
KLASIFIKASI KOMENTAR PADA APLIKASI SHOPEE DENGAN FITUR BAG OF WORD DAN ALGORITMA NAÏVE BAYES

Darma Juang¹⁾, Muhammad Iqbal²⁾
Program Studi Magister Teknologi Informasi
Universitas Pembangunan Panca Budi, Medan
Jl. Jenderal Gatot Subroto KM. 4,5 Sei Sikambing, Medan, Sumatera Utara
email : darma.juang55@gmail.com¹⁾, muhammadiqbalpb@gmail.com¹⁾

Abstrak

Analisis sentimen merupakan metode klasifikasi yang digunakan untuk mengelompokkan opini yang terkandung dalam sebuah teks. Opini terbagi tiga yaitu opini positif, negatif, dan netral. Analisis sentimen sering digunakan untuk mengetahui opini masyarakat terhadap E-commerce, seperti Shopee, melalui ulasan di komentar aplikasi. Penulis menggunakan fitur Bag of Words untuk merepresentasikan data teks ke dalam bentuk vektor yang dapat diolah oleh algoritma machine learning. Algoritma yang digunakan adalah naive bayes, yang dikenal memiliki kemampuan yang baik dalam menangani data yang kompleks dan menghasilkan prediksi yang akurat. Hasil dari penelitian ini menunjukkan bahwa model yang dibangun memiliki akurasi sebesar 81 %. Hasil ini menunjukkan bahwa pendekatan yang digunakan cukup efektif dalam menganalisis sentimen pengguna terhadap aplikasi Shopee. Penelitian ini memberikan kontribusi signifikan dalam penerapan machine learning untuk analisis teks dan membuka peluang untuk pengembangan lebih lanjut dalam bidang ini.

Kata Kunci: komentar, Naïve Bayes, Bag Of Word

1. Pendahuluan

Analisis sentimen adalah teknik yang digunakan untuk mengelompokkan opini dalam sebuah teks ke dalam tiga kategori, yaitu positif, negatif, dan netral. Teknik ini banyak diterapkan pada data berupa komentar, ulasan produk, dan media sosial, di mana pengguna bebas menyampaikan pendapat mereka. Analisis sentimen merupakan bagian dari text mining yang bertujuan untuk mengekstraksi opini, emosi, dan penilaian seseorang terhadap suatu topik dengan menggunakan pemrosesan bahasa alami (Cahyaningtyas et al., 2021). Beberapa penelitian mengenai analisis sentiment sudah pernah dilakukan pada beberapa media sosial seperti twitter atau x di mana topiknya tentang kenaikan harga BBM menggunakan metode ekstraksi fitur Bag Of Word untuk algoritma sendiri menggunakan Support Vector Machine mendapatkan hasil akurasi 65% (Darmawan et al., 2023). yang menggunakan analisis sentimen yang data set berasal dari komentar pada salah satu toko beras di aplikasi Shopee menggunakan metode bag of word dan algoritma naive bayes mendapatkan akurasi sebesar 76% (Sugiarto et al., 2023).

Shopee merupakan salah satu e-commerce yang terkenal di Indonesia. Pada kuartal ketiga tahun 2023, Shopee menempati posisi pertama sebagai e-commerce dengan jumlah pengunjung bulanan terbanyak di Indonesia. Menurut data dari databoks Shopee mencatatkan 273 juta pengunjung pada periode tersebut [1]. Ulasan positif dari pelanggan dapat meningkatkan persepsi calon pembeli dan mendorong minat mereka untuk membeli produk. Sebaliknya, ulasan negatif dapat menyebabkan calon pembeli membatalkan niat mereka untuk melakukan pembelian (Sugiarto et al., 2023).

Pada penelitian sebelumnya mengenai analisis sentimen menggunakan algoritma random forest atau bukan sudah pernah dilakukan dengan menggunakan data set sendiri bersumber dari ulasan marketplace selain itu ada juga penelitian yang dilakukan dengan mengambil data dari sosial media seperti twitter seperti pada penelitian yang bertujuan untuk identifikasi ujaran kebencian di media sosial twitter menggunakan metode backpropagation neural network berbasis lexicon base featured dan bag of word mendapatkan akurasi sebesar 78,81% (Munir et al., 2018). Selain dari media sosial ada juga data set yang didapatkan dari portal berita seperti pada penelitian Komparasi Ekstraksi Fitur dalam Klasifikasi Teks Multilabel Menggunakan Algoritma Machine Learning di mana fitur bag of word menggunakan algoritma decision tree mendapatkan akurasi terbesar pada split data 80-20

Oleh karena itu penelitian ini akan menggunakan bag of word sebagai fitur ekstraksi kata menggunakan metode naive bayes untuk melakukan proses klasifikasinya

2. Landasan Teori

Preprocessing

Preprocessing merupakan tahapan awal dalam mengolah data input sebelum memasuki proses tahapan utama dari metode Latent Semantic Analysis (LSA). Preprocessing text dilakukan untuk tujuan penyeragaman dan kemudahan pembacaan serta proses LSA. Preprocessing terdiri dari beberapa tahapan. Adapun tahapan

preprocessing berdasarkan (Triawati, 2009) , yaitu: case folding,tokenizing / parsing, filtering, stemming. Berikut penjelasan empat tahapan dalam proses preprocessing adalah sebagai berikut.

1. Case Folding.

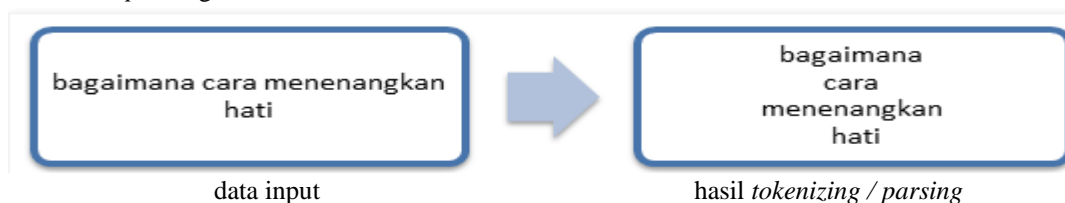
Case folding merupakan tahapan proses yang mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap *delimiter* (pembatas) (Triawati, 2009).Contoh penggunaan case folding adalah sebagai berikut.



Gambar 1. Proses Case Folding

2. Tokenizing

Tahap *tokenizing / parsing* adalah tahap pemotongan kata berdasarkan tiap kata yang menyusunya (Triawati, 2009). Selain itu, *spasi* digunakan untuk memisahkan antar kata.



Gambar 2. Proses Tokenizing

3. Filtering

Tahap filtering adalah tahap mengambil kata - kata penting dari hasil tokenizing. Proses filtering dapat menggunakan algoritma stoplist (membuang kata yang kurang penting) atau wordlist (menyimpan kata penting). Stoplist / stopword adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan bag-of-words. Contoh stopword adalah “yang”, “dan”, “di”, “dari” dan lain – lain. (Triawati, 2009).



Gambar 3. Proses filtering

Penelitian Terkait

Beberapa penelitian sebelumnya yang menggunakan Naive Bayes untuk analisis sentimen.

No	Studi / Penelitian	Data / Domain / Fokus	Ringkasan & Temuan
1	Sentiment Analysis of JNE User Perception using Naïve Bayes Classifier Algorithm	Ulasan pengguna layanan JNE dari Google Play (1.876 review)	Mengklasifikasikan review menjadi positif/negatif dengan NB — cocok untuk analisis opini pelanggan. Jurnal UPN Veteran Yogyakarta
2	Optimization of Naïve Bayes Classifier Method Using TF-IDF for Sentiment Analysis	Review publik terhadap RSUD (data dari Google Maps)	Preprocessing + fitur TF-IDF + NB → model dengan performa cukup baik (Precision positif 83%, F1-score ~0.91) Jurnal UMB

2	Analisis Sentimen Masyarakat Berdasarkan Komentar Kerja Sama Tiktok Shop dan Tokopedia di Instagram Menggunakan Naïve Bayes	Komentar Instagram atas kerja sama e-commerce (641 komentar)	Menggunakan NB untuk mengukur respon pengguna terhadap partnership; hasil terdiri dari distribusi sentimen + metrik evaluasi. Jurnal ITB Semarang
3	Penerapan Naïve Bayes untuk Analisis Sentimen terhadap Mobil Listrik	Komentar YouTube tentang mobil listrik di Indonesia	NB digunakan untuk mengklasifikasi sentimen pengguna; setelah teknik SMOTE, akurasi naik signifikan. Jurnal Nasional UMP
4	Analisis Sentimen Komentar YouTube terkait Pembahasan eSIM menggunakan Naïve Bayes dan Random Forest	Komentar YouTube tentang layanan eSIM (324 komentar)	NB berhasil dengan akurasi tinggi (~98.52%) dalam klasifikasi sentimen; cocok sebagai referensi langsung jika kamu mau gunakan data YouTube. BSI Journal
5	Analisis Sentimen pada Komentar YouTube tentang Insiden Ojek Online dan Brimob menggunakan Naïve Bayes	Komentar YouTube (1.121 komentar) pada video berita insiden	Klasifikasi ke 3 kelas sentimen: positif, netral, negatif — hasil menunjukkan sebagian besar netral/negatif; memberi gambaran tantangan pada data opini publik. Jurnal UNW

Bag of Word

Metode ini merupakan salah satu cara yang cukup sederhana dalam memproses datateks dengan mengubahnya menjadi vektor angka agar dapat diproses oleh komputer. Pada dasarnya, metode ini hanya menghitung frekuensi kemunculan kata dalam keseluruhan dokumen yang diproses (Tri Putra et al., n.d.)

Dalam notasi matematis, Bag of Words (BOW) dapat direpresentasikan dengan cara berikut: jika d adalah sebuah dokumen dan v adalah kumpulan kata atau kosakata dari keseluruhan dokumen, maka representasinya adalah sebagai berikut:

$$\text{BoW}(d) = [\text{count}(w_1, d), \text{count}(w_2, d), \dots, \text{count}(w_n, d)]$$

Algoritma Naive Bayes

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Definisi lain mengatakan Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. [2]

Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu. Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Naive Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan (Pattekari, 2012)

Persamaan dari teorema naïve bayes adalah

$$P(H|X) = P(x|H) \cdot P(H)$$

$P(X)$

Dimana :

X : Data dengan class yang belum diketahui

H : Hipotesis data yang merupakan suatu class spesifik

$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi X (posteriori probabilitas)

$P(H)$ = Probabilitas Hipoteses H (prior probabilitas)

$P(X|H)$ = Probabilitas X berdasarkan kondisi pada hipotesis

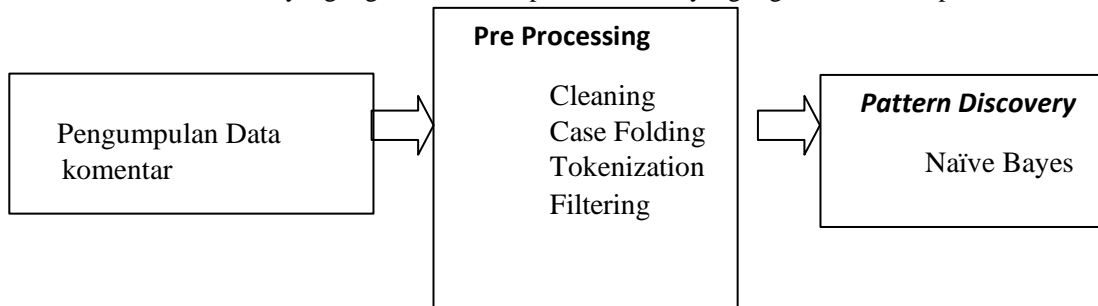
$H P(X)$ = Probabilitas X

3. Metode Penelitian

Dalam pembahasan di jurnal ini, penulis ingin menganalisis komentar kata dengan menggunakan metode naive bayes. Hasil penelitian yang diharapkan dengan metode naive bayes dalam memberikan nilai pengukuran dalam menganalisis komentar berdasarkan kemunculan teks. Penelitian dilakukan menggunakan komentar berbahasa indonesia tersimpan dalam bentuk document csv. Data yang digunakan diambil diambil dari shopee. Teknik pengumpulan data yang digunakan peneliti dalam pengumpulan data adalah Mengumpulkan literatur, jurnal, paper, dan bacaan-bacaan lainnya yang berhubungan dengan algoritma klasifikasi . Mengumpulkan data penelitian yang diperoleh secara online dari shopee dataset.

Proses Analisis Komentar pada Dokumen

Berikut ini adalah metode yang digunakan untuk proses analisis yang digunakan dalam penelitian ini.



Gambar 5. Proses Analisis komentar pada Dokumen

Pada tahapan ini yang dilakukan yaitu case folding, mengubah semua huruf dalam teks menjadi huruf kecil. Kemudian dilakukan proses parsing. Parsing yang digunakan adalah parsing sederhana yaitu memecah sebuah teks menjadi kumpulan kata-kata tanpa memperhatikan keterkaitan antar kata dan peran atau kedudukannya dalam kalimat. Hasil yang diperoleh yaitu karakter huruf saja. Selanjutnya proses yang dilakukan yaitu penghilangan stopword. Pada saat selesai melakukan parsing, sistem akan melakukan pengecekan ke dalam daftar stopword. Jika kata merupakan stopword maka kata itu akan dibuang baik angka dan simbol

4. Hasil

Dalam melaksanakan penelitian, peneliti melaksanakan penelitian menggunakan *Google Collaboratory* sebagai alat untuk menggunakan bahasa pemrograman python, berikut hasil dan pembahasan yang telah peneliti kaji meliputi beberapa tahapan yang diuraikan dibawah ini :

1. Install Library

```

1. Install Library google play scraper untuk mengambil data dengna teknik scrapper

1 pip install google-play-scraper
Requirement already satisfied: google-play-scraper in /usr/local/lib/python3.12/dist-packages (1.2.7)
  
```

2. Import Library

```

2. Import Library

1 # library scrapping dataset via playstore
2 from google_play_scraper import sort, reviews
3
4 # library untuk menggunakan dataframe dari python
5 import pandas as pd
6
7 # library untuk ekstraksi fitur menggunakan Bag Of word
8 from sklearn.feature_extraction.text import CountVectorizer
9
10 # library naive bayes classifier
11 from sklearn.feature_extraction.text import CountVectorizer
12 from sklearn.model_selection import train_test_split
13 from sklearn.naive_bayes import MultinomialNB
14 from sklearn.metrics import accuracy_score, classification_report
15
16 # visualisasi prediksi menggunakan seaborn
17 import matplotlib.pyplot as plt
18 import seaborn as sns
19 from sklearn.metrics import confusion_matrix, classification_report
20
  
```

3. Scrapping Data Ulasan Shoope di *Playstore*

	content	score	lowercase
0	lama paketnya udah 5 hari di tempat yang sama ...	1	lama paketnya udah 5 hari di tempat yang sama ...
1	semoga semakin sukses dalam sejarah	5	semoga semakin sukses dalam sejarah
2	terkadang ada pengantarnya yg kurang amanah & ...	3	terkadang ada pengantarnya yg kurang amanah & ...
3	mantap gas poll	5	mantap gas poll
4	Dari dulu pake shopee gak kecewa, eh sekarang ...	1	dari dulu pake shopee gak kecewa, eh sekarang ...
...
4995	Aplikasinya bagus!	5	aplikasinya bagus!
4996	aplikasi yang sangat bagus buat berbelanja barang	5	aplikasi yang sangat bagus buat berbelanja barang
4997	belanja lebih mudah	5	belanja lebih mudah
4998	saya member platinum. dalam 1 tahun dari ratus...	4	saya member platinum. dalam 1 tahun dari ratus...
4999	membantu bngt,tp kadang ada ongkir yg mhl bngt	5	membantu bngt,tp kadang ada ongkir yg mhl bngt

5000 rows X 3 columns

c. Remove Angka

	content	score	lowercase	remove_angka
0	lama paketnya udah 5 hari di tempat yang sama ...	1	lama paketnya udah 5 hari di tempat yang sama ...	lama paketnya udah hari di tempat yang sama mulu
1	semoga semakin sukses dalam sejarah	5	semoga semakin sukses dalam sejarah	semoga semakin sukses dalam sejarah
2	terkadang ada pengantarnya yg kurang amanah & ...	3	terkadang ada pengantarnya yg kurang amanah & ...	terkadang ada pengantarnya yg kurang amanah & ...
3	mantap gas poll	5	mantap gas poll	mantap gas poll
4	Dari dulu pake shopee gak kecewa, eh sekarang ...	1	dari dulu pake shopee gak kecewa, eh sekarang ...	dari dulu pake shopee gak kecewa, eh sekarang ...
...
4995	Aplikasinya bagus!	5	aplikasinya bagus!	aplikasinya bagus!
4996	aplikasi yang sangat bagus buat berbelanja barang	5	aplikasi yang sangat bagus buat berbelanja barang	aplikasi yang sangat bagus buat berbelanja barang
4997	belanja lebih mudah	5	belanja lebih mudah	belanja lebih mudah
4998	saya member platinum. dalam 1 tahun dari ratus...	4	saya member platinum. dalam 1 tahun dari ratus...	saya member platinum. dalam tahun dari ratus...
4999	membantu bngt,tp kadang ada ongkir yg mhl bngt	5	membantu bngt,tp kadang ada ongkir yg mhl bngt	membantu bngt,tp kadang ada ongkir yg mhl bngt

5000 rows X 4 columns

d. Remove Simbol

	content	score	lowercase	remove_angka	remove_simbol
0	lama paketnya udah 5 hari di tempat yang sama ...	1	lama paketnya udah 5 hari di tempat yang sama ...	lama paketnya udah hari di tempat yang sama mulu	lama paketnya udah hari di tempat yang sama mulu
1	semoga semakin sukses dalam sejarah	5	semoga semakin sukses dalam sejarah	semoga semakin sukses dalam sejarah	semoga semakin sukses dalam sejarah
2	terkadang ada pengantarnya yg kurang amanah & ...	3	terkadang ada pengantarnya yg kurang amanah & ...	terkadang ada pengantarnya yg kurang amanah & ...	terkadang ada pengantarnya yg kurang amanah & ...
3	mantap gas poll	5	mantap gas poll	mantap gas poll	mantap gas poll
4	Dari dulu pake shopee gak kecewa, eh sekarang ...	1	dari dulu pake shopee gak kecewa, eh sekarang ...	dari dulu pake shopee gak kecewa, eh sekarang ...	[dari, dulu, pake, shopee, gak, kecewa, eh, sekarang, ...]
...
4995	Aplikasinya bagus!	5	aplikasinya bagus!	aplikasinya bagus!	[aplikasinya, bagus]
4996	aplikasi yang sangat bagus buat berbelanja barang	5	aplikasi yang sangat bagus buat berbelanja barang	aplikasi yang sangat bagus buat berbelanja barang	[aplikasi, yang, sangat, bagus, buat, berbelanja, barang]
4997	belanja lebih mudah	5	belanja lebih mudah	belanja lebih mudah	[belanja, lebih, mudah]
4998	saya member platinum. dalam 1 tahun dari ratus...	4	saya member platinum. dalam 1 tahun dari ratus...	saya member platinum. dalam tahun dari ratus...	[saya, member, platinum, dalam, tahun, dari, ratus, ...]
4999	membantu bngt,tp kadang ada ongkir yg mhl bngt	5	membantu bngt,tp kadang ada ongkir yg mhl bngt	membantu bngt,tp kadang ada ongkir yg mhl bngt	[membantu, bngt, tp, kadang, ada, ongkir, yg, mhl, bngt]

5000 rows X 5 columns

e. Tokenizing

	content	score	lowercase	remove_angka	remove_simbol	tokenizing
0	lama paketnya udah 5 hari di tempat yang sama ...	1	lama paketnya udah 5 hari di tempat yang sama ...	lama paketnya udah hari di tempat yang sama mulu	lama paketnya udah hari di tempat yang sama mulu	[lama, paketnya, udah, hari, di, tempat, yang, sama, mulu]
1	semoga semakin sukses dalam sejarah	5	semoga semakin sukses dalam sejarah	semoga semakin sukses dalam sejarah	semoga semakin sukses dalam sejarah	[semoga, semakin, sukses, dalam, sejarah]
2	terkadang ada pengantarnya yg kurang amanah & ...	3	terkadang ada pengantarnya yg kurang amanah & ...	terkadang ada pengantarnya yg kurang amanah & ...	terkadang ada pengantarnya yg kurang amanah & ...	[terkadang, ada, pengantarnya, yg, kurang, amanah, &]
3	mantap gas poll	5	mantap gas poll	mantap gas poll	mantap gas poll	[mantap, gas, poll]
4	Dari dulu pake shopee gak kecewa, eh sekarang ...	1	dari dulu pake shopee gak kecewa, eh sekarang ...	dari dulu pake shopee gak kecewa, eh sekarang ...	[dari, dulu, pake, shopee, gak, kecewa, eh, sekarang, ...]	[dari, dulu, pake, shopee, gak, kecewa, eh, se..., ...]
...
4995	Aplikasinya bagus!	5	aplikasinya bagus!	aplikasinya bagus!	aplikasinya bagus!	[aplikasinya, bagus]
4996	aplikasi yang sangat bagus buat berbelanja barang	5	aplikasi yang sangat bagus buat berbelanja barang	aplikasi yang sangat bagus buat berbelanja barang	aplikasi yang sangat bagus buat berbelanja barang	[aplikasi, yang, sangat, bagus, buat, berbelanja, barang]
4997	belanja lebih mudah	5	belanja lebih mudah	belanja lebih mudah	belanja lebih mudah	[belanja, lebih, mudah]
4998	saya member platinum. dalam 1 tahun dari ratus...	4	saya member platinum. dalam 1 tahun dari ratus...	saya member platinum. dalam tahun dari ratus...	saya member platinum. dalam tahun dari ratus...	[saya, member, platinum, dalam, tahun, dari, ratus, ...]
4999	membantu bngt,tp kadang ada ongkir yg mhl bngt	5	membantu bngt,tp kadang ada ongkir yg mhl bngt	membantu bngt,tp kadang ada ongkir yg mhl bngt	membantu bngt,tp kadang ada ongkir yg mhl bngt	[membantu, bngt, tp, kadang, ada, ongkir, yg, mhl, bngt]

5000 rows X 6 columns

f. Labelling

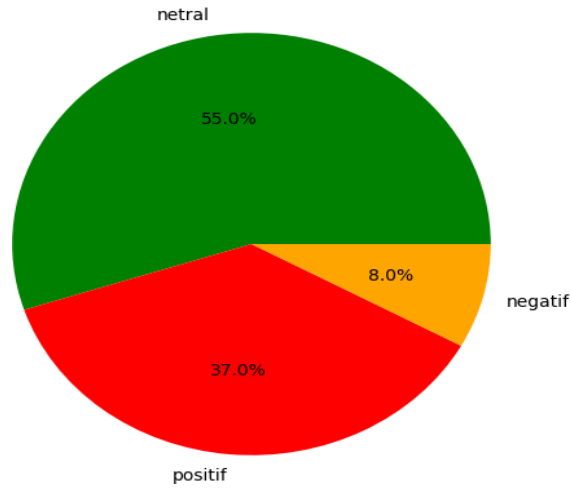
	content	score	lowercase	remove_angka	remove_simbol	tokenizing	labelling
0	lama paketnya udah 5 hari di tempat yang sama ...	1	lama paketnya udah 5 hari di tempat yang sama ...	lama paketnya udah hari di tempat yang sama mulu	lama paketnya udah hari di tempat yang sama mulu	[lama, paketnya, udah, hari, di, tempat, yang, sama, mulu]	netral
1	semoga semakin sukses dalam sejarah	5	semoga semakin sukses dalam sejarah	semoga semakin sukses dalam sejarah	semoga semakin sukses dalam sejarah	[semoga, semakin, sukses, dalam, sejarah]	positif
2	terkadang ada pengantarnya yg kurang amanah & ...	3	terkadang ada pengantarnya yg kurang amanah & ...	terkadang ada pengantarnya yg kurang amanah & ...	terkadang ada pengantarnya yg kurang amanah & ...	[terkadang, ada, pengantarnya, yg, kurang, amanah, &]	negatif
3	mantap gas poll	5	mantap gas poll	mantap gas poll	mantap gas poll	[mantap, gas, poll]	positif
4	Dari dulu pake shopee gak kecewa, eh sekarang ...	1	dari dulu pake shopee gak kecewa, eh sekarang ...	dari dulu pake shopee gak kecewa, eh sekarang ...	[dari, dulu, pake, shopee, gak, kecewa, eh, sekarang, ...]	[dari, dulu, pake, shopee, gak, kecewa, eh, se..., ...]	positif
...
4995	Aplikasinya bagus!	5	aplikasinya bagus!	aplikasinya bagus!	aplikasinya bagus!	[aplikasinya, bagus]	positif
4996	aplikasi yang sangat bagus buat berbelanja barang	5	aplikasi yang sangat bagus buat berbelanja barang	aplikasi yang sangat bagus buat berbelanja barang	aplikasi yang sangat bagus buat berbelanja barang	[aplikasi, yang, sangat, bagus, buat, berbelanja, barang]	positif
4997	belanja lebih mudah	5	belanja lebih mudah	belanja lebih mudah	belanja lebih mudah	[belanja, lebih, mudah]	netral
4998	saya member platinum. dalam 1 tahun dari ratus...	4	saya member platinum. dalam 1 tahun dari ratus...	saya member platinum. dalam tahun dari ratus...	saya member platinum. dalam tahun dari ratus...	[saya, member, platinum, dalam, tahun, dari, ratus, ...]	netral
4999	membantu bngt,tp kadang ada ongkir yg mhl bngt	5	membantu bngt,tp kadang ada ongkir yg mhl bngt	membantu bngt,tp kadang ada ongkir yg mhl bngt	membantu bngt,tp kadang ada ongkir yg mhl bngt	[membantu, bngt, tp, kadang, ada, ongkir, yg, mhl, bngt]	netral

5000 rows X 7 columns

7. Labelling Data

	content	score	lowercase	remove_angka	remove_titahai	tokenizing	label
0	suka	5	suka	suka	suka	[suka]	netral
1	lama paketrnya udah 5 hari di tempat yang sama...	1	lama paketrnya udah 5 hari di tempat yang sama...	lama paketrnya udah hari di tempat yang sama...	lama paketrnya udah hari di tempat yang sama...	[lama, paketrnya, udah, hari, di, tempat, yang, ...]	netral
2	semoga semakin sukses dalam sejarah	5	semoga semakin sukses dalam sejarah	semoga semakin sukses dalam sejarah	semoga semakin sukses dalam sejarah	[semoga, semakin, sukses, dalam, sejarah]	positif
3	terkadang ada pengantarnya yg kurang amanah & ...	3	terkadang ada pengantarnya yg kurang amanah & ...	terkadang ada pengantarnya yg kurang amanah & ...	terkadang ada pengantarnya yg kurang amanah & ...	[terkadang, ada, pengantarnya, yg, kurang, amanah, &, ...]	negatif
4	mantap gas poli	5	mantap gas poli	mantap gas poli	mantap gas poli	[mantap, gas, poli]	positif
...
4995	mantap	5	mantap	mantap	mantap	[mantap]	positif
4996	Aplikasinya bagus!	5	aplikasinya bagus!	aplikasinya bagus!	aplikasinya bagus!	[aplikasinya, bagus]	positif
4997	aplikasi yang sangat bagus buat berbelanja barang	5	aplikasi yang sangat bagus buat berbelanja barang	aplikasi yang sangat bagus buat berbelanja barang	aplikasi yang sangat bagus buat berbelanja barang	[aplikasi, yang, sangat, bagus, buat, berbelanja, barang]	positif
4998	belanja lebih mudah	5	belanja lebih mudah	belanja lebih mudah	belanja lebih mudah	[belanja, lebih, mudah]	netral
4999	saya member platinum, dalam 1 tahun dari ratus...	4	saya member platinum, dalam 1 tahun dari ratus...	saya member platinum, dalam tahun dari ratus...	saya member platinum dalam tahun dari ratus...	[saya, member, platinum, dalam, tahun, dari, r, ...]	netral

Persentase Label Komentar



	mentah	bersih	label
0	suka	[suka]	netral
1	lama paketrnya udah 5 hari di tempat yang sama ...	[lama, paketrnya, udah, hari, di, tempat, yang, ...]	netral
2	semoga semakin sukses dalam sejarah	[semoga, semakin, sukses, dalam, sejarah]	positif
3	terkadang ada pengantarnya yg kurang amanah & ...	[terkadang, ada, pengantarnya, yg, kurang, ama...]	negatif
4	mantap gas poli	[mantap, gas, poli]	positif
...
4995	mantap	[mantap]	positif
4996	Aplikasinya bagus!	[aplikasinya, bagus]	positif
4997	aplikasi yang sangat bagus buat berbelanja barang	[aplikasi, yang, sangat, bagus, buat, berbelan...]	positif
4998	belanja lebih mudah	[belanja, lebih, mudah]	netral
4999	saya member platinum, dalam 1 tahun dari ratus...	[saya, member, platinum, dalam, tahun, dari, r, ...]	netral

komparasi sebelum dan sesudah preprocessing serta labelling

8. Fitur Ekstraksi (Bag Of Word)

```

13. Extraksi Fitur use (Bag Of Word)

1 teks = dataset["bersih"].astype(str).tolist()
2
3 vectorizer = CountVectorizer()
4 X = vectorizer.fit_transform(teks)
5
6 fitur = vectorizer.get_feature_names_out()
7
8 print(fitur)
9 print(X.toarray())

*** ['aa' 'aamin' 'abang' ... 'zlama' 'zonk' 'zosh']
[[0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 ...
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]]
    
```

9. Naïve Bayes Classifier

```

14. Naive Bayes Classifier

349] 1 # split data x dan y
      2 y = dataset['label']
      3 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

350] 1 # split data
      2 model = MultinomialNB()
      3 model.fit(X_train, y_train)

***
- MultinomialNB
  MultinomialNB()

```

10. Evaluasi Model

```

15. Prediksi dan Evaluasi Model

77] 1 y_pred = model.predict(X_test)
      2 akurasi = accuracy_score(y_test, y_pred)
      3
      4 akurasi = (akurasi * 100)
      5
      6 print("Akurasi Klasifikasi:", akurasi, "\t")
      7 print("Satuan persen : ", akurasi, "%")
      8 print("\t")
      9
      10 print(classification_report(y_test, y_pred))
      11

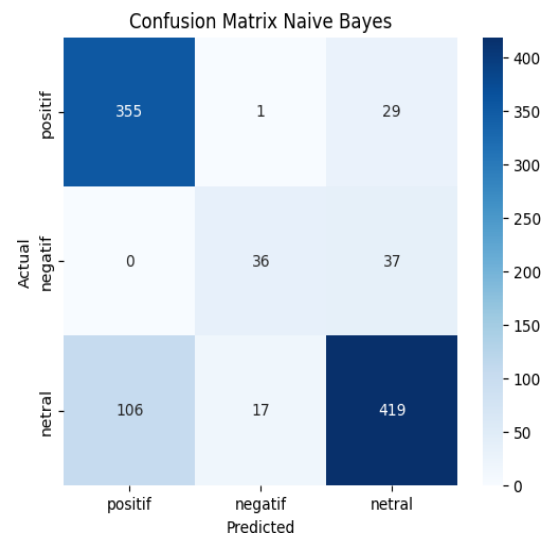
*** Akurasi Klasifikasi: 81.0
     Satuan persen : 81.0 %

              precision    recall  f1-score   support

negatif      0.67       0.49       0.57         73
netral       0.86       0.77       0.82        542
positif      0.77       0.92       0.84        385

accuracy          0.77         0.73         0.81       1000
macro avg         0.77         0.73         0.74       1000
weighted avg      0.81         0.81         0.81       1000

```



5. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, diperoleh 5.000 data ulasan yang di scrapping dari playstore menggunakan library “google_play_scraper”, data tersebut diproses ke tahapan pre-processing yang meliputi (menghapus kolom, mengecilkan huruf (lowercase), menghapus angka, menghapus simbol, tokenizing, serta labelling data). Dalam penelitian ini, pembobotan kata menggunakan Bag Of Word. Dan hasil klasifikasi ulasan menggunakan algoritma Naïve Bayes Classifier diperoleh hasil akurasi mencapai 81% dan menurut batas ambang minimal kategori baik dalam klasifikasi yaitu 75% . maka, model ini terkategori baik dalam mengklasifikasi ulasan komentar.

6. Daftar Pustaka

- [1] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiarsari, “Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode Decision Tree Berbasis SMOTE,” *Aiti*, vol. 18, no. 2, pp. 173–184, 2021, doi: 10.24246/aiti.v18i2.173-184.
- [2] B. Bustami, “Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi,” *TECHSI - J. Tek. Inform.*, vol. 5, no. 2, 2013, doi: 10.29103/techsi.v5i2.154.
- [3] Sugiarto, D., Sari, S., Ariwibowo, A. B., Nabilah Putri, F., Mulya, D., Aulia, T., & Zaki, A. N. (2023). *PERBANDINGAN KINERJA KLASIFIKASI SENTIMEN ULASAN PRODUK PEMBELIAN BERAS DI MARKETPLACE SHOPEE*. 17(1). <https://doi.org/10.47111/JTI>
- [4] Sugiarto, D., Utami, E., & Yaqin, A. (n.d.). *Perbandingan Kinerja Model TF-IDF dan BOW untuk Klasifikasi Opini Publik Tentang Kebijakan BLT Minyak Goreng*.
- [5] Rusdi Rahman, M., & Febri Diansyah, A. (2024). Sentiment Analysis on the Shopee Application on Playstore Using the Random Forest Classification Method. *Inform : Jurnal Ilmiah Bidang Teknologi Informasi Dan Komunikasi*, 9(1). <https://doi.org/10.25139/inform.v9i1.5465>
- [6] Muara Sains, J., Ilmu Kesehatan, dan, Trisari Harsanti Putri, W., & Hendrowati, R. (2018). *PENGGALIAN TEKS DENGAN MODEL BAG OF WORDS TERHADAP DATA TWITTER*. 2(1), 129–138.
- [7] Utama, H., & Masruro, A. (2022). *Analisis Sentimen pada Twitter menggunakan Word Embedding dengan Pendekatan Word2Vec*.
- [8] Rusdi Rahman, M., & Febri Diansyah, A. (2024). Analisis Sentimen pada Aplikasi Shopee di Playstore

- Menggunakan Metode Klasifikasi Random Forest. *Jurnal Ilmiah Bidang Teknologi Informasi Dan Komunikasi*, 9(1). <https://doi.org/10.25139/inform.v9i1.5465>
- [9] Suswadi, S., & Erkamim, Moh. (2023). Sentiment Analysis of Shopee App Reviews Using Random Forest and Support Vector Machine. *ILKOM Jurnal Ilmiah*, 15(3), 427–435. <https://doi.org/10.33096/ilkom.v15i3.1610.427-435>
- [10] Saepudin, A., Faqih, A., & Dwilestari, G. (n.d.). *Perbandingan Algoritma Klasifikasi Support Vector Machine, Random Forest dan Logistic Regression Pada Ulasan Shopee* (Vol. 18, Issue 1).
- [11] Rozy, F., Rangkuti, S., Fauzi, M. A., Sari, Y. A., Dewi, E., & Sari, L. (2018). *Analisis Sentimen Opini Film Menggunakan Metode Naïve Bayes dengan Ensemble Feature dan Seleksi Fitur Pearson Correlation Coefficient* (Vol. 2, Issue 12). <http://j-ptiik.ub.ac.id>
- [12] Khomsah, S., & Agus Sasmito Aribowo. (2020). Text-Preprocessing Model Youtube Comments in Indonesian. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(4), 648–654. <https://doi.org/10.29207/resti.v4i4.2035>
- [13] Munir, M. M., Fauzi, M. A., & Perdana, R. S. (2018). *Implementasi Metode Backpropagation Neural Network berbasis Lexicon Based Features dan Bag of Words Untuk Identifikasi Ujaran Kebencian Pada Twitter* (Vol. 2, Issue 10). <http://j-ptiik.ub.ac.id>
- [14] Kusuma Pangkasidhi, M., Novianus Palit, H., & Gunawan, A. (n.d.). *Analisis Sentimen Mahasiswa di Surabaya Terhadap Pelayanan Vaksinasi COVID-19 Menggunakan Beberapa Classifier*.
- [15] Cahyaningtyas, C., Nataliani, Y., & Widiyari, I. R. (2021). Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE. *AITI: Jurnal Teknologi Informasi*, 18(Agustus), 173–184.
- [16] Muara Sains, J., Ilmu Kesehatan, dan, Trisari Harsanti Putri, W., & Hendrowati, R. (2018). *PENGGALIAN TEKS DENGAN MODEL BAG OF WORDS TERHADAP DATA TWITTER*. 2(1), 129–138.