

PENERAPAN METODE K-NEAREST NEIGHBOR (KNN) DALAM ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI SHOPEE PADA GOOGLE PLAY STORE

Fauzia Win Salsabila¹, Tumini², Agmawarnida³

fauziawinsalsabila@gmail.com¹, kemuningijo@gmail.com², agmawarnida2013@gmail.com³

Universitas Panca Sakti Bekasi

ABSTRAK

Kemajuan teknologi telah mendorong pertumbuhan pengguna e-commerce, termasuk Shopee yang saat ini menjadi salah satu platform belanja online terbesar di Indonesia. Banyaknya ulasan pengguna pada aplikasi Shopee di Google Play Store menghasilkan kumpulan pendapat dalam jumlah besar. Pendapat-pendapat tersebut dapat menjadi bahan pertimbangan penting, baik bagi perusahaan dalam meningkatkan kualitas layanan maupun bagi masyarakat dalam mengambil keputusan berbelanja. Namun, banyaknya jumlah ulasan membuat proses analisis secara manual menjadi kurang efektif dan memakan waktu. Karena itu, penelitian ini menggunakan cara otomatis untuk mengetahui apakah komentar pengguna Shopee cenderung positif, negatif atau netral. Metode yang digunakan adalah algoritma K-Nearest Neighbor (KNN) dengan data analisis sebanyak 100.000 ulasan. Tahapan penelitian ini dilakukan melalui beberapa tahapan. Pertama, data ulasan pengguna Shopee diambil dari Kaggle. Kedua, dilakukan preprocessing teks yang meliputi pembersihan data, normalisasi, tokenisasi, penghapusan kata tidak penting (stopword removal), dan stemming. Selanjutnya, kata-kata diberi bobot menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Proses klasifikasi dilakukan dengan algoritma K-Nearest Neighbor (KNN) menggunakan beberapa variasi nilai k. Hasil penelitian menunjukkan bahwa sebagian besar ulasan bersifat positif 77,2%, sedangkan ulasan negatif mencapai 18,5% dan ulasan netral sebesar 4,3%. Model KNN konsisten menghasilkan akurasi sebesar 77% pada berbagai skenario pembagian data latih dan data uji. Performa terbaik ditunjukkan pada ulasan positif dengan nilai precision sebesar 0,79, recall 0,95-0,96 dan f1-score 0,87. Namun, performa pada ulasan negatif dan netral masih rendah karena adanya ketidak seimbangan jumlah data. Kesimpulan dari penelitian ini adalah metode K-Nearest Neighbor (KNN) cukup baik untuk mengenali ulasan positif, tetapi masih perlu perbaikan agar bisa lebih akurat membedakan ulasan negatif dan netral. Hasil ini bisa membantu pengembang aplikasi dalam menyempurnakan sistem analisis, serta menjadi bahan pertimbangan bagi pelaku bisnis e-commerce untuk meningkatkan layanan.

Kata Kunci: Analisis Sentimen, Shopee, K-Nearest Neighbor, TF-IDF, Google Play.

ABSTRACT

Technological advancements have driven the growth of e-commerce users, including Shopee, which is currently one of the largest online shopping platforms in Indonesia. The large number of user reviews on the Shopee application in the Google Play Store generates a vast collection of opinions. These opinions can serve as important considerations, both for companies in improving service quality and for the public in making purchasing decisions. However, the large volume of reviews makes manual analysis inefficient and time-consuming. Therefore, this study applies an automated approach to determine whether Shopee user comments tend to be positive, negative, or neutral. The method used is the K-Nearest Neighbor (KNN) algorithm with a dataset of 100,000 reviews. The research was conducted through several stages. First, Shopee user review data was obtained from Kaggle. Second, text preprocessing was carried out, including data cleaning, normalization, tokenization, stopword removal, and stemming. Next, words were weighted using the Term Frequency-Inverse Document Frequency (TF-IDF) method. The classification process was performed using the K-Nearest Neighbor (KNN) algorithm with several variations of the k value. The results of the study show that most of the reviews were positive (77.2%), while negative reviews accounted for 18.5% and neutral reviews 4.3%. The KNN model consistently achieved an accuracy

of 77% across various scenarios of training and testing data splits. The best performance was shown in positive reviews with a precision of 0.79, recall of 0.95–0.96, and an f1-score of 0.87. However, the performance on negative and neutral reviews remained low due to data imbalance. The conclusion of this study is that the K-Nearest Neighbor (KNN) method is fairly effective in recognizing positive reviews, but further improvements are needed to achieve higher accuracy in distinguishing negative and neutral reviews. These findings can assist application developers in enhancing automated analysis systems and serve as a consideration for e-commerce businesses in improving service quality.

Keywords: Sentiment Analysis, Shopee, K-Nearest Neighbor, TF-IDF, Google Play.

PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah mendorong pertumbuhan industri di berbagai aspek kehidupan, salah satunya adalah dalam kegiatan jual beli. *E-commerce* menjadi salah satu sektor berkembangnya teknologi industri dalam bertransformasi secara digital karena memberikan kemudahan dalam melakukan pembelian secara online. Masyarakat tidak lagi dibatasi oleh lokasi fisik dalam membeli barang atau jasa, sehingga mereka dapat mengakses berbagai produk hanya melalui aplikasi smartphone. Salah satu *e-commerce* yang mengalami pertumbuhan signifikan salah satunya adalah Shopee.

Shopee didirikan oleh Sea Group pada tahun 2015, Shopee telah berhasil menjadi pemimpin *platform e-commerce* di Asia Tenggara, termasuk di Indonesia. Shopee banyak menarik pengguna setiap harinya dikarenakan memiliki banyak fitur seperti fitur gratis ongkos kirim, *voucher* belanja, shopeepay dan shopee memiliki *event* di setiap bulanya seperti tanggal kembar yaitu *event* 11.11 dan 12.12, banyaknya diskon yang ditawarkan Shopee, membuat para pengguna disukai oleh semua kalangan. Berdasarkan data dari berbagai sumber seperti iPrice dan similar *website*, Shopee secara konsisten menjadi *platform* belanja online dengan jumlah terbanyak di Indonesia.

Dalam *e-commerce* ulasan pengguna atau *review* memainkan peranan penting. Ulasan menjadi acuan utama bagi calon pembeli untuk mengevaluasi kualitas produk, layanan penjualan, dan pengalaman belanja secara keseluruhan. Beberapa riset menunjukkan bahwa banyak sekali keputusan pembelian konsumen dipengaruhi oleh pendapat pengguna lain. Oleh karena itu, hasil ulasan menjadi aset berharga dalam memberikan *insight* langsung dari konsumen. Namun, semakin bertambahnya jumlah pengguna dan transaksi, jumlah ulasan yang masuk ke *platform* shopee juga semakin bertambah besar. Sehingga, menganalisis masing-masing ulasan secara manual sangatlah tidak efisien, memakan waktu dan rentan bias. Oleh karena itu, diperlukan sistem otomatis yang mampu mengklasifikasikan ulasan-ulasan ke dalam kategori sentimen seperti positif, negatif dan netral.

Secara umum, teknik klasifikasi teks memiliki beberapa opsi, seperti Algoritma *K-Nearest Neighbor*, *Naive Bayes*, dan *Support Vector Machine*. Analisis sentimen adalah metode otomatis pemrosesan data teks untuk mengidentifikasi sudut pandang atau pendapat mengenai subjek atau objek tertentu, seperti orang, kelompok, produk, atau apa pun dalam kumpulan data yang diteliti. Analisis adalah teknik untuk mengidentifikasi, mengekstrak, atau mendeteksi informasi emosional dari teks, terutama dalam konteks opini atau sentimen pengguna. Tujuan utama dari analisis ini adalah untuk memahami perasaan atau pendapat yang diungkapkan oleh pengguna melalui ulasan mereka di Shopee. Salah satu solusi yang mungkin dapat digunakan adalah menggunakan metode klasifikasi berbasis teks, yaitu analisis sentimen. Dalam penelitian ini, metode yang digunakan adalah *K-Nearest Neighbor* (KNN). Menggunakan jarak terdekat atau kemiripan dengan item, pendekatan *K-Nearest Neighbor* menggunakan data latih untuk mengklasifikasikan objek. Dalam penggunaannya, algoritma KNN memiliki beberapa kelebihan, di antaranya sederhana dan mudah dipahami, non-parametrik, mudah dalam penyesuaian model serta tahan terhadap *noise*. KNN

merupakan algoritma klasifikasi yang bekerja berdasarkan kedekatan jarak antara data yang di uji dengan data yang sudah di labeli. Meskipun termasuk algoritma sederhana, KNN terbukti sangat efektif pada teks pendek seperti ulasan pengguna.

Tujuan penelitian ini adalah menerapkan metode KNN untuk menganalisis sentimen pengguna aplikasi Shopee yang tersedia di Google Play Store. Hasil dari penelitian ini diharapkan dapat memberikan wawasan umum terhadap persepsi pengguna dan menjadi dasar pengambilan keputusan bagi pengembang aplikasi dan pelaku bisnis.

METODE

Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksperimen klasifikasi teks. Tahapan penelitian ini mencakup pengumpulan data ulasan aplikasi Shopee, dilanjutkan dengan *preprocessing* data, ekstraksi fitur menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF), serta penerapan algoritma *K-Nearest Neighbor* (KNN) untuk proses klasifikasi. Kinerja model dievaluasi menggunakan sejumlah metrik pengukurannya yaitu akurasi, *precision*, *recall*, dan *f1-score*. Selanjutnya dilakukan visualisasi dan asosiasi kata.

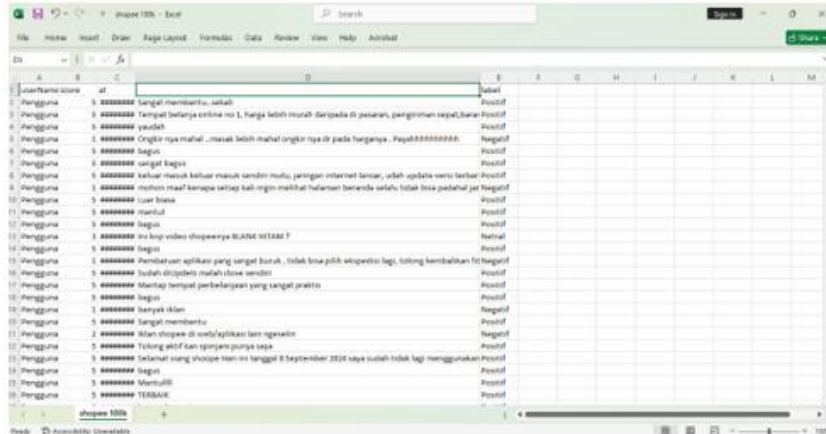
HASIL DAN PEMBAHASAN

Penarikan Data

Penarikan data dalam penelitian ini tidak dilakukan dengan teknik *web scraping*, melainkan menggunakan dataset sekunder yang telah tersedia pada platform Kaggle. Dataset tersebut berisi data aplikasi Shopee dari Google Play Store dalam format file *csv*. Setelah di unduh dari Kaggle dataset diunggah ke GitHub agar dapat diakses secara publik. Selanjutnya, tautan *raw* file dari GitHub dicantumkan di Google Colab sehingga dataset dapat diproses lebih lanjut menggunakan bahasa pemrograman Python dengan bantuan Pustaka *pandas*.

Langkah-langkah pengolahan data yang dilakukan adalah sebagai berikut :

1. Mengunduh dataset dari Kaggle aplikasi Shopee dari Google Play Store diperoleh melalui Kaggle dalam format *csv*.
2. Mengunggah dataset ke GitHub file *csv* yang telah diunduh dari Kaggle di unggah ke repository GitHub sehingga dapat diakses menggunakan tautan *raw* file dari GitHub.
3. Mengimpor Pustaka Python yang diperlukan pustaka *pandas* digunakan dalam penelitian ini untuk membaca, menampilkan, dan melakukan manipulasi data.
4. Membaca dataset dari GitHub menggunakan *pandas* dataset dimuat ke dalam struktur *DataFrame* dengan memanfaatkan fungsi *head()*, melihat jumlah baris dan kolom dengan fungsi *shape*, serta memeriksa tipe data setiap atribut dengan fungsi *info()*.
5. Melakukan eskplorasi awal dataset tahap ini dilakukan pemeriksaan awal terhadap dataset seperti menampilkan beberapa baris data pertama dengan fungsi *head()*, melihat jumlah baris dan kolom dengan fungsi *shape*, serta memeriksa tipe data setiap atribut dengan fungsi *info()*.
6. Melakukan pra-pemrosesan data tahap ini dilakukan pembersihan data (*data cleaning*), termasuk penghapusan data duplikat, penanganan nilai (*missing values*), serta transformasi data agar siap digunakan pada tahap analisis berikutnya.



Gambar 1. Penarikan Data

a. Pelabelan

Data ulasan yang diperoleh disatukan dalam tabel yang memuat *username*, *score*, *at*, *content* dan penambahan kolom label. Proses pelabelan ulasan didasarkan pada *score rating* yang terdapat 5 nilai yaitu 1-5 dimana nilai 1,2 dikategorikan sebagai ulasan negatif (1), dan nilai 3 dikategorikan sebagai netral (2) dan nilai 4, 5 di kategorikan sebagai ulasan positif (3). Adapun hasil dari pelabelan dapat dilihat pada Tabel 1

Tabel 1. Pelabelan Data Ulasan

USERNAME	SCORE	CONTENT	LABEL
PENGGUNA GOOGLE	5	Sangat membantu..sekali	3
PENGGUNA GOOGLE	5	Tempat belanja online no. 1 harga lebih murah daripada di pasaran pengiriman.Cepat barangnya realpict ada harga ada kualitas	3
PENGGUNA GOOGLE	5	Yaudah	3
:	:	:	:
PENGGUNA GOOGLE	1	Ongkirnya mahal..masak lebih mahal ongkirnya dr pada harganya payahhh	1

```
mapping = {1: "Negatif", 2: "Negatif", 3: "Netral", 4: "Positif", 5: "Positif"}
df['label'] = df['score'].map(mapping)
```

Gambar 2. Script Labeling

Script di atas digunakan untuk membaca pelabelan data, ini dilakukan secara otomatis dengan menggunakan script Python melalui fungsi `map()`. Pada proses ini, setiap nilai yang terdapat pada kolom *score* dipetakan ke dalam label sentimen sesuai dengan kategori yang telah ditetapkan. Dengan cara tersebut setiap ulasan teks (*content*) memperoleh label sentimen yang merepresentasikan pandangan atau persepsi pengguna berdasarkan skor yang diberikan.

	userName	score	at	content	label
0	Pengguna Google	5	08-09-2024 04:37	Sangat membantu..sekali	Positif
1	Pengguna Google	5	08-09-2024 04:37	Tempat belanja online no 1, harga lebih murah ...	Positif
2	Pengguna Google	5	08-09-2024 04:37	yaudah	Positif
3	Pengguna Google	1	08-09-2024 04:37	Ongkir nya mahal ..masak lebih mahal ongkir ny...	Negatif
4	Pengguna Google	5	08-09-2024 04:36	bagus	Positif

Gambar 3. Hasil Pelabelan

b. *Preprocessing Data*

Data ulasan yang sudah telah dilabeli kemudian di *preprocessing text* meliputi proses *cleaning*, *slang/normalisasi*, *tokenizing*, *stopword*, *stemmed*. Hasil sebelum dan setelah *preprocessing* teks dapat dilihat pada tabel berikut :

Tabel 2. Contoh Sebelum dan Sesudah *Preprocessing*

Sebelum <i>Preprocessing</i>	Setelah <i>Preprocessing</i>
Sangat membantu..sekali	Membantu sekali
Tempat belanja online no 1, harga lebih murah daripada di pasaran, pengiriman cepat, barangnya realpict ada harga ada kualitas, saran banyakin voucher potongan harga dan gretong	belanja online harga murah pasar pengiriman cepat barangnya realpict harga kualitas saran voucher potong harga gratis
:	:
Ongkir nya mahal ..masak lebih mahal ongkir nya dr pada harganya . Payahhhhhhhhhh	ongkos kirim mahal masak mahal ongkos kirim harga payahhhhhhhhhh

Proses *preprocessing text* yang dilakukan meliputi:

- a) Proses penginputan data menggunakan Python library pandas dengan script berikut :

```
import pandas as pd

# Gunakan wget untuk mendownload file dari Github (gunakan link raw)
!wget "https://raw.githubusercontent.com/salsabila2403/skripsi/main/shopee%20100k.csv" -O "shopee 100k.csv"

# Baca sheet dari file Excel
df = pd.read_csv("content/shopee 100k.csv")
```

Gambar 4. Script Input Data

Script diatas digunakan untuk pengolahan data, ini dilakukan dengan memanfaatkan bahasa pemrograman Python serta pustaka pandas untuk membaca sekaligus mengelola dataset. Data yang digunakan berbentuk file Csv yang tersimpan pada repository GitHub. Proses pengunduhan dilakukan menggunakan perintah *wget* dengan memanfaatkan tautan *raw file* GitHub sehingga dataset dapat diperoleh secara langsung.

- b) Proses *cleaning* berupa menghilangkan simbol atau tanda baca dalam dokumen seperti @,.,”;/!&*. Hasil *cleaning* dapat dilihat pada tabel berikut :

Tabel 3. Sebelum dan Sesudah *Cleaning*

Sebelum <i>Cleaning</i>	Setelah <i>Cleaning</i>
Sangat membantu	sangat membantu
Iklan shopee di web/aplikasi lain ngeselin	iklan shopee di web/aplikasi lain ngeselin
Tolong aktif kan spinjam punya saya	tolong aktif kan spinjam punya saya
:	:
Selamat siang shoope Hari ini tanggal 8 September 2024 saya sudah tidak lagi menggunakan	selamat siang shoope hari ini tanggal september saya sudah tidak lagi menggunakan shoope karna mau saya uninstall

Sebelum <i>Cleaning</i>	Setelah <i>Cleaning</i>
shooper Karna mau saya uninstall	



- c) Normalisasi berupa memperbaiki kata yang tidak baku menjadi baku. Hasil dapat dilihat pada tabel berikut :

Tabel 4. Sebelum dan Sesudah *Slang*/Normalisasi

Sebelum Normalisasi	Setelah Normalisasi
Apk yg bagus	aplikasi yang bagus
Tingkat kan pelayanan	tingkat kan pelayanan
Membantu berbelanja, keamanan kenyamanan lumayan sangat baik	membantu berbelanja keamanan kenyamanan lumayan sangat baik
:	:
Suka bangeut pake aplikasi shopee tegas menindak pada seller nakal. Jadi tempat berbelanja aman dan terpercaya	suka bangeut pakai aplikasi shopee tegas menindak pada seller nakal jadi tempat berbelanja aman dan terpercaya

id	username	score	at	context	label	cleaning
0	Pengguna Google	5	2024-09-08 04:37:00	Sangat membantu, sekali	Positif	sangat membantusekali
1	Pengguna Google	5	2024-09-08 04:37:00	Tempat belanja online no 1, harga lebih murah ...	Positif	tempat belanja online no harga lebih murah da...
2	Pengguna Google	5	2024-09-08 04:37:00	yaudah	Positif	yaudah
3	Pengguna Google	1	2024-09-08 04:37:00	ongkir nya mahal , masak lebih mahal ongkir ny...	Negatif	ongkir nya mahal masak lebih mahal ongkir nya ...
4	Pengguna Google	5	2024-09-08 04:36:00	bagus	Positif	bagus
...
99995	Lestari Citra lestari	5	2024-06-21 17:09:00	terbaik banget selalu dapat vocer gratis ongkir	Positif	terbaik banget selalu dapat vocer gratis ongkir
99996	Kumalasari Kasih kumata	5	2024-06-21 17:03:00	Pelayanan sangat sigap dan cepat	Positif	pelayanan sangat sigap dan cepat
99997	Zuhdi Yusandi pratama	5	2024-06-21 17:02:00		Like	like
99998	Arifin Ciputra	5	2024-06-21 17:02:00	Aplikasi nya sangat membantu bgt dan mudah dim...	Negatif	aplikasi nya sangat membantu bgt dan mudah dim...
99999	アジ	1	2024-06-21 17:01:00	ah shopee asu, sap masuk apk langsung masuk k...	Negatif	ah shopee asu sap masuk apk langsung masuk ka...

Gambar 5. Hasil dari *Cleaning* dan Normalisasi

- d) *Tokenizing* berupa pemecah kalimat menjadi unit-unit kata individu. Hasil dapat dilihat pada tabel berikut :

Tabel 5. Sebelum dan Sesudah *Tokenizing*

Sebelum <i>Tokenizing</i>	Setelah <i>Tokenizing</i>
Apk yg bagus	['aplikasi', 'yang', 'bagus']
Tingkat kan pelayanan	['tingkat', 'kan', 'pelayanan']
Membantu berbelanja, keamanan kenyamanan lumayan sangat baik	['membantu', 'berbelanja', 'keamanan', 'kenyamanan', 'lumayan', 'sangat', 'baik']
:	:
Suka bangeut pake aplikasi shopee tegas menindak pada seller nakal. Jadi tempat berbelanja aman dan terpercaya	['suka', 'bangeut', 'pakai', 'aplikasi', 'shopee', 'tegas', 'menindak', 'pada', 'seller', 'nakal', 'jadi', 'tempat', 'berbelanja', 'aman', 'dan', 'terpercaya']

- e) *Stopword* berupa penghapusan kata yang tidak penting. Hasil dapat dilihat pada tabel berikut:

Tabel 6. Sebelum dan Sesudah *Stopword*

Sebelum <i>Stopword</i>	Setelah <i>Stopword</i>
Shopee skrg dah gk sebgus dlu., ekspedisi ny lelet.. Kecewa	['shopee', 'gk', 'sebagus', 'ekspedisi', 'lambat', 'kecewa']
Untuk saat ini shopee cukup bagus.. Saran untuk shope ya tolong jika penjual yg pengemasan selama 3/4hari untuk pelanggan itu sudah tidak wajar.. jd tolong beri peringatan tuk penjual .. dan jika	['shopee', 'bagus', 'saran', 'shopee', 'iya', 'penjual', 'pengemasan', 'konsumen', 'wajar', 'peringatan', 'tuk', 'penjual', 'penjual', 'aktif']

Sebelum <i>Stopword</i>	Setelah <i>Stopword</i>
penjual tersebut tidak aktif selama berbulan bulan harap d'nonaktifkan atau blokir trima kasih	'berbulan', 'harap', 'dnonaktifkan', 'blokir', 'terima', 'kasih', 'kasih']
Harus nya shopee jangan egois dan mementingkan pelanggan saja. Perihal retur pengembalian barang harus nya ada pihak ketiga atau penengah sebagai hakim yg adil. Ahir" ini saya sering jadi korban penipuan pembeli. Retur barang dengan barang bekas .	['shopee', 'egois', 'mementingkan', 'konsumen', 'perihal', 'retur', 'pengembalian', 'barang', 'penengah', 'hakim', 'adil', 'korban', 'penipuan', 'pembeli', 'retur', 'barang', 'barang', 'bekas']
:	:
Shopee skrg dah gk sebgus dlu., expedisi ny lelet.. Kecewa	['shopee', 'gk', 'sebagus', 'expedisi', 'lambat', 'kecewa']

- f) *Stemmed* berupa kata di ubah menjadi bentuk dasar dengan menghapus awalan dan akhiran. Hasil dapat dilihat pada tabel berikut :

Tabel 7. Hasil Sebelum dan Sesudah *Stemmed*

Sebelum <i>Stemmed</i>	Setelah <i>Stemmed</i>
Shopee untuk sekarang ini masih terdepan dan terpercaya dibandingkan dengan yang lain	shopee depan percaya banding
Update tampilan barunya lama sekali 🤔😞	baru tampil baru
Sejauh ini udah ga ngelag apk nya, tpi berat bgt penyimpanannya, kalo bisa dikurangim fitur2 yg tidak perlu, soalnya bikin hp cpet panas kalo buka apk ini	lelet aplikasi banget simpan dikurangim fitur bikin seluler cpet panas buka aplikasi
:	:
Shopee untuk sekarang ini masih terdepan dan terpercaya dibandingkan dengan yang lain	shopee depan percaya banding

- g) Menyimpan dataset pada file baru

```
# Simpan ke CSV
df.to_csv('hasil_preprocessing.csv', index=False)
```

Gambar 6. Proses Akhir *Processing*

Gambar diatas memperlihatkan potongan script yang digunakan dalam menyimpan dataset hasil preprocessing ke dalam file baru. Penyimpanan ini dilakukan agar data yang telah melalui tahap pelabelan, *cleaning*, normalisasi, *tokenizing*, *stopword*, dan *stemmed* dapat disimpan secara terstruktur sehingga siap digunakan pada proses analisis selanjutnya.

- h) Pembobotan Kata

Setelah *preprocessing text*, hasil dari proses tersebut akan dihitung bobot katanya menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). Adapun data ilustrasi yang digunakan dapat dilihat pada tabel berikut:

Tabel 8. Data Ilustrasi

Teks	Label	Dokumen
barangnya bagus dan murah	Positif	D1
Pengiriman cepat sekali	Positif	D2
Kualitas jelek, barang rusak	Negatif	D3

:	:	:
Packing rapi, harga murah	?	D4

Data yang diperoleh selanjutnya akan dilakukan proses pembobotan setiap kata pada dokumen. Proses pembobotan pertama yang dilakukan adalah *Term Frequency* (TF). Hasil dapat dilihat pada tabel berikut:

Tabel 9. Pembobotan TF

Term	Perhitungan TF				DF	N/DF
	TF					
	D1	D2	D3	D4		
bagus	1				1	4
Barangnya	1				1	4
Dan	1				1	4
Murah	1			1	2	2
Cepat		1			1	4
Barang			1		1	4
Jelek			1		1	4
Kualitas			1		1	4
Harga				1	1	4
Packing				1	1	4
JUMLAH	4	1	3	3		

Setelah itu dilakukan proses normalisasi pada nilai TF dan melakukan pembobotan *Invers Document Frequency* (IDF). Hasilnya dapat dilihat pada tabel berikut :

Tabel 10. Normalisasi Pembobotan Tf dan IDF

Normalisasi Nilai TF dan Perhitungan IDF				
TF				IDF
D1	D2	D3	D4	
0.25				0.602
0.25				0.602
0.25				0.602
0.25			0.333	0.301
	1			0.602
		0.333		0.602
		0.333		0.602
		0.333		0.602
			0.333	0.602
			0.333	0.602

Kemudian yang terakhir melakukan pembobotan TF-IDF. Hasilnya dapat dilihat pada tabel berikut:

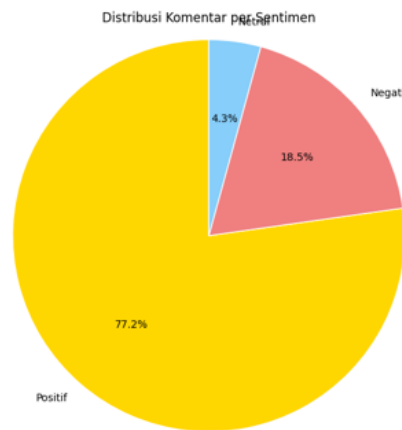
Tabel 11 Pembobotan TF-IDF

Pembobotan Nilai TF dan Perhitungan IDF				
TF				IDF
D1	D2	D3	D4	
0.1505	0	0	0	
0.1505	0	0	0	
0.1505	0	0	0	
0.075	0	0	0.100	
0	0.602	0	0	
0	0	0.200	0	
0	0	0.200	0	
0	0	0.200	0	

0	0	0	0.200
0	0	0	0.200

Deskripsi Analisis Sentimen Pengguna Terhadap Kualitas Aplikasi E-Commerce Shopee

Analisis deskriptif menyajikan gambaran secara umum terkait ulasan pengguna aplikasi Shopee. Dataset di ambil dari Kaggle berasal dari pihak ke tiga yang sudah di lakukan *scrapping* data dari google play store sebanyak 100.000 ulasan. Data yang telah di peroleh kemudian dilakukan proses pelabelan dengan tiga kategori yaitu “Positif”, “Netral”, dan “Negatif”. Presentasi hasil pelabelan dari data disajikan pada gambar berikut :



Gambar 7. Presentasi Pelabelan

Berdasarkan gambar diatas memperlihatkan ulasan pengguna Shopee berdasarkan kategori sentimen. Ulasan yang di berikan dikelompokkan ke dalam tiga jenis sentimen, yaitu positif, negatif, dan netral. Dari hasil visualisasi terlihat bahwa sebagian besar ulasan 77,2 % masuk ke dalam kategori positif yang menandakan mayoritas pengguna merasa puas serta memiliki pengalaman yang baik ketika menggunakan Shopee. Kemudian 18,5% ulasan dikategorikan sebagai negatif yang menunjukkan adanya kritik atau rasa ketidakpuasan dari sebagian pengguna. Sementara itu, sentimen netral hanya mencapai 4,3% sehingga jumlahnya jauh lebih sedikit dibandingkan dua kategori lainnya. Dari visualisasi ini dapat dilihat bahwa sentimen positif lebih dominan dan menegaskan layanan aplikasi secara keseluruhan diterima dengan baik meskipun masih terdapat sebagian kecil komentar yang bersifat kritis.

Hasil Analisis Sentimen Data Ulasan Aplikasi Shopee Pada Google Play Store

Data yang diperoleh sebanyak 100.000 ulasan dari Kaggle kemudian dilakukan proses klasifikasi menggunakan metode K-NN. Sebelum melakukan klasifikasi data dibagi dahulu menjadi data *training* dan data *testing*. Adapun pembagian data yang digunakan adalah 80%:20%, 70%: 30%, dan 60%:40%. Hasil klasifikasi metode K-NN menggunakan pembagian data disajikan dengan menggunakan tabel *confusion matrix*. Tabel *confusion matrix* dapat dilihat sebagai berikut:

Tabel 12. Confusion Training dan Data Testing 80%:20%

Label Asli	Label Prediksi			Total
	Negatif	Netral	Positif	
Negatif	639	42	3066	3747
Netral	93	15	768	876
Positif	636	61	14680	15377
Total	1368	118	18514	20000

Tabel 13. Confusion Data Training dan Testing 70%:30%

Label Prediksi				
Label Asli	Negatif	Netral	Positif	Total
Negatif	939	47	4578	5564
Netral	155	9	1129	1293
Positif	1066	58	22019	23143
Total	2160	114	27726	30000

Tabel 14. Confusion Data Training dan Data Testing 60%:40%

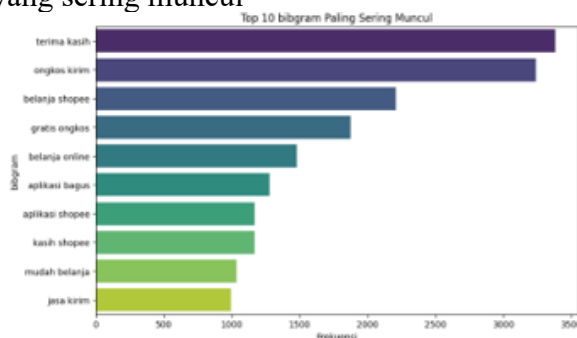
Label Prediksi				
Label Asli	Negatif	Netral	Positif	Total
Negatif	1121	48	6278	7447
Netral	200	13	1524	1737
Positif	1230	81	29505	30816
Total	2551	142	37307	40000

Pada tabel diatas menjelaskan hasil pengujian metode *K-Nearest Neighbor* (KNN) pada tiga skenario pembagian data (80:20, 70:30, 60:40) menunjukkan bahwa tingkat akurasi keseluruhan relatif stabil pada angka 77% sehingga variasi rasio data training dan testing tidak berdampak signifikan terhadap performa model.

Jika ditinjau lebih mendalam melalui matrix evaluasi, terlihat bahwa kinerja model sangat baik pada kelas positif (precision 0.79, recall 0.95-0.96, f1-score 0.87), yang disebabkan oleh dominasi jumlah data pada kelas tersebut. Sebaliknya, performa pada kelas negatif masih rendah (f1-score 0.22-0.25) dan pada kelas netral hampir tidak terdeteksi (f1-score 0.01-0.03). Hal ini menunjukkan adanya ketidak seimbangan kelas (*class imbalance*) yang membuat model lebih bias terhadap ulasan positif.

Visualisasi Data

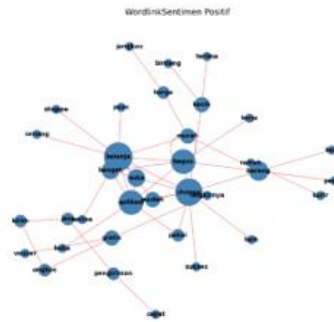
1. Visualisasi kata yang sering muncul



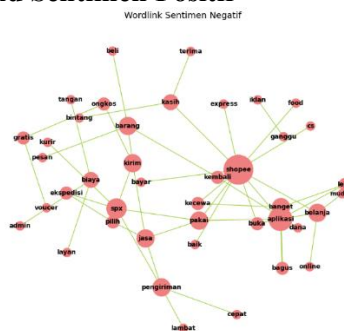
Gambar 8. Diagram Visualisasi Kata Yang Sering Muncul

Gambar diatas menampilkan diagram dengan frekuensi kemunculan paling tertinggi dalam ulasan pengguna aplikasi Shopee. Diagram ini merupakan gabungan dua kata yang sering muncul secara berurutan dalam teks, sehingga dapat menggambarkan konteks umum yang sering digunakan pengguna. Berdasarkan hasil visualisasi yang paling dominan adalah kata ini “terima kasih”, “ongkos kirim”, “belanja shopee”, dan “gratis ongkos”. Visualisasi ini menunjukkan bahwa pengguna banyak mengekspresikan rasa puas terhadap *e-commerce*.

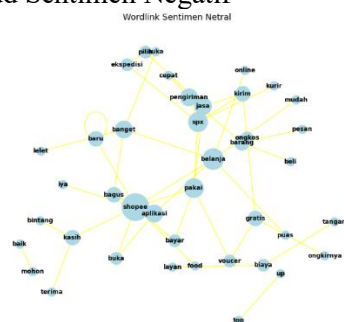
2. WordCloud Sentimen



Gambar 9. WordCloud Sentimen Positif



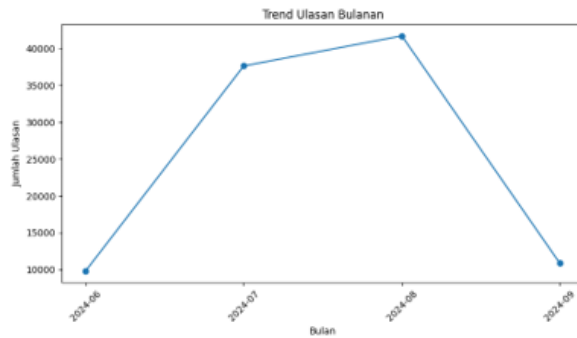
Gambar 10. WordCloud Sentimen Negatif



Gambar 11. WordCloud Sentimen Netral

Pada gambar di atas adalah hasil visualisasi *WordCloud* dan *WordLink* memperlihatkan kecenderungan kosakata yang berbeda pada tiap kategori sentimen sehingga dapat memberikan gambaran lebih jelas mengenai persepsi pengguna aplikasi Shopee. Pada ketiga kategori sentimen menunjukkan bahwa ulasan positif lebih didominasi oleh kata-kata seperti “terimakasih”, “berbelanja”, dan “bagus” yang menggambarkan kepuasan pengguna. Ulasan negatif banyak memuat kata “kurir”, “lambat”, dan “kecewa” yang menandakan keluhan terutama pada layanan pengiriman. Sedangkan ulasan netral didominasi kata “aplikasi”, “barang”, dan “pakai” yang cenderung bersifat informatif tanpa ekspresi emosional. Secara keseluruhan hasil analisis menunjukkan bahwa mayoritas pengguna memiliki persepsi positif terhadap aplikasi Shopee terutama dalam hal kemudahan berbelanja dan harga ongkos kirim.

3. Trend Ulasan Bulanan



Gambar 12. Trend Ulasan Bulanan Pengguna

Tabel 15. Trend Ulasan Bulanan Pengguna

	Bulanan	Jumlah
0	2024-06	9814
1	2024-07	37633
2	2024-08	41718
3	2024-09	10835

Pada gambar dan tabel di atas menunjukkan tren ulasan pengguna Shopee dari bulan Juni hingga September 2024. Pada bulan Juni 2024 jumlah ulasan masih relatif rendah sekitar 9814 lalu terjadi kenaikan secara signifikan pada bulan Juli 2024 dengan jumlah ulasan melonjak menjadi sekitar 37633, ulasan tertinggi terjadi pada bulan Agustus 2024 dengan jumlah lebih dari 41718 ulasan. Namun, pada bulan September 2024 jumlah ulasan turun drastis hingga hanya sekitar 10835 ulasan.

Pembahasan

Berdasarkan hasil analisis sentimen yang telah dilakukan terhadap 100.000 ulasan pengguna aplikasi Shopee di Google Play Store diketahui bahwa sebagian besar ulasan 77,2% termasuk dalam kategori positif. Hal ini menjelaskan tingginya tingkat kepuasan pengguna terhadap layanan Shopee baik dalam aspek kemudahan berbelanja harga yang bersaing maupun fitur pendukung seperti gratis ongkir dan voucher. Temuan ini konsisten dengan meningkatnya popularitas Shopee sebagai salah satu *e-commerce* terbesar di Indonesia. Meski demikian, terdapat 18,5% ulasan yang bernuansa negatif umumnya terkait kendala pengiriman, performa aplikasi, dan layanan pelanggan. Sementara itu, ulasan netral hanya mencapai 4,3% dan lebih banyak bersifat informatif tanpa menunjukkan emosi yang kuat.

Klasifikasi dengan algoritma *K-Nearest Neighbor* (KNN) pada berbagai rasio pembagian data menunjukkan akurasi 77%. Model memberikan performa terbaik pada ulasan positif (*precision* 0,79 : *recall* 0,95-0,96 : *f1-score* 0,87) tetapi masih rendah pada kelas negatif (*f1-score* 0,22-0,25) dan hampir tidak akurat pada kelas netral (*f1-score* 0,01-0,03). Hal ini dipengaruhi oleh ketidak seimbangan data dimana ulasan positif jauh lebih dominan dibandingkan dua kelas lainnya.

Visualisasi data melalui *Word Frequency* dan *Word Cloud* memperlihatkan bahwa ulasan positif di dominasi kata “terima kasih”, “bagus”, dan “belanja”, sedangkan ulasan negatif menekankan pada kata “kurir”, “lambat”, dan “kecewa”. Untuk ulasan netral, kata yang muncul lebih bersifat deskriptif seperti “aplikasi”, “barang”, dan “pakai”.

Secara keseluruhan temuan ini menunjukkan bahwa Shopee berhasil membangun citra positif di mata penggunanya, meskipun aspek pengiriman dan kecepatan layanan masih perlu ditingkatkan. Dari sisi metode, KNN cukup efektif untuk mengklasifikasi ulasan

positif namun masih diperlukan pendekatan tambahan seperti *oversampling* pada kelas minoritas atau pengguna algoritma lain untuk meningkatkan kinerja klasifikasi pada kelas negatif dan netral. Dengan demikian penelitian ini memberikan gambaran menyeluruh mengenai persepsi pengguna terhadap Shopee sekaligus membuktikan potensi penerapan *machine learning* khususnya KNN dalam analisis sentimen berbasis teks.

KESIMPULAN

Hasil penelitian terhadap 100.000 ulasan pengguna Shopee di Google Play Store menunjukkan bahwa sebagian besar ulasan termasuk kategori positif 77,2% yang mencerminkan kepuasan pengguna terhadap kemudahan transaksi, harga produk, serta fitur promosi seperti gratis ongkir dan voucher. Ulasan negatif 18,5% umumnya berhubungan dengan masalah pengiriman, kinerja aplikasi, dan layanan pelanggan, sedangkan ulasan netral 4,3% didominasi oleh komentar biasa tanpa kecenderungan emosional tertentu.

Kinerja algoritma *K-Nearest Neighbor* (KNN) pada berbagai rasio data training dan testing (80:20, 70:30, 60:40) menunjukkan akurasi konsisten sebesar 77%. Model paling efektif dalam mengklasifikasikan ulasan positif (*precision* 0,79 : *recall* 0,95-0,96 : *f1-score* 0,87), namun masih rendah pada ulasan negatif (*f1-score* 0,01-0,03). Kondisi ini dipengaruhi oleh ketidak seimbangan jumlah data antar kelas.

Hasil visualisasi *word frequency* dan *wordcloud* turut memperkuat temuan tersebut, di mana ulasan positif menonjolkan kata “terima kasih”, “bagus”, dan “belanja”, ulasan negatif banyak memuat kata “kurir”, “lambat”, dan “kecewa”, sementara ulasan netral di dominasi kata “aplikasi”, “barang”, dan “pakai”. Hasil penelitian ini menunjukkan bahwa pengguna masih merasa, perlu adanya perbaikan pada aspek pengiriman dan kecepatan layanan

DAFTAR PUSTAKA

- Adzani, L., Informatika, P. S., Teknik, F., Informatika, D. A. N., Bina, U., & Informatika, S. (2024). *SHOPEE DI APLIKASI PLAY STORE NAIVE BAYES* Diajukan untuk memenuhi salah satu syarat kelulusan Progam Sarjana.
- Amanatulloh, S. A., Septyawan, R. A., Rizal, S., Maulana, S. D., & Wafiroh, R. H. (2024). *Colab Dalam Materi Ajar Phyton*. 9(2), 51–58.
- Pinasty, W. A., & Habib, M. A. F. (2024). Pengaruh Harga, Promosi, dan Ulasan Produk terhadap Keputusan Pembelian Melalui Marketplace Shopee pada Mahasiswa Jurusan Manajemen Bisnis Syariah Angkatan 2019-2022. *TRILOGI: Jurnal Ilmu Teknologi, Kesehatan, Dan Humaniora*, 5(1), 9–19. <https://doi.org/10.33650/trilogi.v5i1.7299>
- Sabrina, A. E., & Yasin, M. (2025). Clustering Data Konsumen E-Commerce Menggunakan Algoritma K-Means Dan Dataset Kaggle. *Journal of Computer Science and Informatics Engineering*, 4(2), 96–109. <https://doi.org/10.55537/cosie.v4i2.1123>
- Sukriadi, S., Ismail, I., & Andzar, A. M. (2023). Penerapan Text Mining Dalam Klasifikasi Judul Skripsi Yang Diusulkan Mahasiswa Menggunakan Metode Naïve Bayes. *Jurnal Ilmiah Sistem Informasi Dan Teknik Informatika (JISTI)*, 6(2), 184–196. <https://doi.org/10.57093/jisti.v6i2.174>
- Wahyu Istalama Firdaus, A. (2021). Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan). *Jurnal JUPITER*, 13(1), 66.
- Yahya. (2022). *Data Mining*. CV Jejak.
- Zuhdi, H. N., & Prasetyo, B. (2025). *Sentiment Analysis On Ipusnas Application Reviews In Google Play Store Using Naive Bayes Classifier Analisis Sentimen pada Ulasan Aplikasi iPusnas di Google Play Store Menggunakan Naive Bayes Classifier*. 5(1), 12–19. <https://journal.irpi.or.id/index.php/ijirse/article/view/1846>