

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil Pengambilan Data (Data Acquisition)

4.1.1 Hasil Pengambilan Data

Proses pengambilan data (web scraping) berhasil dilakukan menggunakan pustaka google-play-scraper terhadap ulasan aplikasi TikTok pada platform Google Play Store. Proses ini menargetkan 5.000 data ulasan, namun pada rentang waktu filter yang ditetapkan (1 Januari 2026 s.d. 30 Juni 2026) hanya tersedia 482 data ulasan yang berhasil dihimpun beserta tiga atribut utamanya, yaitu teks ulasan (content), skor bintang (rating), dan tanggal pengiriman ulasan (date). Selisih antara target dan hasil aktual ini terjadi karena volume ulasan riil aplikasi TikTok pada rentang tanggal tersebut memang tidak mencapai 5.000. Cuplikan lima data teratas hasil scraping disajikan pada Tabel 4.1.1 berikut.

Tabel 4.1.1 Cuplikan Data Mentah Hasil Scraping

No	Ulasan	Rating	Tanggal
1	Good	5	2026-06-30 23:35:38
2	bagus sekali	5	2026-06-30 15:08:04
3	puas banget	5	2026-06-30 12:06:49
4	Saya sangat suka	5	2026-06-30 05:22:42
5	hai tiktok, maaf aq perbarui ulasan nya, dri yg bintang 5 jdi cuma 2...	2	2026-06-29 17:17:09

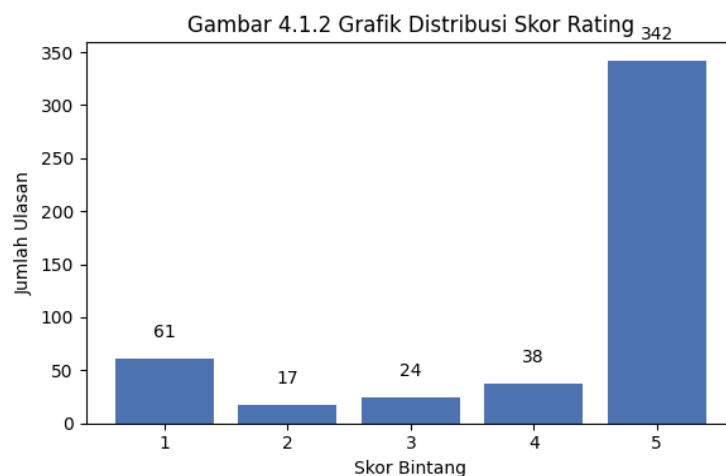
Total keseluruhan data yang berhasil ditarik pada tahap ini adalah 482 baris data. Jumlah ini berada di bawah target 5.000 baris data yang ditetapkan pada BAB III (3.2), karena volume ulasan riil aplikasi TikTok pada rentang tanggal 1 Januari– 30 Juni 2026 memang hanya sebanyak itu.

4.1.2 Deskripsi Karakteristik Skor Bintang Dataset

Sebelum dilakukan pelabelan, dilakukan analisis deskriptif terhadap distribusi skor bintang pada 482 data mentah untuk memahami karakteristik awal dataset.

Tabel 4.1.2 Deskripsi Distribusi Skor Bintang Data Mentah (482 Data)

Skor Bintang	Jumlah	Persentase (%)
1	61	12,66%
2	17	3,53%
3	24	4,98%
4	38	7,88%
5	342	70,95%



Gambar 4.1.2 Grafik Distribusi Skor Rating

Berdasarkan Tabel dan Gambar 4.1.2, data ulasan didominasi oleh Bintang 5 sebanyak 342 ulasan (70,95%), menunjukkan tingginya tingkat kepuasan mayoritas pengguna terhadap aplikasi TikTok. Pola ini konsisten dengan temuan pada dataset sebelumnya, bahkan dengan proporsi yang sedikit lebih tinggi. Namun demikian, terlihat pula fenomena polarisasi sentimen, ditandai dengan jumlah Bintang 1 sebanyak 61 ulasan (12,66%) yang menempati posisi kedua terbanyak, mengindikasikan adanya

sebagian pengguna yang mengalami masalah teknis atau ketidakpuasan signifikan, meskipun jarak proporsinya terhadap Bintang 5 jauh lebih besar dibandingkan pada dataset sebelumnya. Sementara itu, rating menengah (Bintang 2, 3, dan 4) memiliki porsi yang jauh lebih kecil, masing-masing di bawah 8%.

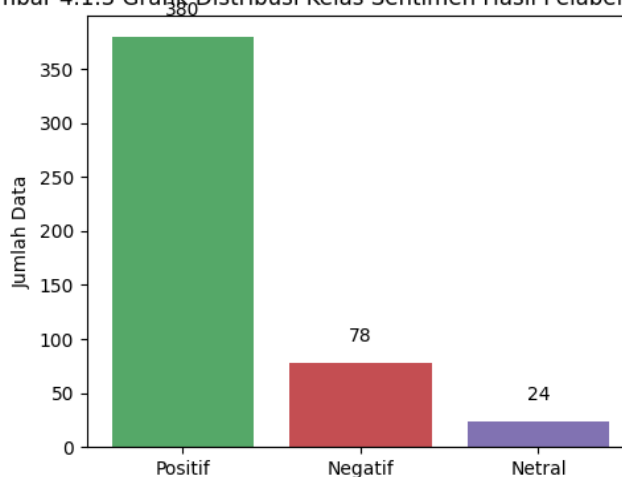
4.1.3 Hasil Pelabelan Otomatis (Distant Supervision)

Label sentimen dibuat secara otomatis berdasarkan skor bintang menggunakan teknik distant supervision, dengan aturan: Bintang 4-5 dipetakan menjadi kelas Positif, Bintang 3 menjadi kelas Netral, dan Bintang 1-2 menjadi kelas Negatif.

Tabel 4.1.3 Distribusi Kelas Sentimen Hasil Pelabelan Otomatis

Kelas Sentimen	Kriteria Skor Bintang	Jumlah	Persentase (%)
Positif	Bintang 4 dan 5	380	78,84%
Negatif	Bintang 1 dan 2	78	16,18%
Netral	Bintang 3	24	4,98%
Total	1 - 5	482	100,00%

Gambar 4.1.3 Grafik Distribusi Kelas Sentimen Hasil Pelabelan Otomatis



Gambar 4.1.3 Grafik Distribusi Kelas Sentimen Hasil Pelabelan Otomatis

Hasil pelabelan otomatis menunjukkan dominasi kelas Positif sebesar 78,84% (380 data), diikuti kelas Negatif sebesar 16,18% (78 data), dan kelas Netral yang hanya menyumbang 4,98% (24 data) dari keseluruhan dataset. Ketimpangan proporsi ini mengindikasikan bahwa dataset bersifat tidak seimbang (imbalanced), khususnya pada kelas Netral yang jumlah sampelnya jauh lebih sedikit dibandingkan dua kelas lainnya, bahkan dengan jumlah sampel absolut yang sangat kecil (24 data) akibat ukuran dataset keseluruhan yang juga jauh lebih kecil dari target awal. Kondisi ini menjadi salah satu tantangan utama yang perlu diantisipasi pada tahap pelatihan

4.2 Prapemrosesan Teks dan Pelabelan

Tahap prapemrosesan bertujuan membersihkan teks ulasan dari noise sebelum diproses oleh model. Berikut adalah contoh transformasi teks pada setiap tahapan menggunakan salah satu data ulasan yang mengandung variasi kata tidak baku, yaitu ulasan yang diawali “hai tiktok, maaf aq perbarui ulasan nya...”, karena kalimat “apk ini sangat keren” pada dataset lama tidak lagi muncul pada dataset hasil scraping baru.

4.2.1 Case Folding

Tahap case folding menyeragamkan seluruh huruf dalam teks ulasan menjadi huruf kecil (lowercase).

Tabel 4.2.1 Contoh Hasil Transformasi Tahap Case Folding

Tahap	Teks
Sebelum	Good
Sesudah	good

4.2.2 Cleansing

Tahap cleansing membersihkan teks dari URL, angka, tanda baca, dan karakter non-ASCII (emoji) menggunakan RegEx.

Tabel 4.2.2 Contoh Hasil Transformasi Tahap Cleansing

	Teks
Sebelum	bagus sekali 👍

Sesudah	bagus sekali
---------	--------------

4.2.3 Normalization (Normalisasi Kata)

Tahap normalization mengubah kata tidak baku/gaul menjadi bentuk baku menggunakan kamus normalisasi.

Tabel 4.2.3 Contoh Hasil Transformasi Tahap Normalization

Tahap	Teks
Sebelum	<i>hai tiktok maaf aq perbarui ulasan nya dri yg bintang jdi cuma karena sekarang tiktok makin eror suka tiba gak bisa di buka dan cara mulihin nya harus uninstal dli bru instal lgi klo ssekali sih oke ini jdi tiap hari kan mumet pdahal dri smua sosmed saya pling suka tiktok tolong pencerahannya tiktok</i>
Sesudah	<i>hai tiktok maaf aq perbarui ulasan nya dari yang bintang jadi cuma karena sekarang tiktok makin eror suka tiba tidak bisa di buka dan cara mulihin nya harus uninstal dli bru instal lgi klo ssekali sih oke ini jadi tiap hari kan mumet padahal dari smua sosmed saya pling suka tiktok tolong pencerahannya tiktok</i>

4.2.4 Tahap Stopword Removal

Tahap stopwords removal membuang kata umum yang tidak membawa muatan sentimen menggunakan pustaka Sastrawi.

Tabel 4.2.4 Contoh Hasil Transformasi Tahap Stopword Removal

Tahap	Teks
Sebelum	saya sangat suka
Sesudah	sangat suka

Pada tahap ini, kata “saya” yang tergolong kata ganti orang tanpa muatan sentimen berhasil dihilangkan, menyisakan kata-kata inti yang lebih relevan secara sentimen: “sangat suka”.

Setelah keseluruhan pipeline prapemrosesan diterapkan pada 482 data, sebanyak 444 data tersisa dengan teks yang valid (tidak kosong), sementara sisanya dieliminasi karena hanya berisi karakter yang seluruhnya hilang pada tahap cleansing (misalnya ulasan yang seluruhnya berupa emoji atau angka).

4.3 Pembagian Dataset

Dataset bersih (444 data) dibagi menggunakan metode Holdout dengan rasio 80:20, serta menerapkan teknik Stratified Random Sampling agar proporsi kelas sentimen tetap terjaga pada kedua subset data.

Tabel 4.3 Rincian Komposisi Pembagian Dataset Sentimen (Rasio 80:20)

Kelas Sentimen	Data Latih (80%)	Data Uji (20%)
Positif	277	69
Negatif	61	16
Netral	17	4
Total	355	89

Hasil pembagian menghasilkan 355 data latih dan 89 data uji. Proporsi masing-masing kelas sentimen pada data latih dan data uji tetap konsisten dengan proporsi pada dataset awal, membuktikan teknik stratified sampling berhasil diterapkan dengan baik meskipun ukuran dataset keseluruhan jauh lebih kecil dari target awal.

4.4 Tokenisasi dan Word Embedding IndoBERT

4.4.1 Tokenisasi

Teks bersih diubah menjadi token menggunakan tokenizer bawaan IndoBERT (indobenchmark/indobert-base-p1) dengan metode WordPiece, panjang maksimum token (max_length) sebesar 64. Contoh hasil tokenisasi pada data latih pertama, yaitu kalimat “love”, menghasilkan token: [CLS], love, [SEP], diikuti oleh token [PAD] hingga mencapai panjang 64.

IndoBERT berhasil direduksi menjadi vektor fitur padat berdimensi 128 melalui operasi konvolusi dan pooling, siap diteruskan ke lapisan klasifikasi akhir.

4.6 Skenario Pengujian dan Parameter Model

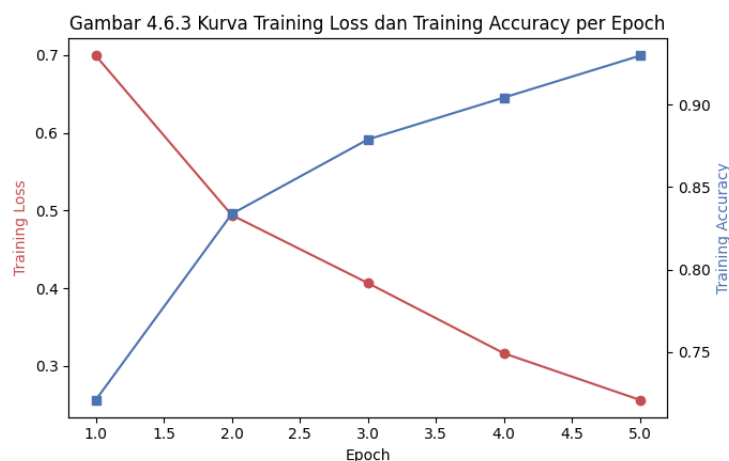
Model dilatih menggunakan konfigurasi hyperparameter sebagai berikut: optimizer AdamW, learning rate 2×10^{-5} , fungsi kerugian Cross-Entropy Loss, ukuran batch 32, dan jumlah epoch sebanyak 5.



Gambar 4.6.1 Skenario Pengujian dan Parameter Model

Tabel 4.6 Log Hasil Pelatihan Model per Epoch

Epoch	Training Loss	Training Akurasi
1	0,6992	72,11%
2	0,4941	83,38%
3	0,4067	87,89%
4	0,3163	90,42%
5	0,2563	92,96%



Gambar 4.6.2 Kurva Training Loss dan Training Accuracy per Epoch

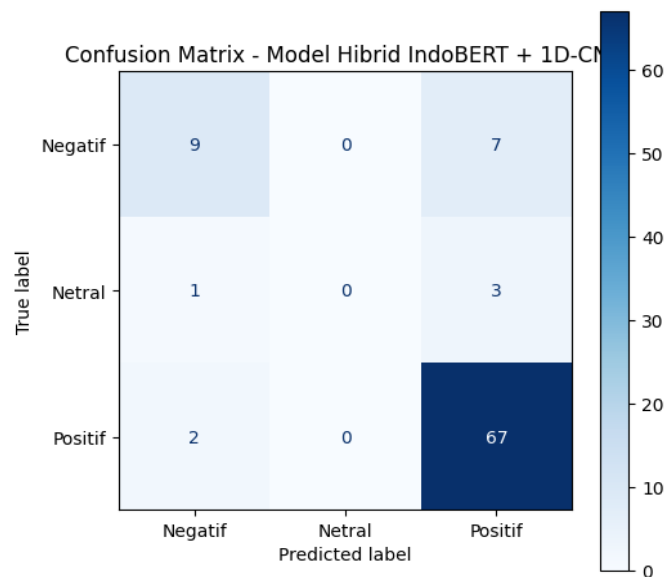
Grafik menunjukkan tren training loss yang terus menurun secara konsisten dari 0,6992 pada epoch pertama menjadi 0,2563 pada epoch kelima, diiringi peningkatan training akurasi dari 72,11% menjadi 92,96%. Tren ini mengindikasikan proses pembelajaran model berjalan stabil tanpa gejala divergensi, dan performa model terus membaik seiring bertambahnya epoch pada data latih, meskipun jumlah data latih jauh lebih sedikit (355 data) dibandingkan rencana awal.

4.7 Hasil Pengujian dan Evaluasi Kinerja Model

Model yang telah dilatih kemudian diuji terhadap 89 data uji yang telah diisolasi sejak tahap dataset splitting. Hasil evaluasi disajikan pada Tabel 4.7 berikut.

Tabel 4.7 Laporan Klasifikasi (Classification Report) pada Data Uji

Kelas	Precision	Recall	F1-Score	Support
Negatif	0,75	0,56	0,64	16
Netral	0,00	0,00	0,00	4
Positif	0,87	0,97	0,92	69
Akurasi			0,85	89
Macro Avg	0,54	0,51	0,52	89
Weighted Avg	0,81	0,85	0,83	89



Gambar 4.7 Confusion Matrix Model Hibrid IndoBERT + 1D-CNN

Secara keseluruhan, model hibrid IndoBERT + 1D-CNN mencatatkan nilai akurasi sebesar 85,39%, presisi sebesar 80,94%, recall sebesar 85,39%, dan F1-Score sebesar 82,71% pada data uji. Model menunjukkan performa yang sangat baik dalam mengklasifikasikan kelas Positif (F1-Score 0,92) dan cukup baik pada kelas Negatif (F1-Score 0,64). Namun demikian, model menunjukkan performa yang gagal total dalam mengklasifikasikan kelas Netral, dengan precision, recall, dan F1-Score yang seluruhnya bernilai 0,00 dari 4 data uji berlabel Netral, artinya model tidak berhasil menebak satu pun data Netral dengan benar. Hasil ini konsisten dengan temuan pada tahap pelabelan (4.1.3) yang menunjukkan proporsi kelas Netral yang sangat kecil (hanya 4,98% dari keseluruhan data, atau setara 17 data latih dan 4 data uji saja), sehingga model sama sekali kekurangan contoh yang cukup untuk mempelajari pola kelas ini. Fenomena ini menjadi bukti empiris yang lebih tegas atas dampak class imbalance yang telah diprediksi sebelumnya — bahkan lebih parah dibandingkan perkiraan awal karena ukuran dataset keseluruhan yang jauh di bawah target (482 dari target 5.000 data) turut memperkecil jumlah sampel Netral secara absolut. Temuan ini memperkuat urgensi saran penanganan ketidakseimbangan data (class-weighted loss maupun oversampling) yang telah diusulkan pada BAB V (5.2) untuk penelitian selanjutnya, dan mengindikasikan perlunya penambahan volume data (misalnya dengan memperluas rentang tanggal scraping) agar kelas Netral memiliki representasi yang memadai.

