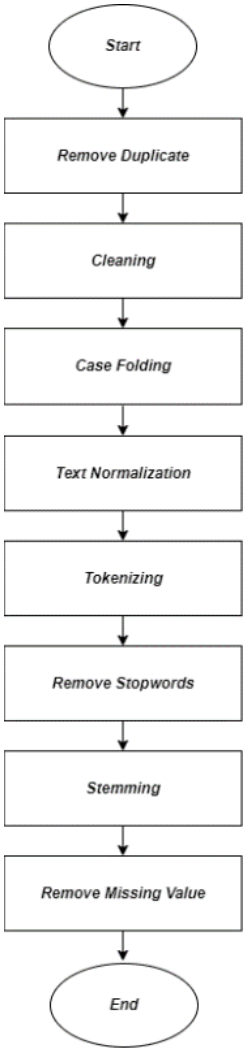
Jumlah dataset yang digunakan pada penelitian ini yaitu 25.000 *record*. Jumlah ini diambil berdasarkan distribusi rating atau score. Jadi setiap rating atau score (1-5) diambil per 5.000 record. Sehingga Jumlah data yang dihasilkan adalah 25.000 *records* yang terdiri dari rating 1-5 di *platform Google play store*dengan jumlah kolom dataset sebesar 13 diantaranya *reviewId, userName, userImage, content, score, thumbsUpCount, reviewCreatedVersion, at, replyContent, repliedAt, appVersion, sortOrder, dan appId*. Adapun dataset yang diambil yaitu ulasan pengguna *game* eFootbal pada tanggal 12 November 2023 sampai 21 Juni 2024.

3.2. Data Exploration

Tahap selanjutnya yaitu memahami setiap kolom/*feature* yang ada pada dataset, adapun terdapat 13 kolom/*feature* yang ada pada dataset diantaranya *reviewId, userName, userImage, content, score, thumbsUpCount, reviewCreatedVersion, at, replyContent, repliedAt, appVersion, sortOrder, dan appId*. Dari 13 kolom/*feature* yang ada pada dataset digunakan 4 kolom/*feature* yaitu *reviewId, content, score, dan thumbsUpCount.*

3.3. Text Preprocessing

Tahap awal preprocessing teks bertujuan untuk membersihkan dataset dari gangguan agar lebih terstruktur. Proses ini menggunakan modul NLTK dan library Sastrawi di *Python*. Langkah-langkah yang dilakukan meliputi:



Gambar 2. Tahap Text Preprocessing

a. *Remove Duplicate:* Menghapus data ulasan yang duplikat berdasarkan *reviewId* agar tidak ada ulasan ganda dari pengguna yang sama.
b. *Cleaning:* Menghilangkan karakter tidak valid seperti tanda baca, angka, simbol, dan emoji untuk memudahkan pengolahan data.
c. *Case folding:* Menyeragamkan semua teks menjadi huruf kecil.
d. *Text normalization:* Memperbaiki kata yang salah, *typo*, singkatan, dan kata yang tidak sesuai konteks.
e. *Tokenizing:*Memisahkan kata-kata dalam kalimat.
f. *Remove stopwords:* Menghapus kata-kata yang sering muncul tetapi tidak penting, seperti kata penghubung atau kata ganti.
g. *Stemming:* Mengubah kata-kata menjadi bentuk dasar dengan menghapus imbuhan.
h. *Remove missing value:* Menghapus ulasan yang kosong atau hanya berisi karakter khusus setelah proses pembersihan, karena ulasan tersebut akan bernilai *NaN*.

3.4. Labeling

Dalam penelitian ini, proses pelabelan ditentukan berdasarrkan nilai rating atau score pada ulasan. Dengan asumsi nilai (1-2) adalah negatif, nilai (3) adalah netral, dan nilai (4-5) adalah positif.

3.5. Modeling

Tahap pemodelan melibatkan input data yang telah dipersiapkan dari tahap pengumpulan data ke dalam model. Adapun pada tahap pemodelan terdapat 2 skenario yaitu skenario 1 model IndoBERT + *ThumbsUp* only adalah proses pemodelah menggunakan IndoBERT dengan dataset yang digunakan hanya mencakup ulasan yang memiliki nilai thumsUp minimal 1, yang berarti ulasan-ulasannya dianggap relevan atau mendapat dukungan dari pengguna lain dan skenario 2 model IndoBERT digunakan dengan seluruh dataset digunakan tanpa mempertimbangkan nilai *thumbsUp*. Selanjutnya meng-*handle imbalanced* jumlah labelnya menggunakan *random oversampling* agar jumlah persebaran jumlah labelnya menjadi seimbang atau sama. Setelah itu, Proses pembagian dataset memisahkan data menjadi tiga bagian: *data train*, *data validation*, dan *data test.* Pembagian ini penting untuk menentukan data yang akan digunakan dalam pelatihan dan pengujian model. Proporsi pembagian data pada penelitian ini yaitu 80:10:10. Dimana pembagiannya adalah 80% untuk *data train*, 10% untuk *data validation*, dan 10% untuk *data test*.

Selanjutnya, dalam pelatihan model IndoBERT, diperlukan definisi parameter yang berbeda untuk penelitian ini. *Hyperparameter* yang digunakan dalam penelitian ini yaitu *Batch size* 32, *epoch* 10, dan *learning rate* 0,000003.

3.6. Evaluation

Proses evaluasi model bertujuan untuk mengukur seberapa baik model dalam melakukan klasifikasi pada berbagai kelas. Salah satu metode evaluasi yang digunakan dalam penelitian ini adalah *confusion matrix*, yang merupakan alat pengukuran kinerja model pada tugas klasifikasi *machine learning* dengan *output* berupa dua kelas atau lebih. Lalu melihat *grafik loss function* untuk menilai apakah model mengalami *overfitting* atau tidak. Selanjutnya, pada tahap evaluasi ini, hasil prediksi analisis sentimen untuk setiap kalimat dalam dataset dievaluasi. Nilai akurasi tertinggi yang diperoleh selam proses pelatihan model sebelumnya dijadikan sebagai akurasi model. Terdapat empat kinerja evaluasi yang digunakan, yaitu *f1-score*, *precision*, *recall*, dan *accuracy* yang sering digunakan dalam tugas klasifikasi teks dan analisis sentimen.

Setelah mengidentifikasi hasil sentimen positif, negatif, dan netral visualisasi dengan *wordcloud* akan dilakukan untuk menampilkan kata-kata yang sering muncul dalam ulasan sentimen.

4. HASIL DAN PEMBAHASAN

4.1. Scraping Data

Sumber data yang digunakan dalam penelitian ini berasal dari ulasan pengguna *game* eFootball di *Google Play Store*. Data diambil menggunakan teknik

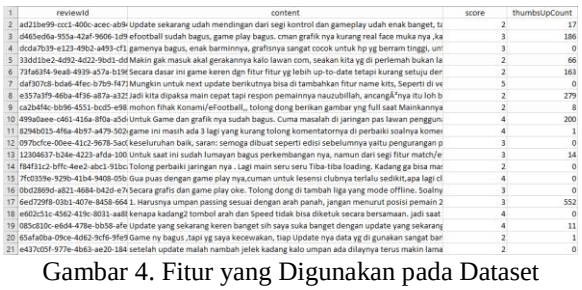
crawling data dengan *library google-play-scraper*, yang memungkinkan peneliti untuk mengumpulkan ulasan secara otomatis dari *platform* tersebut. Rentang tanggal pengambilan data adalah dari 12 November 2023 hingga 21 Juni 2024.

Jumlah dataset ditentukan dengan mengambil masing-masing rating (1-5) sebanyak 5.000 atau dengan jumlah total 25.000 data. Akan tetapi, jumlah data yang berhasil dikumpulkan dalam penelitian ini sebanyak 20.927 ulasan. Hal ini bisa terjadi karena keterbatasan data yang tersedia pada rentang tanggal pengambilan datanya. Data diambil berdasarkan dua kategori utama: ulasan terbaru (NEWEST) dan ulasan paling relevan (MOST_RELEVANT). Data ini kemudian diproses dan dibersihkan untuk memastikan kualitas data yang akan digunakan dalam analisis sentimen. Data ini digunakan sebagai input untuk model indoBERT dalam proses klasifikasi sentimen.

Gambar 3. Hasil Scraping Data

4.2. Data Exploration

Terdapat banyak atribut hasil dari *scraping* dalam *dataframe*. Adapun dalam penelitian ini, fitur yang akan digunakan adalah *reviewId, content, score*, dan *thumbsUpCount*. Kolom *reviewId* akan digunakan untuk mencari nilai duplikat pada ulasan, kolom *content* akan digunakan untuk pemodelan sentimen, kolom *score* akan digunakan untuk menentukan label sentimen (positif, netral, dan negatif), dan kolom *thumbsUpCount* akan digunakan untuk validasi data ulasan, dimana semakin banyak jumlah like/dukungan maka ulasana tersebut dianggap semakin relevan.



	reviewId	content	score	thumbsUpCount
2	ad21be99-ccc1-400c-acec-ab9	Update sekarang udah mendingan dari segi kontrol dan gameplay udah enak banget, tz	2	17
3	d465ed6a-955a-42af-9606-1d9	efootball sudah bagus, game play bagus. cman grafik nya kurang real face muka nya ,ka	3	186
4	dcda7b39-e123-49b2-a493-cf1	gamenya bagus, enak barminnya, grafisnya sangat cocok untuk hp yg berram tinggi, unt	3	0
5	33dd1be2-4d92-4d22-9bd1-dd	Makin gak masuk akal gerakannya kalo lawan com, seakan kita yg di perlemah bukan la	2	66
6	73fa63f4-9ea8-4939-a57a-b19	Secara dasar ini game keren dgn fitur fitur yg lebih up-to-date tetapi kurang setuju der	2	163
7	daf307b8-bda6-4fec-b7b9-f47	Mungkin untuk next update berikutnya bisa di tambahkan fitur name kits, Seperti di ve	5	0
8	e357a3f9-46ba-4f36-a87a-a32	Jadi kita dipaksa main cepat tapi respon pemainnya nauzubillah, ancangÃ¨nya itu loh b	2	279
9	ca2b4f4c-bb96-4551-bcd5-e98	mohon tlhak Konami/eFootball,, tolong dong berikan gambar yng full saat Mainkannya	2	8
10	499a0aee-c461-416a-8f0a-a5d	Untuk Game dan grafik nya sudah bagus. Cuma masalah di jaringan pas lawan penggun	4	200
11	8294b015-4f6a-4b97-a479-502	game ini masih ada 3 lagi yang kurang tolong komentatornya di perbaiki soalnya komer	4	1
12	097bcfce-00ee-41c2-9678-5ac	keseluruhan baik, sarani: semoga dibuat seperti edisi sebelumnya yaitu pengurangan p	3	0
13	12304637-b24e-4223-afda-100	Untuk saat ini sudah lumayan bagus perkembangan nya, namun dari segi fitur match/er	3	14
14	f84f31c2-bffc-4ee2-abc1-91bc	Tolong perbaiki jaringan nya . Lagi main seru seru Tiba-tiba loading. Kadang ga bisa mar	2	0
15	7fc0359e-929b-41b4-9408-05b	Gua puas dengan game play nya,cuman untuk lesensi clubnya terlalu sedikit,apa lagi cl	4	0
16	0bd2869d-a821-4684-b42d-e7x	Secara grafis dan game play oke. Tolong dong di tambah liga yang mode offline. Soalny	3	0
17	6ed729f8-03b1-407e-8458-664	1. Harusnya umpan passing sesuai dengan arah panah, jangan menurut posisi pemain 2	3	552
18	e602c51c-4562-419c-8031-aa8	kenapa kadang2 tombol arah dan Speed tidak bisa diketuk secara bersamaan. Jadi saat	4	0
19	085c810c-e6d4-478e-bb58-afe	Update yang sekarang keren banget sih saya suka banget dengan update yang sekarang	4	11
20	65af3d0a-09ce-4d62-9cf6-9fe9	Game ny bagus ,tapi yg saya kecewakan, tiap Update nya data yg di gunakan sangat ban	2	1
21	e437c05f-977e-4b63-ae20-184	setelah update malah nambah jelek kadang kalo umpan ada di laynya terus makin lama	2	0

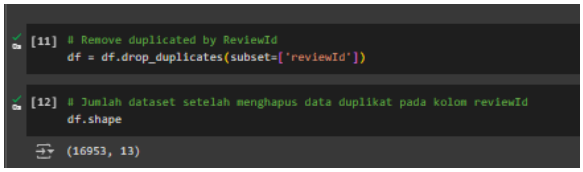
Gambar 4. Fitur yang Digunakan pada Dataset

4.3. Text Preprocessing

Text Preprocessing adalah proses untuk membersihkan dan memformat data teks agar siap digunakan dalam model pembelajaran mesin. Tahap yang dilakukan yaitu r*emove duplicated, cleaning, case folding, text nomarlization, tokenizing, remove stopword*, dan *stemming*. Berikut adalah penjelasan langkah-langkah yang dilakukan:

a. *Remove Duplicated*

Pada tahap ini, dataset dibersihkan dari ulasan yang duplikat berdasarkan kolom “reviewId”. Duplikasi data dapat menyebabkan bias dalam model, oleh karena itu perlu dihapus.



```
[11] # Remove duplicated by ReviewId
     df = df.drop_duplicates(subset=['reviewId'])

[12] # Jumlah dataset setelah menghapus data duplikat pada kolom reviewId
     df.shape

     (16953, 13)
```

Gambar 5. Proses Remove Duplicated

Pada Gambar 5 menampilkan jumlah data setelah menghapus nilai data yang memiliki nilai duplikat berdasarkan kolom reviewId. Dari jumlah data awal sebanyak 20.927, setelah dilakukan *remove duplicated* datanya menjadi 16.953 data. Artinya terdapat 3.974 data ulasan yang *duplicated*.

b. *Cleaning*

Langkah ini mencakup penghapusan karakter unik seperti tanda baca, angka, emoji, dan karakter non-alfabet lainnya. Hal ini bertujuan untuk memastikan bahwa teks hanya berisi kata-kata yang bermakna.

Tabel 2. Tahap Cleaning Pada Ulasan

Sebelum *Cleaning*	**Setelah *Cleaning***
Ternyata selama ini saya main efootball ngelag karna device saya yg kentang 😂 dan sekarang sudah ganti HP dgn spek yg lebih ganas jadi mantap main nya rata kanan 😎	Ternyata selama ini saya main efootball ngelag karna device saya yg kentang dan sekarang sudah ganti HP dgn spek yg lebih ganas jadi mantap main nya rata kanan

Proses *cleaning* adalah langkah penting dalam *preprocessing* teks yang membantu dalam meningkatkan kualitas dan kebersihan data teks. Dari tabel yang diberikan, terlihat jelas bagaimana semua karakter unik dan simbol yang tidak diperlukan dihapus, yang berkontribusi pada penyederhanaan dan standarisasi data sebelum proses pemodelan lebih lanjut.

c. *Case Folding*

Pada tahap ini, semua teks ulasan dikonversi menjadi huruf kecil.

Tabel 3. Tahap *Case Folding*

Sebelum *Case Folding*	**Setelah *Case Folding***
Ternyata selama ini saya main efootball ngelag karna device saya yg kentang dan sekarang sudah ganti HP dgn spek yg lebih ganas jadi mantap main nya rata kanan	**ternyata** selama ini saya main efootball ngelag karna device saya yg kentang dan sekarang sudah ganti hp dgn spek yg lebih ganas jadi mantap main nya rata kanan

Hal ini dilakukan untuk memastikan konsistensi dan menghindari perlakuan berbeda terhadap huruf besar dan huruf kecil.

d. *Text Normalization*

Normalisasi teks bertujuan untuk mengubah kata-kata yang tidak baku menjadi kata yang baku dan juga kata-kata yang salah atau atau tidak sesuai. Proses ini biasanya melibatkan penggantian slang, singkatan, atau kata-kata tidak resmi dengan bentuk yang lebih umum.

Tabel 4. Tahap *Text Nomalization*

Sebelum *Text Normalization*	**Setelah *Text Normalization***
ternyata selama ini saya main efootball ngelag **karna** device saya **yg** kentang dan sekarang sudah ganti hp dgn spek **yg** lebih ganas jadi mantap main nya rata kanan	ternyata selama ini saya main efootball ngelag **karena** device saya **yang** kentang dan sekarang sudah ganti hp dengan spek **yang** lebih ganas jadi mantap main nya rata kanan

Proses normalisasi teks adalah langkah penting dalam *preprocessing* teks yang membantu meningkatkan konsistensi dan standarisasi data teks.

e. *Tokenizing*

Tokenisasi adalah proses memecah teks menjadi unit-unit yang lebih kecil seperti kata atau frasa. Tokenisasi memudahkan proses analisis teks dengan memisahkan setiap kata dalam teks.

Tabel 5. Tahap *Tokenizing*

Sebelum *Tokenizing*	**Setelah *Tokenizing***
ternyata selama ini saya main efootball ngelag karena device saya yang kentang dan sekarang sudah ganti hp dengan spek yang lebih ganas jadi mantap main nya rata kanan	[‘ternyata’, ‘selama’, ‘ini’, ‘saya’, ‘main’, ‘efootball’, ‘ngelag’, ‘karena’, ‘device’, ‘saya’, ‘yang’, ‘kentang’, ‘dan’, ‘sekarang’, ‘sudah’, ‘ganti’, ‘hp’, ‘dengan’, ‘spek’, ‘yang’, ‘lebih’, ‘ganas’, ‘jadi’, ‘mantap’, ‘main’, ‘nya’, ‘rata’, ‘kanan’]

Proses *tokenizing* membantu dalam mempersiapkan teks untuk analisis lebih lanjut atau untuk dimasukkan ke dalam model *machine learning*.

f. *Remove Stopword*

Dalam penelitian ini, proses penghapusan *stopwords* dilakukan menggunakan *library* Sastrawi yang khusus menangani teks berbahasa Indonesia.

Tabel 6. Tahap *Remove Stopword*

Sebelum *Remove Stopwords*	**Setelah *Remove Stopwords***
[‘ternyata’, ‘selama’, **‘ini’**, **‘saya’**, ‘main’, ‘efootball’, ‘ngelag’, **‘karena’**, ‘device’, **‘saya’**, **‘yang’**, ‘kentang’, **‘dan’**, ‘sekarang’, **‘sudah’**, ‘ganti’, ‘hp’, **‘dengan’**, ‘spek’, **‘yang’**, ‘lebih’, ‘ganas’, ‘jadi’, ‘mantap’, ‘main’, ‘nya’, ‘rata’, ‘kanan’]	ternyata selama main efootball ngelag device kentang sekarang ganti hp spek lebih ganas jadi mantap main nya rata kanan

Penghapusan stopwords membantu dalam mengurangi kebisingan dalam data, sehingga algoritma analisis sentimen dapat lebih efektif dalam mengidentifikasi makna dari teks yang diberikan.

g. *Stemming*

Pada penelitian ini, proses *stemming* dilakukan menggunakan *library* Sastrawi, yang merupakan salah satu library pemrosesan bahasa alami untuk bahasa Indonesia.

Tabel 7. Tahap Stemming

Sebelum *Stemming*	**Setelah *Stemming***
ternyata selama main efootball ngelag device kentang sekarang ganti hp spek lebih ganas jadi mantap main nya rata kanan	**nyata lama** main efootball ngelag device kentang sekarang ganti hp spek lebih ganas jadi mantap main nya rata kanan

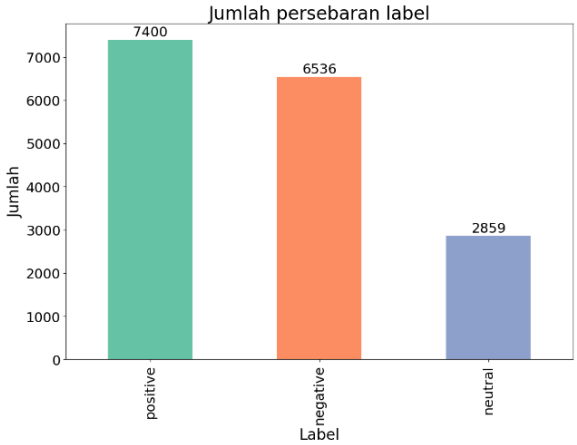
Proses *Stemming* membantu mengurangi jumlah kata yang perlu dianalisis dengan menggabungkan variasi kata ke dalam satu bentuk dasar.

h. *Remove Missing Value*

Setelah melewati tahapan *stemming* dilakukan kembali proses penghapusan data ulasan yang tidak memiliki nilai atau NaN. Hal ini bisa terjadi karena pada ulasan yang memiliki karakter khusus seperti nilai angka, tanda baca, dan emoji saja akan otomatis terhapus setelah melewati proses *text preprocessing* sehingga ulasannya menjadi missing value. Adapun setelah melakukan pengecekan terdapat 158 ulasan dengan nilai NaN dan akan dihapus. Dengan demikian jumlah datanya sekarang menjadi 16.795.

4.4. Labeling

Pada penelitian ini label yang digunakan ada 3 label yaitu *negative, neutral* dan *positive.* Adapun penentuan labelnya berdasarkan rating/*score* pada ulasan. Dengan rincian score 1 hingga 2 untuk label *negative*, score 3 untuk label *neutral*, dan score 4 hingga 5 untuk label *positive*.



Gambar 6. Jumlah Persebaran Label

Gambar 5 menunjukkan persebaran rating ulasan untuk *game* eFootball. Sebanyak 7400 ulasan diberi label positif (rating 4-5), menunjukkan mayoritas pengguna puas dengan *game* ini. Ada 6536 ulasan negatif (rating 1-2), mencerminkan sejumlah besar pengguna yang kurang puas. Sementara itu, 2859 ulasan diberi label netral (rating 3), jumlah yang paling sedikit di antara ketiga kategori, menunjukkan pengguna dengan pendapat yang tidak terlalu positif maupun negatif.

4.5. Modeling

Pada proses pemodelan akan dilakukan menggunakan dua skenario.

4.6. IndoBERT + *thumbsUp Only*

Dataset yang digunakan pada saat pemodelan diambil berdasarkan ulasan yang hanya memiliki nilai *thumbsUp* (minimal 1). Nilai *thumbsUp* dianggap sebagai indikator relevansi ulasan, di mana ulasan dengan nilai *thumbsUp* yang lebih tinggi dianggap lebih berguna atau didukung oleh pengguna lain. Terdapat 4026 ulasan yang memiliki nilai *thumbsUp* dengan rincian, 2283 ulasan berlabel *negative*, 1229 ulasan berlabel *positive* dan 514 ulasan berlabel *neutral*.

4.7. IndoBERT

Skenario kedua menggunakan seluruh dataset tanpa memperhatikan nilai *thumbsUp*. Dataset ini terdiri dari 16.795 ulasan, dengan 7.400 ulasan berlabel *positive*, 6.536 ulasan berlabel *negative*, dan 2.859 ulasan berlabel *neutral*.

Kedua skenario ini dirancang untuk mengevaluasi apakah relevansi ulasan berdasarkan nilai *thumbsUp* mempengaruhi kinerja model IndoBERT dalam klasifikasi sentimen. Skenario pertama mempertimbangkan relevansi ulasan dengan menggunakan nilai *thumbsUp*, sementara skenario kedua melibatkan seluruh ulasan tanpa memperhatikan relevansi. Hasil dari kedua skenario ini akan digunakan untuk menilai kinerja model berdasarkan metrik seperti *accuracy*, *precision*, *recall*, *f1-score* dan performa lainnya, serta untuk memahami pengaruh relevansi ulasan terhadap hasil analisis sentimen.

4.8. *Random Oversampling*

Selanjutnya yaitu menerapkan *random oversampling* karena pada kedua skenario terlihat persebaran label tidak seimbangan (*imbalance data*) sehingga perlu diseimbangkan. Berdasarkan persebaran labelnya ulasan dengan label *positive* dan *negative* cenderung mendominasi, sementara ulasan *neutral* lebih sedikit. Oleh karena itu, perlu dilakukan penanganan ketidakseimbangan label seperti *random oversampling* pada label yang kurang dominan (*neutral*) untuk memastikan model yang dilatih memiliki performa yang baik dan tidak bias terhadap label mayoritas (*positive* dan *negative*). Pada skenario 1, karena label *negative* memiliki jumlah terbanyak, yaitu 2283 ulasan, maka jumlah ulasan dengan label *positive* dan *neutral* dinaikkan hingga masing-masing mencapai 2283 ulasan melalui proses *random oversampling*. Demikian pula pada skenario 2, karena label positive memiliki jumlah terbanyak, yaitu 7400 ulasan, maka jumlah ulasan dengan label *negative* dan *neutral* juga dinaikkan hingga masing-masing mencapai 7400 ulasan.

4.9. *Split Dataset*

Proses pembagian data pada penelitian ini terbagi menjadi tiga yaitu data train, data validation, dan data test. Dengan proporsi 80:10:10 dimana 80% untuk *data train*, 10% untuk *data validation*, dan 10% untuk *data test*.

4.10. *Training Model*

Pada penelitian ini model klasifikasi yang digunakan yaitu IndoBERT. Model IndoBERT adalah bagian dari *library transformer* yang tersedia secara *open source* di *Hugging Face*, sebuah *platform open source* yang memainkan peran penting dalam berbagai aplikasi pemrosesan bahasa alami (Natural Language Processing). Adapun *pre-trained* model yang digunakan yakni IndoBERT base p-1.

Langkah pertama dalam klasifikasi sentimen dengan IndoBERT adalah proses *embedding*. Tahap ini melibatkan beberapa langkah, seperti memasukkan data, mengubah kalimat menjadi token untuk setiap kata, menambahkan token khusus seperti [CLS] di awal kalimat dan [SEP] di akhir kalimat untuk memisahkan kalimat, serta menyisipkan token [PAD] agar panjang teks di setiap *batch* seragam.

Setelah itu, model melakukan *encoding*, di mana setiap kata dikonversi menjadi vektor numerik berdasarkan indeks kosakata dari model *pre-trained* IndoBERT. Vektor tersebut kemudian diproses melalui jaringan IndoBERT yang terdiri dari 12 lapisan *Encoder* dengan mekanisme *multi-head self-attention*, yang memberikan bobot pada setiap token dalam kalimat. Hasil akhirnya adalah vektor *output* untuk setiap token, yang kemudian dilewatkan ke lapisan klasifikasi untuk menghasilkan nilai *logits*, yang diubah menjadi probabilitas kelas melalui fungsi *softmax*.

Setelah dilakukan *pre-trained* IndoBERT, selanjutnya menyesuaikan *hyperparameter* agar kinerja model menjadi optimal. Dalam membangun *classifier* ini, digunakan beberapa *hyperparameter* penting seperti dropout 0,1 untuk mencegah *overfitting*, *learning rate* 0,000003 untuk mengontrol kecepatan pembelajaran, dan *batch size* 32. Model dilatih selama 10 *epoch*, di mana pemilihan jumlah *epoch* didasarkan pada pengujian awal yang menunjukkan bahwa setelah 10 *epoch*, model tidak mengalami penurunan performa atau *underfitting*. *Early stopping* juga diterapkan untuk memonitor performa pada data validasi dan menghentikan

pelatihan jika tidak ada peningkatan, mencegah *overfitting*.

Optimizer yang digunakan adalah Adam, yang menggabungkan kelebihan dari AdaGrad dan RMSProp, mampu menyesuaikan *learning rate* secara adaptif untuk mempercepat konvergensi dan meningkatkan kinerja model. Berikut merupakan proses pemodelan data ulasan eFootball pada tabel 8.

Tabel 8. Proses pemodelan Data Ulasan eFootball

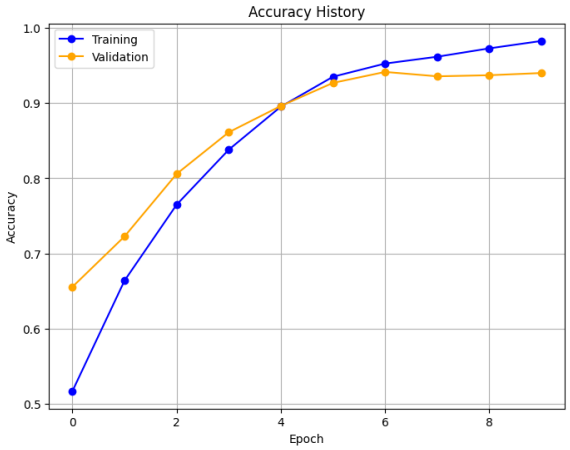
Proses	Hasil
Input	[‘[CLS]’, ‘nyata’, ‘lama’, ‘main’, ‘ef’, ‘##oo’, ‘##tb’, ‘##all’, ‘ngel’, ‘##ag’, ‘device’, ‘kentang’, ‘sekarang’, ‘ganti’, ‘hp’, ‘spek’, ‘lebih’, ‘ganas’, ‘jadi’, ‘mantap’, ‘main’, ‘nya’, ‘rata’, ‘kanan’, ‘[SEP]’]
Token ID	[2, 3325, 985, 2724, 990, 3255, 17234, 1742, 10293, 82, 14921, 9408, 747, 4068, 2109, 9278, 216, 13790, 472, 7424, 2724, 1107, 5652, 2846, 3]
Attention Mask	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]

4.11. Evaluation

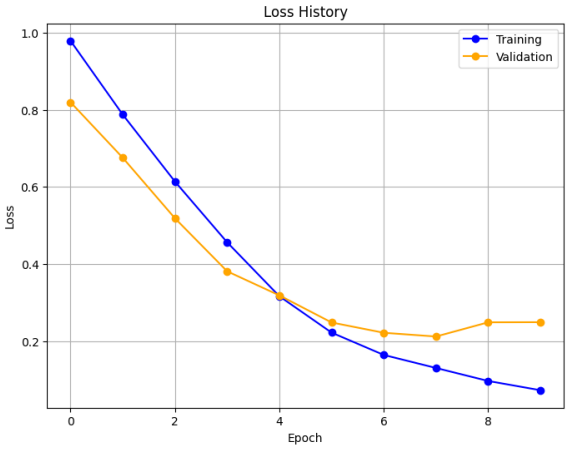
Pada tahap evaluasi, dilakukan analisis terhadap performa model IndoBERT menggunakan beberapa metrik seperti grafik *accuracy-loss function, confusion matrix,* dan visualisasi *wordcloud*.

4.12. Grafik *Accuracy vs Loss*

Untuk mengevaluasi konsistensi kinerja model yang dibangun, kita akan melihat seberapa besar perbedaan antara akurasi dan loss pada *data validation*. Dengan jumlah *epoch* yang ditentukan sebanyak 10 *epoch* dan juga menerapkan fungsi *early stopping* untuk mencegah model mengalami *overfitting.*

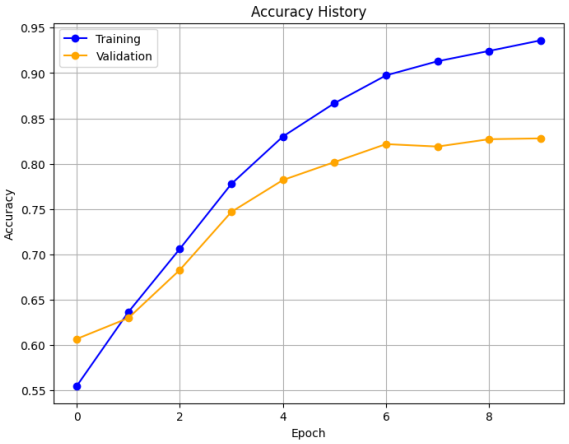


Gambar 7. Grafik Akurasi Skenario 1 (IndoBERT + *thumbsUp Only*)

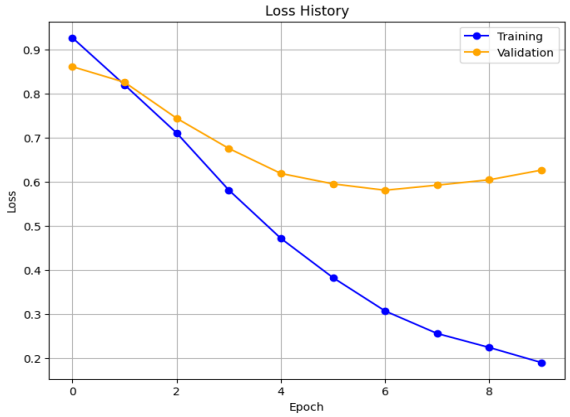


Gambar 8. Grafik Loss Skenario 1 (IndoBERT + *thumbsUp Only*)

Didapatkan hasil berdasarkan Gambar 7 dan Gambar 8 terlihat adanya peningkatan akurasi dan penurunan nilai loss pada skenario 1 (IndoBERT + *thumbsUp Only*) selama 10 *epoch*. Pada data pelatihan, akurasi meningkat dari 52% di *epoch* pertama menjadi 98% di *epoch* kesepuluh. Sementara itu, pada data validasi, akurasi bertambah dari 66% di *epoch* pertama menjadi 94% di *epoch* kesepuluh, dengan selisih akurasi antara data pelatihan dan validasi sebesar 4%. Nilai *loss function* pada data pelatihan menurun dari 97,95% di *epoch* pertama menjadi 7% di *epoch* kesepuluh. Untuk data validasi, nilai loss function turun dari 82% menjadi 24,88% di *epoch* kesepuluh. Penurunan *loss function* ini menunjukkan bahwa model IndoBERT dapat mempelajari dan memprediksi data baru dengan lebih baik, serta tetap konsisten tanpa mengalami masalah *overfitting*. Meskipun pada *epoch* kesembilan nilai loss mengalami kenaikan tetapi kenaikannya tidak signifikan dan kembali menurun pada *epoch* kesepuluh, sehingga tidak mengaktifkan fungsi *early stopping*.



Gambar 9. Grafik Akurasi Skenario 2 (IndoBERT)

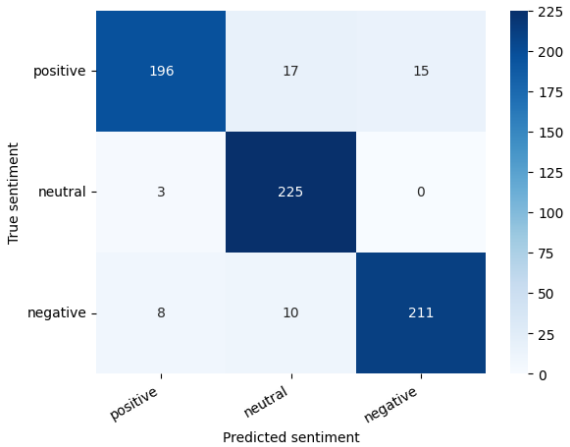


Gambar 10. Grafik *Loss* Skenario 2 (IndoBERT)

Didapatkan hasil berdasarkan Gambar 9 dan Gambar 10 terlihat adanya peningkatan dan penurunan nilai loss pada skenario 2 (IndoBERT) selama 10 epoch. Pada data pelatihan, akurasi meningkat dari 55% di *epoch* pertama menjadi 94% di *epoch* kesepuluh. Sementara itu, pada data validasi, akurasi bertambah dari 61% di *epoch* pertama menjadi 83% di *epoch* kesepuluh, dengan selisih akurasi antara data pelatihan dan validasi sebesar 11%. Nilai *loss function* pada data pelatihan menurun dari 92,61% di *epoch* pertama menjadi 18,94% di *epoch* kesepuluh. Untuk data validasi, nilai loss function turun dari 86% menjadi 62,59% di *epoch* kesepuluh. Penurunan *loss function* ini menunjukkan bahwa model IndoBERT dapat mempelajari dan memprediksi data baru dengan lebih baik, serta tetap konsisten tanpa mengalami masalah *overfitting*. Meskipun pada *epoch* ketujuh sampai kesepuluh nilai *loss* mengalami mengalami peningkatan, tetapi tidak mengaktifkan fungsi *early stopping* karena peningkatannnya tidak signifikan.

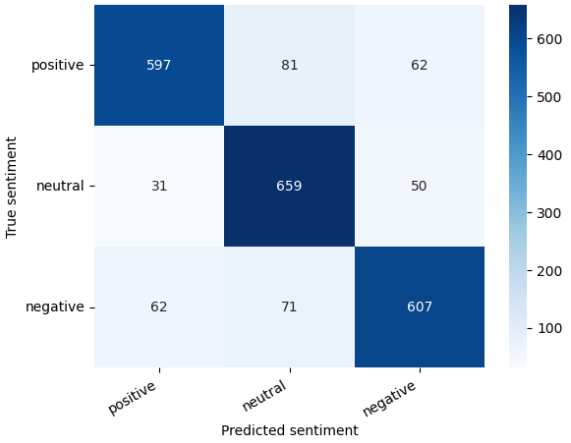
4.13. *Confusion Matrix*

Tahap *confusion matrix* digunakan untuk mengukur kinerja model yang telah diterapkan. Hasil *confusion matrix* pada skenario 1 (IndoBERT + *thumbsUp Only*) dan skenario 2 (IndoBERT) adalah sebagai berikut.



Gambar 11. Hasil *Confusion Matrix* pada Skenario 1 (IndoBERT + *thumbsUp Only*)

Gambar 11 menunjukkan hasil dari 228 ulasan *positive*, model mengklasifikasikan 196 dengan benar (*True Positive*). Namun, 17 ulasan salah diklasifikasikan sebagai *neutral* (*False Neutral*) dan 15 sebagai *negative* (*False Negative*). Lalu dari 228 ulasan *neutral*, model mengklasifikasikan 225 dengan benar (*True Neutral*). Namun, 3 ulasan salah diklasifikasikan sebagai *positive* (*False Positive*). Begitu juga dari 229 ulasan *negative*, model mengklasifikasikan 211 dengan benar (*True Negative*). Namun, 8 ulasan salah diklasifikasikan sebagai *positive* (*False Positive*) dan 10 sebagai neutral (*False Neutral*).



Gambar 12. Hasil *Confusion Matrix* pada Skenario 2 (IndoBERT)

Gambar 12 menunjukkan hasil dari 740 ulasan *positive*, model mengklasifikasikan 587 dengan benar (*True Positive*). Namun, 81 ulasan salah diklasifikasikan sebagai *neutral* (*False Neutral*) dan 62 sebagai *negative* (*False Negative*). Lalu dari 740 ulasan *neutral*, model mengklasifikasikan 659 dengan benar (*True Neutral*). Namun, 31 ulasan salah diklasifikasikan sebagai *positive* (*False Positive*) dan 50 sebagai negative (*False Negative*). Begitu juga dari 740 ulasan *negative*, model mengklasifikasikan 607 dengan benar (*True Negative*). Namun, 62 ulasan salah diklasifikasikan sebagai *positive* (*False Positive*) dan 71 sebagai neutral (*False Neutral*).

	precision	recall	f1-score	support
positive	0.95	0.86	0.90	228
neutral	0.89	0.99	0.94	228
negative	0.93	0.92	0.93	229
accuracy			0.92	685
macro avg	0.92	0.92	0.92	685
weighted avg	0.92	0.92	0.92	685

Gambar 13. Hasil Performa Model pada Skenario 1 (IndoBERT + *thumbsUp Only*)

Berdasarkan Gambar 13 didapatkan hasil performa model pada skenario 1, menunjukkan performa yang baik dalam mengklasifikasikan sentimen positif, netral, dan negatif. Berdasarkan hasil

evaluasi, model memiliki *precision* sebesar 95% untuk sentimen positif, 89% untuk netral, dan 93% untuk negatif, yang menunjukkan tingkat ketepatan prediksi yang tinggi. *Recall* model mencapai 86% untuk positif, 99% untuk netral, dan 92% untuk negatif, menandakan model mampu mengidentifikasi mayoritas ulasan dengan benar di setiap kategori. *F1-score* masing-masing kategori adalah 90% untuk positif, 94% untuk netral, dan 93% untuk negatif, yang mencerminkan keseimbangan antara *precision* dan *recall*.

Akurasi keseluruhan model mencapai 92%, dengan *macro average* dan *weighted average precision*, *recall*, serta *f1-score* sebesar 92%, menunjukkan konsistensi performa di semua kategori. Jumlah data yang hampir merata untuk setiap kategori memastikan distribusi data yang seimbang, sehingga model dapat menangkap pola dengan baik tanpa bias yang signifikan.

	precision	recall	f1-score	support
positive	0.87	0.81	0.83	740
neutral	0.81	0.89	0.85	740
negative	0.84	0.82	0.83	740
accuracy			0.84	2220
macro avg	0.84	0.84	0.84	2220
weighted avg	0.84	0.84	0.84	2220

Gambar 14. Hasil Performa Model pada Skenario 2 (IndoBERT)

Berdasarkan Gambar 14 didapatkan hasil performa model pada skenario 2, menunjukkan performa yang baik dengan akurasi keseluruhan sebesar 84%. Nilai *precision* untuk sentimen positif, netral, dan negatif masing-masing adalah 87%, 81%, dan 84%, menunjukkan tingkat ketepatan prediksi yang cukup tinggi. *Recall* untuk ketiga kategori adalah 81% untuk positif, 89% untuk netral, dan 82% untuk negatif, menandakan bahwa model mampu mengidentifikasi mayoritas ulasan dengan benar. *F1-score* untuk semua kategori berkisar antara 83% hingga 85%, menunjukkan keseimbangan yang baik antara *precision* dan *recall*.

Macro average dan *weighted average* untuk precision, recall, dan F1-score masing-masing adalah 0,84, menandakan performa yang konsisten di semua kategori.

4.14. *Report Classifier*

Tabel 9. Hasil Pemodelan pada Skenario 1

Skenario 1 (IndoBERT + *thumbsUp* Only)			
Accuracy	***Precision***	***Recall***	***F1-score***
92%	92%	92%	92%

Tabel 10. Hasil Pemodelan pada Skenario 2

Skenario 2 (IndoBERT)			
Accuracy	***Precision***	***Recall***	***F1-score***
84%	84%	84%	84%

Berdasarkan hasil pemodelan dengan menguji dua skenario yang sudah dilakukan, didapatkan bahwa hasil yang ada pada skenario 1 (IndoBERT + *thumbsUp Only*) menjadi hasil terbaik untuk diterapkan pada proses pemodelan klasifikasi data ulasan *game* eFootball menggunakan algoritma IndoBERT. Karena memiliki performa akurasi yang lebih baik dibandingkan dengan skenario 2 yaitu 92% pada data *testing*. Hasil *confusion matrix* juga menunjukkan performa stabil pada metrik lain seperti *precision*, *recall*, dan *f1-score*, yang menunjukkan bahwa model tidak hanya akurat tapi juga seimbang dalam memprediksi kelas positif, netral, dan negatif. Dengan hasil precision sebesar 92%, recall 92% dan f1-score 92%. Hal ini membuktikan bahwa performa model yang menggunakan validasi *thumbsUp* memiliki hasil yang lebih bagus dikarenakan data ulasan dengan validasi *thumbsUp* memiliki relevansi yang lebih baik terhadap rating yang diberikan. Semakin banyak tinggi jumlah *thumbsUp* pada ulasan, semakin relavan ulasan tersebut dianggap sesuai dengan ratingnya.

Dari hasil penelitian ini dapat disimpulkan bahwa penggunaan validasi *thumbsUp* pada model IndoBERT mampu meningkatkan performa model secara signifikan, dengan perbedaan akurasi mencapai 8% lebih tinggi dibandingkan tanpa menggunakan validasi *thumbsUp*.

4.15. *Visualisasi Wordcloud*

Wordcloud adalah representasi visual dari kata-kata yang paling sering muncul dalam teks. Dalam analisis sentimen ulasan *game* eFootball, *wordcloud* digunakan untuk menampilkan kata-kata dominan berdasarkan kategori sentimen (positif, netral, negatif). Prosesnya melibatkan pengolahan teks, pemilihan kata kunci, dan visualisasi. Selain mengukur performa model sentimen, *wordcloud* memberikan wawasan lebih dalam tentang kata-kata dan isu utama yang diungkapkan pengguna, sehingga menjadi alat visualisasi yang efektif untuk analisis sentimen teks.

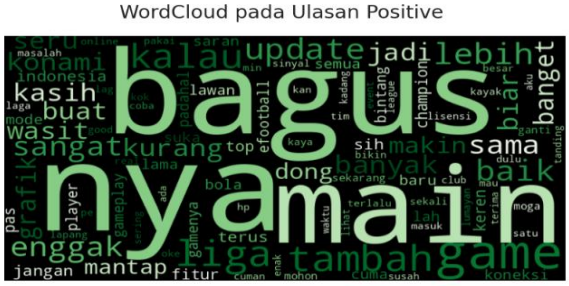
Gambar 15. *Wordcloud* Ulasan Negatif *Game* eFootball

Berdasarkan Gambar 15 menunjukkan *wordcloud* untuk ulasan dengan label sentimen negatif. Kata-kata yang sering muncul dalam ulasan

negatif mencerminkan berbagai masalah dan keluhan yang dialami oleh pengguna saat bermain *game eFootball*. Beberapa kata yang paling dominan dalam *wordcloud* tersebut antara lain "*game*", "main", "update", "enggak", "koneksi", "bagus", "malah", "kalah", "parah", "kurang", "susah", "*online*", "hp", "download", "lambat", "wasit", "lag", "lama", "jelek", "*game*play", "bug", dan "delay". Dari hasil analisis ini, dapat disimpulkan bahwa pengguna *game eFootball* banyak mengeluhkan tentang performa *game*, masalah koneksi, pembaruan yang tidak memuaskan, serta adanya *bug* dan *lag* dalam permainan. Pengembang *game* dapat menggunakan informasi ini untuk fokus pada perbaikan aspek-aspek yang sering dikeluhkan oleh pengguna, guna meningkatkan pengalaman bermain *game* yang lebih baik dan memuaskan.

Gambar 16. *Wordcloud* Ulasan Netral *Game* eFootball

Berdasarkan Gambar 16 menunjukkan *wordcloud* untuk ulasan dengan label sentimen *neutral*. Kata-kata yang sering muncul dalam ulasan netral mencerminkan pandangan yang seimbang dari pengguna. Beberapa kata yang paling dominan dalam *wordcloud* tersebut antara lain "main", "*game*", "bagus", "baik", "update", "koneksi", "liga", "enggak", "tambah", "lama", "bug", "grafik", dan "fitur". Dari analisis ini, dapat disimpulkan bahwa pengguna mengakui aspek-aspek positif seperti grafis dan fitur, tetapi juga menyoroti area yang memerlukan perbaikan, seperti masalah konektivitas dan *bug* dalam *game*. Hal ini memberikan pandangan yang lebih holistik tentang bagaimana *game* ini diterima oleh pengguna yang tidak memiliki kecenderungan sangat positif atau sangat negatif.



Gambar 17. *Wordcloud* Ulasan Positif *Game* eFootball

Berdasarkan Gambar 17 menunjukkan *wordcloud* untuk ulasan dengan label sentimen *positive*. Kata-kata yang sering muncul dalam ulasan positif mencerminkan bahwa pengguna sangat puas dengan *game* ini. Beberapa kata yang paling dominan dalam *wordcloud* tersebut antara lain "bagus", "main", "*game*", "baik", "update", "liga", "lebih", "tambah", "sangat", "grafik", "mantap", "wasit", "makin", "top", "event", "bintang", "sekarang", dan "keren". Dari analisis ini, ulasan positif terhadap *game eFootball* di *Google play store*menunjukkan bahwa pengguna sangat menghargai berbagai aspek dari *game* ini, terutama dari segi *gameplay*, pembaruan, fitur liga, grafis, dan berbagai event. Kata-kata yang sering muncul menggambarkan tingkat kepuasan yang tinggi dan pengalaman bermain yang nyaman dan menyenangkan. Hal ini memberikan pandangan bahwa *game eFootball* berhasil memenuhi harapan banyak pengguna dan memberikan pengalaman bermain yang baik.

5. KESIMPULAN DAN SARAN

Penelitian ini mengklasifikasikan sentimen ulasan pada *game* eFootball menggunakan model IndoBERT. Data dikumpulkan melalui *crawling* dari *Google Play Store* menggunakan *library google-play-scraper*, dengan rentang waktu 12 November 2023 hingga 21 Juni 2024, menghasilkan 20.997 ulasan. Proses pembersihan data dilakukan menggunakan NLTK dan Sastrawi. Label sentimen (positif, netral, negatif) ditentukan berdasarkan rating: 1-2 untuk negatif, 3 untuk netral, dan 4-5 untuk positif. Untuk mengatasi ketidakseimbangan data, digunakan metode *random oversampling*. Pemodelan dilakukan dalam dua skenario: Skenario 1 (IndoBERT + *thumbsUp*) dan Skenario 2 (IndoBERT). Dataset dibagi menjadi 80% *training*, 10% *validation*, dan 10% *testing*. *Hyperparameter* yang digunakan adalah *batch size* 32, *dropout* 0,1, dan *learning rate* 0,000003. Berdasarkan penelitian ini, penggunaan validasi *thumbsUp* terbukti meningkatkan performa model sentimen secara signifikan. Pada skenario dengan validasi *thumbsUp*, model menghasilkan akurasi yang lebih tinggi dibandingkan skenario tanpa validasi *thumbsUp*. Hal ini mengindikasikan bahwa ulasan dengan validasi *thumbsUp* memiliki relevansi yang lebih baik dan dapat memperkuat proses klasifikasi sentimen. Dengan akurasi 92% pada skenario pertama dan 84% pada skenario kedua, jelas bahwa validasi *thumbsUp* memberikan kontribusi besar terhadap kualitas hasil prediksi model IndoBERT dalam menganalisis sentimen ulasan eFootball di *Google Play Store*.

Dalam penelitian selanjutnya, disarankan untuk melakukan eksperimen tambahan dengan variasi dataset yang lebih besar dan dengan validasi *thumbsUp* yang lebih ketat untuk mendapatkan hasil yang lebih mendalam. Selain itu, penerapan metode lain dalam menangani ketidakseimbangan data, seperti SMOTE atau metode lain, dapat diujikan untuk membandingkan hasilnya. Pengembang *game* juga

dapat memanfaatkan hasil penelitian ini untuk lebih memahami umpan balik pengguna dan meningkatkan kualitas *game* eFootball berdasarkan ulasan yang relevan.

DAFTAR PUSTAKA

[1] V. A. Dihni, "Jumlah Gamers Indonesia Terbanyak Ketiga di Dunia," databoks.katadata.co.id. Accessed: Aug. 10, 2024. [Online]. Available: https://databoks.katadata.co.id/teknologi-telekomunikasi/statistik/950b8ba78451f97/jumlah-gamers-indonesia-terbanyak-ketiga-di-dunia

[2] Kementerian Komunikasi dan Informasi. "Topang Ekonomi Digital, Kominfo Dukung Pengembangan Industri Gim," kominfo.go.id. Accessed: Aug. 10, 2024. [Online]. Available: https://www.kominfo.go.id/content/detail/45061/siaran-pers-no-474hmkominfo102022-tentang-topang-ekonomi-digital-kominfo-dukung-pengembangan-industri-gim/0/siaran_pers#:~:text=Data%20Peta%20Ekosistem%20Industri%20Game,53%2C4%20juta%20PC%20gamers.

[3] Statista, "Mobile Games – Indonesia," statista.com. Accessed: Aug. 11, 2024. [Online]. Available: https://www.statista.com/outlook/dmo/digital-media/video-games/mobile-games/indonesia

[4] Gusrananda. (2023). PERILAKU FANATISME DAN KONSUMTIF GAME ONLINE EFOOTBALL. (Tugas Akhir). Universitas Islam Indonesia, Yogyakarta

[5] S. Adhikary "FC Mobile vs eFootball 2024 Mobile: Which game should play in 2023?" sportskeeda.com. Accessed: Aug. 13, 2024. [Online]. Available: https://www.sportskeeda.com/esports/fc-mobile-vs-eFootball-2024-mobile-which-game-play-2023.

[6] Data AI, "Top Apps – Download & Revenue Rank," data.ai.com. Accessed: Aug. 13, 2024. [Online]. Available: https://www.data.ai/intelligence/top-apps/downloads-revenue/overview?date=!(%272024-09-15%27,%272024-09-21%27)&device_code=android-all&country_code=ID&unified_category_id=800014&granularity=daily&top-apps.breakdown=unified

[7] Gameopedia, "Sentiment Analysis for Games | Opinion Mining | Gameopedia," gameopedia.com, Accessed: Oct. 12, 2024. [Online]. Available: https://www.gameopedia.com/our-services/sentiment-analysis-for-games/.

[8] L. F. S. Britto dan L. D. S. Pacífico, "Evaluating Video Game Acceptance in Game Reviews using Sentiment Analysis Techniques," *Proceedings of SBGames*, Recife, Brazil, pp. 399–402, Nov. 2020.

[9] Y. Yu, D. T. Dinh, B. H. Nguyen, F. Yu, dan V. N. Huynh, "Understanding Mobile Game Reviews Through Sentiment Analysis: A Case Study of PUBGm," *SpringerLink*, vol. 11, no. 3, pp. 400–402, 2023. [Online]. Available: https://link.springer.com/. [Accessed: Oct. 12, 2024].

[10] S. Chakraborty, I. Mobin, A. Roy, and M. H. Khan, "Rating Generation of Video Games using Sentiment Analysis and Contextual Polarity from Microblog," in *Proc. Int. Conf. Comput. Tech. Electron. Mech. Syst. CTEMS*, 2018, pp. 157–161, doi: 10.1109/CTEMS.2018.8769149.

[11] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," *IEEE Access*, vol. 8, pp. 23522–23530, 2020, doi: 10.1109/ACCESS.2020.2969854.

[12] H. W. Kang, K. K. Chye, Z. Y. Ong, and C. W. Tan, "The Science of Emotion: Malaysian Airlines Sentiment Analysis using BERT Approach," *Int. Conf. Digital Transformation Appl. ICDXA*, 2021, doi: 10.56453/icdxa.2021.1013.

[13] I. Septian and I. L. Kharisma, "Implementasi Metode Bidirectional Encoder Representations from Transformers (BERT) untuk Analisis Sentimen Komentar Pengguna Aplikasi Dana di Instagram," in *Seminar Nasional Rekayasa, Sains dan Teknologi*, vol. 3, 2023.

[14] A. S. P. Braja and A. Kodar, "Implementasi Fine-Tuning BERT untuk Analisis Sentimen terhadap Review Aplikasi PUBG Mobile di Google Play Store," *J. Inform. Merdeka Pasuruan*, vol. 7, no. 3, pp. 120, 2023, doi: 10.51213/jimp.v7i3.779.

[15] R. Feldman and J. Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, 2007.

[16] G. Salton, *Automatic Text Processing: Transformation, Analysis, and Retrieval of Information by Computer*. Addison Wesely, 1988.

[17] V. Mohan, "Text Mining Research: A Survey," *Int. J. Innov. Res. Comput. Commun. Eng.*, 2016, doi: 10.15680/IJIRCCE.2016.

[18] H. Liu, I. Chatterjee, M. Zhou, X. S. Lu, and A. Abusorrah, "Aspect-Based Sentiment Analysis: A Survey of Deep Learning Methods," *IEEE Trans. Comput. Social Syst.*, vol. 7, no. 6, pp. 1358–1375, 2020, doi: 10.1109/TCSS.2020.3033302.

[19] Q. T. Ain et al., "Sentiment Analysis Using Deep Learning Techniques: A Review," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 6, 2017.

[20] A. Ligthart, C. Catal, and B. Tekinerdogan, "Systematic Reviews in Sentiment Analysis: A

Tertiary Study," *Artif. Intell. Rev.*, vol. 54, no. 7, pp. 4997–5053, 2021, doi: 10.1007/s10462-021-09973-3.

[21] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *CoRR*, abs/1810.04805, 2018. [Online]. Available: http://arxiv.org/abs/1810.04805.

[22] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," 2020. [Online]. Available: http://arxiv.org/abs/2011.00677.

[23] Z. Lan et al., "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations," 2020. [Online]. Available: https://arviv.org/abs/1909.11942.

[24] A. Y. Triyanto and R. Kusumaningrum, "Implementasi Teknik Sampling untuk Mengatasi Imbalanced Data pada Penentuan Status Gizi Balita dengan Menggunakan Learning Vector Quantization," *J. Sist. Inform.*, vol. 19, pp. 39–50, 2017.

[25] R. D. Fitriani, H. Yasin, and F. Statistika, "Penanganan Klasifikasi Kelas Data Tidak Seimbang dengan Random Oversampling pada Naive Bayes," *J. Statistika Sains dan Matematika*, vol. 10, no. 1, pp. 11–20, 2021.

[26] D. Osinga, *Deep Learning Cookbook: Practical Recipes to Get Started Quickly*. O'Reilly Media, Inc., 2018.

[27] M. S. Wibawa, "Pengaruh Fungsi Aktivasi, Optimisasi, dan Jumlah Epoch Terhadap Performa Jaringan Saraf Tiruan," *J. Sist. dan Inform.*, vol. 11, no. 2, 2017.

[28] C. C. Aggarwal, "Neural Networks and Deep Learning," Spinger, 2018.

[29] I. Düntsch and G. Gediga, "Confusion Matrices and Rough Set Data Analysis," *J. Phys. Conf. Ser.*, vol. 1229, no. 1, 2019, doi: 10.1088/1742-6596/1229/1/012055.

[30] A. J. Gallego, J. Calvo-Zaragoza, J. J. Valero-Mas, and J. R. Rico-Juan, "Clustering-based K-Nearest Neighbor Classification for Large-Scale Data with Neural Codes Representation," *Pattern Recognit.*, vol. 74, pp. 531–543, 2018, doi: 10.1016/j.patcog.2017.09.038.

[31] M. R. Maarif, "Content Analysis on Twitter Users Interaction within First 100 Days of Jakarta's New Government by Using Text Mining," *J. Pekommas*, vol. 3, no. 2, pp. 137, 2018, doi: 10.30818/jpkm.2018.2030203.

[32] A. Fahmi, I. Ramadhan, and P. Studi, "Analisis Sentiment Masyarakat Selama Bulan Ramadhan dalam Menghadapi Pandemi Covid-19," *J. Inform. dan Sist. Inform.*, vol. 1, no. 1, pp. 608–617, 2020.