

PENERAPAN ALGORITMA *K-NEAREST NEIGHBOR* (K-NN) UNTUK ANALISIS SENTIMEN TERHADAP DATA ULASAN APLIKASI *E-COMMERCE* LAZADA PADA *GOOGLE PLAYSTORE*

Zulkifli Rais, Muhammad Kasim Aidid*, Asti Dewi Putri

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Makassar, Indonesia

Kata Kunci : Sentiment
Analysis, K-Nearest Neighbor
(K-NN), *E-Commerce*, Lazada

Abstrak:

Classification is the process of grouping objects based on their characteristics. Various classification methods have been employed, ranging from manual grouping to using technology as an aid in the process. One commonly used classification method is the K-Nearest Neighbour (K-NN) algorithm. K-NN predicts the class of data based on the majority class of its nearest neighbours. The novelty of this research lies in using the K-NN method on the case of Lazada application user sentiment on the Google Play Store. In this study, the review classification used is positive and negative labels. Additionally, three accuracy comparisons between training and testing data were used: 80%: 20%, 70%: 30%, and 60%: 40%. Based on the research results from the classification process of Lazada application user reviews on the Google Play Store, an accuracy of 87.00% was obtained for the training and testing data comparison of 80% : 20%.

1. Pendahuluan

Pada umumnya *data mining* sangat berpengaruh dalam proses menemukan informasi dari sekumpulan data yang berupa pengetahuan yang selama ini prosesnya tidak diketahui secara manual (Banjarsari dkk., 2016). Konsep dasar data mining adalah mengidentifikasi informasi yang tersembunyi dalam suatu *database* (Luvia dkk., 2017). *Data mining* melibatkan penggunaan berbagai teknik seperti *machine learning*, pengenalan pola, manajemen basis data, statistik, dan teknik visualisasi untuk mengeksplorasi dan menggali informasi yang berharga dari kumpulan data besar (WPR dkk., 2021).

Klasifikasi merupakan proses pengelompokkan objek berdasarkan karakteristik yang dimilikinya. Berbagai metode klasifikasi yang dilakukan mulai dari pengelompokkan secara manual hingga menggunakan teknologi sebagai bantuan dalam proses tersebut (Wibawa dkk., 2018). Dalam penelitian ini algoritma yang akan digunakan adalah *K- Nearest Neighbor* (K-NN).

Algoritma *K-Nearest Neighbor* (K-NN) adalah metode yang melakukan klasifikasi berdasarkan kategori dengan mencari k objek terdekat dalam data *testing* yang serupa dengan objek diamati pada data *training* (Yolanda dkk., 2021). Metode ini menggunakan pendekatan yang relatif mudah namun memiliki tingkat akurasi yang tinggi (Dewi dkk., 2022).

Analisis sentimen merupakan komponen dari klasifikasi teks yang bertujuan untuk mengekstrak informasi yang bersifat personal atau berdasarkan pendapat dari sebuah teks yang berbentuk bahasa alami, opini, dan sentimen yang kemudian dapat ditindaklanjuti dalam pengambilan keputusan (Rais dkk., 2022). Selain itu, analisis sentimen mampu

* Corresponding author.

E-mail address: kasimaidid@unm.ac.id



memahami, mengolah data, mengekstrak secara otomatis dengan menggunakan komputasi yang sering dikaitkan dengan proses analisis data sosial media sehingga mendapatkan hasil ulasan dari pengguna yang beragam, dari saran positif hingga keluhan yang negatif (Mufidah dkk., 2022).

Terkait dengan hal tersebut, ulasan dari pengguna memiliki peran yang sangat penting untuk mengetahui persepsi konsumen terhadap suatu produk atau layanan (Akmali dkk., 2024). Selain itu, ulasan dari pengguna juga digunakan sebagai alat yang paling efektif dan efisien untuk mendapatkan informasi terhadap suatu produk atau layanan (Liu, 2012).

Aplikasi Lazada digunakan sebagai aplikasi e-commerce yang dapat diakses melalui web dan aplikasi android (Play Store). E-commerce merupakan transaksi jual-beli yang memanfaatkan teknologi elektronik, memudahkan kita untuk membeli barang dan jasa secara online tanpa perlu datang ke toko secara langsung. Lazada telah menjadi salah satu platform e-commerce terbesar dan populer dengan jutaan pengguna aktif dan memiliki pilihan produk yang beragam (Kurnia, 2023). Google Play Store adalah layanan konten digital milik google yang terdapat produk-produk seperti aplikasi, game, buku, musik/lagu dll. Selain berbelanja produk online, tersedia juga fitur penilaian bagi pengguna untuk memberikan ulasan (review) mengenai kelebihan atau kekurangan dalam penggunaan aplikasi Lazada.

Review dari pengguna aplikasi Lazada biasanya mencakup usulan atau saran positif dan kritik negatif yang bisa memengaruhi prospek pelanggan, baik secara langsung maupun tidak langsung. Biasanya, kapasitas data ulasan pengguna aplikasi Lazada yang dipublikasikan di *Google Play Store* terus meningkat, ini mengakibatkan kesulitan bagi perusahaan untuk mengumpulkan informasi terperinci dari semua ulasan yang tersedia.

Penelitian sebelumnya menunjukkan bahwa algoritma K-Nearest Neighbor (K-NN) telah digunakan secara luas dalam analisis sentimen dan klasifikasi teks. Hasanah (2024) membandingkan model analisis sentimen Shopee di Google Play Store menggunakan K-NN, yang menghasilkan akurasi sebesar 83%. Kusuma & Cahyono (2023) juga meneliti analisis sentimen e-commerce dan menemukan bahwa kombinasi K-NN dengan NLP efektif dalam tugas ini. Muniar dkk. (2020) menggunakan K-NN untuk klasifikasi dokumen berita online dan menunjukkan kinerjanya yang baik dalam klasifikasi otomatis. Pratama & Suryono (2025) melaporkan bahwa K-NN dengan nilai K=9 mencapai akurasi tertinggi 93% dalam pengujian dan evaluasi. Sementara itu, Riyanto & Santoso (2024) menggunakan model Support Vector Machine (SVM) untuk klasifikasi sentimen ulasan aplikasi Lazada dan memperoleh akurasi 84,33%. Berdasarkan hasil-hasil tersebut, penelitian ini menggunakan metode K-NN untuk menganalisis sentimen dalam kasus yang sama.

Dalam penelitian ini akan melakukan klasifikasi data ulasan sentimen pada aplikasi *e-commerce* Lazada pada *google play store*.

2. Tinjauan Pustaka

2.1 Klasifikasi

Klasifikasi merupakan proses dalam pengembangan model yang mengelompokkan suatu objek berdasarkan atributnya. Proses klasifikasi data biasanya dimulai dari membangun aturan klasifikasi tertentu yang menggunakan data *training* sebagai tahapan pembelajaran dan pengujian sebagai data *testing* (Raharjo & Winarko, 2014). Salah satu metode klasifikasi yang sering digunakan adalah *K-Nearest Neighbor* (K-NN).

K-Nearest Neighbor (K-NN) adalah teknik yang digunakan untuk mengklasifikasikan objek berdasarkan data uji coba yang memiliki jarak terdekat atau kesamaan ciri yang paling banyak dengan objek tersebut (Syahid dkk., 2016). Jarak kedekatan melalui interval dapat dihitung menggunakan *Euclidean Distance* yang dituliskan dalam persamaan 2.1 berikut (Wafiyah dkk., 2017):

$$d_{(i,j)} = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2} \quad (2.1.1)$$

Keterangan:

$d_{(i,j)}$ = jarak antara dokumen ke-i dengan dokumen ke-j

x_{ip} = kata ke-p pada kolom ke-i

x_{jp} = kata ke-p pada kolom ke-j

Sebagai indikasi aturan sifat klasifikasi untuk mengetahui tingkat akurasi oleh sistem dengan metode K-NN perlu dilakukannya *Confusion Matrix*. Hal ini mencakup jumlah elemen yang dikelompokkan dengan benar atau tidak benar untuk setiap kelas. Salah satu keuntungan dari *confusion matrix* ini adalah memudahkan dalam mengevaluasi *confusion* dua kelas. Untuk setiap contoh dalam set pengujian, akan dibandingkan kelas sebenarnya dengan kelas klasifikasinya. Contoh positif/negatif yang diklasifikasikan dengan benar oleh *classifier* disebut *True Positive/True Negative*,

sedangkan contoh positif/negatif yang salah diklasifikasikan disebut *False Negative/False Positive* dapat dilihat pada Tabel 2.1.1:

Tabel 2.1.1 *Confusion Matrix*

Klasifikasi	Klasifikasi Sebenarnya	
	Negatif	Positif
Aktual Klasifikasi	Negatif	True Negative (TN)
	Positif	False Positive (FP)
		False Negative (FN)
		True Positive (TP)

Keterangan:

True Positive (TP) : total data hasil klasifikasi positif dan klasifikasi sebenarnya adalah positif

False Positive (FP) : total data hasil klasifikasi positif namun klasifikasi sebenarnya adalah negatif

False Negative (FN) : total data hasil klasifikasi negatif namun klasifikasi sebenarnya adalah positif

True Negative (TN) : total data hasil klasifikasi negatif dan klasifikasi sebenarnya adalah negatif

Untuk menghitung akurasi menggunakan *confusion matrix*, dapat dituliskan pada persamaan 2.2 berikut ini (Deolika & Taufiq Luthfi, 2019):

$$\text{Akurasi} = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.1.2)$$

2.2 Text Mining

Text mining merupakan teknik yang digunakan untuk mengklasifikasikan, mengelompokkan, mengekstrak informasi, dan melakukan pencarian informasi atau data besar. Proses penting dalam teknik ini adalah untuk mengidentifikasi kata-kata yang mencerminkan isi dokumen, kemudian menganalisis hubungan antara dokumen-dokumen tersebut. Adapun beberapa proses yang dilakukan dalam teknik ini adalah:

1) *Preprocessing Text*

Dalam tahap *preprocessing* melibatkan serangkaian tindakan untuk membersihkan, mengatur, dan menyesuaikan data sehingga dapat dipergunakan dengan efisien dalam proses analisis atau pemodelan (Tabassum & Patil, 2020). Beberapa proses yang dilakukan pada tahap ini yaitu:

a. *Case folding*

Case folding digunakan untuk penyeragaman bentuk huruf atau kata kedalam bentuk huruf kecil.

b. *Punctuation removal*

Punctuation removal merupakan proses menghilangkan simbol atau tanda baca pada dokumen teks.

c. *Tokenizing*

Tokenizing merupakan proses memecah kalimat menjadi kata-kata menjadi lebih berarti dan bermakna.

d. *Stopword removal*

Filtering atau *Stopword removal* merupakan tahap pengambilan kata-kata yang penting dari hasil *tokenizing*.

e. *Stemming*

Dalam pengolahan kata, *stemming* digunakan untuk mendapatkan kata dasar dari kata yang telah mengalami imbuhan. Ini dilakukan dengan asumsi bahwa kata-kata tersebut sebenarnya memiliki arti dan makna yang sama.

2) *Pembobotan Kata (Term Weighting)*

Pembobotan kata adalah sebuah teknik yang digunakan untuk memberikan bobot atau nilai pada setiap kata yang ada dalam suatu dokumen. Pembobotan ini mencerminkan seberapa pentingnya suatu kata dalam menentukan kelas atau kategori dalam dokumen tersebut (Deolika & Luthfi, 2019). Terdapat beberapa metode yang digunakan dalam pembobotan kata (*term weighting*) diantaranya adalah:

a. *Document Frequency* (DF) istilah yang merujuk pada kata-kata yang sering muncul dalam dokumen, namun kurang memberikan informasi yang berarti, contohnya seperti “dan”, “di”, “atau”, “bisa”, dan “merupakan” (Nurjanah, dkk., 2017).

b. *Term Frequency* (TF) menunjukkan adanya jumlah kemunculan kata dalam suatu dokumen teks. Perhitungan bobot TF dituliskan dalam persamaan 2.4 berikut (Nurjanah, dkk., 2017):

$$W_{tf_{t,d}} = \begin{cases} 1 + \log_{10} tf_{t,d} & \text{jika } tf_{t,d} > 0 \\ 0, & \text{lainnya} \end{cases} \quad (2.2.1)$$

Keterangan:

$W_{tf_{t,d}}$: bobot kata t pada dokumen d

$tf_{t,d}$: jumlah kemunculan kata t pada dokumen d

- c. *Invers Document Frequency* (IDF) menunjukkan seberapa sering sebuah kata muncul dalam seluruh teks dokumen. Kata yang jarang muncul pada keseluruhan dokumen teks memiliki nilai IDF lebih besar dibandingkan dengan kata yang sering muncul. Perhitungan bobot IDF dituliskan dalam persamaan 2.5 berikut (Nurjanah, dkk., 2017):

$$idf_t = \log^{10} \left(\frac{N}{df_{(t)}} \right) \quad (2.2.2)$$

Keterangan:

N : Jumlah dokumen teks

idf_t : nilai IDF

$df_{(t)}$: jumlah dokumen yang mengandung kata yang sering muncul

- d. *Term Frequency-Invers Document Frequency* (TF-IDF) merupakan hasil dari menggabungkan nilai *Term Frequency* (TF) dan nilai *Invers Document Frequency* (IDF) ketika menghitung pembobotannya. Persamaan pembobotan kata TF-IDF dituliskan dalam persamaan 2.6 berikut (Nurjanah, dkk., 2017):

$$W_{t,d} = W_{tf_{t,d}} \times idf_t \quad (2.2.3)$$

Keterangan:

$W_{tf_{t,d}}$: nilai TF

idf_t : nilai IDF

2.3 Aplikasi Lazada

Lazada merupakan perusahaan e-commerce asal Singapura yang didirikan oleh Rocket Internet pada tahun 2011. Sebagai pelopor ekosistem e-commerce di Asia Tenggara, Lazada menyediakan platform marketplace yang didukung oleh berbagai layanan, termasuk pemasaran digital, analisis data, dan logistik. Lazada menawarkan beragam produk, seperti elektronik, pakaian, peralatan rumah tangga, serta produk kesehatan dan kecantikan. Selain itu, platform ini menyediakan berbagai opsi pembayaran, termasuk transfer bank, dompet digital (Dana, OVO), kredit Lazada, serta cicilan melalui aplikasi pihak ketiga seperti Kredivo dan Akulaku (Putri dkk., 2022).

2.4 Google Play Store

Google Play Store merupakan sebuah platform digital yang dimiliki oleh Google, menyediakan berbagai konten seperti musik, buku, aplikasi, permainan, dan pemutar media. Salah satu aplikasi yang tersedia di *Google Play Store* adalah Lazada. Terdapat fitur penilaian yang memungkinkan pelanggan memberikan ulasan mengenai kelebihan dan kekurangan pada aplikasi Lazada. Ulasan tersebut mencakup pendapat pengguna tentang kualitas aplikasi, yang bisa menjadi pertimbangan bagi calon pengguna sebelum mengunduh dan menggunakan aplikasi tersebut (Saputra dkk., 2019).

3. Metode penelitian

3.1 Jenis Penelitian

Bentuk penelitian yang akan dilakukan adalah penelitian eksploratif. Penelitian eksploratif bertujuan untuk mengumpulkan informasi baru tentang suatu topik tertentu dengan tujuan memahami permasalahan yang sedang atau akan muncul. Penelitian ini akan menggali informasi mengenai analisis sentimen terhadap data ulasan aplikasi e-commerce Lazada pada *Google Play Store* yang diklasifikasikan menggunakan algoritma *K-Nearest Neighbor* (K-NN).

3.2 Sumber Data

Data yang digunakan dalam penelitian adalah data sekunder, yaitu berjumlah 1000 data yang diambil dari *Google Play Store* berupa ulasan pengguna aplikasi Lazada dalam rentang waktu selama tahun 2024.

3.3 Prosedur Penelitian

Prosedur yang dilakukan dalam penelitian ini adalah:

1. Mengumpulkan informasi dan referensi yang dibutuhkan dalam penelitian
2. Mengumpulkan data ulasan pengguna aplikasi Lazada pada *Google Play Store*

3. Melakukan klasifikasi menggunakan algoritma K-Nearest Neighbor
4. Membuat interpretasi hasil klasifikasi
5. Membuat kesimpulan
6. Menyusun laporan hasil peneliti

3.4 Teknik Analisis Data

Teknik analisis data dalam penelitian ini adalah sebagai berikut:

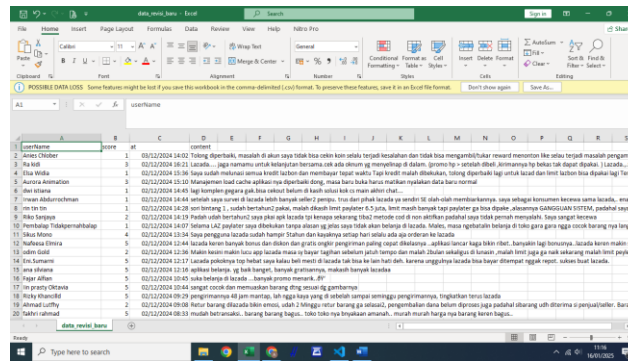
- I. Untuk mengetahui deskripsi analisis sentimen pengguna terhadap kualitas aplikasi *e-commerce* Lazada, adapun teknik analisis yang dilakukan yaitu:
 1. Melakukan pengambilan data ulasan pengguna aplikasi Lazada pada Google *Play Store*
 2. Melakukan proses pelabelan pada data
 3. Melakukan *preprocessing* data yang terdiri dari 5 langkah yaitu:
 - a) *Case folding*, proses mengubah huruf kapital menjadi huruf kecil
 - b) *Punctuation removal*, proses menghilangkan simbol dan tanda baca
 - c) *Tokenization*, proses mengubah kalimat menjadi kata
 - d) *Stopword removal*, proses menghilangkan kata sambung, kata ganti orang ketiga, karakter dan angka
 - e) *Stemming*, proses menghilangkan imbuhan kata
 4. Melakukan pembobotan kata yang terdiri dari 4 metode yaitu:
 - a) *Document Frequency* (DF), kata-kata yang sering muncul dalam dokumen
 - b) *Term Frequency* (TF), jumlah kemunculan kata dalam suatu dokumen teks
 - c) *Invers Document Frequency* (IDF), seberapa sering sebuah kata muncul dalam seluruh teks dokumen
 - d) *Term Frequency-Invers Document Frequency* (TF-IDF), gabungan nilai TF dan nilai IDF ketika menghitung pembobotannya
 5. Membagi data *training* dan data *testing*
 6. Membangun model klasifikasi data *training* menggunakan algoritma K-Nearest Neighbor
 7. Menerapkan model pada data *testing*
- II. Untuk mengetahui hasil klasifikasi algoritma K-NN dengan menganalisis sentimen data ulasan aplikasi Lazada pada Google *Play Store*, teknik analisis yang dilakukan yaitu:
 1. Menganalisis hasil klasifikasi
 2. Menghitung akurasi metode K-NN menggunakan confusion matrix pada persamaan 2.2
 3. Interpretasi hasil analisis
 4. Penarikan kesimpulan

4. Hasil dan Pembahasan

4.1 Penarikan Data

Penarikan data dilakukan menggunakan Google Colab. Pada proses pengumpulan data ini, digunakan teknik **web scrapping** untuk memperoleh data dari google *play store*. Berikut adalah langkah-langkah yang dilakukan dalam teknik *web scrapping* data di google *playstore*:

- 1) Langkah pertama yang dilakukan adalah menginstal *library python scrapper* yang akan di jalankan di Google Colab serta mengimpor library dalam Google Collab.
- 2) Tahap kedua arahkan scrapping sesuai dengan *url id* Lazada di Google *Play Store* serta mensortir data ulasan yang akan diambil dengan yang relevan sebanyak 1000 ulasan.
- 3) Kemudian tahap ketiga mencetak data dalam bentuk tabel yang telah ditarik. Setelah itu, mengkategorikan beberapa data menjadi 4 bagian, yaitu 'userName', 'score', 'at', 'content' dalam bentuk data frame
- 4) Dan terakhir menyimpan ulasan aplikasi menjadi data frame serta disimpan ke dalam file CSV. Berikut adalah hasil *scrapping* data ulasan aplikasi Lazada dapat dilihat pada Gambar 4.1.1:



Gambar 4.1.1 Penarikan Data

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 kategori yaitu 1-5 dimana nilai 1, 2, dan 3 dikategorikan sebagai ulasan Negatif (0) dan nilai 4 dan 5 dikategorikan sebagai ulasan Positif (1). Adapun hasil dari pelabelan dapat dilihat pada Tabel 4.1.1

Tabel 4.1.1 Pelabelan Data Ulasan

User Name	Score	Content	Label
Dwi Novitasari	5	saya sangat senang belanja di lazada,,kurir2nya sangat baik serta pengirimannya cepat,,semoga makin sukses lazada	1
Namikula Mmn	3	Lazada yang Sekarang, Sering telat pengirimannya, estimasi pengiman menurut Saya terlalu lama, produk yang saya pesan juga, Sering ga sesuai	0
Joel joel	5	pengiriman cepat dan mudah mencari barang yang mau di pesan	1
:	:	:	:
Barbara	1	percuma ada fitur optimasi otomatis produk, biarpun skor sudah sangat baik tetap aja produk gak muncul di halaman pencarian / utama	0

Hasil yang diperoleh dari data ulasan yang telah dilabeli kemudian di *preprocessing text* meliputi proses *case folding*, *punctuation removal*, *stopward removal*, *tokenazing*, dan *stemming*. Hasil sebelum dan setelah *preprocessing text* dapat dilihat pada Tabel 4.1.2 berikut:

Tabel 4.1.2 Ulasan Sebelum dan Setelah Preprocessing

Sebelum Preprocessing	Setelah Preprocessing
saya sangat senang belanja di lazada,,kurir2nya sangat baik serta pengirimannya cepat,,semoga makin sukses lazada	senang belanja lazada kurir kirim cepat semoga sukses
Lazada yang Sekarang, Sering telat pengirimannya, estimasi pengiman menurut Saya terlalu lama, produk yang saya pesan juga, Sering ga sesuai	telat kirim kirim produk pesan sesuai
pengiriman cepat dan mudah mencari barang yang mau di pesan	kirim cepat mudah cari pesan
:	:
percuma ada fitur optimasi otomatis produk, biarpun skor sudah sangat baik tetap aja produk gak muncul di halaman pencarian / utama	fitur optimasi otomatis produk skor produk muncul laman cari utama

Setelah *preprocessing text*, hasil dari proses tersebut akan dihitung bobot katanya menggunakan kata *Term Frequency - Inverse Document Frequency* (TF-IDF). Adapun data ilustrasi yang digunakan dapat dilihat pada Tabel 4.1.3

Tabel 4.1.3 Data Ilustrasi

Teks	Label	Dokumen
senang belanja lazada kurir kirim cepat semoga sukses	Positif	D1
telat kirim kirim produk pesan sesuai	Negatif	D2
kirim cepat mudah cari pesan	Positif	D3
:	:	:
fitur optimasi otomatis produk skor produk muncul laman cari utama	?	D4

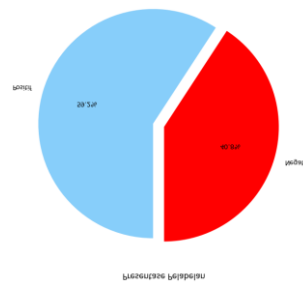
Data yang diperoleh selanjutnya akan dilakukan proses pembobotan (TF-IDF). Hasilnya dapat dilihat pada Tabel 4.1.4 berikut:

Tabel 4.1.4 Pembobotan TF-IDF

Hasil Pembobotan TF-IDF			
D1	D2	D3	D4
0,075	0	0	0
0,075	0	0	0
0,075	0	0	0
0,075	0	0	0
0,016	0,042	0,025	0
0,038	0	0,06	0
0,075	0	0	0
0,075	0	0	0
0	0,101	0	0
0	0,05	0	0,038
0	0,05	0,06	0
0	0,101	0	0
0	0	0,12	0
0	0	0,06	0,038
0	0	0	0,075
0	0	0	0,075
0	0	0	0,075
0	0	0	0,075
0	0	0	0,075
0	0	0	0,075

4.2 Deskripsi Analisis Sentimen Pengguna Terhadap Kualitas Aplikasi *E-Commerce* Lazada

Analisis deskriptif menyajikan gambaran secara umum terkait ulasan pengguna aplikasi Lazada. *Scrapping* data yang diperoleh dari google *play store* sebanyak 1000 ulasan. Data yang telah di peroleh kemudian dilakukan proses pelabelan dengan dua kategori yaitu “Positif” dan “Negatif”. Presentasi hasil pelabelan dari data disajikan pada Gambar 4.2.1:

**Gambar 4.2.1** Presentasi Pelabelan

Berdasarkan Gambar 4.1 dapat dilihat bahwa presentasi ulasan aplikasi Lazada pada google *play store* mengandung ulasan Positif lebih banyak dibanding ulasan Negatif. Dari 1000 data ulasan, sebanyak 59,2% ulasan Positif dan sebanyak 40,8% ulasan Negatif.

4.3 Hasil Analisis Sentimen Data Ulasan Aplikasi Lazada Pada Google Playstore

Data yang telah diperoleh sebanyak 1000 ulasan dari google *play store* 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%.

Berdasarkan proses klasifikasi menggunakan metode K-NN yang diuji pada 15 nilai k , Tabel 4.2.1 akan menunjukkan akurasi berdasarkan hasil pengujian nilai k :

Tabel 4.2.1 Hasil Perbandingan Akurasi Nilai k

Nilai k	Perbandingan 80% : 20%	Perbandingan 70% : 30%	Perbandingan 60% : 40%
1	42.50%	42.67%	38.75%
3	73.50%	72.00%	68.25%
5	82.50%	78.67%	74.50%
7	85.00%	79.67%	79.25%
9	84.00%	81.33%	81.25%
11	86.00%	81.00%	80.75%
13	87.00%	83.67%	81.00%
15	86.50%	83.67%	83.00%

Dapat dilihat pada tabel diatas, bahwa pada perbandingan 80%:20% mempunyai nilai k terbaik terdapat pada $k = 13$, sedangkan pada perbandingan 70%:30% mempunyai nilai k terbaik terdapat pada $k = 13$, dan pada perbandingan 60%:40% mempunyai nilai k terbaik terdapat pada $k = 15$. Setelah didapatkan nilai k terbaik, kemudian dilakukan kembali uji akurasi *presisi*, *recall*, dan *f1-score* dengan perbandingan data *training* dan *testing* 80% : 20%, 70% : 30%, dan 60% : 40%. Dapat dilihat pada Tabel 4.2.2 hasil akurasi *presisi*, *recall*, dan *f1-score*:

Tabel 4.2.2 Hasil *Presisi*, *Recall*, dan *F1-Score*

Perbandingan Data <i>Training</i> dan <i>Testing</i> (%)	k	<i>Presisi</i>		<i>Recall</i>		<i>F-score</i>	
		0	1	0	1	0	1
80% : 20%	13	91.00%	82.00%	86.00%	88.00%	88.00%	85.00%
70% : 30%	13	87.00%	78.00%	84.00%	82.00%	85.00%	80.00%
60% : 40%	15	90.00%	74.00%	81.00%	86.00%	85.00%	80.00%

4.4 Pembahasan

Dalam proses klasifikasi ulasan positif dan negatif pengguna aplikasi Lazada pada google *play store*, penelitian ini menggunakan metode klasifikasi K-Nearest Neighbor (K-NN). Berdasarkan hasil prediksi klasifikasi K-NN

- Hasil *preprocessing text* sebanyak 1000 ulasan, dengan jumlah ulasan beropini Positif dengan presentase 59.2% dan ulasan yang beropini Negatif memiliki presentase 40.8%. Dapat dilihat bahwa dataset memiliki distribusi yang hampir seimbang, meskipun terdapat sedikit lebih banyak ulasan beropini Positif dibandingkan ulasan beropini Negatif.
- Hasil dari penerapan algoritma K-Nearest Neighbor (K-NN) dalam mengklasifikasikan data ulasan pengguna aplikasi Lazada menjadi kelas positif dan negatif dengan perbandingan data latih dan data uji sebesar 80% : 20% diperoleh hasil klasifikasi ulasan sentimen dengan tingkat akurasi sebesar nilai $k = 13$ sebesar 87.00%.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, maka penulis memberikan beberapa saran, yaitu:

- 1) Melakukan klasifikasi teks pada analisis sentimen menggunakan metode lain seperti metode SVM, Naïve Bayes dan lainnya.
- 2) Melakukan klasifikasi teks pada analisis sentimen menggunakan gabungan 2 metode atau lebih.
- 3) Menambah limit maksimal batas ulasan pengguna diatas 1000 ulasan.

Referensi

- Alessandro Riyanto, V., & Budi Santoso, D. (2024). Penerapan Model Support Vector Machine Pada Klasifikasi Sentimen Ulasan Aplikasi Lazada. *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)*, 9(1), 178–184. <https://tunasbangsa.ac.id/ejurnal/index.php/jurasik>
- Banjarsari, M. A., Budiman, I., & Farmadi, A. (2016). Penerapan K-Optimal Pada Algoritma Knn Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan Ip Sampai Dengan Semester 4. *Klik-Kumpulan Jurnal Ilmu Komputer*, 2(2), 159–173.
- Book, . (2017). *Data Mining K-Nearest Neighbor*. <https://www.researchgate.net/publication/321804055>
- Deolika, A., & Taufiq Luthfi, E. (2019). Analisis Pembobotan Kata Pada Klasifikasi Text Mining. *Jurnal Teknologi Informasi*, 3(2).
- Dewi, S. P., Nurwati, N., & Rahayu, E. (2022). Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor. *Building of Informatics, Technology and Science (BITS)*, 3(4), 639–648. <https://doi.org/10.47065/bits.v3i4.1408>
- Dirjen, S. K., Riset, P., Pengembangan, D., Dikti, R., Saputra, S. A., Rosiyadi, D., Gata, W., Husain, S. M., Komputer, M. I., Komputer, I., Nusa, S., Kramat, M., Penelitian, P., Lipi, I., & Tangerang, U. M. (2017). Terakreditasi SINTA Peringkat 2 Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization. *masa berlaku mulai*, 1(3), 377–382.
- Habib Kusuma, I., & Cahyono, N. (2023). *Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor*. 8(3).
- Hasanah, K. (2024). Comparison of Sentiment Analysis Model for Shopee Comments on Google Play Store. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 13(1), 21–30. <https://doi.org/10.32736/sisfokom.v13i1.1916>
- Ilham, A., & Pramusinto, W. (2023). *3 rd Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) 30 Agustus 2023-Jakarta* (Vol. 2, Nomor 2).
- Infalkar, S., & Sharma, J. (2015). A survey on text mining-techniques and application. *International Journal of Research In Science & Engineering*, 24, 1–14.
- Kurnia, W. (2023). *Sentimen Analisis Aplikasi E-Commerce Berdasarkan Ulasan Pengguna Menggunakan Algoritma Stochastic Gradient Descent*. 4(1), 138–143. <https://doi.org/10.33365/jtsi.v4i2.2561>

- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Luvia, Y. S., Windarto, A. P., Solikhun, S., & Hartama, D. (2017). Penerapan Algoritma C4. 5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di Amik Tunas Bangsa. *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, 1(1), 75–79.
- Maulana, R., Hertaryawan, P., Raihan, M., Santoso, I., Tinggi, S., Komputer, I., Karya Informatika, C., Radin, J., Ii, I., Sawit, D., Pengadilan, S., Agama, T., Jakarta, D., Timur, J., Teknologi, U., Jakarta, M., Minangkabau, J., 60 -Manggarai, N., Selatan, J., & Id., R. A. (t.t.). *Komparasi Algoritma Naive Bayes Dan K-Nearest Neighbor Pada Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi Tokopedia*. 17(2). <https://doi.org/10.47111/JTI>
- Mufidah, F. S., Winarno, S., Alzami, F., Udayanti, E. D., & Sani, R. R. (2022). Analisis Sentimen Masyarakat Terhadap Layanan ShopeeFood Melalui Media Sosial Twitter Dengan Algoritma Naïve Bayes Classifier. *JOINS (Journal of Information System)*, 7(1), 14–25. <https://doi.org/10.33633/joins.v7i1.5883>
- Nurjanah, W. E., Setya Perdana, R., & Fauzi, M. A. (2017). *Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet* (Vol. 1, Nomor 12). <http://j-ptiik.ub.ac.id>
- Pimpalkar, A. P., & Retna Raj, R. J. (2020). Influence of Pre-Processing Strategies on the Performance of ML Classifiers Exploiting TF-IDF and BOW Features. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, 9(2), 49–68. <https://doi.org/10.14201/adcaij2020924968>
- Putri, A. J., Syafira, A. S., Purbaya, M. E., & Purnomo, D. (2022). Analisis Sentimen E-Commerce Lazada pada Jejaring Sosial Twitter Menggunakan Algoritma Support Vector Machine. *Jurnal TRINISTIK: Jurnal Teknik Industri, Bisnis Digital, dan Teknik Logistik*, 1(1), 16–21. <https://doi.org/10.20895/trinistik.v1i1.447>
- Raharjo, S., & Winarko, E. (2014). Klusterisasi, Klasifikasi Dan Peringkasan Teks Berbahasa Indonesia. *Universitas Gunadarma-Depok*, 8.
- Rais, Z., Said, R. N., & Ruliana, R. (2022). Text Classification on Sentiment Analysis of Marketplace SHOPEE Reviews On Twitter Using K-Nearest Neighbor (KNN) Method. *JINAV: Journal of Information and Visualization*, 3(1), 1–8. <https://doi.org/10.35877/454ri.jinav1389>
- Rizzo Irfan, M., & Ali Fauzi, M. (2018). *Analisis Sentimen Kurikulum 2013 pada Twitter menggunakan Ensemble Feature dan Metode K-Nearest Neighbor* (Vol. 2, Nomor 9). <http://j-ptiik.ub.ac.id>
- Sihombing, P. R., & Yuliaty, I. F. (2021). Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia. *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 20(2), 417–426. <https://doi.org/10.30812/matrik.v20i2.1174>
- Syahid, D., & Nursantika, D. (2016). *Sistem Klasifikasi Jenis Tanaman Hias Daun Philodendron Menggunakan Metode K-Nearest Neighbor (Knn) Berdasarkan Nilai Hue, Saturation, Value (Hsv)*. 1.
- Tabassum, A., & Patil, R. R. (2020). A Survey on Text Pre-Processing & Feature Extraction Techniques in Natural Language Processing. *International Research Journal of Engineering and Technology*. www.irjet.net
- Wibawa, A. P., Guntur, M., Purnama, A., Akbar, M. F., & Dwiyanto, F. A. (2018). Metode-metode klasifikasi. *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, 3(1).
- WPR, A. A., Rozi, F., & Sukmana, F. (2021). Prediksi penjualan produk unilever menggunakan metode k-nearest neighbor. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 6(1), 155–160.

- Yolanda, I., Fahmi, H., & Penerapan, [. (t.t.). *JIKOMSI [Jurnal Ilmu Komputer dan Sistem Informasi]*15 Penerapan Data Mining Untuk Prediksi Penjualan Produk Roti Terlaris Pada PT.Nippon Indosari Corpindo Tbk Menggunakan Metode K-Nearest Neighbor. 3(3), 9.
- Yulia Muniar, A., & Ria Lestari, K. (t.t.). *Muniar, Penerapan Algoritma K-Nearest Neighbor pada Pengklasifikasian Dokumen Berita Online Penerapan Algoritma K-Nearest Neighbor Pada Pengklasifikasian Dokumen Berita Online.*
- Zhong, N., Li, Y., & Wu, S.-T. (2010). Effective pattern discovery for text mining. *IEEE transactions on knowledge and data engineering*, 24(1), 30–44.