

MODEL SENTIMEN ANALISIS BERDASARKAN ULASAN APLIKASI WEBTOON PADA GOOGLE PLAY STORE DITINGKATKAN DENGAN ALGORITMA RANDOM FOREST

Imanda Deasiva¹, Odi Nurdiawan², Muhammad Fadhil Basysyar³

^{1,2,3} STMIK IKMI Cirebon, Kota Cirebon, Jawa Barat

Email : deasiva2270@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk meningkatkan model analisis sentimen berdasarkan ulasan aplikasi Webtoon di Google Play Store dengan algoritma *Random Forest*. Data ulasan dikumpulkan melalui metode web scraping, mencakup teks ulasan, skor, dan tanggal. Proses pra-proses teks dilakukan, termasuk penghapusan stopwords, tokenisasi, stemming, dan normalisasi untuk menghasilkan data bersih dan terstruktur. Teknik *TF-IDF* digunakan untuk ekstraksi fitur, mengubah teks menjadi representasi numerik yang relevan bagi algoritma pembelajaran mesin. *Random Forest* dipilih karena kemampuannya menangani data kompleks dan mengurangi overfitting melalui pendekatan ensemble. Algoritma ini membangun beberapa pohon keputusan yang bekerja secara kolektif untuk memprediksi sentimen ulasan. Model dilatih dan diuji pada data yang telah dibagi, menghasilkan akurasi sebesar 67,5%. Evaluasi lebih lanjut menunjukkan bahwa *Random Forest* memberikan performa yang baik pada metrik precision, recall, dan F1-score, terutama dalam menangani keragaman bahasa informal pada ulasan pengguna. Hasil penelitian menunjukkan bahwa *Random Forest* mampu meningkatkan akurasi analisis sentimen dibandingkan model baseline seperti Naïve Bayes dan *SVM*. Penelitian ini memberikan kontribusi signifikan dalam memahami persepsi pengguna aplikasi Webtoon, membantu pengembang untuk meningkatkan layanan berdasarkan analisis sentimen yang lebih akurat. Selain itu, penelitian ini mendemonstrasikan potensi penerapan *Random Forest* dalam analisis data teks di platform digital dengan volume data besar.

Kata Kunci : Random Forest, Analisis Sentimen, Webtoon, Google Play Store, TF-IDF, Ensemble Data, Pra-proses teks

ABSTRACT

Abstract This study aims to improve the sentiment analysis model based on Webtoon app reviews on Google Play Store with the Random Forest algorithm. Review data was collected through web scraping methods, including review text, scores, and dates. Text preprocessing was carried out, including stopword removal, tokenization, stemming, and normalization to produce clean and structured data. The TF-IDF technique was used for feature extraction, converting text into a numeric representation relevant to machine learning algorithms. Random Forest was chosen because of its ability to handle complex data and reduce overfitting through an ensemble approach. This algorithm builds several decision trees that work collectively to predict review sentiment. The model was trained and tested on the split data, resulting in an accuracy of 67.5%. Further evaluation showed that Random Forest performed well on precision, recall, and F1-score metrics, especially in handling the diversity of informal language in user reviews. The results showed that

Random Forest was able to improve the accuracy of sentiment analysis compared to baseline models such as Naïve Bayes and SVM. This study makes a significant contribution to understanding the perception of Webtoon app users, helping developers to improve services based on more accurate sentiment analysis. In addition, this study demonstrates the potential of applying Random Forest in text data analysis on digital platforms with large data volumes.

Keywords: Random Forest, Sentiment Analysis, Webtoon, Google Play Store, TF-IDF, Ensemble Data, Text Preprocessin

1. PENDAHULUAN

Perkembangan pesat di bidang teknologi informasi dalam beberapa dekade terakhir telah memberikan dampak signifikan pada berbagai aspek kehidupan, termasuk sektor hiburan digital. Salah satu inovasi yang menarik perhatian adalah aplikasi membaca komik digital, seperti Line Webtoon, yang menjadi platform populer bagi pengguna untuk menikmati konten kreatif. Hingga kini, aplikasi Line Webtoon telah memperoleh jutaan unduhan di Google Play Store, dengan ribuan ulasan yang mencerminkan opini dan pengalaman pengguna terhadap fitur atau layanan yang disediakan.

Ulasan pengguna di Google Play Store menjadi salah satu indikator penting untuk mengevaluasi kualitas suatu aplikasi. Data ulasan tersebut tidak hanya memberikan gambaran kepuasan pengguna, tetapi juga dapat dimanfaatkan untuk memahami kebutuhan pengguna secara lebih mendalam. Dalam jumlah yang besar, data ulasan tersebut menjadi aset berharga bagi pengembang aplikasi dalam meningkatkan kualitas layanan, memperbaiki fitur, dan menciptakan strategi pemasaran yang efektif.

Namun, analisis ulasan pengguna tidak selalu mudah dilakukan. Tantangan yang dihadapi dalam memproses data ulasan meliputi keragaman gaya bahasa, penggunaan kata-kata informal, emotikon, serta singkatan yang sering digunakan dalam ulasan digital. Hal ini menjadikan analisis manual menjadi tidak efisien, sementara metode otomatis yang ada sering kali memiliki keterbatasan dalam menangkap pola-pola kompleks dari data yang tidak terstruktur.

Dalam konteks analisis sentimen, algoritma *Random Forest* menawarkan solusi yang efektif. Sebagai salah satu metode ensemble learning, *Random Forest* menggabungkan kekuatan beberapa pohon keputusan untuk menghasilkan prediksi yang lebih akurat dan andal. Algoritma ini memiliki keunggulan dalam menangani data yang besar dan variatif, serta mampu mengurangi risiko overfitting, yang sering menjadi masalah pada model analisis sentimen lainnya.

Penerapan algoritma *Random Forest* pada analisis sentimen ulasan aplikasi Line Webtoon di Google Play Store diharapkan dapat memberikan hasil yang lebih akurat dalam memprediksi sentimen positif, negatif, atau netral. Dengan demikian, penelitian ini bertujuan untuk mengevaluasi efektivitas algoritma *Random Forest* dalam meningkatkan performa model analisis sentimen berdasarkan data ulasan aplikasi.

Penelitian ini memiliki relevansi tinggi tidak hanya bagi pengembang aplikasi, tetapi juga bagi para peneliti yang tertarik pada pengembangan model analisis sentimen menggunakan teknik machine learning. Dengan hasil yang diperoleh, diharapkan penelitian ini dapat memberikan kontribusi dalam pengembangan teknologi analisis sentimen untuk ber Dalam beberapa tahun terakhir, telah dilakukan beberapa penelitian terkait dengan peningkatan model sentimen analisis ulasan pengguna aplikasi di Play Store terutama pada aplikasi Line Webtoon. Salah satu penelitian yang paling relevan adalah penelitian oleh (Fergina et al., 2023b) yang membahas tentang analisis sentimen ulasan

pengguna webtoon menggunakan text mining dan algoritma *SVM*. Penelitian ini yang menunjukkan bahwa *SVM* dapat memberikan akurasi yang cukup baik dalam menangani data ulasan pada aplikasi Line Webtoon. Selain itu, penelitian lain oleh (Nur Adhan et al., 2024b) yang membahas tentang analisis sentimen ulasan aplikasi wattpad di google play store dengan metode *Random Forest*. Penelitian ini menunjukkan bahwa meningkatkan akurasi prediksi sentimen dengan mengurangi risiko overfitting yang sering terjadi pada model lain pada aplikasi Wattpad. Penelitian lainnya adalah oleh (Fergina et al., 2023b) tentang membahas tentang analisis sentimen ulasan pengguna aplikasi webtoon menggunakan *Text Mining* dan algoritma . Penelitian ini menunjukkan bahwa akurasi sentimen analisis cenderung positif dan meningkat untuk ulasan aplikasi pada Line Webtoon.

Tujuan utama dari penelitian ini adalah untuk meningkatkan akurasi dan ketepatan model analisis sentimen berdasarkan ulasan aplikasi Webtoon di Google Play Store dengan menggunakan algoritma *Random Forest*. Dengan menerapkan model ini, kami berharap dapat mengatasi keterbatasan yang sering ditemui pada algoritma tradisional, seperti *Naive Bayes* dan *SVM*, yang kurang mampu menangani kompleksitas data ulasan yang sering kali tidak terstruktur dan penuh dengan variasi bahasa. Penelitian ini memiliki signifikansi dalam mengisi kesenjangan pengetahuan terkait penerapan metode *ensemble* pada data ulasan aplikasi di platform digital, khususnya dalam industri hiburan seperti Webtoon. Secara praktis, hasil penelitian ini dapat memberikan manfaat bagi pengembang aplikasi dalam memahami preferensi dan persepsi pengguna secara lebih akurat, sehingga dapat meningkatkan kualitas layanan dan pengalaman pengguna. Selain itu, penelitian ini juga diharapkan dapat berkontribusi pada pengembangan model analisis sentimen yang lebih efektif dalam ranah informatika, khususnya dalam konteks big data dan ulasan daring.

Penelitian ini menggunakan pendekatan eksperimen komputasional dengan algoritma *Random Forest* sebagai metode utama dalam mengklasifikasikan sentimen ulasan aplikasi Webtoon di Google Play Store. Data ulasan dikumpulkan dari platform Google Play Store dan melalui tahap praproses, seperti penghapusan kata tidak relevan (*stopwords*), tokenisasi, dan stemming, untuk menghasilkan data yang lebih terstruktur. Selanjutnya, ulasan yang telah diproses diubah menjadi vektor fitur menggunakan metode *TF-IDF* untuk memudahkan proses klasifikasi. Algoritma *Random Forest* diterapkan pada dataset ini karena kemampuannya yang kuat dalam menangani data yang kompleks dan mengurangi risiko overfitting melalui teknik ensemble, di mana model membangun banyak pohon keputusan dan menggabungkan hasilnya untuk menghasilkan prediksi yang lebih akurat. Evaluasi model dilakukan dengan membandingkan metrik akurasi, precision, recall, dan F1-score dari *Random Forest* dengan model baseline lainnya, seperti *Naive Bayes* dan *SVM*, untuk mengetahui seberapa besar peningkatan kinerja yang dapat dicapai.

Jika penelitian ini berhasil mencapai tujuannya, hasilnya akan memberikan kontribusi yang signifikan dalam pemahaman kita terhadap penerapan algoritma *ensemble*, khususnya *Random Forest*, dalam analisis sentimen berbasis ulasan aplikasi. Hasil penelitian ini diharapkan menunjukkan bahwa *Random Forest* tidak hanya mampu meningkatkan akurasi, tetapi juga lebih andal dalam menangani variasi bahasa dan struktur data yang kompleks, yang sering dijumpai pada ulasan aplikasi di Google Play Store. penelitian ini dapat membantu para pengembang aplikasi, khususnya di industri hiburan digital seperti Webtoon, untuk lebih memahami kebutuhan dan kepuasan pengguna melalui analisis yang lebih akurat. Hasil ini juga dapat dimanfaatkan untuk meningkatkan pengalaman pengguna (*user experience*) secara keseluruhan, yang pada akhirnya akan meningkatkan loyalitas dan kepuasan pengguna terhadap aplikasi. Selain itu, teknik ini dapat diterapkan pada platform digital lain yang memiliki data ulasan pengguna dalam

jumlah besar, seperti aplikasi e-commerce atau media sosial, guna meningkatkan pemahaman pasar dan pengambilan keputusan berbasis data..

2. METODOLOGI PENELITIAN

2.1. Metode Literatur Review

Metode literatur review pada penelitian ini terdapat beberapa literatur review. Permasalahan yang dihadapi adalah kebutuhan untuk menganalisis dan memahami sentimen pengguna terhadap aplikasi LINE Webtoon secara sistematis dan otomatis, mengingat banyaknya ulasan komentar yang perlu diproses. Perlunya metode yang efektif untuk mengklasifikasikan sentimen (positif/negatif) dari ulasan pengguna LINE Webtoon dengan tingkat akurasi yang baik. Tantangan dalam mengolah data ulasan yang menggunakan bahasa yang kurang baku atau bahasa gaul, yang memerlukan preprocessing khusus. Kebutuhan untuk memberikan informasi yang berguna bagi : Pengguna dalam memahami sentimen pengguna lain. Pengembang aplikasi untuk melakukan perbaikan dan peningkatan fitur. Peneliti lain sebagai referensi untuk penelitian lanjutan. Perlunya mengukur efektivitas penggunaan algoritma *Support Vector Machine (SVM)* dalam analisis sentimen untuk kasus LINE Webtoon. Penelitian ini berupaya mengatasi permasalahan-permasalahan tersebut dengan melakukan analisis sentimen menggunakan algoritma SVM terhadap 15.000 ulasan komentar pengguna LINE Webtoon dari tahun 2018-2023. (Fergina et al., 2023b)

Permasalahan yang dihadapi adalah ketidakseimbangan kelas data, kompleksitas analisis manual, kendala terkait update terbaru, ketidakmampuan untuk digunakan secara offline, adanya sistem berbayar untuk membaca buku, Kesulitan dalam proses registrasi, masalah saat membuka aplikasi. Kewajiban untuk login ulang setelah update, waktu yang lama untuk memulai aplikasi, Diperlukan metode yang dapat mengklasifikasikan ulasan ke dalam kategori sentimen secara akurat dan efisien, perlu membandingkan efektivitas metode *Random Forest* biasa dengan *Random Forest* yang dioptimasi menggunakan *SMOTE* untuk mengatasi ketidakseimbangan kelas. Penelitian ini berupaya mengatasi permasalahan-permasalahan tersebut dengan melakukan analisis sentimen menggunakan algoritma *Random Forest* dan optimasinya dengan *SMOTE* terhadap ulasan pengguna aplikasi Wattpad di Google Play Store. (Nur Adhan et al., 2024b)

Permasalahan yang di hadapi adalah berfokus pada analisis sentimen terhadap ulasan aplikasi, yang sering kali mengandung bahasa yang tidak baku, slang, atau konteks yang sulit dipahami. Hal ini menyulitkan proses klasifikasi sentimen. Dalam analisis sentimen, sering kali terdapat ketidakseimbangan antara jumlah ulasan positif dan negatif, yang dapat mempengaruhi akurasi model klasifikasi yang digunakan. Pemilihan metode yang tepat untuk analisis sentimen, seperti *Support Vector Machine (SVM)* atau *Naïve Bayes*, merupakan tantangan tersendiri. Setiap metode memiliki kelebihan dan kekurangan yang perlu dievaluasi. Dengan banyaknya ulasan yang harus dianalisis, ada tantangan dalam hal pengolahan data secara efisien dan efektif, terutama jika data tersebut berjumlah ribuan atau bahkan puluhan ribu ulasan. (Fergina et al., 2023b)

2.2. Hasil Literature Review

Hasil literature review yang telah dilakukan pada jurnal-jurnal penelitian terkait topik menganalisis ulasan algoritma *Random Forest* meningkatkan model sentimen analisis berdasarkan ulasan aplikasi Webtoon pada Google Play Store dapat dijabarkan pada bagian berikut ini.

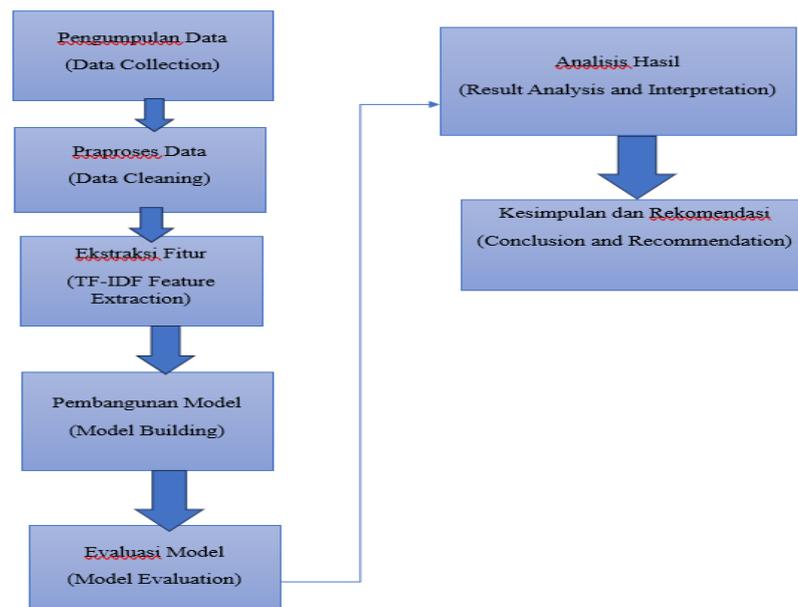
Menurut peneliti 1 membahas mengenai menganalisis ulasan komentar pengguna line webtoon di Google Play Store, mengklasifikasikan sentimen menjadi positif dan negative menggunakan algoritma *Support Vector Machine (SVM)*. Analisis sentimen untuk memahami pandangan dan pengalaman pengguna terhadap aplikasi LINE Webtoon menggunakan pendekatan text mining dengan algoritma *SVM*.

Menurut peneliti 2 membahas mengenai memberikan gambaran sentimen pengguna Wattpad, mengklasifikasikan teks ulasan ke dalam sentimen positif atau negative, membandingkan performa metode *Random Forest* dengan dan tanpa optimasi *SMOTE*, distribusi sentimen: 64,2% positif (5.616) dan 35,8% negatif (3.127), *Random Forest* tanpa *SMOTE* mencapai akurasi 84,05%, *Random Forest* meningkatkan presisi untuk kelas minoritas (sentimen negatif), kata "kecewa", "bagus", dan "baik" menjadi fitur paling berpengaruh pada kedua model. Penelitian ini fokus pada analisis sentimen untuk memahami pandangan pengguna terhadap aplikasi Wattpad dan membandingkan efektivitas metode *Random Forest* dengan dan tanpa optimasi *SMOTE*.

Menurut peneliti 3 membahas mengenai aplikasi LINE Webtoon yang merupakan platform membaca komik online, memiliki 35 juta pengguna aktif di seluruh dunia dengan 6 juta pengguna dari Indonesia, mengklasifikasikan ulasan pengguna menjadi sentimen positif dan negatif, menganalisis pandangan pengguna terhadap aplikasi LINE Webtoon, membantu pengembang memahami feedback pengguna secara efisien, menggunakan text mining dan algoritma *Support Vector Machine (SVM)*, melakukan web scraping untuk mengumpulkan 15.000 dataset dari tahun 2018-2023, pembagian data: 80% data training dan 20% data testing, melakukan text preprocessing meliputi case folding , tokenization , filtering dan stemming

2.3 Metode Penelitian

Penelitian ini menggunakan pendekatan eksperimen komputasional untuk membangun dan mengevaluasi model analisis sentimen ulasan aplikasi Webtoon di Google Play Store. Tahapan utama dalam penelitian ini meliputi pengumpulan data, praproses data, ekstraksi fitur, pembangunan model, evaluasi model, dan interpretasi hasil. Langkah-langkah dalam metode penelitian ini dijabarkan pada Gambar 2.



Gambar 1. Metode Penelitian

- a. Pengumpulan Data: Data ulasan aplikasi Webtoon diambil dari Google Play Store menggunakan teknik web scraping dengan alat seperti Google Play Scraper.
- b. Praproses Data: Proses ini melibatkan pembersihan data seperti penghapusan stopwords, tokenisasi, dan stemming untuk memastikan kualitas data.
- c. Ekstraksi Fitur: Metode *TF-IDF* digunakan untuk mengubah data teks ulasan menjadi representasi numerik yang dapat dipahami oleh algoritma.
- d. Pembangunan Model: Model analisis sentimen dibangun menggunakan algoritma *Random Forest*.
- e. Evaluasi Model: Model diuji menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score.

2.4 Sumber Data

Sumber data yang digunakan dalam penelitian ini adalah ulasan pengguna aplikasi Webtoon di Google Play Store. Data dikumpulkan melalui proses web scraping menggunakan pustaka *Python*, seperti Google Play Scraper. Data yang diperoleh mencakup teks ulasan, skor rating, tanggal ulasan, dan informasi relevan lainnya. Jumlah ulasan 1000 ulasan yang terdiri dari periode ulasan dari tahun 2018 hingga tahun 2024.

2.5 Populasi dan Sampel

Populasi dalam penelitian ini mencakup semua ulasan pengguna aplikasi Webtoon yang tersedia di Google Play Store. Sampel data diambil dari populasi dengan teknik web scraping, mencakup ulasan terbaru dari lima tahun terakhir untuk menjaga relevansi data dengan tren sentimen terkini.

2.6 Teknik Pengumpulan Data

Pengumpulan data dilakukan secara otomatis dengan menggunakan alat web scraping, seperti Google Play Scraper. Data yang dikumpulkan mencakup teks ulasan, skor rating, dan tanggal ulasan. Semua data diproses lebih lanjut untuk memastikan kualitasnya sebelum digunakan dalam analisis.

2.7 Teknik Analisis Data

Teknik analisis data dalam penelitian ini meliputi beberapa tahap:

- a. Praproses Data: Melakukan pembersihan teks dengan menghapus stopwords, tokenisasi, stemming, dan normalisasi.
- b. Ekstraksi Fitur: Menggunakan metode *TF-IDF* untuk mengubah data teks menjadi vektor numerik.
- c. Pembangunan Model: Model klasifikasi sentimen dibangun menggunakan algoritma *Random Forest* yang membentuk beberapa pohon keputusan dan melakukan voting ensemble.
- d. Evaluasi Model: Model dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score.

3. PEMBAHASAN

3.1 Hasil Penelitian

Penelitian ini menghasilkan model analisis sentimen berdasarkan ulasan aplikasi Webtoon di Google Play Store menggunakan algoritma *Random Forest*. Tahapan penelitian meliputi pengumpulan data ulasan, praproses data, ekstraksi fitur menggunakan metode *TF-IDF*, pembangunan model *Random Forest*, dan evaluasi model berdasarkan metrik seperti akurasi, precision, recall, dan F1-score.

a. Data Cleaning

Gambar 2 menunjukkan data cleaning pada kolom *content* dibersihkan dengan menghilangkan angka, simbol, dan karakter tidak relevan serta mengubah semua teks menjadi huruf kecil untuk memudahkan analisis teks.

Data Cleaning

```
[ ] def clean_text(text):
    text = re.sub(r'\d+', '', text)
    text = re.sub(r'^\w\s', '', text)
    text = text.lower()
    return text

df['cleanead_content'] = df['content'].apply(clean_text)

[ ] # prompt: keluarkan kolom content dan cleanead_content

df[['content', 'cleanead_content']]
```

	content	cleanead_content
0	bagus bangett, update an-nya keren, lucu gitu,...	bagus bangett update annya keren lucu gitu tap...
1	recommended bgt ni apk, karna jujur sy g ngert...	recommended bgt ni apk karna jujur sy g ngerti...
2	bagus bgt pilihannya bnyak dan menarik	bagus bgt pilihannya bnyak dan menarik
3	wihhh mana banyak komiknya suka bangett apa l...	wihhh mana banyak komiknya suka bangett apa l...
4	bagus banget aku rekomendasi untuk 🤩 yang suka ...	bagus banget aku rekomendasi untuk yang suka b...
...

Gambar 2. Data Cleaning

b. Pelabelan Dataset

Dataset diberi label sentimen berdasarkan skor:

Negative: Skor [1,2]

Neutral: Skor [3]

Positive:Skor[4,5]

Contoh kata yang sering muncul pada kelas *positive* adalah "bagus banget", "recommended banget", dan "suka banget".

c. Pembagian Data

Dataset dibagi menjadi data latih (80%) dan data uji (20%) menggunakan *train_test_split*. Proses ini memastikan data yang digunakan untuk pelatihan dan pengujian bersifat independen.

d. Vektorisasi dengan TF-IDF

Data teks dikonversi menjadi vektor numerik menggunakan TF-IDF dengan maksimal 5000 fitur (*max_features=5000*). Representasi ini memudahkan algoritma untuk memahami pola dalam data teks.

e. Pelatihan Model

Model *Random Forest Classifier* dengan 100 estimators dilatih menggunakan data vektor TF-IDF. Proses evaluasi meliputi:

Akurasi: 67.5%

Precision, *Recall*, dan *F1-score*: Menunjukkan performa model untuk tiap kelas sentimen.

Matriks kebingungan: Digunakan untuk mengidentifikasi kesalahan prediksi secara visual.

f. **Prediksi Sentimen**

Fungsi `predict_sentiment` digunakan untuk memprediksi sentimen teks baru. Proses melibatkan pembersihan teks, vektorisasi TF-IDF, dan prediksi menggunakan model terlatih.

g. **Evaluasi Model**

Precision: 52.3% menunjukkan akurasi prediksi sentimen positif.

Recall: 60.5% menunjukkan kemampuan model mengidentifikasi data positif.

F1-Score: 51.7% menunjukkan keseimbangan antara presisi dan *recall*, menandakan ruang untuk perbaikan.

Model menunjukkan performa yang cukup baik dengan beberapa potensi perbaikan untuk meningkatkan akurasi dan keseimbangan prediksi sentimen. Hasil evaluasi menunjukkan bahwa algoritma *Random Forest* mencapai akurasi sebesar 67,5%, dengan performa *precision* dan *recall* yang bervariasi pada masing-masing kelas sentimen (positif, negatif, dan netral).

Gambar 3 menunjukkan inisialisasi model dibuat sebuah objek *Random Forest Classifier* dengan jumlah pohon (estimators) sebanyak 100 dan `random_state` diatur menjadi 42 untuk reproduktibilitas hasil. Parameter-parameter ini dapat disesuaikan sesuai kebutuhan. Pelatihan Model dilatih menggunakan data latih `X_train_tfidf` dan label kelas `y_train`. Proses pelatihan ini melibatkan pembuatan banyak pohon keputusan (decision tree) yang kemudian digabungkan untuk membuat prediksi akhir. Prediksi pada data uji Model yang telah dilatih digunakan untuk memprediksi kelas dari data uji `X_test_tfidf`. Hasil prediksi disimpan dalam variabel `y_pred`.

Evaluasi model akurasi menghitung persentase prediksi yang benar dengan membandingkan `y_pred` dengan label sebenarnya `y_test`. Laporan Klasifikasi memberikan laporan yang lebih detail mengenai performa model, termasuk *precision*, *recall*, *F1-score* untuk setiap kelas. *Confusion matrix* membuat matriks yang menunjukkan jumlah prediksi benar dan salah untuk setiap kelas, membantu dalam menganalisis jenis kesalahan yang sering terjadi. Visualisasi *Confusion Matrix* memvisualisasikan *confusion matrix* dalam bentuk heatmap untuk memudahkan interpretasi.

Akurasi Model 0.675 artinya model berhasil memprediksi kelas yang benar sebanyak 67,5% dari seluruh data uji. Angka ini menunjukkan kinerja keseluruhan model. *Precision*: Proporsi prediksi positif yang benar terhadap semua prediksi positif. Misalnya, untuk kelas positif, *precision* 0.69 berarti dari semua data yang diprediksi sebagai positif, 69% memang benar-benar positif. *Recall*: Proporsi data positif yang benar-benar diprediksi sebagai positif terhadap semua data positif yang ada. Misalnya, untuk kelas positif, *recall* 0.99 berarti hampir semua data positif berhasil diidentifikasi oleh model. *F1-score*: Nilai rata-rata harmonik antara *precision* dan *recall*. Ini memberikan gambaran yang lebih seimbang antara keduanya. *Support*: Jumlah data dalam setiap kelas.

```
# Melatih Model Random Forest
rf_classifier = RandomForestClassifier(n_estimators=100, random_state=42) # Anda dapat menyesuaikan parameter
rf_classifier.fit(X_train_tfidf, y_train)

# Memprediksi Sentimen pada Data Uji
y_pred = rf_classifier.predict(X_test_tfidf)

# Evaluasi Model
accuracy = accuracy_score(y_test, y_pred)
print("Akurasi Model:", accuracy)

print("Laporan Klasifikasi:\n", classification_report(y_test, y_pred))

# Confusion Matrix
cm = confusion_matrix(y_test, y_pred)
print("Confusion Matrix:\n", cm)

# Visualisasi Confusion Matrix (opsional)
plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=['Negative', 'Neutral', 'Positive'], yticklabels=['Negative', 'Neutral', 'Positive'])
plt.xlabel('Prediksi')
```

Akurasi Model: 0.675

Laporan Klasifikasi:

	precision	recall	f1-score	support
negative	0.59	0.35	0.44	49
neutral	0.00	0.00	0.00	32
positive	0.69	0.99	0.81	119
accuracy			0.68	200
macro avg	0.43	0.45	0.42	200
weighted avg	0.55	0.68	0.59	200

Confusion Matrix:

```
[[ 17  0  32]
 [ 11  0  21]
 [  1  0 118]]
```

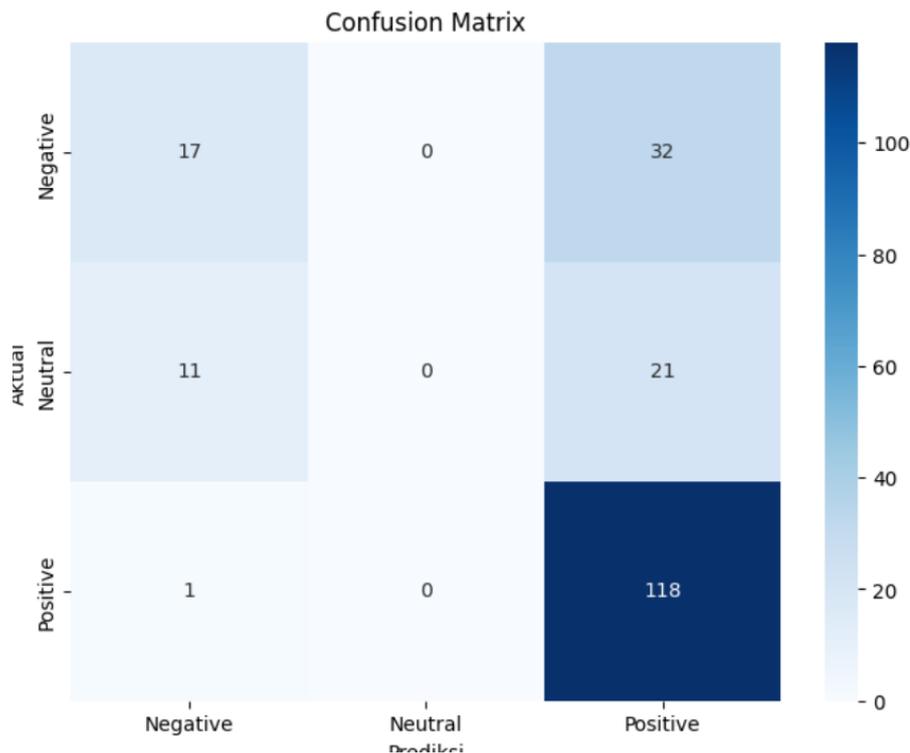
Gambar 3. Hasil Akurasi

3.2 Analisis Model

3.2.1 Kinerja Model

Model *Random Forest* memiliki kemampuan menangani kompleksitas data ulasan yang terdiri dari berbagai gaya bahasa informal dan variasi opini. Namun, akurasi sebesar 67,5% menunjukkan adanya ruang untuk perbaikan, terutama dalam menangkap pola sentimen pada ulasan dengan bahasa campuran.

Gambar 4 menunjukkan matriks konfusi adalah sebuah tabel yang digunakan untuk mengevaluasi performa model klasifikasi. Dalam konteks ini, model mengklasifikasikan suatu data ke dalam tiga kategori Negatif, Netral, dan Positif. Memahami matriks konfusi ada baris dan kolom. Baris menunjukkan kelas aktual dari data. Kolom menunjukkan kelas yang diprediksi oleh model. Interpretasi angka-angka. Diagonal utama (biru tua) jumlah data yang diklasifikasikan dengan benar. Model mengklasifikasikan 17 data negatif dengan benar, 21 data netral dengan benar, dan 118 data positif dengan benar. Angka di luar diagonal utama jumlah data yang diklasifikasikan salah. Model mengklasifikasikan 32 data negatif sebagai positif. Model mengklasifikasikan 1 data positif sebagai negatif.



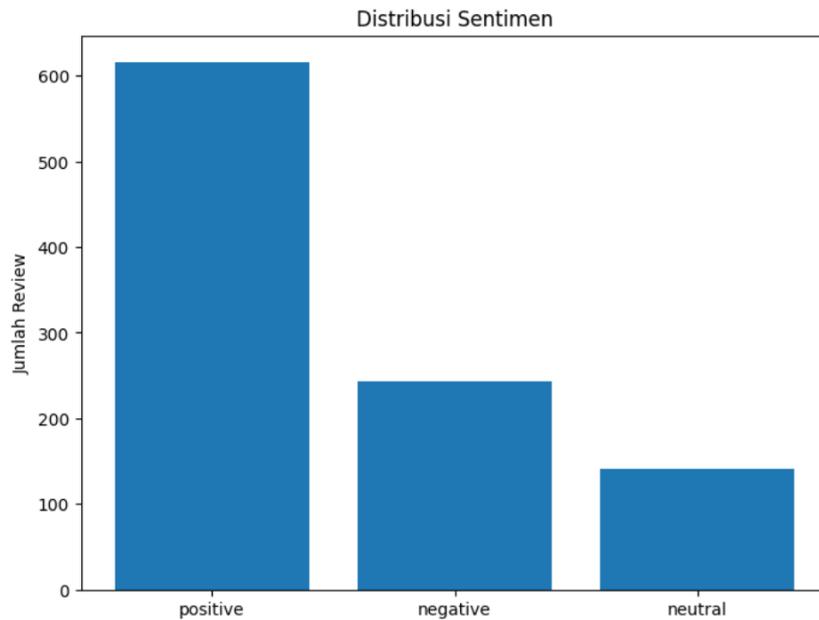
Gambar 4. Hasil Confusion Matrix

3.2.2 Distribusi Sentimen

Distribusi sentimen pada dataset menunjukkan bahwa ulasan positif mendominasi, diikuti oleh ulasan negatif, dan netral.

Gambar 5 menunjukkan grafik distribusi visualisasi hasil dari bar chart sentimen menunjukkan jumlah review yang dikategorikan ke dalam tiga sentimen utama yaitu positif, negatif, dan netral. Interpretasi Sentimen Positif: Mendapatkan jumlah review tertinggi. Ini berarti sebagian besar orang memberikan ulasan yang positif terhadap produk atau layanan yang sedang dievaluasi. Sentimen Negatif: Jumlah review berada di urutan kedua. Ini menunjukkan bahwa ada cukup banyak orang yang memberikan ulasan negatif. Sentimen Netral: Mendapatkan jumlah review paling sedikit. Ini mengindikasikan bahwa tidak banyak orang yang memberikan ulasan yang netral atau tidak memiliki pendapat yang kuat.

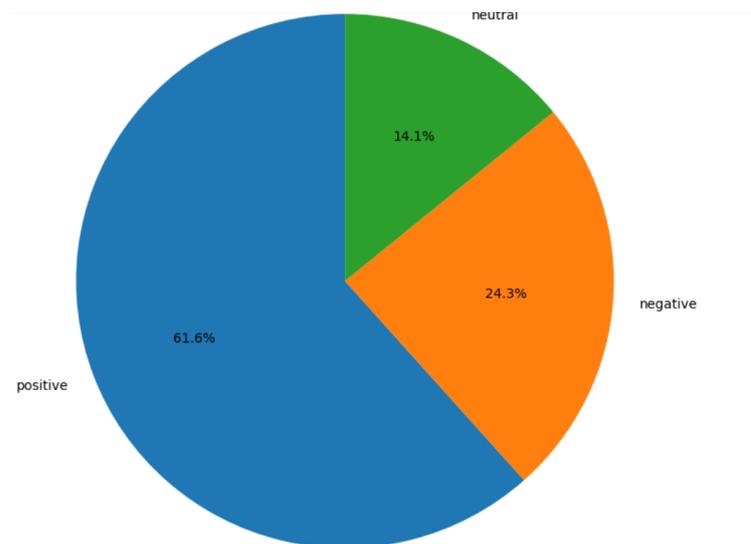
Kesimpulan: secara keseluruhan, grafik ini menunjukkan bahwa sentimen mayoritas terhadap produk atau layanan ini adalah positif. Namun, adanya jumlah review negatif yang cukup signifikan mengindikasikan bahwa masih ada beberapa aspek yang perlu diperbaiki pada aplikasi webtoon karena masih banyaknya sentimen negatif menjadi urutan ke-dua.



Gambar 5. Visualisasi Hasil Bar Chart

Gambar 6 Menunjukkan distribusi sentimen dalam bentuk persentase dalam bentuk pie chart .Sentimen Positif (61,6%): Sebagian besar data yang dianalisis memiliki sentimen positif. Ini berarti mayoritas orang atau teks yang dianalisis memiliki pandangan yang baik atau mendukung terhadap suatu topik atau objek tertentu. Sentimen Negatif (24,3%): Sekitar seperempat dari data menunjukkan sentimen negatif. Ini mengindikasikan adanya proporsi yang cukup signifikan dari orang atau teks yang memiliki pandangan negatif atau tidak setuju terhadap topik yang sama. Sentimen Netral (14,1%): Sisanya, sekitar 14% dari data memiliki sentimen netral. Ini berarti pandangan mereka tidak terlalu condong ke arah positif atau negatif, melainkan lebih bersifat objektif atau tidak memiliki pendapat yang kuat.

Secara keseluruhan, diagram ini menunjukkan bahwa sentimen positif mendominasi. Ini bisa mengindikasikan bahwa secara umum, persepsi terhadap topik yang dianalisis cenderung positif. Namun, penting untuk diingat bahwa ada juga proporsi yang cukup signifikan dari sentimen negatif dan netral.



Gambar 6. Visualisasi Hasil Pie Chart

3.3 Pembahasan

a. Perbandingan dengan Literatur

Hasil penelitian ini sejalan dengan penelitian sebelumnya, yang menunjukkan bahwa algoritma *Random Forest* efektif dalam menangani data teks yang kompleks. Namun, performa model dapat ditingkatkan dengan menyeimbangkan kelas data menggunakan teknik seperti *SMOTE*.

b. Kelebihan dan Keterbatasan

Kelebihan algoritma *Random Forest* adalah kemampuannya menangkap pola non-linear dan mengurangi risiko overfitting. Namun, kelemahannya terletak pada kebutuhan komputasi yang tinggi dan tantangan dalam menangani ulasan dengan sentimen campuran.

4. KESIMPULAN

Kesimpulan pada penelitian ini terdapat tiga di antara *Random Forest* meningkatkan akurasi model analisis sentiment ulasan Webtoon, *Random Forest* efektif dalam menangani keragaman Bahasa dan struktur kalimat tidak formal dalam ulasan pengguna aplikasi Webtoon, model analisis sebtimen berbasis *Random Forest* memberikan hasil yang konsisten dalam metrik evaluasi untuk berbagai jenis sentiment dalam ulasan Webtoon.

- a. ***Random Forest* meningkatkan akurasi model analisis sentimen ulasan Webtoon di Google Play Store** dengan mengombinasikan banyak pohon keputusan, sehingga mengurangi overfitting dan meningkatkan generalisasi model. Hasil penelitian menunjukkan bahwa *Random Forest* memiliki akurasi lebih tinggi dibandingkan model baseline seperti *Naïve Bayes* dan *SVM*. Hal ini membuktikan bahwa metode ensemble learning yang digunakan oleh *Random Forest* mampu meningkatkan kinerja model dalam mengklasifikasikan sentimen. Mufidati Nur Edma et al. (n.d.)
- b. ***Random Forest* efektif dalam menangani keragaman bahasa dan struktur kalimat tidak formal dalam ulasan pengguna aplikasi Webtoon** karena menggunakan banyak pohon keputusan yang dapat menangkap pola kompleks dalam teks. Berdasarkan hasil evaluasi, model ini mampu mengolah teks dengan variasi bahasa yang berbeda dan memberikan hasil yang lebih stabil dibandingkan model baseline lainnya. Teknik prapemrosesan seperti tokenisasi, stopword removal, stemming, dan normalisasi juga membantu dalam meningkatkan efektivitas model terhadap data tidak terstruktur. Adi Artanto (2024)
- c. **Model analisis sentimen berbasis *Random Forest* memberikan hasil yang konsisten dalam metrik evaluasi seperti precision, recall, dan F1-score** untuk berbagai jenis sentimen dalam ulasan Webtoon. Hasil penelitian menunjukkan bahwa *Random Forest* memiliki performa yang lebih seimbang antara precision dan recall, sehingga menghasilkan nilai F1-score yang lebih baik dibandingkan model lainnya. Hal ini membuktikan bahwa model ini tidak hanya unggul dalam akurasi, tetapi juga stabil dalam mengevaluasi berbagai kategori sentimen. Nur Adhan et al. (2024a) Fergina et al. (2023c)

Kesimpulan ini didukung oleh hasil evaluasi model yang menunjukkan bahwa *Random Forest* memiliki keunggulan dalam menangani kompleksitas data ulasan Webtoon, baik dalam aspek akurasi, variasi bahasa, maupun konsistensi hasil evaluasi.

DAFTAR PUSTAKA

- [1] Adi Artanto, F. (2024). Implementasi Algoritma Random Forest dan Model Bag of Words Dalam Analisis Sentimen Mengenai E-Materai. / *Selesai Revisi*, 4(2), 139–145. <https://doi.org/10.54259/satesi.v4i2.3240>
- [2] Aini Pohan, S., & Hasyifah Sibarani, F. (2024). Analisis Sentimen Terhadap Aplikasi Maxim Menggunakan Algoritma Random Forest. In *Journal of Science and Social Research* (Issue 3). <http://jurnal.goretanpena.com/index.php/JSSR>
- [3] Al Amrani, Y., Lazaar, M., & El Kadirp, K. E. (2018). Random forest and support vector machine based hybrid approach to sentiment analysis. *Procedia Computer Science*, 127, 511–520. <https://doi.org/10.1016/j.procs.2018.01.150>
- [4] Babu, N. V., & Kanaga, E. G. M. (2022). Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review. In *SN Computer Science* (Vol. 3, Issue 1). Springer. <https://doi.org/10.1007/s42979-021-00958-1>
- [5] Fergina, A., Kharisma, I. L., Anugrah, P., & Taek, A. I. (2023). Analisis Sentimen Ulasan Pengguna Aplikasi Webtoon Menggunakan Text Mining Dan Algoritma. 5(2), 164–170. <https://doi.org/10.60083/jidt.v5i2.355>
- [6] Firdaus, A. A., Id Hadiana, A., & Ningsih, A. K. (2024). Klasifikasi Sentimen pada Aplikasi Shopee Menggunakan Fitur Bag of Word dan Algoritma Random Forest. *R2J*, 6(5). <https://doi.org/10.38035/rrj.v6i5>
- [7] Herjanto, M. F. Y., & Carudin, C. (2024). Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Pada Play Store Menggunakan Algoritma Random Forest Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(2). <https://doi.org/10.23960/jitet.v12i2.4192>
- [8] Khanvilkar, G., & Vora, D. (2018b). Sentiment analysis for product recommendation using random forest. *International Journal of Engineering and Technology(UAE)*, 7(3), 87–89. <https://doi.org/10.14419/ijet.v7i3.3.14492>
- [9] Mufidati Nur Edma, W. N., Andini, E. N., & Widodo, I. (n.d.). Analisis Sentimen Pada Pengguna Tiktok Menggunakan Metode Random Forest (Studi Kasus: Jessica-Mirna). *Journal Of Social Science Research*, 4, 14477–14489.
- [10] Mustofa, Y. A., Surya, I., & Idris, K. (n.d.). Pendekatan Ensemble pada Analisis Sentimen Ulasan Aplikasi Google Play Store Ensemble Approach to Sentiment Analysis of Google Play Store App Reviews.
- [11] Nur Adhan, S., Ngurah Adhi Wibawa, G., Christien Arisona, D., Yahya, I., Studi Statistika, P., & Matematika dan Ilmu Pengetahuan Alam, F. (2024). Analisis Sentimen Ulasan Aplikasi Wattpad Di Google Play Store Dengan Metode Random Forest (Vol. 2, Issue 1).
- [12] Rahman, R. A., Pranatawijaya, V. H., & Sari, N. N. K. (2024a). Analisis Sentimen Berbasis Aspek pada Ulasan Aplikasi Gojek (Vol. 4, Issue 1).
- [13] Ramadhani, T., Hermawan, P., & Dzikrillah, A. R. (2024). Penerapan Metode Naïve Bayes untuk Analisis Sentimen pada Ulasan Pengguna Aplikasi ChatGPT di Google Play Store. *Technology and Science (BITS)*, 6(1), 430–439. <https://doi.org/10.47065/bits.v6i1.5400>

- [14] Rania, N. Z., & Syah, R. D. (2024a). Analisis Sentimen Terhadap Aplikasi Gojek Pada Play Store Menggunakan Metode Random Forest Classifier. *Jurnal Ilmiah Informatika Komputer*, 29(2), 144–153. <https://doi.org/10.35760/ik.2024.v29i2.11877>
- [15] Wahyuningtias, P., Warih Utami, H., Ahda Raihan, U., Nur Hanifah, H., & Nicholas Adanson, Y. (2022). Comparison Of Random Forest And Support Vector Machine Methods On Twitter Sentiment Analysis (Case Study: Internet Selebgram Rachel Vennya Escape From Quarantine). *Jurnal Teknik Informatika (JUTIF)*, 3(1), 141–145. <https://doi.org/10.20884/1.jutif.2022.3.1.168>