

BAB 4

HASIL PENELITIAN

4.1 RINGKASAN HASIL PENELITIAN

Penelitian ini bertujuan untuk menganalisis sentimen dan tingkat kepuasan pengguna terhadap dua aplikasi *streaming*, yaitu Vidio dan Loklok, berdasarkan rating dan ulasan yang diperoleh di Google Play Store. Data yang dianalisis berjumlah 1.000 ulasan, masing-masing terdiri dari 500 ulasan per aplikasi. Setelah dilakukan klasifikasi sentimen menggunakan algoritma *Naïve Bayes*, ditemukan bahwa mayoritas ulasan pengguna bersifat negatif, yaitu sebesar 86,7% pada Vidio dan 73,3% pada Loklok. Evaluasi performa model menunjukkan bahwa klasifikasi bekerja sangat baik pada aplikasi Loklok (precision 0,98), namun kurang optimal pada aplikasi Vidio (precision 0,42). Selain itu, ditemukan juga ketidaksesuaian antara rating dan isi ulasan, dengan adanya sejumlah ulasan berating tinggi namun diklasifikasikan sebagai sentimen negatif, yaitu sebanyak 17 kasus pada Vidio dan 32 kasus pada Loklok. Hasil-hasil ini menunjukkan bahwa analisis sentimen berbasis teks memberikan wawasan tambahan yang tidak selalu terlihat dari rating numerik saja, serta menyoroti pentingnya pemilihan model yang sesuai dengan karakteristik data ulasan.

4.2 PENGAMBILAN DATA

Pengambilan data dilakukan dengan teknik *web scraping* menggunakan pustaka `google-play-scraper` pada bahasa pemrograman Python. Data yang dikumpulkan berupa ulasan pengguna terhadap dua aplikasi streaming, yaitu Vidio dan Loklok, yang tersedia di platform Google Play Store. Masing-masing aplikasi dikumpulkan sebanyak 500 ulasan terbaru dengan parameter pengambilan `lang='id'` (bahasa Indonesia), `country='id'` (wilayah Indonesia), dan `Sort.NEWEST` untuk mendapatkan ulasan paling baru. Ulasan yang diperoleh awalnya memiliki beberapa atribut seperti `reviewId`, `userName`, `userImage`, `content`, `score`, `thumbsUpCount`, dan `reviewCreatedVersion`. Namun dalam penelitian ini, hanya menggunakan dua atribut sebagai fokus utama penelitian yaitu:

- `content` – teks ulasan dari pengguna, digunakan untuk analisis sentimen, dan

- `score` – rating numerik dari 1 hingga 5, digunakan untuk melihat kecenderungan dan korelasi dengan sentimen.

Berikut ini adalah ada cuplikan kode dalam proses pengambilan data ulasan dan rating dari kedua aplikasi yaitu vidio dan loklok yang menggunakan tools Google Collab:

`ambil_ulasan_loklok(jumlah=500):`

```
def ambil_ulasan_vidio(jumlah_ulasan=500):
    semua_ulasan = []
    token_lanjutan = None

    while len(semua_ulasan) < jumlah_ulasan:
        ulasan, token_lanjutan = reviews(
            'com.vidio.android',
            lang='id',
            country='id',
            sort=Sort.NEWEST,
            count=100,
            continuation_token=token_lanjutan
        )

        semua_ulasan.extend(ulasan)
        if not token_lanjutan:
            break
        sleep(1)
    semua_ulasan = semua_ulasan[:jumlah_ulasan]
    df = pd.DataFrame(semua_ulasan)
    return df
```

```
def ambil_ulasan_loklok(jumlah=500):  
    hasil = []  
    token = None  
  
    while len(hasil) < jumlah:  
        ulasan, token = reviews(  
            'com.bergel.tauros',  
            lang='id',  
            country='id',  
            sort=Sort.NEWEST,  
            count=200,  
            continuation_token=token  
        )  
  
        hasil.extend(ulasan)  
        if not token:  
            break  
        sleep(1)  
    hasil = hasil[:jumlah]  
    df = pd.DataFrame(hasil)  
    return df
```

4.3 Hasil Preprocessing Data

Setiap ulasan pengguna dari Google Play Store telah melalui beberapa tahapan preprocessing untuk siap digunakan dalam penelitian analisis sentimen. Berikut ini setiap alur tahapan beserta ilustrasi hasilnya:

4.3.1 Pembersihan teks (*cleaning*)

Setelah data dikumpulkan, tahap pertama dalam preprocessing adalah *cleaning*. Proses ini dilakukan dengan menghapus URL, angka, tanda baca, huruf kapital, karakter khusus, serta melakukan normalisasi kata tidak baku dan slang serta typo. Berikut ini adalah hasil pembersihan teks dari ulasan pengguna aplikasi Vidio dan Loklok ditunjukkan pada Gambar 4.1 dan 4.2:

	content
0	ngefreeze
1	ini aplikasi vidio kenapa sekarang ga bisa log...
2	Hai untuk pihak developer nya, saya kembali de...
3	bonus premium dari xl ga bisa di claim
4	bagus

	cleaned
0	membeku
1	ini aplikasi vidio kenapa sekarang tidak bisa ...
2	hai untuk pihak pengembang nya saya kembali de...
3	bonus berbayar dari xl tidak bisa di klaim
4	bagus

Gambar 4.1 Data ulasan aplikasi Vidio sebelum dan sesudah proses cleaning

	content
0	terlalu banyak iklan 🙄
1	WOW
2	luar biasa cepet tayang nya dari aplikasi yang...
3	the bestt
4	Terakhir di perbaharui jadi bnyak iklan baru b...

	cleaned
0	terlalu banyak iklan
1	WOW
2	luar biasa cepat tayang nya dari aplikasi yang...
3	terbaik
4	terakhir di perbaharui jadi banyak iklan baru ...

Gambar 4.2 Data ulasan aplikasi Loklok sebelum dan sesudah proses cleaning

4.3.2 Filtering (Penghapusan Stopwords)

Setelah tahap pembersihan teks, dilakukan proses filtering untuk menghapus kata-kata yang dianggap tidak memiliki makna penting dalam analisis sentimen. Tujuan dari filtering adalah menyederhanakan teks dengan menghilangkan kata yang tidak berkontribusi terhadap makna inti ulasan, sehingga hasil analisis menjadi lebih akurat. Berikut ini adalah hasil filtering dari ulasan pengguna aplikasi Vidio dan Loklok ditunjukkan pada Gambar 4.3 dan 4.4.

```

                                content
0                                ngefreeze
1  ini aplikasi vidio kenapa sekarang ga bisa log...
2  Hai untuk pihak developer nya, saya kembali de...
3      bonus premium dari xl ga bisa di claim
4                                bagus

                                filtered
0                                membeku
1  aplikasi vidio kenapa tidak bisa masuk nomor h...
2  pengembang saya mengedit ulasan aplikasi berja...
3      bonus berbayar xl tidak bisa klaim
4                                bagus

```

Gambar 4.3 Data ulasan aplikasi Vidio sebelum dan sesudah proses filter

```

                                content
0                                terlalu banyak iklan 🙄
1                                wow
2  luar biasa cepet tayang nya dari aplikasi yang...
3                                the bestt
4  Terakhir di perbaharui jadi bnyak iklan baru b...

                                filtered
0                                terlalu banyak iklan
1                                wow
2      luar biasa cepat tayang aplikasi yang
3                                terbaik
4  perbaharui banyak iklan buka ada tanda x tidak...

```

Gambar 4.4 Data ulasan aplikasi Vidio sebelum dan sesudah proses filter

4.3.3 Stemming (Safe Stem)

Pengurangan kata ke bentuk dasarnya dilakukan dengan metode stemming. Pendekatan safe stem diterapkan untuk menjaga agar kata-kata penting yang mengandung muatan sentimen tidak hilang atau berubah makna secara drastis. Berikut ini adalah hasil stemming dari ulasan pengguna dari aplikasi Vidio dan Loklok ditunjukkan pada Gambar 4.5 dan 4.6.

	content
0	ngefreeze
1	ini aplikasi vidio kenapa sekarang ga bisa log...
2	Hai untuk pihak developer nya, saya kembali de...
3	bonus premium dari xl ga bisa di claim
4	bagus

	stemmed
0	beku
1	aplikasi vidio kenapa tidak bisa masuk nomor h...
2	kembang saya edit ulasan aplikasi jalan baik t...
3	bonus bayar xl tidak bisa klaim
4	bagus

Gambar 4.5 Data ulasan aplikasi Vidio sebelum dan sesudah proses stemming

	content
0	terlalu banyak iklan 🙄
1	WOW
2	luar biasa cepet tayang nya dari aplikasi yang...
3	the bestt
4	Terakhir di perbaharui jadi bnyak iklan baru b...

	stemmed
0	terlalu banyak iklan
1	WOW
2	luar biasa cepat tayang aplikasi yang
3	baik
4	baharu banyak iklan buka ada tanda x tidak gitu

Gambar 4.5 Data ulasan aplikasi Vidio sebelum dan sesudah proses stemming

4.3.4 Tokenizing

Kalimat hasil preprocessing dipecah menjadi token-token kata sebagai representasi unit analisis yang lebih sederhana dan mudah diproses oleh algoritma klasifikasi. Berikut ini adalah hasil tokenizing dari ulasan pengguna dari aplikasi Vidio dan Loklok ditunjukkan pada Gambar 4.7 dan 4.8.

```

                                content
0                                ngefreeze
1 ini aplikasi vidio kenapa sekarang ga bisa log...
2 Hai untuk pihak developer nya, saya kembali de...
3         bonus premium dari xl ga bisa di claim
4                                bagus

                                tokens
0                                [beku]
1 [aplikasi, vidio, kenapa, tidak, bisa, masuk, ...
2 [kembang, saya, edit, ulasan, aplikasi, jalan,...
3         [bonus, bayar, xl, tidak, bisa, klaim]
4                                [bagus]
```

Gambar 4.7 Data ulasan aplikasi Vidio sebelum dan sesudah proses tokenizing

```

                                content
0                                terlalu banyak iklan 🙄
1                                WOW
2 luar biasa cepet tayang nya dari aplikasi yang...
3                                the bestt
4 Terakhir di perbaharui jadi bnyak iklan baru b...

                                tokens
0                                [terlalu, banyak, iklan]
1                                [wow]
2 [luar, biasa, cepat, tayang, aplikasi, yang]
3                                [baik]
4 [baharu, banyak, iklan, buka, ada, tanda, x, t...
```

Gambar 4.8 Data ulasan aplikasi Loklok sebelum dan sesudah proses tokenizing

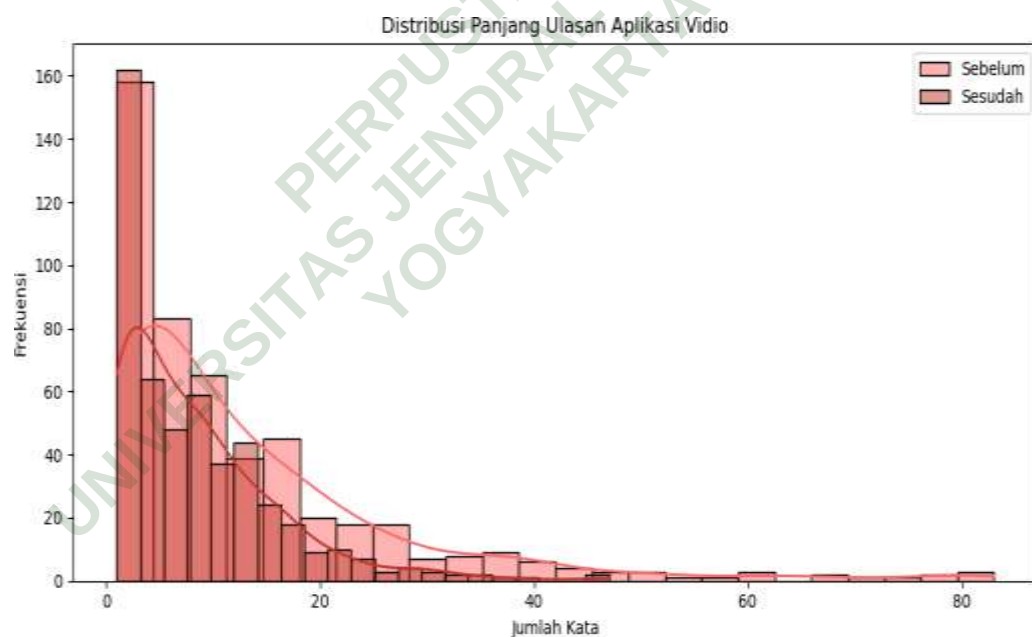
4.3.5 Visualisasi Hasil Preprocessing

Setelah melalui tahapan preprocessing, dilakukan beberapa visualisasi untuk masing-masing aplikasi, yaitu vidio dan loklok. Visualisasi ini bertujuan untuk memberikan gambaran mengenai perubahan data ulasan pengguna sebelum dan sesudah diproses, serta kata dominan yang sering muncul.

A. Aplikasi Vidio

a) Distribusi Panjang Ulasan

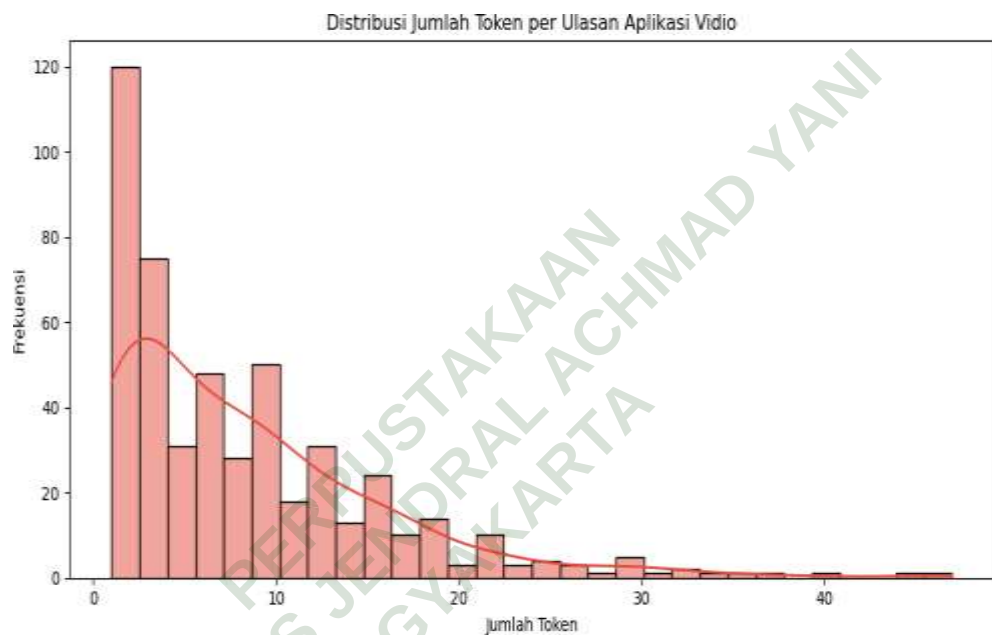
Grafik pada Gambar 4.9 menunjukkan distribusi jumlah kata dalam ulasan pada aplikasi vidio umumnya terdiri dari sedikit kata, yaitu 1 hingga 5 kata, sebelum maupun sesudah preprocessing. Setelah melalui proses pembersihan, filtering, dan stemming, jumlah kata dalam ulasan cenderung berkurang. Hal ini menunjukkan bahwa teks menjadi lebih singkat dan relevan untuk analisis sentimen



Gambar 4.9 Distribusi Panjang Ulasan Aplikasi Vidio Sebelum dan Sesudah Preprocessing

b) Distribusi Jumlah Token Per Ulasan

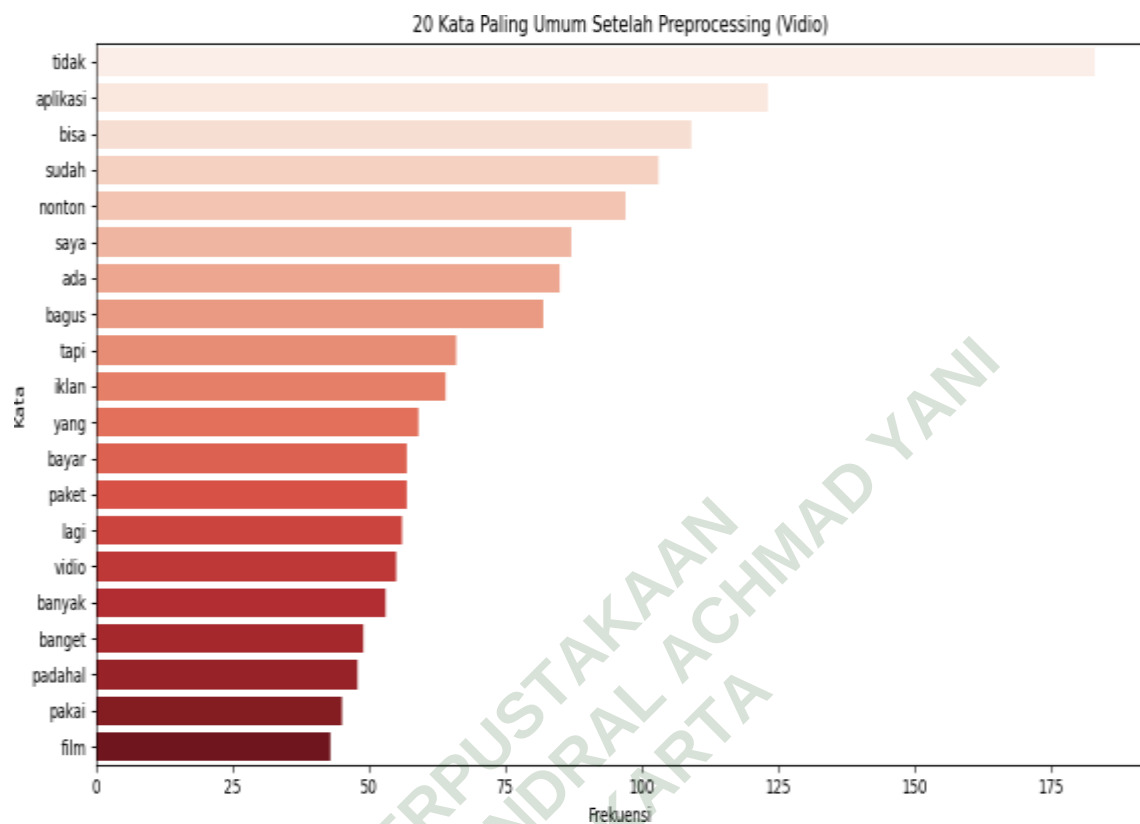
Grafik Gambar 4.10 menunjukkan distribusi jumlah token per ulasan menunjukkan bahwa mayoritas ulasan pengguna aplikasi Vidio terdiri dari 1 hingga 10 token. Hal ini mencerminkan bahwa sebagian besar ulasan bersifat singkat, yang umum dijumpai pada platform seperti Google Play Store.



Gambar 4.10 1 Distribusi Jumlah Token per Ulasan Pengguna Aplikasi Vidio Setelah Preprocessing

c) Distribusi 20 Kata Paling Umum Setelah Preprocessing

Grafik Gambar 4.11 menunjukkan 20 kata yang paling sering muncul dalam ulasan pengguna aplikasi Vidio. Kata-kata seperti “tidak”, “aplikasi”, “bisa”, dan “nonton” muncul dengan frekuensi tinggi. Hal ini mencerminkan isu utama yang dibahas pengguna, terutama terkait pengalaman menonton dan kendala teknis aplikasi.



Gambar 4.11 20 Kata Teratas Hasil Preprocessing Ulasan Aplikasi Vidio

d) Wordcloud Ulasan Vidio

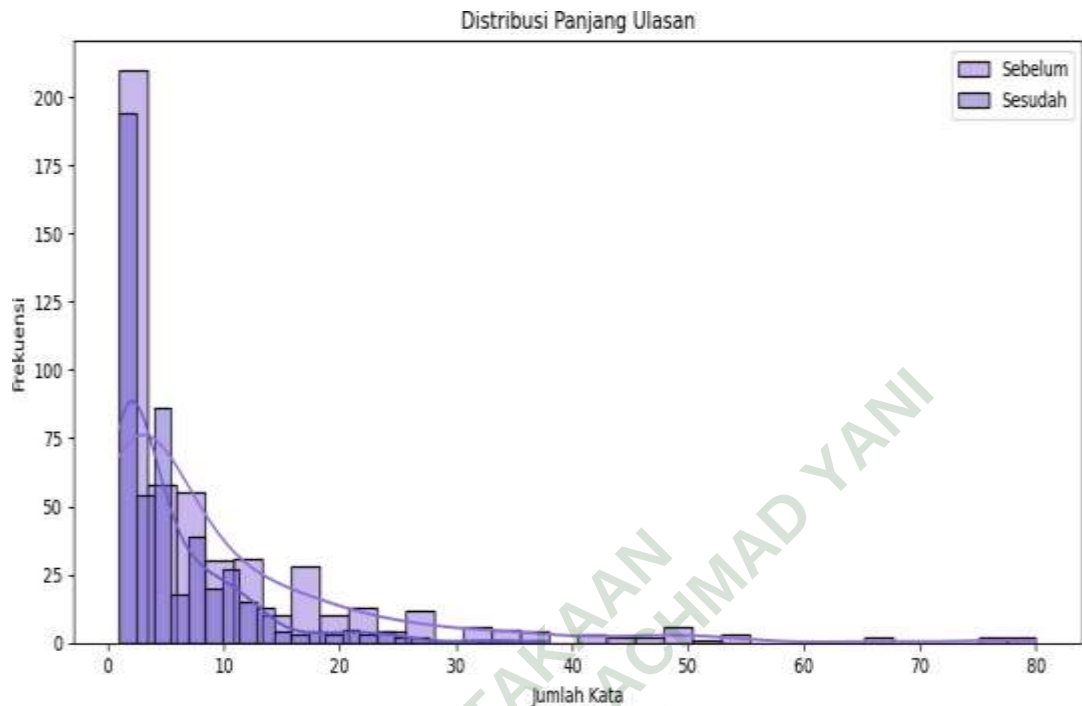
Grafik Gambar 4.12 menunjukkan visualisasi wordcloud dari kata-kata yang paling sering muncul dalam ulasan pengguna aplikasi Vidio setelah proses preprocessing. Ukuran huruf mencerminkan frekuensi kemunculan kata; semakin besar, semakin sering kata tersebut muncul. Kata seperti *aplikasi*, *tidak*, *saya*, *bisa*, dan *nonton* tampak dominan, mencerminkan isu utama yang dibahas pengguna.

Wordcloud Kata-Kata yang Serin

Wordcloud Kata-Kata yang Serin

Wordcloud Kata-Kata yang Serin

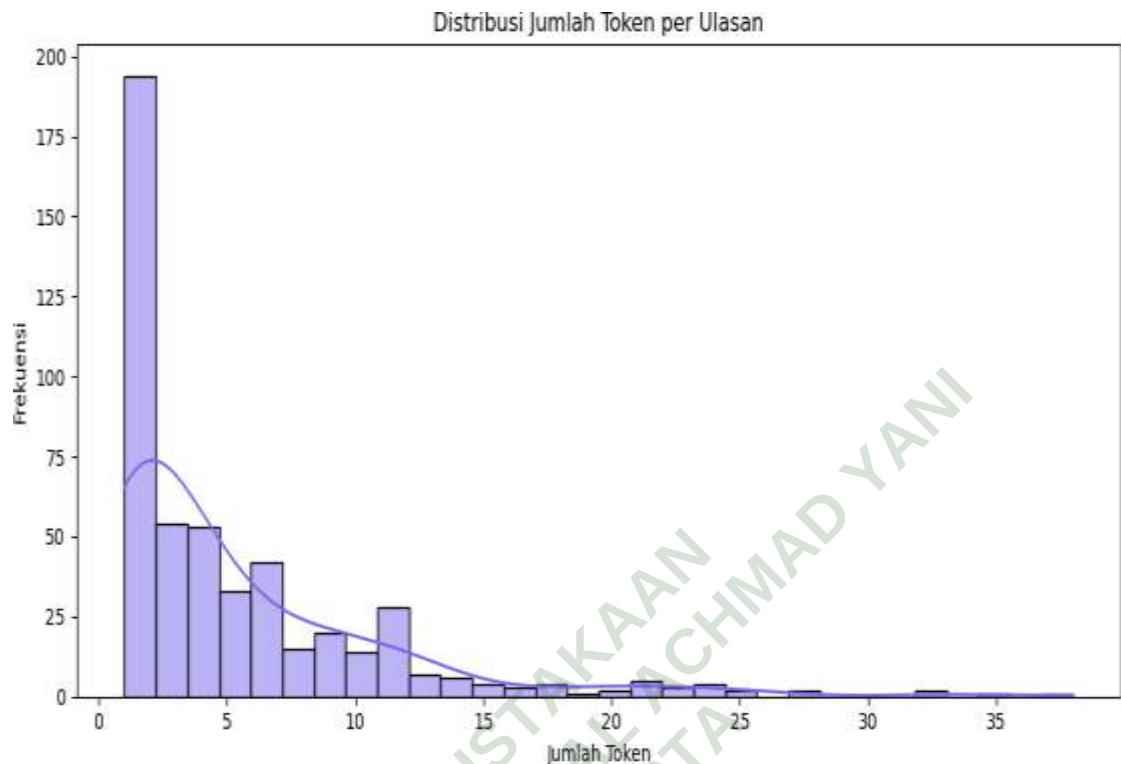
Wordcloud Kata-Kata yang Serin



Gambar 4.13 Distribusi Panjang Ulasan Aplikasi Loklok Sebelum dan Sesudah Preprocessing

b) Distribusi Jumlah Token per Ulasan

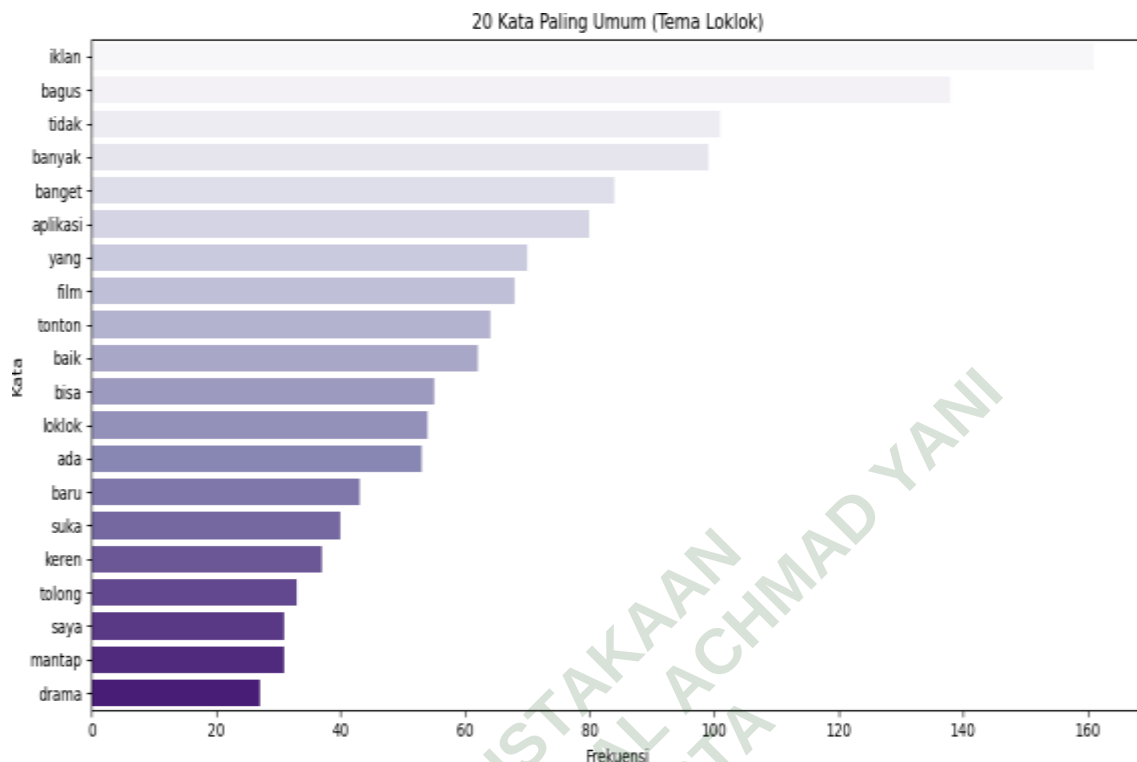
Gambar 4.14 menunjukkan bahwa sebagian besar ulasan hanya terdiri dari sedikit kata yaitu 1 sampai 5 token. Semakin panjang ulasan, jumlahnya semakin sedikit. Pola ini menunjukkan bahwa ulasan pengguna cenderung singkat. Yang merupakan hal umum dijumpai pada platform seperti Google Play Store.



Gambar 4.14 Distribusi Jumlah Token per Ulasan Pengguna Aplikasi Loklok

c) Distribusi 20 Kata Paling Umum Setelah Preprocessing

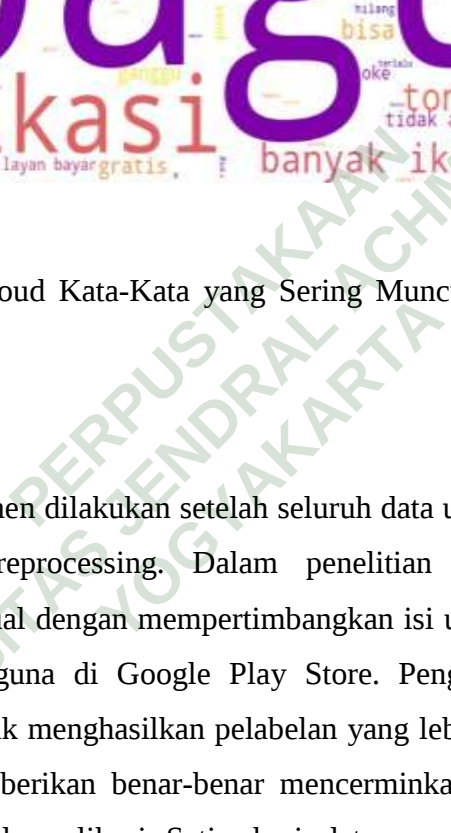
Grafik Gambar 4.15 menunjukkan kata yang paling sering muncul dalam ulasan pengguna aplikasi Loklok. Kata “iklan”, “bagus”, dan “tidak” menjadi yang paling dominan. Ini mencerminkan campuran sentimen positif dan negatif, dengan banyak pengguna menyoroti keberadaan iklan serta kualitas aplikasi.



Gambar 4.15 20 Kata Teratas Hasil Preprocessing Ulasan Aplikasi Loklok

d) Wordcloud Ulasan Loklok

Grafik Gambar 4.16 menunjukkan bahwa kata “iklan” paling menonjol, menandakan banyaknya keluhan pengguna terhadap gangguan iklan. Selain itu, kata-kata seperti “bagus”, “film”, dan “tonton” juga sering muncul. Menggambarkan minat pengguna terhadap konten yang disediakan. Ini mencerminkan campuran sentimen, dengan fokus utama pada kualitas layanan dan pengalaman menonton.



oud Kata-Kata yang Sering Muncu

oud Kata-Kata yang Sering Munc

oud Kata-Kata yang Sering Munc

oud Kata-Kata yang Sering Munc

- oud Kata-Kata yang Sering Munc

oud Kata-Kata yang Sering Munc

utama. Adapun kategori label sentimen dalam penelitian ini dibagi menjadi dua kelas utama:

- Label Positif: Jika ulasan mengandung pujian, kesan menyenangkan, atau menunjukkan kepuasan terhadap aplikasi.
- Negatif: Jika ulasan berisi kritik, keluhan, atau pengalaman yang mengecewakan.

Dengan demikian, seluruh data ulasan telah diberikan label sentimen yang sesuai dan membentuk data terlabeli. Data terlabeli ini selanjutnya digunakan sebagai data latih dan data uji dalam proses pelatihan dan evaluasi model klasifikasi sentimen menggunakan algoritma Naïve Bayes.

4.5 MEMBANGUN MODEL KLASIFIKASI

Setelah data ulasan pengguna aplikasi Vidio dan Loklok melalui proses *preprocessing* dan labelisasi sentimen, langkah selanjutnya adalah membangun model klasifikasi untuk menganalisis sentimen pengguna terhadap masing-masing aplikasi. Tahapan ini mencakup pembagian data, pelatihan model, hingga evaluasi performanya model klasifikasi.

4.5.1 PEMBAGIAN DATA

Dalam penelitian ini menggunakan total 500 data ulasan untuk setiap aplikasi, sehingga terdapat 1.000 data secara keseluruhan. Setiap data ulasan telah melalui proses pembersihan teks (*preprocessing*), dan pelabelan sentimen (positif dan negatif). Untuk membangun model klasifikasi yang akurat, data dibagi menjadi dua bagian utama, yaitu data latih (*training data*), dan data uji (*testing data*).

Pembagian dilakukan dengan proporsi 64% data latih dan 36% data uji. Proporsi ini dipilih agar model mendapatkan cukup informasi dari data pelatihan, namun tetap memiliki data uji yang memadai untuk mengevaluasi performa klasifikasi. Data latih digunakan dalam proses pelatihan model klasifikasi Naïve Bayes, sementara data uji digunakan untuk mengukur seberapa baik model mampu mengklasifikasikan data baru secara akurat. Berikut ini adalah rincian pembagian data untuk masing-masing aplikasi:

Tabel 4.1 Pembagian Data Latih dan Data Uji

Aplikasi	Total Data	Data Latih (64%)	Data Uji (36%)
Vidio	500	320	180
Loklok	500	320	180

Dengan proporsi ini, masing-masing model untuk aplikasi Vidio dan Loklok dilatih menggunakan 320 data dan diuji menggunakan 180 data. Pembagian dilakukan secara acak terkontrol (*random state*) untuk menghindari bias dan memastikan hasil yang dapat direplikasi.

4.5.2 Pelatihan Model

Setelah proses data ulasan dibagi menjadi data latih dan data uji, tahap selanjutnya melakukan pelatihan model klasifikasi guna menganalisis sentimen pengguna terhadap aplikasi vidio dan loklok. Algoritma yang digunakan dalam penelitian ini adalah *Multinomial Naïve Bayes*, yang dikenal efektif untuk klasifikasi teks karena memanfaatkan frekuensi kemunculan kata dalam dokumen. Sebelum proses pelatihan dilakukan, terlebih dahulu dilakukan ekstraksi fitur dari data teks menggunakan metode *TF-IDF (Term Frequency-Inverse Document Frequency)*. Metode ini berfungsi mengubah teks menjadi bentuk numerik dengan mempertimbangkan suatu kata dalam keseluruhan dokumen. Parameter yang digunakan dalam proses TF-IDF meliputi:

- `ngram_range=(1, 3)` untuk menangkap unigram hingga trigram (satu hingga tiga kata berurutan),
- `min_df=2` untuk menghilangkan kata yang jarang muncul,
- `max_df=0.85` untuk mengabaikan kata yang terlalu sering muncul.

Setelah fitur berhasil diekstraksi, model *Multinomial Naïve Bayes*, dilatih menggunakan data latih direpresentasikan dalam bentuk numerik. Melalui proses ini, model belajar dari pola distribusi kata dalam ulasan yang sudah diberi label sentimen positif dan negatif. Pelatihan dilakukan secara terpisah untuk masing-masing aplikasi menggunakan prosedur dan parameter yang sama, sehingga hasil evaluasi model dapat dibandingkan secara adil dan objektif.

Berikut adalah potongan kode utama yang digunakan dalam proses pelatihan model:

```
# Ekstraksi fitur dengan TF-IDF
Vectorizer = TfidfVectorizer(ngram_range=(1,3), min_df=2, max_df=0.85)
X_train_vec = vectorizer.fit_transform(X_train)

# Pelatihan model Naïve Bayes
model = MultinomialNB()
model.fit(X_train_vec, y_train)
```

4.5.3 Pengujian dan Evaluasi Model

Setelah proses pelatihan selesai dilakukan, langkah selanjutnya adalah pengujian dan evaluasi model klasifikasi untuk mengukur performa dalam mengklasifikasikan sentimen pengguna. Pengujian dilakukan dengan menggunakan data uji (testing data) yang telah dipisahkan sebelumnya dari dataset utama. Dalam penelitian ini, proses pengujian dilakukan secara terpisah untuk masing-masing aplikasi, yaitu Vidio dan Loklok. Meskipun dilakukan secara terpisah, seluruh prosedur, parameter, dan metrik evaluasi yang digunakan tetap sama, sehingga hasil yang diperoleh dapat dibandingkan secara adil dan objektif.

Model yang dilatih akan digunakan untuk memprediksi sentimen dari data uji, dan hasil prediksi akan dibandingkan dengan label aktual (labelisasi manual) untuk menghitung metrik evaluasi. Evaluasi kinerja model dilakukan dengan menggunakan empat metrik utama, yaitu:

- **Akurasi (*accuracy*):** Persentase jumlah prediksi yang benar terhadap keseluruhan data uji.
- **Precision:** Proporsi prediksi positif yang benar-benar termasuk ke dalam kelas positif
- **Recall:** proporsi data aktual positif yang berhasil dikenali secara benar oleh model.
- **F1-Score:** Nilai harmonis antara precision dan recall, yang digunakan untuk mengukur keseimbangan kinerja model.

Berikut ini adalah potongan kode yang digunakan untuk melakukan pengujian dan evaluasi model:

```
# Prediksi data uji
y_pred = model.predict(X_test_vec)

# Evaluasi performa model
print("Akurasi:", accuracy_score(y_test, y_pred))
print("Precision:", precision_score(y_test, y_pred,
pos_label='positif'))
print("Recall:", recall_score(y_test, y_pred, pos_label='positif'))
print("F1-Score:", f1_score(y_test, y_pred, pos_label='positif'))
```

Hasil dari pengujian ini akan digunakan untuk mengevaluasi seberapa baik model dalam mengklasifikasikan sentimen berdasarkan data ulasan, serta menjadi dasar untuk membandingkan performa antara Vidio dan Loklok.

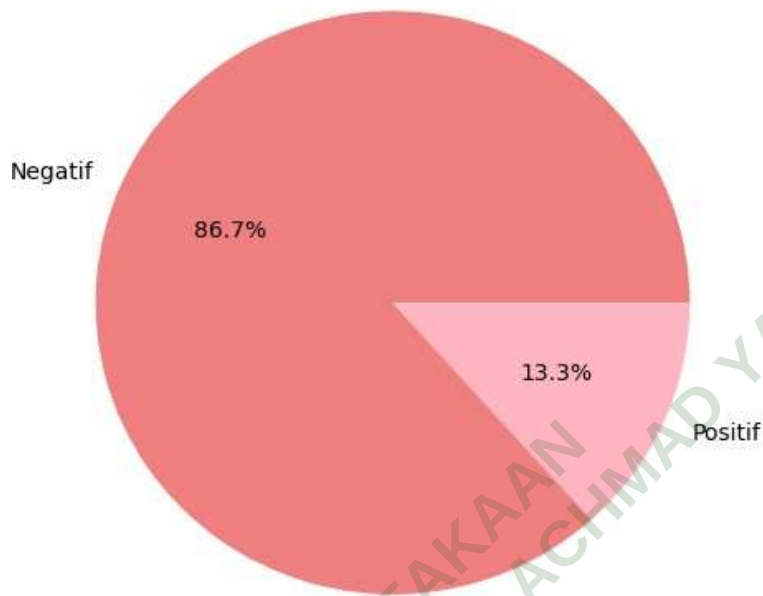
4.6 Evaluasi dan Analisis Hasil Klasifikasi

Setelah model *Naïve Bayes* dilatih dan diuji terhadap data ulasan pengguna aplikasi Vidio dan Loklok, dilakukan evaluasi serta analisis terhadap performa dan hasil klasifikasi. Analisis ini mencakup distribusi antara rating dan prediksi sentimen, serta perbandingan antara kedua model ini.

4.6.1 Distribusi Sentimen

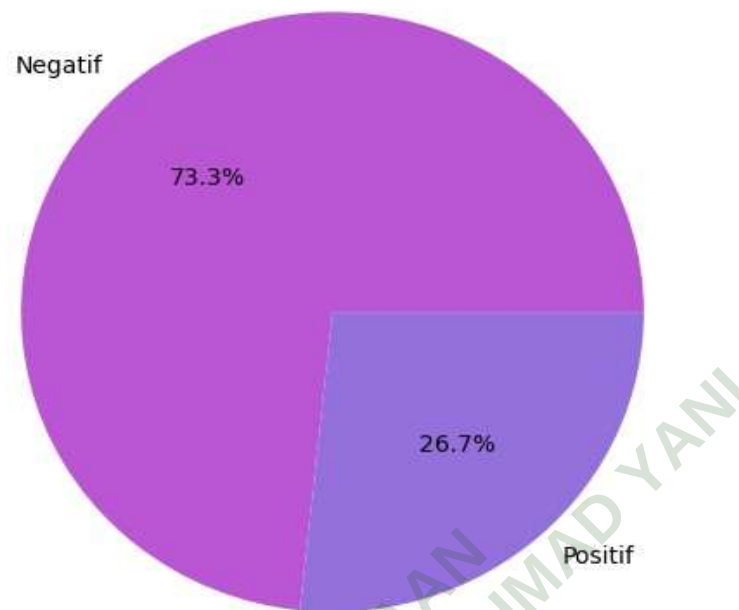
Distribusi sentimen menunjukkan kecenderungan ulasan pengguna terhadap masing-masing aplikasi berdasarkan hasil klasifikasi menggunakan *algoritma Naïve Bayes*. Visualisasi dilakukan dalam bentuk diagram pie (pie chart) agar perbandingan sentimen antara aplikasi Vidio dan Loklok lebih mudah diamati.

Distribusi Sentimen Prediksi – Aplikasi Vidio

**Gambar 4.17** Distribusi Sentimen Ulasan Aplikasi Vidio

Pada aplikasi Vidio, hasil klasifikasi menunjukkan bahwa sentimen negatif mendominasi sebesar 86.7% sedangkan sentimen positif sebesar 13.3%. Hal ini mengindikasikan bahwa sebagian besar pengguna vidio cenderung menyampaikan pengalaman kurang memuaskan terhadap layanan aplikasi tersebut.

Distribusi Sentimen Prediksi – Aplikasi Loklok



Gambar 4.18 1 Distribusi Sentimen Ulasan Aplikasi Loklok

Berbeda dengan Vidio, aplikasi Loklok memiliki distribusi sentimen negatif sebesar 73.3% dan sentimen positif sebesar 26.7%. Meskipun sentimen negatif masih mendominasi, presentase sentimen positif yang lebih tinggi menunjukkan bahwa Loklok mendapatkan respon yang relative lebih baik dari pengguna dibandingkan Vidio.

Tabel 4.2 Ringkasan Distribusi Sentimen

Aplikasi	Sentimen Positif (%)	Sentimen Negatif (%)
Vidio	13.3%	86.7%
Loklok	26.7%	73.3%

Berdasarkan hasil klasifikasi kedua aplikasi cenderung menerima lebih banyak ulasan negatif, tetapi Loklok menunjukkan persepsi pengguna yang sedikit lebih positif dibandingkan Vidio.

Perbedaan ini dapat mencerminkan variasi kualitas layanan, pengalaman pengguna, atau permasalahan teknis yang berbeda pada kedua platform streaming tersebut.

4.6.2 Evaluasi Performa Model Klasifikasi

Evaluasi performa model dilakukan untuk mengetahui seberapa baik algoritma *Naïve Bayes* dalam mengklasifikasikan sentimen ulasan pengguna terhadap aplikasi Vidio dan Loklok. Evaluasi dilakukan menggunakan 4 metrik utama yaitu akurasi, precision, recall, F1-score. Tabel 4.3 menunjukkan hasil evaluasi kedua model.

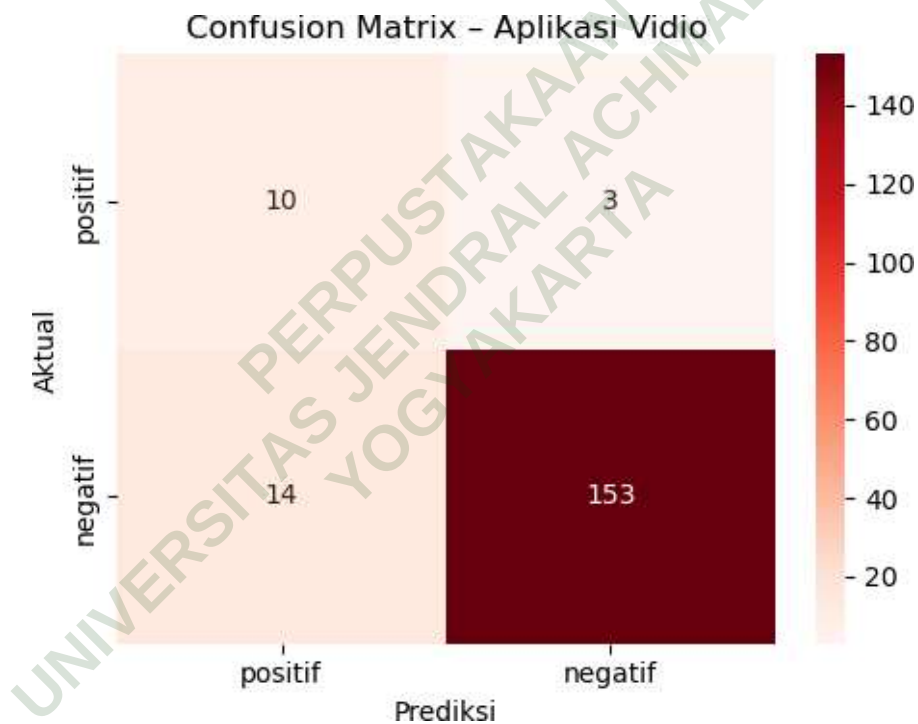
Tabel 4.3 Hasil Evaluasi Kinerja Model Klasifikasi Sentimen

Aplikasi	Akurasi	Precision	Recall	F1-Score
Vidio	0.91	0.42	0.77	0.54
Loklok	0.86	0.98	0.84	0.91

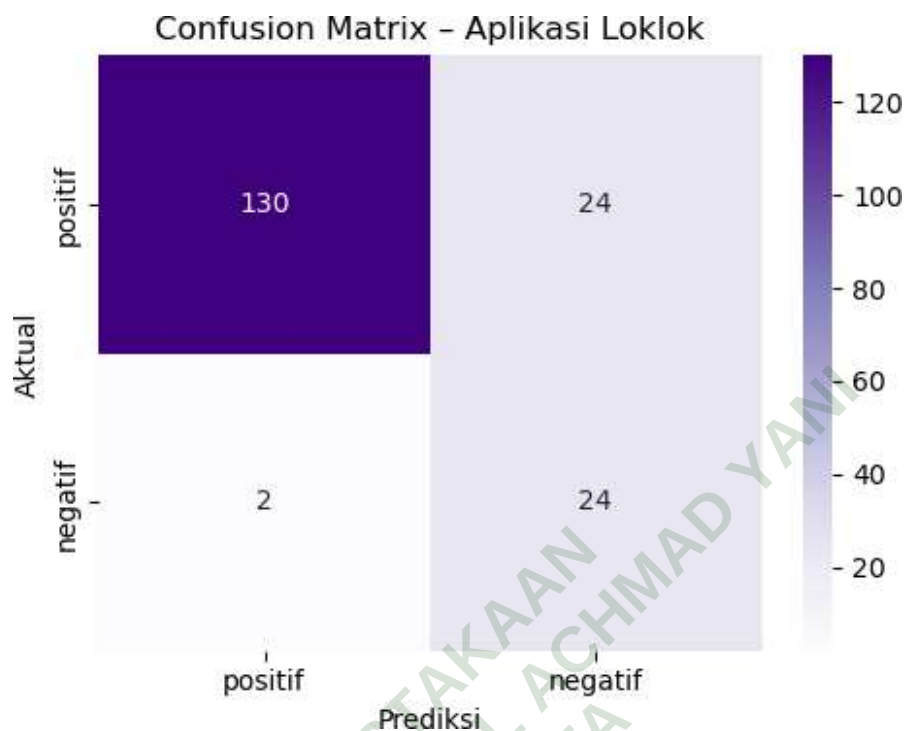
Hasil evaluasi menunjukkan bahwa model klasifikasi untuk aplikasi Loklok memiliki performa yang lebih baik dibandingkan dengan model untuk Vidio, terutama pada metrik *precision* dan *F1-score*. Nilai *precision* sebesar 0.98 pada Loklok mengindikasikan bahwa sebagian besar prediksi positif memang benar-benar positif, dan nilai *F1-score* sebesar 0.91 mencerminkan keseimbangan antara ketepatan dan cakupan prediksi. Sementara itu, model Vidio memang memiliki akurasi yang tinggi (0.91), namun nilai *precision* yang rendah (0.42) menunjukkan bahwa banyak ulasan yang diklasifikasikan sebagai positif secara keliru. Hal ini berdampak pada *F1-score* yang rendah (0.54), yang mengindikasikan adanya ketidakseimbangan performa model dalam mengidentifikasi dan mengklasifikasikan ulasan positif secara akurat. Berdasarkan hasil evaluasi, dapat disimpulkan bahwa tingkat keakuratan metode *Naïve Bayes* cukup bervariasi. Model bekerja dengan sangat baik pada aplikasi Loklok, ditunjukkan oleh nilai *precision* dan *F1-score* yang tinggi. Namun, pada aplikasi Vidio, meskipun akurasinya tinggi, nilai *precision* yang rendah menunjukkan bahwa model masih kurang tepat dalam mengklasifikasikan sentimen secara akurat. Hal ini menunjukkan bahwa efektivitas metode sangat dipengaruhi oleh karakteristik data yang digunakan.

4.6.3 Analisis Confusion Matriks

Confusion Matriks digunakan untuk mengevaluasi kinerja model klasifikasi dalam membedakan antara sentimen positif dan negatif. Matriks ini menampilkan jumlah prediksi yang benar dan salah berdasarkan label aktual dan prediksi model. Dalam penelitian ini, confusion matrix membantu mengidentifikasi sejauh mana model dapat mengenali sentimen dengan akurat, sekaligus menunjukkan adanya kesalahan prediksi, seperti false positive (ulasan negatif yang diklasifikasikan sebagai positif) dan false negative (ulasan positif yang diklasifikasikan sebagai negatif). Visualisasi confusion matrix untuk masing-masing aplikasi disajikan berikut:



Gambar 4.19 Confusion Matrix Aplikasi Vidio

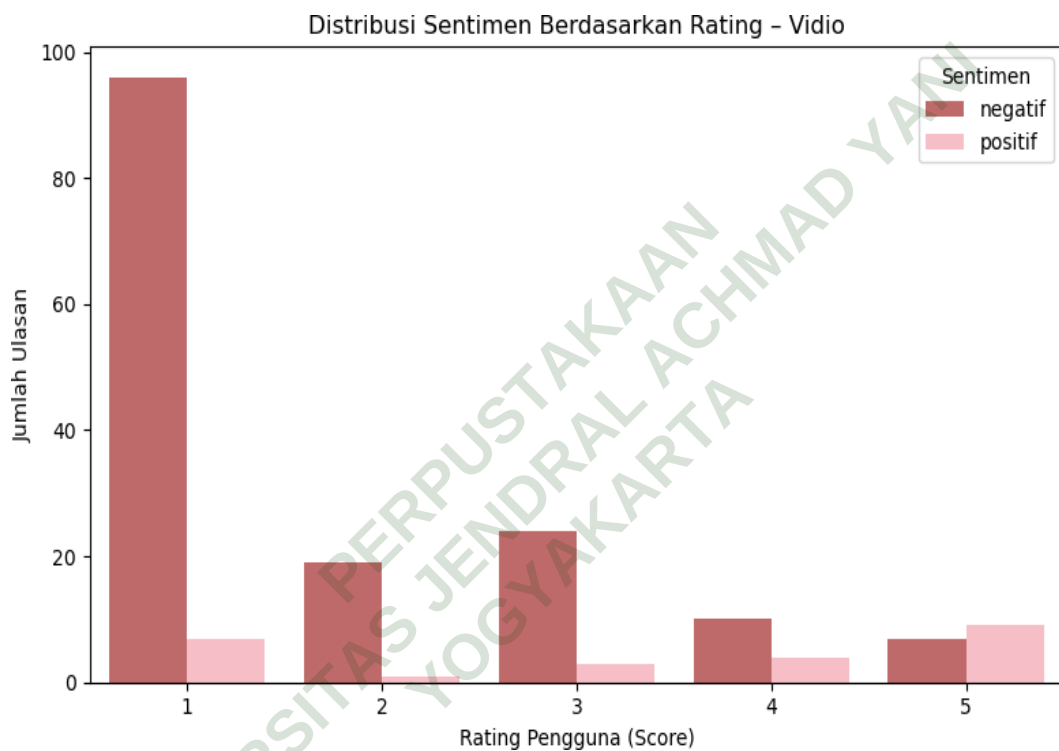


Gambar 4.20 Confusion Matrix Aplikasi Loklok

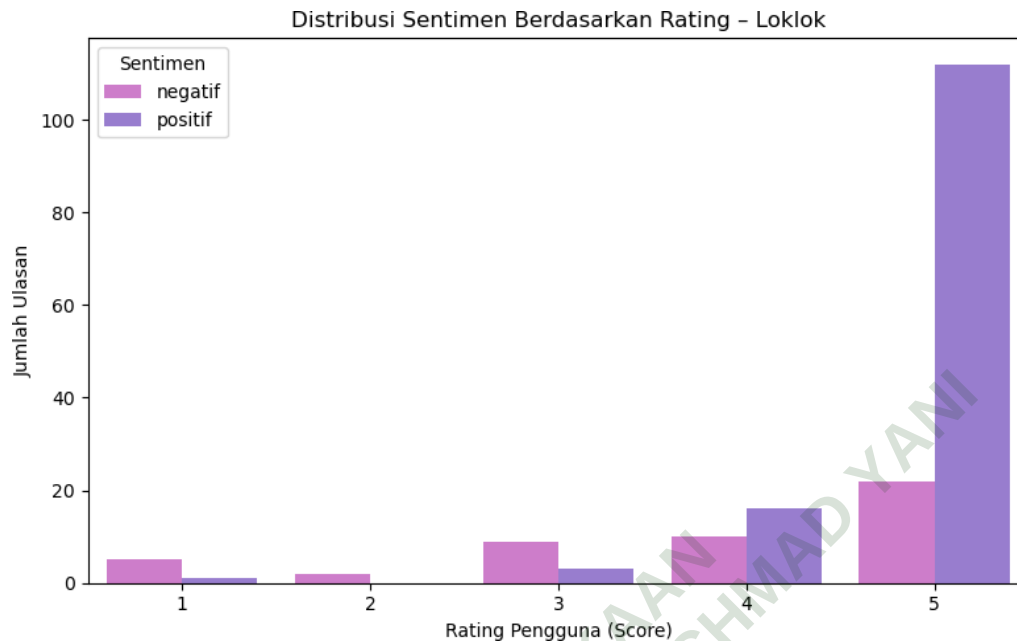
Pada aplikasi Vidio, meskipun model memiliki akurasi tinggi, ditemukan cukup banyak prediksi sentimen positif yang sebenarnya termasuk ulasan negatif. Hal ini tercermin dari nilai *precision* yang rendah. Model cenderung mengklasifikasikan banyak ulasan sebagai positif, meskipun sebagian di antaranya keliru. Sementara pada aplikasi Loklok, model menunjukkan distribusi klasifikasi yang lebih seimbang, dengan kesalahan prediksi yang lebih sedikit. Hal ini sejalan dengan nilai *precision* dan *F1-score* yang tinggi, yang menunjukkan bahwa model lebih akurat dalam membedakan sentimen. Dengan demikian, confusion matrix tidak hanya memberikan gambaran kuantitatif terhadap performa model, tetapi juga memperkuat temuan evaluasi sebelumnya mengenai kekuatan dan kelemahan masing-masing model klasifikasi.

4.6.4 Hubungan Rating dan Sentimen

Hubungan antara rating dan sentimen dianalisis untuk mengetahui apakah terdapat pola yang konsisten antara skor bintang yang diberikan pengguna (1–5) dengan isi sentimen dalam ulasan. Visualisasi dilakukan untuk masing-masing aplikasi, sehingga pola distribusi dapat terlihat secara komparatif.



Gambar 4.21 Distribusi Rating dan Sentimen Aplikasi Vidio

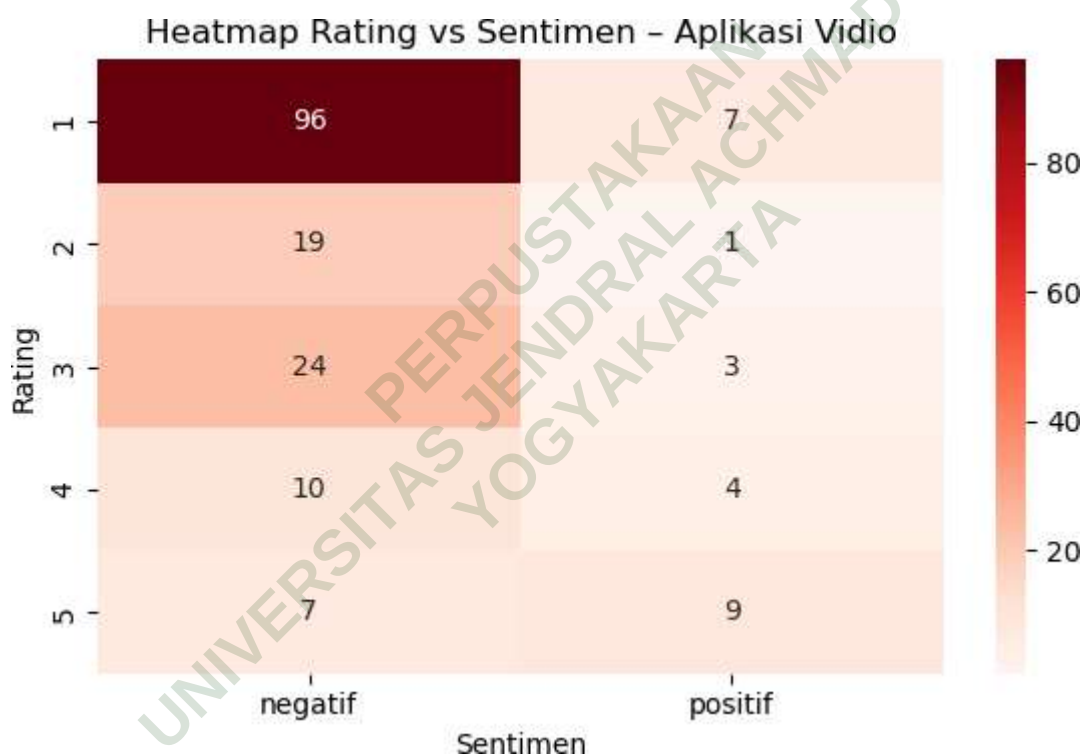


Gambar 4.22 Distribusi Rating dan Sentimen Aplikasi Loklok

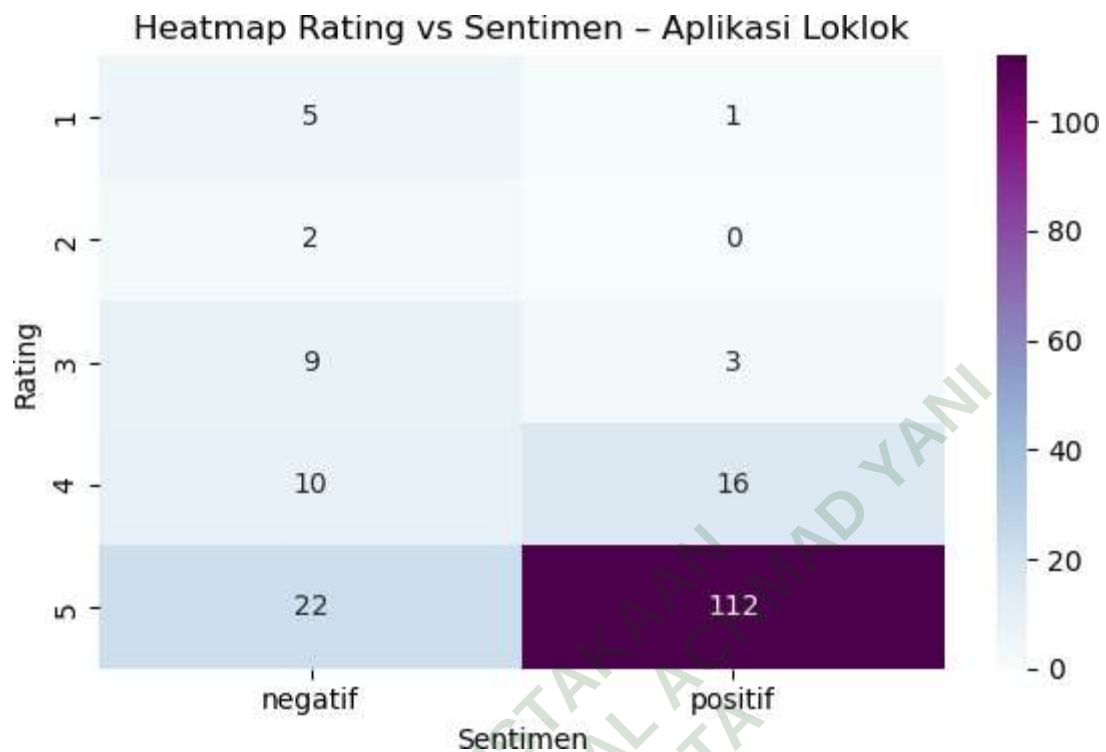
Pada aplikasi Vidio, sentimen negatif mendominasi hampir di semua tingkat rating, termasuk pada skor tinggi seperti 4 dan 5. Bahkan, jumlah ulasan negatif pada rating 1 sangat tinggi dibandingkan positif. Hal ini menunjukkan bahwa meskipun pengguna memberikan rating tinggi, isi ulasan masih banyak yang berisi keluhan, yang dapat mencerminkan adanya *mismatch* atau ketidaksesuaian antara penilaian kuantitatif dan opini verbal pengguna. Sementara itu, pada aplikasi Loklok, distribusi menunjukkan kecenderungan yang lebih sesuai antara rating dan sentimen. Ulasan dengan skor 4 dan 5 didominasi oleh sentimen positif, sedangkan skor 1 hingga 3 lebih banyak mengandung sentimen negatif. Hal ini menunjukkan adanya hubungan yang lebih konsisten antara rating numerik dan isi ulasan pengguna. Dengan demikian, aplikasi Loklok menunjukkan pola korelasi yang lebih kuat antara rating dan sentimen, sedangkan pada aplikasi Vidio, terdapat indikasi bahwa pengguna mungkin memberikan rating tinggi meskipun isi ulasannya mengandung sentimen negatif.

4.6.5 Analisis Mismatch antara Rating dan Sentimen

Analisis mismatch dilakukan untuk mengidentifikasi ketidaksesuaian antara rating tinggi (≥ 4) yang diberikan oleh pengguna dengan sentimen negatif hasil prediksi model. Kasus seperti ini penting untuk dianalisis karena dapat mengindikasikan adanya ulasan yang ambigu, mengandung dua sisi (positif dan negatif), atau ketidaksesuaian antara angka rating dan isi ulasan, yang bisa disebabkan oleh ulasan ambigu, pujian yang disertai kritik, atau perilaku pemberian rating yang tidak konsisten. Visualisasi dalam bentuk heatmap berikut memperlihatkan pola hubungan antara skor rating dan hasil klasifikasi sentimen:



Gambar 4.23 Heatmap Rating vs Sentimen Aplikasi Vidio



Gambar 4.24 Heatmap Rating vs Sentimen Aplikasi Loklok

Berdasarkan adanya mismatch atau ketidaksesuaian antara rating numerik yang diberikan pengguna dengan hasil klasifikasi sentimen model. Pada aplikasi Vidio, terdapat 17 ulasan dengan rating tinggi (4 dan 5) namun diklasifikasikan sebagai sentimen negatif oleh model. Sementara itu, pada aplikasi Loklok, jumlah mismatch tercatat lebih tinggi, yaitu 32 ulasan dengan rating tinggi yang juga diprediksi sebagai sentimen negatif. Hal ini menunjukkan bahwa meskipun pengguna memberikan rating tinggi, isi ulasan mereka mengandung konteks yang mencerminkan ketidakpuasan, kritik halus, atau keluhan tersirat yang tidak secara eksplisit tercermin dalam rating angka. Mismatch ini dapat terjadi karena berbagai faktor, seperti penggunaan sarkasme, kalimat bersifat campuran (*mixed sentiment*), atau ulasan yang mengandung keluhan tetapi tetap memberi rating tinggi karena alasan lain (misalnya loyalitas terhadap aplikasi). Penelitian ini menegaskan pentingnya tidak hanya mengandalkan rating numerik semata dalam menilai kepuasan

pengguna, melainkan juga mempertimbangkan analisis sentimen berbasis teks untuk memperoleh gambaran yang lebih menyeluruh. Tabel 4.4 menunjukkan contoh kasus *mismatch* dari hasil klasifikasi dalam penelitian ini.

Tabel 4.4 Ulasan Rating Tinggi dengan Prediksi Sentimen Negatif (Mismatch)

No	Aplikasi	Rating	Ulasan (Hasil Stem)	Prediksi Sentimen
1	Vidio	5	Iklan kurang banyak kalau bisa menit	Negatif
2	Vidio	4	bagus aplikasi ada banyak drama tapi sayang fitur tidak lengkap	Negatif
3	Loklok	4	asik drama selalu cepat baru iklan jangan terlalu banyak	Negatif
4	Loklok	5	film baru	Negatif

4.6.6 Perbandingann Model Klasifikasi Vidio dan Loklok

Perbandinagan aplikasi Vidio dan Loklok mencakup metrik evaluasi yaitu akurasi precision, recall, F1-score, distribusi sentimen hasil klasifikasi, serta jumlah mismatch antra rating dan prediksi sentimen. Tabel 4.5 menyajikan ringkasan hasil perbandingan.

Tabel 4.5 Perbandingan Model Klasifikasi Aplikasi Vidioo dan Loklok

Aspek Evaluasi	Vidio	Loklok
Akurasi	0.91	0.86
Precision	0.42	0.98
Recall	0.77	0.84
F1-Score	0.54	0.91
Sentimen Dominan	Negatif	Positif
Jumlah Mismatch	17	32

Hasil evaluasi menunjukkan bahwa meskipun model klasifikasi Vidio memiliki akurasi keseluruhan lebih tinggi (91%), namun precision-nya jauh lebih rendah dibandingkan Loklok. Hal ini menandakan bahwa model Loklok lebih konsisten dalam mengklasifikasikan ulasan positif secara akurat, sebagaimana tercermin dalam nilai F1-score yang tinggi (0.91). Selain itu, jumlah mismatch antara rating tinggi dengan prediksi negatif lebih sedikit pada Vidio (17) dibanding Loklok (32), namun hal ini juga dapat dipengaruhi oleh variasi gaya penulisan ulasan, kualitas data, dan persepsi pengguna yang berbeda terhadap masing-masing aplikasi. Secara umum, model klasifikasi Loklok menunjukkan performa yang lebih stabil dan seimbang dalam mendeteksi sentimen pengguna, sementara model Vidio cenderung menghasilkan banyak false positive.