Sentiment Analysis of Shopee App Users on Google Play Store Using the Random Forest Method
Analisis Sentimen Pengguna Aplikasi Shopee Pada Google Play Store Menggunakan Metode Random Forest

Muhammad Zainal Abidin[1)], Mochammad Alfan Rosid[*,2)], Ade Eviyanti[3)]

[1)] Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia
[2)] Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia
[3)] Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia
*Email Penulis Korespondensi: (mochalfanrosid@umsida.ac.id

Abstract. *This study focuses on sentiment analysis to evaluate customer satisfaction with the Shopee app, using comments posted on the Google Play Store as the primary data source. A total of 5,000 comment data were collected over a relevant timeframe, from December 2024 to March 2025. The methodology applied was classification using the Random Forest Classifier algorithm. The analysis results show that the dominant sentiment expressed by users is positive, indicating a good level of satisfaction with the app. The Random Forest model successfully achieved an accuracy of 88%. This figure indicates that the algorithm is quite effective in classifying user comment sentiment. As a key contribution, this study provides up-to-date insights into customer perceptions thanks to the use of very recent data. These findings not only validate the effectiveness of Random Forest in sentiment analysis tasks but also provide valuable information for Shopee to understand user views and make strategic decisions to improve services*.

Keywords - *Sentiment Analysis; Random Forest; Customer Satisfaction; Shopee; Play Store.*

Abstrak. *Penelitian ini berfokus pada analisis sentimen untuk mengevaluasi kepuasan pelanggan terhadap aplikasi Shopee, dengan menggunakan komentar-komentar yang diunggah di Google Play Store sebagai sumber data utama. Sebanyak 5.000 data komentar dikumpulkan dalam rentang waktu yang relevan, yaitu dari Desember 2024 hingga Maret 2025. Metodologi yang diterapkan adalah klasifikasi dengan menggunakan algoritma Random Forest Classifier. Hasil analisis menunjukkan bahwa sentimen dominan yang diekspresikan oleh pengguna adalah positif, yang mengindikasikan tingkat kepuasan yang baik terhadap aplikasi tersebut. Model Random Forest yang digunakan berhasil mencapai nilai akurasi sebesar 88%. Angka ini menunjukkan bahwa algoritma tersebut cukup efektif dalam mengklasifikasikan sentimen komentar pengguna. Sebagai kontribusi utama, penelitian ini menyediakan wawasan terkini mengenai persepsi pelanggan berkat penggunaan data yang sangat baru. Temuan ini tidak hanya memvalidasi efektivitas Random Forest dalam tugas analisis sentimen, tetapi juga memberikan informasi berharga bagi pihak Shopee untuk memahami pandangan pengguna dan membuat keputusan strategis guna meningkatkan layanan.*

Kata Kunci - *Analisis Sentimen; Random Forest; Kepuasan Pelanggan; Shopee; Play Store*

I. PENDAHULUAN

Aplikasi e-commerce seperti Shopee telah menjadi bagian integral dari kehidupan sehari-hari, memfasilitasi aktivitas jual beli secara online dengan beragam fitur menarik. Kepuasan pelanggan menjadi kunci dalam lanskap digital yang kompetitif ini, dipengaruhi oleh kemudahan navigasi, keamanan transaksi, kelengkapan fitur, dan kecepatan layanan (Rizky & Mahfudz, 2022. Ulasan pengguna di Google Play Store menjadi indikator krusial yang mencerminkan pengalaman mereka, meliputi penilaian terhadap fitur, kemudahan penggunaan, kecepatan transaksi, hingga kualitas layanan pelanggan. Analisis sentimen terhadap ulasan ini menjadi sangat penting untuk memahami opini publik dan memperoleh umpan balik berharga, yang kemudian dapat dimanfaatkan untuk perbaikan dan pengembangan aplikasi guna meningkatkan loyalitas pelanggan dan menarik pengguna baru. Data ulasan aplikasi Shopee diperoleh melalui scraping menggunakan API Google-Play-Scraper. Google-Play-Scraper adalah API yang memungkinkan dengan mudah mengekstraksi data informasi aplikasi dan ulasan aplikasi dari Google Play Store tanpa bergantung pada eksternal (Fahmi & Sumarno, 2022). Data yang diekstrak kemudian akan diproses melalui proses text preprocessing. Pendekatan ini melibatkan pengolahan teks semi-terstruktur dari ulasan menjadi data terorganisir, yang kemudian akan diklasifikasikan ke dalam kategori sentimen positif, negatif, atau netral (Arsi et al., 2021). Untuk proses klasifikasi, penelitian ini akan memanfaatkan algoritma Random Forest. Algoritma ini dipilih karena performanya yang unggul dalam klasifikasi data dan kemampuannya menangani data besar serta kompleks tanpa memerlukan banyak penyesuaian model (Amaliah et al., 2022). Proses analisis akan mencakup preprocessing data, ekstraksi fitur, pemodelan, dan evaluasi untuk memastikan akurasi klasifikasi sentimen.

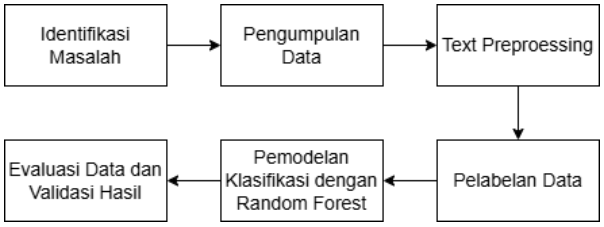
Analisis sentimen, juga dikenal sebagai opinion mining, adalah metode otomatis untuk memahami, mengekstraksi, dan mengolah data teks guna menentukan penilaian positif atau negatif suatu pernyataan (Gifari et al., 2022). Teknik ini krusial untuk mengidentifikasi opini publik terhadap produk, layanan, atau merek, yang dapat menjadi umpan balik vital bagi pengembang. Random Forest merupakan algoritma machine learning supervised yang terbukti efektif dalam klasifikasi teks. Algoritma ini membangun banyak pohon keputusan yang dilatih dengan subset acak data dan fitur, menggabungkan prediksi dari setiap pohon untuk meningkatkan akurasi dan stabilitas model, serta mengurangi overfitting (Mahmuda, 2024).

Penelitian sebelumnya telah banyak meneliti sentimen dan mengklasifikasikan opini pengguna terhadap berbagai objek atau sistem, seringkali memanfaatkan platform media sosial seperti Twitter sebagai sumber data utama karena kemampuannya untuk dilacak dan dievaluasi secara instan (Vonega et al., 2022). Beberapa studi telah menunjukkan bahwa metode Random Forest memiliki performa yang unggul dalam klasifikasi data pada berbagai kasus. Amaliah et al. (2022) dan Mahmudah (2024), secara spesifik menyoroti efektivitas Random Forest dalam klasifikasi teks tradisional. Meskipun demikian, masih terdapat kesenjangan dalam membandingkan sistem penjabaran sentimen menggunakan berbagai pendekatan algoritma, mendorong penelitian ini untuk melakukan perbandingan lebih lanjut.

Penelitian ini bertujuan untuk menganalisis sentimen pengguna yang diungkapkan dalam ulasan aplikasi Shopee di Google Play Store dari Desember 2024 hingga Maret 2025 sebanyak 5.000 data ulasan. Evaluasi yang digunakan yaitu akurasi, recall, presisi dan f1-score. Tujuan utamanya adalah menilai dinamika sentimen masyarakat terhadap aplikasi e-commerce ini. Kebaharuan penelitian ini terletak pada analisis sentimen terkini terhadap pengalaman pengguna Shopee setelah perubahan tren belanja digital, serta upaya membandingkan berbagai prosedur penyelesaian algoritma untuk menemukan pendekatan terbaik dalam memahami opini pengguna. Hasil analisis ini diharapkan memberikan wawasan mendalam dan masukan berguna bagi pengembang aplikasi untuk meningkatkan layanan Shopee.

II. METODE

Penelitian ini akan mengikuti serangkaian langkah sistematis untuk menganalisis sentimen ulasan pengguna aplikasi Shopee di Google Play Store menggunakan metode Random Forest. Tahapan yang dilakukan yaitu identifikasi masalah, pengumpulan data, text preprosessing, pelabelan data, pemodelan klasifikasi dengan Random Forest, dan evaluasi data dan hasil validasi. Diagram alir berikut memberikan gambaran umum prosesnya:



Gambar 1 Alur Penelitian

A. Analisis Sentimen

Analisis sentimen adalah sebuah teknik komputasi untuk menentukan dan mengklasifikasikan polaritas sentimen yang terkandung dalam suatu teks atau dokumen (Wardani et al., 2020). Tujuan utamanya adalah untuk mengkategorikan teks tersebut sebagai positif, negatif, atau netral. Dalam praktiknya, analisis ini sering diterapkan pada data dari jaringan media sosial seperti Twitter untuk mengevaluasi pandangan masyarakat secara luas (Kurniawan & Susanto, 2019). Hal ini juga serupa dengan *opini mining*, di mana fokusnya adalah menambang data tekstual untuk mengekstrak dan menganalisis pendapat yang diungkapkan, baik itu tentang suatu produk, layanan, atau topik spesifik lainnya

B. Random Forest

Random Forest Classifier adalah algoritma klasifikasi ensemble yang berevolusi dari *decision tree*. Setiap pohon keputusan dalam model ini dilatih secara independen menggunakan subset data dan atribut yang dipilih secara acak (Alvanof & Dinata, 2024). Algoritma ini unggul dalam meningkatkan akurasi, bahkan ketika dihadapkan pada data yang tidak lengkap atau memiliki nilai ekstrem (*outliers*) (Siregar et al., 2023). Selain efisiensi dalam pengelolaan data, Random Forest juga mampu mengidentifikasi fitur-fitur yang paling relevan untuk meningkatkan kinerja model klasifikasi secara keseluruhan (Erkamim et al., 2023).

C. Identifikasi Masalah

Penelitian ini dilakukan berdasarkan permasalahan yang telah dijelaskan sebelumnya, yaitu:

a. Penerapan metode Random Forest pada analisis ulasan pengguna di aplikasi Shopee di Play Store, untuk mengklasifikasikan teks ulasan ke dalam label positif, negatif, dan netral.
b. Mengetahui tingkat akurasi metode Random Forest dalam menganalisis sentimen pengguna terhadap aplikasi Shopee di tahun 2024.
c. Memberi hasil analisa berupa saran pengembangan aplikasi mengenai ulasan yang telah dihasilkan untuk meningkatkan kepuasan pelanggan pengguna Shoppe.

D. Pengumpulan Data

Langkah pertama dalam metode penelitian ini adalah pengumpulan data yang diambil dari hasil crawling ulasan pengguna di Play Store terkait aplikasi Shopee (Puspitasari et al., 2023). Proses crawling dilakukan dengan menggunakan kata kunci seperti "Shopee Review", "Shopee User Experience", "Shopee Delivery Service", dan "Shopee Payment Issues". Data yang dikumpulkan mencakup ulasan dari periode pada jangka Desember 2024 sampai Maret 2025. Pengumpulan data pada penelitian ini diimplementasikan dengan menggunakan bahasa pemrograman Python dalam lingkungan Google Colab. Metode crawling yang digunakan tidak memanfaatkan API, namun memanfaatkan library Scrapy untuk mengumpulkan hingga 5.000 ulasan pengguna.

Dalam penelitian ini, data dikumpulkan dari ulasan pengguna aplikasi Shopee di Play Store. Proses pengumpulan data dilakukan menggunakan Python dengan bantuan library Scrapy sebagai alat untuk scraping data terkait opini pengguna terhadap aplikasi Shopee dalam jangka Desember 2024 sampai Maret 2025. Beberapa teknik diterapkan, seperti menghilangkan data duplikat untuk memastikan data yang diperoleh bersih dan relevan. Proses ini juga menyertakan langkah pemilihan atribut penting, seperti teks ulasan pengguna, untuk dianalisis lebih lanjut. Data yang telah terkumpul kemudian disimpan dalam format Excel untuk memudahkan pengolahan dan analisis pada tahap berikutnya.

E. Text Preprocessing

Text Preprocessing dilakukan untuk membentuk kumpulan data yang siap untuk dianalisis. Pada penelitian ini, dilakukan langkah preprocessing dengan empat langkah, yaitu (Fauzan et al., 2021):

a. Data Cleansing yaitu tahap data pada penelitian ini dibersihkan (cleansing) dari anomali semisal jejak tanda baca dan karakter-karakter yang tidak relevan. Proses cleansing ini dimulai dengan mengeliminasi simbol "@", tagar, karakter-karakter asing, dan sebagainya. Kemudian, seluruh huruf besar diubah menjadi huruf kecil (case folding)."
b. Tokenizing merupakan langkah fragmentasi dokumen atau kalimat menjadi unit-unit yang disebut token. Tahap tokenisasi ini menyegmentasi kalimat, kata, simbol, dan entitas penting lainnya.
c. Stopwords Remove merupakan leksikon yang tidak distingtif dalam dokumen. Contohnya, "pun", "akan", "di", "nya", "me", "oleh", "dan", dan lain sebagainya. Eliminasi stopword dilakukan untuk mengoptimalkan performa analisis sentimen.
d. Stemming merupakan teknik normalisasi teks dengan cara mereduksi kata menjadi bentuk dasarnya.

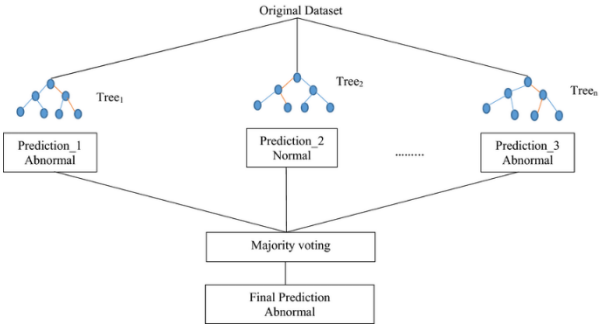
F. Pelabelan Data

Ada dua metode untuk pelabelan data yaitu secara manual dan sentiment menggunakan Bahasa pemrograman python.

a. Pelabelan Manual adalah proses mengklasifikasikan sentimen ulasan secara langsung (Wahyuningsih et al., 2022). Dari 100 ulasan Shopee yang diuji, 60 positif, 25 netral, dan 15 negatif ditemukan. Ini menunjukkan kepuasan pengguna terhadap pengiriman, kualitas produk, dan kemudahan transaksi, meski ada keluhan. Mayoritas positif mengindikasikan Shopee memenuhi ekspektasi belanja *online* pengguna.
b. Pelabelan Phyton digunakan alam penelitian ini dilakukan secara otomatis menggunakan kode Python, mengidentifikasi ulasan sebagai positif, netral, atau negatif[16]. Data ulasan yang telah diproses akan diolah oleh pustaka VADER Sentiment, yang mendukung bahasa Indonesia. Pendekatan berbasis leksikon VADER (Valence Aware Dictionary for Sentiment Reasoning) diterapkan untuk memberikan label. VADER tidak hanya mengukur polaritas, tetapi juga intensitas sentimen. Nilai sentimen VADER berkisar dari -0.05 hingga 0.05: ≥ 0.05 untuk positif, $=0$ untuk netral, dan ≤ -0.05 untuk negatif.

G. Klasifikasi Data

Random Forest diukur untuk optimalisasi analisis sentimen ulasan Shopee. Metode klasifikasi supervised learning ini membangun banyak pohon keputusan dari subset acak data dan fitur. Setiap pohon berkontribusi pada hasil klasifikasi akhir melalui mayoritas suara. Keunggulan Random Forest terletak pada kemampuannya menangani data kompleks, menghasilkan prediksi sentimen yang stabil dan akurat pada ulasan pengguna (Wahyuningsih et al., 2022).



Gambar 2 Random Forest

H. Evaluasi dan Validasi Data

Evaluasi kinerja model Random Forest dilakukan setelah pembagian data terstruktur, dengan mengukur akurasi, presisi, recall, dan F1-Score yang dihitung menggunakan confusion matrix (Fahmi & Sumarno, 2022). Nilai akurasi yang lebih tinggi menunjukkan bahwa algoritma tersebut efektif dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Shopee. Kinerja algoritma ini akan diuji untuk menentukan mana yang lebih baik dalam menganalisis sentimen positif, netral, atau negatif dalam konteks ulasan aplikasi Shopee. Metode evaluasi ini bertujuan untuk memilih algoritma yang paling sesuai untuk pengklasifikasian sentimen pengguna berdasarkan data yang ada. Hasil evaluasi ini akan menjadi dasar untuk menyimpulkan algoritma mana yang paling efektif dalam menganalisis sentimen dalam penelitian ini (Warhiyono et al., 2024).

$$Accuracy = TP + TNTP + FP = TN + FN \quad (3.1)$$

$$Precision = TPTP + FP \quad (3.2)$$

$$Recall = TPTP + FN \quad (3.3)$$

$$f1 - Score = 2TP2TP + FP + FN \quad (3.4)$$

III. HASIL DAN PEMBAHASAN

Hasil dan pembahasan analisis sentiment pada aplikasi Shopee menggunakan metode klasifikasi Random Forest sebagai berikut.

A. Scraping data

Pengumpulan data pada penelitian ini menggunakan API google-play-store dengan mengumpulkan sebanyak 5.000 data ulasan terkait aplikasi Shopee dengan menggunakan skrip berikut ini.

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

Gambar 3 Tahap Coding

Hasil output dari skrip tersebut dapat dilihat pada gambar 4.

reviewId	userName	userImage	content	score	thumbsUp	reviewCre	at
b2986e21	Pengguna G	https://play-lh	Tolong pencerahann	3	76	3.44.26	01/03/2025 08:42
a6580ef6	Pengguna G	https://play-lh	Tolong untuk mempe	3	106	3.44.26	28/02/2025 10:45
0db4d95e	Pengguna G	https://play-lh	Di Shopee banyak pil	4	21	3.44.26	01/03/2025 05:51
fe6851f7-3	Pengguna G	https://play-lh	Bertahun" pakai shop	1	174	3.44.26	28/02/2025 03:57
6b7ae29b	Pengguna G	https://play-lh	Awalnya suka banget	1	118	3.44.26	27/02/2025 22:46
9ab0fbba	Pengguna G	https://play-lh	Saran jangan terlalu b	3	40	3.44.26	28/02/2025 12:58
4c508e9f-(	Pengguna G	https://play-lh	Iklan yang muncul ter	1	49	3.44.26	28/02/2025 05:21
afd7f804-(	Pengguna G	https://play-lh	Kecewa sih. Keamana	1	39	3.44.17	27/02/2025 21:55
521ae359-	Pengguna G	https://play-lh	Tolonglah untuk kine	4	329	3.44.21	23/02/2025 20:54

Gambar 4 Hasil Coding

B. Text preprocessing

Selanjutnya merupakan proses preprocessing, pada tahap ini data yang telah terkumpul akan dilakukan pembersihan data melalui beberapa tahapan yaitu data cleansing, case folding, tokenizing, stopword remove dan stemming (Aditya et al., 2024).

1. Case Folding

Tahapan case folding bertujuan untuk mengkonversikan keseluruhan teks dari huruf kapital menjadi huruf kecil. Tahapan ini membutuhkan *library lower* untuk mengubah huruf kapital yang ada didokumen menjadi huruf kecil (Hidayat et al., 2023). Berikut hasil case folding terdapat pada gambar 5 berikut.

	content	score	clean_review
0	saya gak suka sama kebijakan shopee sekarang s...	1	saya gak suka sama kebijakan shopee sekarang s...
1	Pelayanan penjualnya rerata ramah dan baik, ta...	3	pelayanan penjualnya rerata ramah dan baik, ta...
2	Aplikasinya bagus..cuma agak berat..kalo pas b...	3	aplikasinya bagus..cuma agak berat..kalo pas b...
3	Tidak ada pilihan lagi untuk memilih ekspedisi...	3	tidak ada pilihan lagi untuk memilih ekspedisi...
4	Makin kesini makin kesana Knapa setiap kali ma...	3	makin kesini makin kesana knapa setiap kali ma...

Gambar 5 Case Folding

2. Data Cleansing

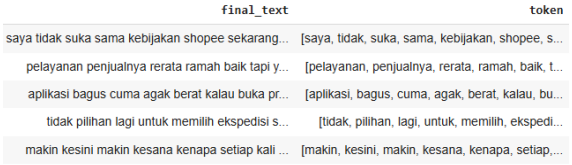
Pada tahapan ini dilakukan penghapusan elemen-elemen yang tidak relevan seperti symbol dan emotikon. Pada tahapan cleansing ini adalah untuk menghapus karakter yang tidak memberikan pengaruh terhadap proses klasifikasi sentiment seperti menghapus tanda baca koma (,), titik (.), hastag (#), mention (@), link dan angka (Erkamim et al., 2023). Tahap ini bertujuan untuk meningkatkan kualitas data dan membersihkan data dari noise[10]. Berikut hasil dari proses data cleansing dapat dilihat pada gambar 6

	content	score	clean_review
0	saya gak suka sama kebijakan shopee sekarang s...	1	saya gak suka sama kebijakan shopee sekarang s...
1	Pelayanan penjualnya rerata ramah dan baik, ta...	3	pelayanan penjualnya rerata ramah dan baik tap...
2	Aplikasinya bagus..cuma agak berat..kalo pas b...	3	aplikasinya bagus cuma agak berat kalo pas buk...
3	Tidak ada pilihan lagi untuk memilih ekspedisi...	3	tidak ada pilihan lagi untuk memilih ekspedisi...
4	Makin kesini makin kesana Knapa setiap kali ma...	3	makin kesini makin kesana knapa setiap kali ma...

Gambar 6 Data Cleansing

3. Tokenizing

Tahapan tokenizing adalah metode untuk melakukan pemisahan kata dalam suatu kalimat dengan tujuan untuk proses analisis teks lebih lanjut (Bin et al., 2020). Pada gambar 7 dibawah ini merupakan hasil dari tahapan tokenizing.

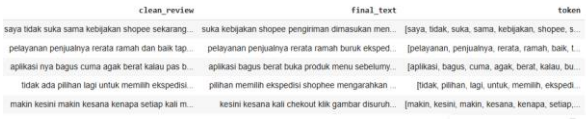


final_text	token
saya tidak suka sama kebijakan shopee sekarang...	[saya, tidak, suka, sama, kebijakan, shopee, s...
pelayanan penjualnya rerata ramah baik tapi y...	[pelayanan, penjualnya, rerata, ramah, baik, t...
aplikasi bagus cuma agak berat kalau buka pr...	[aplikasi, bagus, cuma, agak, berat, kalau, bu...
tidak pilihan lagi untuk memilih ekspedisi s...	[tidak, pilihan, lagi, untuk, memilih, ekspedi...
makin kesini makin kesana kenapa setiap kali ...	[makin, kesini, makin, kesana, kenapa, setiap,...

Gambar 7 Tokenizing

4. Stopword Remove

Berikut merupakan tahapan stopword yaitu membersihkan kata penghubung contohnya saya, lagi, dan, aku, yang, dan imbuhan serta kata sambung lainnya (Dikiyanti et al., 2021). Berikut hasil dari tahapan stopword remove dapat dilihat pada gambar 8 dibawah ini.

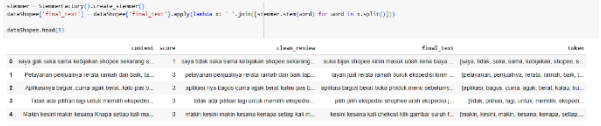


clean_review	final_text	token
saya tidak suka sama kebijakan shopee sekarang...	suka kebijakan shopee pengiriman dimasukan men...	[saya, tidak, suka, sama, kebijakan, shopee, s...
pelayanan penjualnya rerata ramah dan baik tap...	pelayanan penjualnya rerata ramah buruk eksped...	[pelayanan, penjualnya, rerata, ramah, baik, t...
aplikasi nya bagus cuma agak berat kalau pas b...	aplikasi bagus berat buka produk menu sebeluny...	[aplikasi, bagus, cuma, agak, berat, kalau, bu...
tidak ada pilihan lagi untuk memilih ekspedisi...	pilihan memilih ekspedisi shopee mengarahkan ...	[tidak, pilihan, lagi, untuk, memilih, ekspedi...
makin kesini makin kesana kenapa setiap kali m...	kesini kesana kali chekout klik gambar disuruh...	[makin, kesini, makin, kesana, kenapa, setiap,...

Gambar 8 Stopword Remove

5. Stemming

Tahapan Stemming adalah suatu proses pengembalian suatu kata berimbuhan ke dalam bentuk dasarnya (Ardhani et al., 2021). Proses ini menghilangkan awalan, akhiran, sisipan dan confixes (kombinasi dari awalan dan akhiran). Stemming yang digunakan pada proses ini adalah stemming Sastrawi, yang merupakan stemmer pengembangan dari Algoirtma Nazief dan Adriani. Sastrawi sangat bergantung pada kamus kata dasar yang diambil dari kateglo.com dengan perubahan (Khairunnisa et al., 2021).



```
stemmer = StemmerFactory().create_stemmer()
dataShopee['final_text'] = dataShopee['final_text'].apply(lambda x: ' '.join([stemmer.stem(word) for word in x.split()]))

dataShopee.head(5)
```

	content	score	clean_review	final_text	token
0	saya gak suka sama kebijakan shopee sekarang s...	1	saya tidak suka sama kebijakan shopee sekarang...	suka bijak shopee kirim masuk oleh lena biaya ...	[saya, tidak, suka, sama, kebijakan, shopee, s...
1	Pelayanan penjualnya rerata ramah dan baik, ta...	3	pelayanan penjualnya rerata ramah dan baik tap...	layan jual rerata ramah buruk ekspedisi kirim ...	[pelayanan, penjualnya, rerata, ramah, baik, t...
2	Aplikasinya bagus..cuma agak berat..kalo pas b...	3	aplikasi nya bagus cuma agak berat kalau pas b...	aplikasi bagus berat buka produk menu sebeluny...	[aplikasi, bagus, cuma, agak, berat, kalau, bu...
3	Tidak ada pilihan lagi untuk memilih ekspedisi...	3	tidak ada pilihan lagi untuk memilih ekspedisi...	pilih pilih ekspedisi shopee arah ekspedisi j...	[tidak, pilihan, lagi, untuk, memilih, ekspedi...
4	Makin kesini makin kesana Knapa setiap kali ma...	3	makin kesini makin kesana kenapa setiap kali m...	kesini kesana kali chekout klik gambar suruh t...	[makin, kesini, makin, kesana, kenapa, setiap,...

Gambar 9 Stemming

C. Pelabelan Data

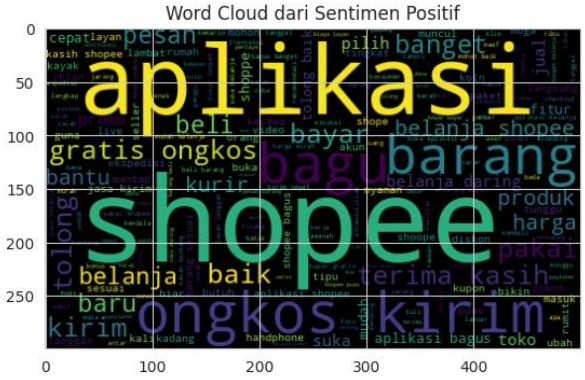
Ulasan masih belum mempunyai sentimen sehingga sulit untuk mencari tahu apakah pengguna memberikan ulasan positif atau negatif (Mukarromah et al., 2021). Proses pemberian sentimen tidak mungkin dilakukan secara manual dengan melihat ulasan secara satu per satu karena membutuhkan waktu yang lama dan memerlukan seorang ahli di bidang bahasa yang dapat menafsirkan ulasan kemudian mengelompokkan ke sentimen positif dan negatif (Dwiki et al., 2021). Maka dari itu, setelah melalui proses text preprosesing kemudian dilakukan tahapan berikutnya yaitu melakukan proses pemberian label sentimen dengan menerapkan metode yang berbasis lexicon atau biasa dikenal dengan lexicon-based method (Hidayati et al., 2021). Kamus yang digunakan adalah InSet Lexicon yang terdiri dari kamus positif dan negatif. Berikut data ditunjukkan pada gambar dibawah ini.

final_text	token	sentimen
suka bijak shopee kirim masuk ubah kena biaya ...	[saya, tidak, suka, sama, kebijakan, shopee, s...	Negatif
layan jual rerata ramah buruk ekspedisi kirim ...	[pelayanan, penjualnya, rerata, ramah, baik, t...	Negatif
aplikasi bagus berat buka produk menu sebelumy...	[aplikasi, bagus, cuma, agak, berat, kalau, bu...	Positif
pilih pilih ekspedisi shophee arah ekspedisi j...	[tidak, pilihan, lagi, untuk, memilih, ekspedi...	Negatif
kesini kesana kali chekout klik gambar suruh f...	[makin, kesini, makin, kesana, kenapa, setiap,...	Positif
...	...	...
akuu suka banget pokok nyaa sukakk banget apli...	[akuu, suka, banget, pokok, nyaa, sukakk, bang...	Positif
bagus rumit cari toko offline cari barang bara...	[bagus, jadi, kita, tidak, perlu, rumit, cari,...	Netral
belanja shoppe muas barg kirim sesuai kirim on...	[belanja, shoppe, sangat, memuaskan, karena, b...	Negatif
shopee keren diskon mudah mudah tingkat diskon...	[sekarang, shopee, keren, banyak, diskonnya, m...	Netral
buka aplikasi shope handphone lelet kayak lamb...	[tiap, buka, aplikasi, shope, pasti, handphone...	Negatif

Gambar 10 Pelabelan Data

D. Visualisasi Data

Pada pengujian ini menggunakan ulasan aplikasi Shopee yang telah dikelompokkan menjadi 3 kelas sentiment yaitu positif, negatif dan netral bertujuan untuk mencari kata yang sering muncul Afdal & Elita, 2022). Dari 5.000 data ulasan, ulasan yang mengandung sentimen positif sebanyak 1.470 ulasan, sentimen negative sebanyak 2.539 ulasan dan sentimen netral sebanyak 991 ulasan. Analisis ini dilakukan menggunakan wordcloud agar dapat menampilkan kata yang sering muncul pada ulasan (Laurenz & Sediyoni, 2021). Pada sentimen positif kata yang sering muncul pada dilihat pada gambar 11.

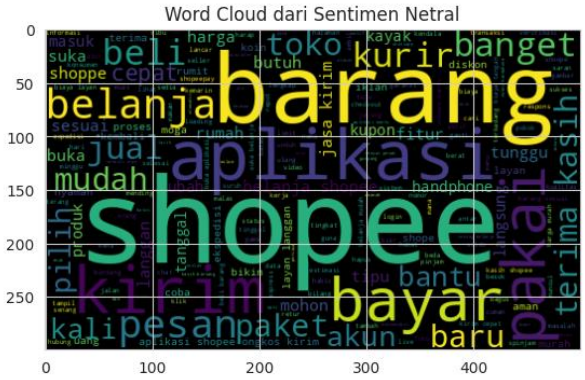


Gambar 11 Sentimen Positif

Pada Gambar 11 didapatkan beberapa kata yang sering muncul yaitu “shopee”, “aplikasi”, “gratis”, “ongkos”, “barang”, “kirim”, “suka”, “bagus”, belajar”, dan “beli”. Kata tersebut merupakan kata yang sering digunakan dalam memberikan ulasan pengguna terkait aplikasi shopee di *google play store*. Hal ini mengartikan bahwa pengguna merasa puas terhadap pelayanan aplikasi shopee dikarenakan memberikan kemudahan, gratis ongkir dan membantu pengguna dalam belanja barang kebutuhan. Selanjutnya, kata yang sering muncul pada kategori negative dapat dilihat pada Gambar 12

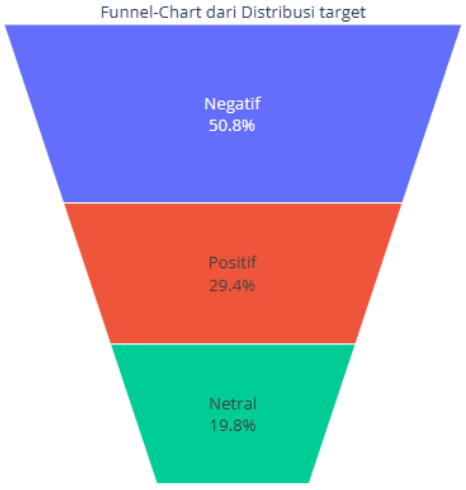
Gambar 12 Sentimen Negatif

Pada Gambar 12 terdapat beberapa kata yang sering muncul yaitu “shopee”, “barang”, “kirim”, “aplikasi”, “kecewa”, “tolong”, “lambat”, “pesan”, “kurir”, “buruk”, “rugi” dan “paket”. Hal ini menunjukan bahwa kata “aplikasi” pada ulasa negatif memberikan informasi tentang keluhan pengguna terhadap aplikasi Shopee yang berat karena aplikasi ini memerlukan memori yang besar, aplikasi lambat saat mengaksesnya meskipun internet lancar dirasa karena masih berhubungan dengan memori tadi, serta aplikasi bermasalah dan sering logout dirasa karena terdapat masalah dari sistem aplikasi Shopee itu sendiri. Lalu, kata ulasan yang sering muncul pada kategori netral dapat dilihat pada Gambar 13



Gambar 13 Sentimen Netral

Pada Gambar 13 terdapat beberapa kata yang sering muncul yaitu “shopee”, “barang”, “bayar, “pakai”, “aplikasi”, “belanja”, “pesan” ,“kurir”, “mudah”, “kirim”, “banget” dan “cepat” menunjukan bahwa pengguna sering bertanya atau memberikan masukan dan saran terkait beberapa kata tersebut. Sedangkan kata “mudah” merepresentasikan bahwa pengguna merasa mudah dalam mengakses aplikasi shopee.



Gambar 14 Grafik Sentimen Netral

E. Pembobotan TF-IDF

Setelah melalui proses preprocessing, data kemudian ditransformasikan menjadi representasi vector dengan menggunakan metode TF-IDF, representasi vector yang dihasilkan dari transformasi tersebut akan diproses oleh model. Nilai TF-IDF dari sebuah kata merupakan kombinasi dari nilai *tf* dan nilai *idf* dalam perhitungan bobot metode TF-IDF ini menggabungkan dua konsep yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen dan inverse frekuensi dokumen yang mengandung kata tersebut. Berikut merupakan script dari metode TF-IDF pada Gambar 15

```
# TF-IDF
from sklearn.feature_extraction.text import TfidfVectorizer

tfidf_vectorizer = TfidfVectorizer()
X_tfidf = tfidf_vectorizer.fit_transform(dataShopee['final_text'])
```

Gambar 15 TF-IDF

F. Algoritma Random Forest

Pada tahapan klasifikasi Random Forest ini dilakukan pembagian data, yaitu data training (latih) dan data testing (test). Perbandingan antara data testing dan training adalah 80:20. Tahap selanjutnya melakukan klasifikasi sentiment menggunakan Random Forest. Pada tahapan pengujian menggunakan confusion matrix untuk mengetahui informasi tentang nilai-nilai yang diprediksi dan hasil yang sebenarnya, biasa digunakan untuk perhitungan accuracy, recall, precision, dan f1-score[28]. Confusion matrix menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji salah diklasifikasikan. Berikut hasil pengujian menggunakan confusion matrix dapat dilihat pada Gambar 16

```
Classification Report for Random Forest (Tuned):
              precision    recall  f1-score   support

     Negatif       0.91      0.93      0.92       499
      Netral       0.84      0.89      0.86       194
     Positif       0.90      0.84      0.87       307

    accuracy                           0.89      1000
   macro avg       0.88      0.88      0.88      1000
weighted avg       0.89      0.89      0.89      1000
```

Gambar 16 Algoritma Random Forest

Dari hasil penelitian, disimpulkan bahwa model ini dapat melakukan analisis sentimen pada dataset model ini dapat melakukan analisis sentiment pada dataset ulasan pengguna aplikasi Shopee dengan akurasi sebesar 88%. Hal ini menandakan bahwa menggunakan algoritma random forest dapat menghasilkan hasil analisis sentiment yang baik.

G. Evaluasi Model

Untuk mengevaluasi model yang telah di buat, digunakan metode *confusion matrix* dan *cross validation score*. Confusion matrix digunakan untuk menunjukkan jumlah hasil klasifikasi yang benar dan salah dari model. Sedangkan, *cross validation score* digunakan untuk mengukur seberapa baik model yang dibuat dalam memprediksi data yang belum dilihat sebelumnya (Hidayati et al., 2021).

H. Visualisasi

Pada penelitian ini, dilakukan visualisasi data menggunakan *wordcloud* untuk menganalisis frekuensi tentang kata-kata yang paling dominan dalam ulasan pengguna, sehingga memudahkan pemahaman mengenai preferensi dan pengalaman pengguna terkait dengan pengguna e-commerce (Kurniasih & Seseno, 2022). Visualisasi *wordcloud* ini membantu dalam mengidentifikasi kata-ata kunci yang paling signifikan dan memperoleh wawasan yang berguna dalam analisis sentiment pada dataset tersebut.

VII. SIMPULAN

Dari hasil penelitian ini, dapat disimpulkan bahwa metode klasifikasi Random Forest mampu menghasilkan nilai akurasi yang cukup tinggi, yaitu sebesar 88%. Meskipun begitu, akurasi ini masih bisa ditingkatkan lebih lanjut dengan menggunakan data yang lebih baik atau lebih bersih. Peningkatan akurasi maksimum juga dapat dicapai dengan menambahkan lebih banyak data ke dalam set pelatihan atau dengan mencoba metode klasifikasi lain. Data yang digunakan dalam penelitian ini terdiri dari 5.000 komentar yang dikumpulkan dari aplikasi Shopee di Play Store.

Adapun saran mengenai ide untuk memperluas penelitian yang serupa adalah penentuan data training dapat mempengaruhi hasil klasifikasi, maka dari itu untuk penelitian diharapkan menambahkan atribut atau data training

yang lebih lengkap karena penentuan tingkat akurat dapat dibentuk oleh data training. Karena data training dapat mempengaruhi hasil klasifikasi. Selain itu, pada saat pelabelan data juga ketika menggunakan vader sentiment juga mempunyai pengaruh ketika kata dalam bahasa Indonesia diberi palabelan dalam menyatakan sebuah sentimen pada kalimat.

UCAPAN TERIMA KASIH

Saya mengucapkan terima kasih yang sebesar-besarnya kepada Dr. Mochammad Alfan Rosid, S.Kom., M.Kom., Dr. Ade Eviyanti, Skom Mkom, selaku pembimbing utama, atas bimbingan, masukan, dan dukungannya dalam menyelesaikan penelitian ini. Penelitian ini didukung oleh kedua orang tua saya dan Universitas Muhammadiyah Sidoarjo melalui program tugas akhir 2025. Kami sangat berterima kasih atas segala dukungan yang diberikan.

REFERENSI

[1] Rizky, I. and Mahfudz, A. (2022) 'Sebagai Variabel Intervening (Studi Pada Pengguna Shopee Di Kota Semarang)', Diponegoro J. Manag., 11(1), pp. 1–13. Tersedia pada: https://ejournal3.undip.ac.id/index.php/djom/index.

[2] Putra, R.F. and Sumarno (2022) 'Aplikasi Sistem Pakar Perencanaan Investasi Pasar Modal Menggunakan Metode Forward Chaining Berbasis Website', Procedia Eng. Life Sci., 3(December).

[3] Arsi, P., Wahyudi, R. and Waluyo, R. (2021) 'Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacana Pindah Ibu Kota Indonesia', J. RESTI (Rekayasa Sist. dan Teknol. Informasi), 5(2), pp. 231–237. doi: 10.29207/resti.v5i2.2698.

[4] Amaliah, S., Nusrang, M. and Aswi, A. (2022) 'Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng', VARIANSI J. Stat. Its Appl. Teach. Res., 4(3), pp. 121–127. doi: 10.35580/variansiunm31.

[5] Gifari, O.I. et al. (2022) 'Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine', J. Inf. Technol., 2(1), pp. 36–40. doi: 10.46229/jifotech.v2i1.330.

[6] Mahmuda, S. (2024) 'Implementasi Metode Random Forest pada Kategori Konten Kanal Youtube', J. Jendela Mat., 2(01), pp. 21–31. doi: 10.57008/jjm.v2i01.633.

[7] Vonega, D.A., Fadila, A. and Kurniawan, D.E. (2022) 'Analisis Sentimen Twitter Terhadap Opini Publik Atas Isu Pencalonan Puan Maharani dalam PILPRES 2024', J. Appl. Informatics Comput., 6(2), pp. 129–135. doi: 10.30871/jaic.v6i2.4300.

[8] Wardani, N.S., Prahutama, A. and Kartikasari, P. (2020) 'Analisis Sentimen Pemindahan Ibu Kota Negara Dengan Klasifikasi Naïve Bayes Untuk Model Bernoulli Dan Multinomial', J. Gaussian, 9(3), pp. 237–246. doi: 10.14710/j.gauss.v9i3.27963.

[9] Kurniawan, I. and Susanto, A. (2019) 'Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019', Eksplora Inform., 9(1), pp. 1–10. doi: 10.30864/eksplora.v9i1.237.

[10] Alvanof, M.M. and Dinata, R.K. (2024) 'Penerapan Algoritma Random Forest dalam Deteksi dan Klasifikasi Ransomware', 5(2).

[11] Siregar, A.P. et al. (2023) 'Implementasi Algoritma Random Forest Dalam Klasifikasi Diagnosis Penyakit Stroke', J. Penelit. Rumpun Ilmu Tek., 2(4), pp. 155–164. doi: 10.55606/juprit.v2i4.3039.

[12] Erkamim, M. et al. (2023) 'Komparasi Algoritme Random Forest dan XGBoosting dalam Klasifikasi Performa UMKM', J. Sist. Inf. Bisnis, 13(2), pp. 127–134. doi: 10.21456/vol13iss2pp127-134.

[13] Puspitasari, R., Findawati, Y. and Rosid, M.A. (2023) 'Sentiment Analysis of Post-Covid-19 Inflation Based on Twitter Using the K-Nearest Neighbor and Support Vector Machine Classification Methods', J. Tek. Inform., 4(4), pp. 669–679. doi: 10.52436/1.jutif.2023.4.4.801.

[14] Fauzan, A. et al. (2021) 'Pengembangan Aplikasi Virtual Tour sebagai Media Pengenalan Lingkungan Kampus PENS berbasis Website', J. Teknol. Terpadu, 7(1), pp. 23–30. doi: 10.54914/jtt.v7i1.341.

[15] Wahyuningtias, P. et al. (2022) 'COMPARISON OF RANDOM FOREST AND SUPPORT VECTOR MACHINE METHODS ON TWITTER SENTIMENT ANALYSIS (CASE STUDY : INTERNET SELEBGRAM RACHEL VENNYA ESCAPE FROM QUARANTINE) PERBANDINGAN METODE RANDOM FOREST DAN SUPPORT VECTOR MACHINE PADA ANALISIS SENTIMEN TWITT', J. Tek. Inform., 3(1), pp. 141–145.

[16] Mukarramah, R., Atmajaya, D. and Ilmawan, L.B. (2021) 'Performance comparison of support vector machine (SVM) with linear kernel and polynomial kernel for multiclass sentiment analysis on twitter', Ilk. J. Ilm., 13(2), pp. 168–174. doi: 10.33096/ilkom.v13i2.851.168-174.

[17] Warjiyono et al. (2024) 'Analisa Prediksi Harga Jual Rumah Menggunakan Algoritma Random Forest Machine Learning', JURSISTEKNI (Jurnal Sist. Inf. dan Teknol. Informasi), 6(2), pp. 416–423.

[18] Aditya, M.F.R., Lutvi, N. and Indahyanti, U. (2024) 'Prediksi Penyakit Hipertensi Menggunakan Metode Decison Tree dan Random Forest', J. Ilm. Komputasi, 23(1), pp. 9–16. doi: 10.32409/jikstik.23.1.3503.
[19] Hidayat, H., Sunyoto, A. and Al Fatta, H. (2023) 'Klasifikasi Penyakit Jantung Menggunakan Random Forest Clasifier', J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan), 7(1), pp. 31–40. doi: 10.47970/siskom-kb.v7i1.464.
[20] Bin, A., Alkatiri, M. and Nasution, A.N.S. (2020) 'OPINI PUBLIK TERHADAP PENERAPAN NEW NORMAL DI MEDIA SOSIAL TWITTER', J. Strateg. Commun., 11(1).
[21] Dikiyanti, T.D., Rukmi, A.M. and Irawan, M.I. (2021) 'Sentiment analysis and topic modeling of BPJS Kesehatan based on twitter crawling data using Indonesian Sentiment Lexicon and Latent Dirichlet Allocation algorithm', J. Phys. Conf. Ser., 1821(1). doi: 10.1088/1742-6596/1821/1/012054.
[22] Ardhani, B.A., Chamidah, N. and Saifudin, T. (2021) 'Sentiment Analysis Towards Kartu Prakerja Using Text Mining with Support Vector Machine and Radial Basis Function Kernel', J. Inf. Syst. Eng. Bus. Intell., 7(2), p. 119. doi: 10.20473/jisebi.7.2.119-128.
[23] Khairunnisa, S., Adiwijaya, A. and Al Faraby, S. (2021) 'Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)', J. Media Inform. Budidarma, 5(2), p. 406. doi: 10.30865/mib.v5i2.2835.
[24] Dwiki, A. et al. (2021) 'Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN', J. Tek. Inform. dan Sist. Inf., 8(2), pp. 636–646.
[25] Hidayati, N., Suntoro, J. and Setiaji, G.G. (2021) 'Perbandingan Algoritma Klasifikasi untuk Prediksi Cacat Software dengan Pendekatan CRISP-DM', J. Sains dan Inform., 7(2), pp. 117–126. doi: 10.34128/jsi.v7i2.313.
[26] Afdal, M. and Elita, L.R. (2022) 'Penerapan Text Mining Pada Aplikasi Tokopedia Menggunakan Algoritma K-Nearest Neighbor', J. Ilm. Rekayasa dan Manaj. Sist. Inf., 8(1), pp. 78–87.
[27] Laurensz, B. and Sediyono, E. (2021) 'Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19', J. Nas. Tek. Elektro dan Teknol. Inf., 10(2), pp. 118–123. doi: 10.22146/jnteti.v10i2.1421.
[28] Pamungkas, B., Purbaya, M.E. and K, D.J.A. (2021) 'Analisis Sentimen Twitter Menggunakan Metode Support Vector Machine (SVM) pada', J. Informatics, Inf. Syst. Softw. Eng. Appl., 3(2), pp. 10–20.
[29] Kurniasih, U. and Suseno, A.T. (2022) 'Analisis Sentimen Terhadap Bantuan Subsidi Upah (BSU) pada Kenaikan Harga Bahan Bakar Minyak (BBM)', J. Media Inform. Budidarma, 6(4), pp. 2335–2340. doi: 10.30865/mib.v6i4.4958.
[30] Herdiyani, T.C. and Zailani, A.U. (2022) 'Sentiment Analysis Terkait Pemindahan Ibu Kota Indonesia Menggunakan Metode Random Forest Berdasarkan Tweet Warga Negara Indonesia', J. Teknol. Sist. Inf., 3(2), pp. 154–165. doi: 10.35957/jtsi.v3i2.2920.
[31] Rania, N.Z. and Syah, R.D. (2024) 'Analisis Sentimen Terhadap Aplikasi Gojek Pada Play Store Menggunakan Metode Random Forest Classifier', J. Ilm. Inform. Komput., 29(2), pp. 144–153. doi: 10.35760/ik.2024.v29i2.11877.
[32] Alita, D. and Isnain, A.R. (2020) 'Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier', J. Komputasi, 8(2), pp. 50–58. doi: 10.23960/komputasi.v8i2.2615.

Conflict of Interest Statement:
The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.