

PENERAPAN METODE RANDOM FOREST DALAM MENGANALISIS SENTIMEN PENGGUNA APLIKASI CAPCUT DI GOOGLE PLAY STORE

Ayu Sagita¹, Ahmad Faqih², Gifthera Dwilestari³, Bambang Siswoyo⁴, Denni Pratama⁵

^{1,2,4} Program Studi Teknik Informatika (S1), STMIK IKMI Cirebon

³ Program Studi Sistem Informasi (S1), STMIK IKMI Cirebon

⁵ Program Studi Komputerisasi Akuntansi (D3), STMIK IKMI Cirebon

Jl. Perjuangan No10B-Majasem, Kesambi Karya Mulya, Kota Cirebon, Jawa Barat Indonesia

ayusagita706@gmail.com

ABSTRAK

Pengguna aplikasi CapCut semakin meluas di kalangan pengguna smartphone. Namun, tanggapan pengguna terhadap aplikasi ini sangat bervariasi. Selain itu, perkembangan terus-menerus dalam menambah fitur dan kemampuan pengeditan telah membawa kompleksitas penggunaan aplikasi ini semakin meningkat. Penelitian ini bertujuan untuk menganalisis sentimen pengguna CapCut melalui ulasan yang terdapat di Google Play Store menggunakan metode Random Forest. Metode ini dipilih untuk mengidentifikasi dan mengelompokkan tanggapan pengguna ke dalam kategori positif dan negatif secara efisien. Dataset yang digunakan dalam penelitian ini mencakup 5000 ulasan, mencerminkan keragaman pendapat dari pengguna yang berpartisipasi aktif. Tahapan preprocessing data melibatkan proses cleaning, case folding, tokenisasi, stopword removal, dan lemmatisasi untuk memastikan kualitas data yang baik sebelum dilakukan analisis sentimen. Selanjutnya, pembobotan kata dilakukan dengan metode *TF-IDF* untuk memberikan nilai bobot pada kata-kata yang mempengaruhi sentimen pengguna. Hasil penelitian menunjukkan bahwa metode Random Forest memberikan tingkat akurasi yang tinggi dalam menganalisis sentimen pengguna CapCut, dengan akurasi sebesar 86%, presisi 89%, recall 81%, dan f1-score 85%. Penelitian ini memberikan pemahaman lebih lanjut terkait tanggapan pengguna terhadap aplikasi CapCut, serta menegaskan keberhasilan metode Random Forest dalam menangani analisis sentimen pada dataset ulasan pengguna di Google Play Store.

Kata kunci: Analisis Sentimen, Aplikasi CapCut, Google Play Store, Metode Random Forest

1. PENDAHULUAN

Berdasarkan situs *Media Indonesia*, CapCut menduduki peringkat pertama sebagai aplikasi pengedit video yang paling populer. Aplikasi ini memiliki beragam kelebihan, termasuk fitur transisi animasi dan efek berbasis kecerdasan buatan. Dalam peringkat 10 besar aplikasi yang paling banyak diunduh berdasarkan situs *CNN Indonesia*, CapCut menempati posisi ke-6 dengan total unduhan mencapai 25 juta, terdiri dari 14 juta di Play Store dan 11 juta di App Store. Aplikasi ini menonjolkan kemudahan antarmuka dan penggunaannya, yang memiliki peran penting dalam menarik minat pengguna [1].

Namun, seiring berkembangnya fitur dan kemampuan pengeditan video, kompleksitas penggunaan CapCut semakin meningkat. Banyak pengguna mengalami kesulitan dalam memahami aspek-aspek lanjutan, seperti efek khusus, transisi yang rumit, dan penerapan filter tingkat tinggi. Terlebih lagi, berbagai pilihan yang tersedia dalam aplikasi ini memberikan kontrol yang luas, tetapi dapat membingungkan bagi mereka yang memiliki keterbatasan pengalaman dalam pengeditan video. Analisis sentimen dapat menjadi langkah penting untuk memahami persepsi dan kebutuhan pengguna secara detail, memberikan informasi berharga kepada individu melalui pemrosesan data yang tidak terstruktur dalam sebuah dataset [2]. Analisis sentimen terhadap ulasan dapat dijalankan dengan

memanfaatkan metode atau algoritma analisis sentimen [3].

Penelitian mengenai analisis sentimen terhadap aplikasi pengedit video sudah pernah dilakukan pada paper [4], dengan fokus pada penggunaan TikTok sebagai platform media sosial yang populer di Indonesia. Metode yang digunakan *BM25* sebagai pembobot kata dan *Improved K-NN* sebagai algoritma klasifikasi sentimen, dan fitur seleksi dengan *Chi-Square*, dengan rasio fitur 50% dan nilai *k* sebesar 20, sistem mencapai hasil terbaik dengan precision 70,03%, recall 67,22%, akurasi 83,33%, dan f-measure 66,26%. Paper [5] menerapkan metode Random Forest dengan fokus pada dua kelas sentimen (positif dan negatif), menunjukkan akurasi sebesar 90%, presisi 90%, recall 91%, dan F1-Score 89%, dengan pengaturan optimal menggunakan 80 pohon keputusan dan kedalaman pohon sebesar 50. Pada paper [6] model Random Forest memiliki kinerja terbaik dengan akurasi 97,16% dan skor *AUC* 0,996.

Data ulasan dan rating pengguna CapCut di Google Play Store akan digunakan sebagai sumber data. Metode Random Forest dipilih sebagai metode klasifikasi untuk mengevaluasi hasil ulasan. Hasil analisis akan membantu dalam pemahaman perasaan dan umpan balik. Pendekatan ini dipilih karena Random Forest menggabungkan prediksi dari berbagai decision tree dalam satu model, sehingga mengurangi kemungkinan *overfitting*. *Overfitting*

adalah situasi dimana model sangat bergantung pada satu dataset, sehingga ketika digunakan pada dataset lain hasilnya bisa sangat berbeda [7]. Algoritma Random Forest telah menunjukkan kinerja yang sangat baik dalam mengidentifikasi kelemahan dalam dataset Google Play Store [8].

Berdasarkan uraian diatas, maka peneliti mengusulkan studi analisis sentimen terhadap ulasan aplikasi CapCut dengan menerapkan metode Random Forest.

2. TINJAUAN PUSTAKA

2.1. Ulasan

Ulasan merupakan sebuah evaluasi atau pandangan yang diberikan individu atau pengguna suatu produk, layanan, tempat, atau topik tertentu. Ulasan dapat diberikan untuk berbagai tujuan, seperti memberikan umpan balik kepada penyedia layanan atau produk, memberikan panduan kepada pengguna lain dalam proses pengambilan keputusan pembelian, atau hanya berbagi pengalaman. Secara khusus, ulasan online memberikan informasi tentang penilaian yang biasanya dinyatakan dalam bentuk rating, sementara pengalaman diungkapkan melalui komentar atau saran [9].

2.2. Analisis Sentimen

Analisis sentimen merupakan subdisiplin dalam text mining yang mulai mendapatkan perhatian luas sekitar tahun 2013. Analisis sentimen adalah sebuah pendekatan komputasional untuk menganalisis opini, sentimen, dan ekspresi emosional yang disampaikan dalam bentuk teks [10]. Menurut [11], analisis sentimen terdiri dari empat kategori utama, yaitu :

- Analisis sentimen berbasis peringkat, yang digunakan untuk mengartikan penilaian bintang dalam ulasan, dimana 5 bintang menunjukkan sentimen sangat positif dan 1 bintang menunjukkan sentimen sangat negatif.
- Deteksi emosi, yang bertujuan untuk mengidentifikasi emosi, seperti kebahagiaan, kemarahan, kesedihan, dan frustrasi yang muncul dalam teks.
- Analisis sentimen berdasarkan aspek, yang digunakan untuk menentukan aspek atau fitur dalam teks yang memiliki sentimen positif, negatif, atau netral.
- Analisis sentimen multibahasa, yang mampu mengenali dan mengidentifikasi bahasa yang digunakan dalam teks sesuai dengan pilihan bahasa yang ditentukan

2.3. Preprocessing

Preprocessing merupakan tahap penting dalam pengelolaan data, terutama dalam kerangka analisis data dan teks. Langkah ini dilakukan untuk menghilangkan gangguan dalam data dan mengorganisir data menjadi struktur yang lebih teratur dan mudah diolah [6]. Berikut tahapan dalam preprocessing menurut [12] yaitu :

- Cleaning**
Cleaning adalah proses menghapus atribut yang tidak memiliki pengaruh signifikan dalam proses klasifikasi.
- Case Folding**
Case folding adalah proses mengubah huruf kapital menjadi huruf kecil untuk menghindari masalah case sensitif dalam proses klasifikasi.
- Tokenisasi**
Tokenisasi adalah proses memecah kalimat utuh menjadi unit-unit kecil yang lebih terstruktur.
- Stopword Removal**
Stopword removal adalah proses penghapusan kata-kata yang tidak memiliki makna atau pengaruh signifikan dalam kalimat.
- Lemmatisasi**
Lemmatisasi adalah proses mengubah kata-kata ke dalam bentuk dasar atau kata dasar yang disebut sebagai "lemma", sehingga kata-kata dengan makna serupa dapat dipresentasikan secara konsisten. Lemmatisasi memiliki kesamaan dengan stemming, hanya saja proses lemmatisasi tidak menggunakan library "Sastrawi", melainkan library 'nlp_id'.

2.4. Pembobotan Kata

Pemberian bobot kata dalam dokumen teks menggunakan metode *TF-IDF* melibatkan langkah-langkah berikut [13] :

- Perhitungan Term Frequency (TF)**
TF adalah ukuran yang menghitung frekuensi atau banyaknya kemunculan kata dalam suatu dokumen teks.
Rumusnya adalah :
$$tf_{t,d} = \{1 + \log_{10} tf_{t,d}, \text{if } tf_{t,d} > 0\}$$
- Inverse Document Frequency (IDF)**
IDF adalah ukuran yang menghitung seberapa sering kata tertentu muncul dalam seluruh koleksi dokumen.

Rumusnya adalah : $idf_t = \log_{10} \left(\frac{N}{df(t)} \right)$

2.5. Random Forest

Random Forest adalah algoritma yang digunakan untuk mengklasifikasikan data dalam volume besar, melibatkan penggabungan beberapa pohon keputusan dan melatihnya menggunakan sampel data yang ada [5]. Algoritma ini terdiri dari sejumlah pohon keputusan yang disebut sebagai "Decision Tree". Masing-masing pohon keputusan memiliki node-node yang digunakan untuk membagi keputusan. Random Forest melakukan prediksi dalam analisis sentimen dengan menggabungkan hasil dari setiap pohon keputusan (Indarbensyah dan Rochmawati, 2021). Menurut [13] proses dalam membuat pohon keputusan, diantaranya :

- Menghitung jumlah pohon yang akan dibuat
- Menghitung tingkat kemurnian gini

- c. Menghitung keuntungan informasi pada setiap pertanyaan selama proses pemisahan
- d. Nilai keuntungan informasi tertinggi menjadi akar.

Mengulangi langkah 2-4 untuk setiap pohon hingga mencapai jumlah pohon yang ditentukan.

Beberapa keunggulan penggunaan Algoritma Random Forest menurut [14], meliputi:

- a. Tingkat akurasi yang tinggi pada hasil prediksi.
- b. Sensitivitas terhadap gangguan atau noise (outliers), memungkinkan penanganan data yang tidak biasa.
- c. Kemudahan dalam implementasi dan paralelisasi, memungkinkan pemrosesan yang efisien pada sistem terdistribusi.
- d. Proses pengolahan data yang lebih cepat dibandingkan dengan teknik boosting dan bagging.

2.6. Evaluasi Model

Evaluasi model merupakan proses penting dalam penilaian kinerja suatu model atau algoritma dalam pemrosesan data. Evaluasi model dapat memberikan wawasan tentang seberapa baik atau buruk model tersebut dalam menyelesaikan tugas yang diberikan, seperti klasifikasi, regresi, atau tugas lainnya. Confusion Matrix adalah suatu alat evaluasi dalam bentuk matriks yang digunakan untuk mengevaluasi kinerja model klasifikasi. Menurut (Raffi dkk., 2023) Confusion Matrix memiliki empat jenis kejadian yang meliputi :

- *TP* (True Positive) : data yang berhasil terklasifikasi sebagai positif dengan benar.
- *TN* (True Negative) : data yang berhasil terklasifikasi sebagai negatif dengan benar.
- *FP*(False Positive) : data yang keliru terklasifikasi sebagai positif.
- *FN* (False Negative) : data yang keliru terklasifikasi sebagai negatif.

Dengan menggunakan nilai yang terdapat dalam Confusion Matrix, maka dapat menghitung berbagai matrix evaluasi berikut [12] :

1. Akurasi (Accuracy), mengukur sejauh mana model dapat mengklasifikasikan data dengan benar, dapat dihitung menggunakan rumus sebagai berikut :

$$\text{accuracy} = \frac{(TP+TN)}{TP+TN+FP+FN}$$

2. Presisi (Precision), mengukur seberapa tepat model dalam memprediksi kelas positif, dapat dihitung menggunakan rumus sebagai berikut :

$$\text{presisi} = \frac{TP}{(TP+FP)}$$

3. Recall, mengukur seberapa baik model dalam mengidentifikasi data kelas positif berdasarkan semua data actual yang seharusnya positif, dapat dihitung menggunakan rumus sebagai berikut :

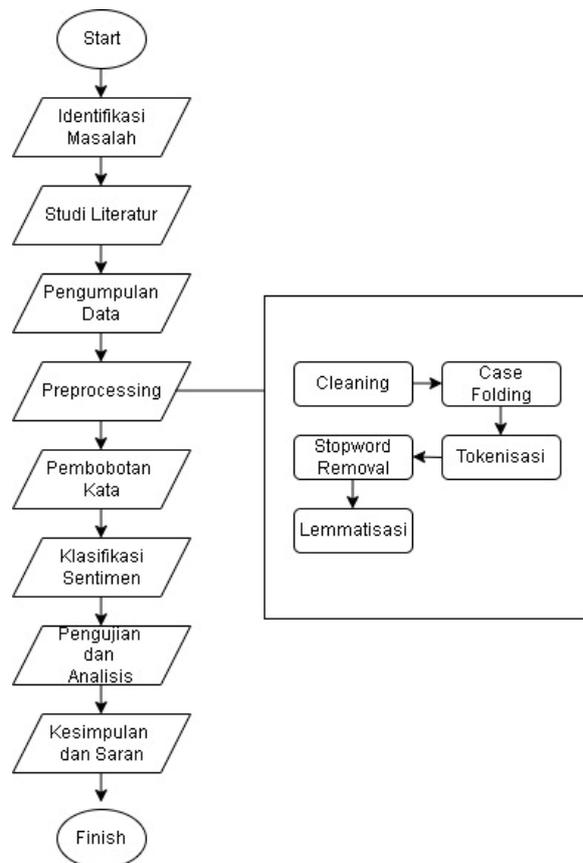
$$\text{recall} = \frac{TP}{(TP+FN)}$$

4. F1-Score, matrix yang memberikan keseimbangan antara presisi dan recall, dapat dihitung menggunakan rumus sebagai berikut :

$$\text{F1-Score} = 2 \times \frac{(\text{Presisi} \times \text{Recall})}{(\text{Presisi} + \text{Recall})}$$

3. METODE PENELITIAN

Penelitian ini menggunakan algoritma Random Forest dalam melakukan klasifikasi sentimen terhadap ulasan pengguna aplikasi CapCut di Google Play Store. Terdapat 9 tahapan penelitian, yaitu dimulai dari identifikasi masalah, studi literatur, pengumpulan data, pelabelan, preprocessing, pembobotan kata, klasifikasi sentimen, pengujian dan analisis, serta penarikan kesimpulan dan saran. Berikut diagram alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Metode Penelitian

3.1. Pengumpulan Data

Penelitian ini melakukan pengumpulan data dengan menggunakan tools python (Google Colab) melalui teknik web scraping. Data yang diambil berisi kolom username, tanggal, rating/score, dan ulasan pada aplikasi CapCut di google play store. Dataset yang dimiliki berjumlah 5000 data ulasan mulai dari tanggal 1 Maret hingga 25 Agustus 2023.

3.2. Pelabelan

Proses pelabelan dilakukan berdasarkan nilai rating yang diberikan pengguna aplikasi CapCut dan hanya fokus pada dua label kelas saja, yaitu positif dan negatif. Tidak menggunakan label kelas netral,

mengingat ulasan cenderung mengandung kata-kata sentimen positif dan negatif

3.3. Preprocessing

Pada penelitian ini, proses yang dilakukan dalam tahapan preprocessing yaitu cleaning, case folding, tokenisasi, stopwords removal, dan lemmatisasi

3.4. Pembobotan Kata

Penelitian ini menggunakan *TF-IDF* (Term Frequency – Inverse Document Frequency) dalam pembobotan kata. Dimana bobot yang diperoleh berdasarkan frekuensi yang muncul pada setiap ulasan.

3.5. Klasifikasi Sentimen

Pada tahap ini, dilakukan analisis untuk menentukan kategori sentimen dari setiap ulasan. Metode yang digunakan dalam penelitian ini adalah klasifikasi sentimen menggunakan algoritma Random Forest.

3.6. Evaluasi Model

Pada tahap ini, kinerja model Random Forest akan dievaluasi menggunakan metrik, seperti akurasi, presisi, recall, dan F1-Score.

3.7. Visualisasi

Visualisasi seperti pie chart dan word cloud dapat digunakan untuk mempresentasikan temuan-temuan penting dari analisis sentimen secara lebih jelas.

4. HASIL DAN PEMBAHASAN

Penerapan metode Random Forest dalam menganalisis sentimen pengguna aplikasi CapCut, dimulai dengan tahap pengumpulan data, dilakukan melalui teknik scraping dengan mengambil 5000 ulasan sebagai sampel.

```
from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    'com.limon.lvoverseas',
    lang='id',
    country='id',
    sort=Sort.MOST_RELEVANT,
    count=5000,
    filter_score_with=None
)
```

Gambar 2. Proses Scraping Ulasan

| | user:time | score | at | content |
|---|-----------------|-------|---------------------|---|
| 0 | Chaity Ajah | 4 | 2023-08-25 13:13:15 | Waktu du siiihhh biasa aja ya buat ngedit nya ... |
| 1 | Indra Pratama | 2 | 2023-08-25 13:09:59 | Maaf ya ini kok sekarang capcut jadi eror ya e... |
| 2 | Dede Taufiq | 4 | 2023-08-25 13:01:21 | Good job developer ternyata gw belum update0Y'... |
| 3 | Riska Yulandani | 5 | 2023-08-25 12:53:53 | Aplikasi nya lumayan lah aku suka tapi kenapa ... |
| 4 | Yana khaira | 2 | 2023-08-25 12:48:29 | Kenapa ya capcut aneh banget akhir' ini, ngebu... |

Gambar 3. Hasil Scraping

Berikutnya, proses pelabelan dilakukan berdasarkan rating yang diberikan pengguna, dimana penelitian ini menginterpretasikan penilaian bintang dalam ulasan [11].

Kemudian, proses preprocessing dilakukan dengan langkah-langkah seperti cleaning, tahap pembersihan data yang seksama termasuk eliminasi karakter khusus, tautan, atau format yang tidak diinginkan.

| | score | content | sentiment |
|---|-------|---|-----------|
| 0 | 4 | Waktu du siiihhh biasa aja ya buat ngedit nya ... | positif |
| 1 | 2 | Maaf ya ini kok sekarang capcut jadi eror ya e... | negatif |
| 2 | 4 | Good job developer ternyata gw belum update0Y'... | positif |
| 3 | 5 | Aplikasi nya lumayan lah aku suka tapi kenapa ... | positif |
| 4 | 2 | Kenapa ya capcut aneh banget akhir' ini, ngebu... | negatif |

Gambar 4. Hasil Pelabelan

Selanjutnya case folding, data disesuaikan dengan mengubahnya menjadi huruf kecil untuk menghindari ambiguitas antara huruf besar dan huruf kecil.

Tabel 1. Hasil Proses Cleaning dan Case Folding

| Sebelum | Sesudah |
|---|--|
| Aplikasi yang sangat berguna bagi editor pemula dengan banyak fitur2 yang tersedia dapat menghasilkan editan yang di Inginkan | aplikasi yang sangat berguna bagi editor pemula dengan banyak fitur yang tersedia dapat menghasilkan editan yang di inginkan |
| Aplikasi nya bagus ya, tapi memakan banyak memori doang minusnya di situ, mohon memperkecil data aplikasi:) | aplikasi nya bagus ya tapi memakan banyak memori doang minusnya di situ mohon memperkecil data aplikasi |

Proses tokenisasi dilakukan untuk memisahkan setiap kata atau frasa menjadi token terpisah.

Tabel 2. Hasil Proses Tokenisasi

| Sebelum | Sesudah |
|---|---|
| Aplikasi yang sangat berguna bagi editor pemula dengan banyak fitur2 yang tersedia dapat menghasilkan editan yang di Inginkan | ['aplikasi', 'yang', 'sangat', 'berguna', 'bagi', 'editor', 'pemula', 'dengan', 'banyak', 'fitur', 'yang', 'tersedia', 'dapat', 'menghasilkan', 'editan', 'yang', 'di', 'inginkan'] |
| Aplikasi nya bagus ya, tapi memakan banyak memori doang minusnya di situ, mohon memperkecil data aplikasi:) | ['aplikasi', 'nya', 'bagus', 'ya', 'tapi', 'memakan', 'banyak', 'memori', 'doang', 'minus', 'di', 'situ', 'mohon', 'memperkecil', 'data', 'aplikasi'] |

Kemudian stopwords removal, penghapusan kata-kata umum (stopword) yang tidak memberikan kontribusi signifikan pada sentimen.

Tabel 3. Hasil Proses Stopword Removal

| Sebelum | Sesudah |
|---|---|
| Aplikasi yang sangat berguna bagi editor pemula dengan banyak fitur2 yang tersedia dapat menghasilkan editan yang di Inginkan | aplikasi berguna editor pemula fitur tersedia menghasilkan editan |
| Aplikasi nya bagus ya, tapi memakan banyak memori doang | aplikasi bagus memakan memori |

| Sebelum | Sesudah |
|---|---------------------------------------|
| minusnya di situ, mohon memperkecil data aplikasi:) | minus mohon memperkecil data aplikasi |

Langkah selanjutnya adalah lemmatisasi, dimana kata-kata diubah ke bentuk dasar untuk memastikan konsistensi dalam representasi kata.

Tabel 4. Hasil Proses Lemmatisasi

| Sebelum | Sesudah |
|---|---|
| Aplikasi yang sangat berguna bagi editor pemula dengan banyak fitur2 yang tersedia dapat menghasilkan editan yang di Inginkan | aplikasi guna editor mula fitur sedia hasil edit |
| Aplikasi nya bagus ya, tapi memakan banyak memori doang minusnya di situ, mohon memperkecil data aplikasi:) | aplikasi bagus makan memori minus mohon kecil data aplikasi |

Setelah tahap preprocessing selesai, dilakukan proses pembobotan kata menggunakan *TF-IDF*. Pembobotan kata dilakukan untuk memberikan nilai numerik pada setiap kata dalam ulasan, memperhitungkan tingkat pentingnya kata tersebut dalam konteks sentimen.

Proses berikutnya melibatkan pembagian dataset menjadi dua bagian utama, dimana 90% dari dataset dijadikan sebagai data latih dan 10% sebagai data uji. Tujuan dari pembagian ini adalah untuk melatih model menggunakan sebagian besar data yang tersedia dan menguji kinerja model pada data yang belum pernah dilihat sebelumnya, dengan harapan mengurangi resiko *overfitting* [7]

Proses konstruksi model Random Forest dimulai dengan mengimpor struktur melalui *library scikit-learn* dan menyimpan arsitektur dengan nama 'rfc' untuk kemudahan akses. Model tersebut kemudian dilatih menggunakan data latih, memungkinkan model untuk menyesuaikan diri dengan pola sentimen dalam dataset.

```

from sklearn.ensemble import RandomForestClassifier
rfc = RandomForestClassifier()
rfc.fit(x_train, y_train)
y_predict_rfc = rfc.predict(x_test)
predict_1 = rfc.predict(x_test)
    
```

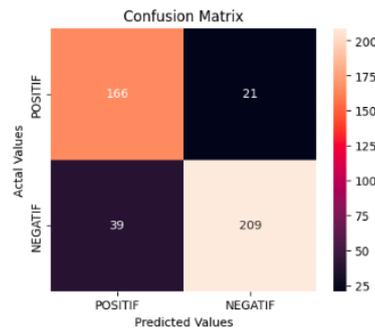
Gambar 5. Code Model Random Forest Classifier

Hasil prediksi dari data uji disimpan dengan nama 'predict_1', memungkinkan peneliti untuk mengevaluasi sejauh mana model dapat memprediksi sentimen ulasan pada data yang belum pernah dilihat sebelumnya.

Proses evaluasi model menjadi langkah penting selanjutnya, dimana matrix evaluasi seperti akurasi, presisi, recall, dan f1-score digunakan [12].

Hasil penelitian menunjukkan bahwa metode Random Forest memberikan tingkat akurasi yang tinggi dalam menganalisis sentimen pengguna

CapCut, dengan akurasi sebesar 86%, presisi 89%, recall 81%, dan f1-score 85%.



Gambar 6. Hasil Confusion Matrix

Dari hasil tersebut, Confusion Matrix juga memberikan informasi tambahan. Terdapat 166 True Positive (TP), yang merupakan ulasan positif yang diklasifikasikan dengan benar. Sebanyak 209 True Negative (TN) menunjukkan ulasan negatif yang diklasifikasikan dengan benar. Disisi lain, terdapat 21 False Positive (FP) yang menunjukkan ulasan negatif yang keliru diklasifikasikan sebagai positif, dan 39 False Negative (FN) yang menunjukkan ulasan positif yang keliru diklasifikasikan sebagai negatif.

```

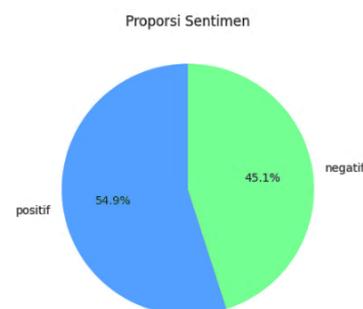
from sklearn.metrics import classification_report
print("RFC Classification Report :\n", classification_report(y_test, predict_1))
    
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| negatif | 0.81 | 0.89 | 0.85 | 187 |
| positif | 0.91 | 0.84 | 0.87 | 248 |
| accuracy | | | 0.86 | 435 |
| macro avg | 0.86 | 0.87 | 0.86 | 435 |
| weighted avg | 0.87 | 0.86 | 0.86 | 435 |

Gambar 7. Hasil Classification Report

Gambar diatas menunjukkan hasil classification report yang memberikan informasi lebih lanjut tentang kinerja model. Dari laporan klasifikasi tersebut, dapat disimpulkan bahwa kemampuan pengklasifikasian sudah sangat baik, dengan tingkat akurasi keseluruhan mencapai 86%, keseimbangan antara precision dan recall mencerminkan kinerja model yang memuaskan.

Dalam proses penelitian ini, penggunaan visualisasi memegang peranan penting dalam mempresentasikan informasi dengan cara yang lebih terstruktur dan mudah dipahami.



Gambar 8. Pie Chart Dari label Ulasan

Visualisasi perbandingan antara jumlah label ‘positif’ dan ‘negatif’ dilakukan melalui pie chart, seperti terlihat pada Gambar 4. 7 yang menunjukkan bahwa label ‘positif’ lebih dominan dengan presentase 54,9%, sedangkan label ‘negatif’ memiliki presentase 45,1%. Analisis ini memberikan informasi tentang sebaran sentimen dalam dataset setelah proses pelabelan.



Gambar 9. Word Cloud Positif

Berdasarkan gambar diatas, dapat diamati bahwa kata-kata yang muncul pada sentimen positif melibatkan istilah-istilah yang bersifat positif terkait dengan pengalaman pengguna. Contohnya, kata "bagus" menunjukkan tingkat kepuasan atau kualitas yang baik, sementara kata "capcut" dan "aplikasi" menunjukkan fokus pada aplikasi yang mungkin digunakan untuk mengedit video. Selain itu, kata-kata seperti "edit", "video", "template", dan "keren" menunjukkan minat atau aktivitas pengguna dalam mengedit video dengan menggunakan aplikasi tersebut.



Gambar 10. Word Cloud Negatif

Disisi lain, pada sentimen negatif terdapat kata-kata seperti "update" dan "tolong" yang memberikan petunjuk mengenai masalah atau ketidakpuasan pengguna. Kata "ga" sering digunakan sebagai singkatan dari "tidak" atau memberikan kesan negatif terhadap suatu hal. Kata "update" mungkin merujuk pada pembaruan aplikasi yang mungkin tidak disukai atau menimbulkan masalah. Kata "tolong" bisa menunjukkan adanya permohonan bantuan atau ketidakpuasan terhadap sesuatu yang memerlukan perbaikan.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil menerapkan metode Random Forest dalam menganalisis sentimen pengguna aplikasi CapCut. Setiap langkah, dari pengumpulan data hingga evaluasi model dilakukan secara sistematis dan didasarkan pada landasan teori yang kuat. Hasil evaluasi menggunakan Classification Report dan Confusion Matrix menunjukkan akurasi sebesar 86%, presisi 89%, recall 81%, dan f1-score 85%. Hasil tersebut menandakan kelayakan

model dalam mengklasifikasikan sentimen. Random Forest terbukti menjadi pilihan yang efektif dengan keunggulannya dalam mengatasi overfitting dan data yang tidak lengkap. Pada penelitian selanjutnya, diharapkan dapat menggunakan dataset ulasan terkini dengan jumlah volume yang lebih besar dan pengambilan data dalam kurun waktu tertentu (misal 1 tahun), serta membandingkan hasil pengujian menggunakan algoritma klasifikasi lain untuk memperoleh hasil yang lebih baik.

DAFTAR PUSTAKA

[1] N. M. Hutabarat, “Penggunaan Aplikasi Edit Video Dalam Materi Seni Musik Pada Pembelajaran Secara Daring,” *Quaerite Verit. J. Pendidik.*, vol. 1, no. 2, hal. 81–90, 2022.

[2] E. D. S. Mulyani, A. T. Hidayatulloh, dan T. Agustawan, “Analisis Sentimen Ulasan Produk Pada Top Brand Produk Masker Di Tokopedia Menggunakan Naive Bayes,” *Semin. Nas. Corisindo*, hal. 287–292, 2022, [Daring]. Tersedia pada: <https://corisindo.stikom-bali.ac.id/penelitian/index.php/semnas/article/view/58%0Ahttps://corisindo.stikom-bali.ac.id/penelitian/index.php/semnas/article/download/58/44>

[3] P. B. Sihotang, F. D. B. Sitanggang, N. Azriansyah, dan E. Indra, “Penerapan Natural Language Processing Untuk Analisis Sentimen Terhadap Aplikasi Streaming,” *J. Ilm. Betrik*, vol. 14, no. 2, hal. 273–282, 2023.

[4] A. Purnamawati, M. N. Winarto, dan M. Mailasari, “Analisis Sentimen Aplikasi TikTok Menggunakan Metode BM25 dan Improved K-NN Fitur Chi-Square,” *J. Komtika (Komputasi dan Inform.*, vol. 7, no. 1, hal. 97–105, 2023, [Daring]. Tersedia pada: <https://journal.unimma.ac.id/index.php/komtika/article/view/8938>

[5] C. G. Indrayanto, D. E. Ratnawati, dan B. Rahayudi, “Analisis Sentimen Data Ulasan Pengguna Aplikasi MyPertamina di Indonesia pada Google Play Store menggunakan Metode Random Forest,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. j-ptiik.ub.ac.id, hal. 1131–1139, 2023. [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/12390/5636>

[6] E. Fitri, Y. Yuliani, S. Rosyida, dan W. Gata, “Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine,” *J. Transform.*, vol. 18, no. 1, hal. 71–80, 2020, [Daring]. Tersedia pada: <http://journals.usm.ac.id/index.php/transformatika/article/view/2317>

[7] M. A. A. Jihad, A. Adiwijaya, dan W. Astuti, “Analisis Sentimen Terhadap Ulasan Film Menggunakan Algoritma Random Forest,” *e-*

- Proceeding Eng.*, vol. 8, no. 5, hal. 10153–10165, 2021.
- [8] S. Sudianto, P. Wahyuningtias, H. W. Utami, U. A. Raihan, H. N. Hanifah, dan Y. N. Adanson, “Comparison Of Random Forest And Support Vector Machine Methods On Twitter Sentiment Analysis (Case Study: Internet Selebgram Rachel Vennya Escape From Quarantine),” *J. Tek. Inform.*, vol. 3, no. 1, hal. 141–145, 2022.
- [9] B. B. Baskoro, I. Susanto, dan S. Khomsah, “Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR),” *J. Informatics Inf. Syst. Softw. Eng. Appl.*, vol. 3, no. 2, hal. 21–29, 2021, doi: 10.20895/INISTA.V3.
- [10] D. Alita dan A. Rahman, “Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier,” *J. Komputasi*, vol. 8, no. 2, hal. 50–58, 2020, [Daring]. Tersedia pada: <https://jurnal.fmipa.unila.ac.id/komputasi/article/view/2615>
- [11] F. A. Larasati, D. E. Ratnawati, dan B. T. Hanggara, “Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 9, hal. 4305–4313, 2022, [Daring]. Tersedia pada: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11562>
- [12] M. Raffi, A. Suharso, dan I. Maulana, “Analisis Sentimen Ulasan Aplikasi Binar Pada Google Play Store Menggunakan Algoritma Naive Bayes,” *J. Inf. Technol. Comput. Sci.*, vol. 6, no. 1, hal. 238–462, 2023, doi: 10.31539/intecom.v6i1.6117.
- [13] H. C. Morama, D. E. Ratnawati, dan I. Arwani, “Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 4, hal. 1702–1708, 2022, [Daring]. Tersedia pada: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/10908>
- [14] P. P. E. Indarbensyah dan N. Rochmawati, “Penerapan N-Gram menggunakan Algoritma Random Forest dan Naive Bayes Classifier pada Analisis Sentimen Kebijakan PPKM 2021,” *J. Informatics Comput. Sci.*, vol. 2, no. 4, hal. 235–244, 2021, [Daring]. Tersedia pada: <https://ejournal.unesa.ac.id/index.php/jinacs/article/view/39845>
- [15] CNN Indonesia. “TikTok Jadi Raja Aplikasi Dunia, Kalahkan Instagram Hingga WhatsApp”. *cnnindonesia.com*, 19 September 2023. [Daring]. Tersedia pada: <https://www.cnnindonesia.com/teknologi/20230918143850-185-1000458/tiktok-jadi-raja-aplikasi-dunia-kalahkan-instagram-hingga-whatsapp>. [Akses 10 November 2023]
- [16] Pricillia.” Aplikasi Edit Video di HP Terbaik dan Gratis, Cocok untuk Pemula dan Profesional”. *mediaindonesia.com*, 20 Oktober 2023. [Daring]. Tersedia pada: <https://mediaindonesia.com/humaniora/622903/6-aplikasi-edit-video-di-hp-terbaik-dan-gratis-cocok-untuk-pemula-dan-profesional>. [Akses 10 November 2023]