

BAB 4

HASIL PENELITIAN

4.1 RINGKASAN HASIL PENELITIAN

Bab ini menyajikan hasil prediksi sentimen ulasan produk Maula Hijab di Tokopedia menggunakan metode Naïve Bayes. Data ulasan diperoleh melalui studi pustaka dan teknik *web scraping*.

4.2 HASIL PENELITIAN

Selama periode 9 Mei hingga 6 Juni 2024, data ulasan pelanggan Maula Hijab dikumpulkan dari Tokopedia.com/hijabmaula. Data yang mencakup periode 6 bulan di tahun 2024 ini akan dianalisis untuk memahami sentimen pelanggan terhadap produk Maula Hijab.

1. Hasil Pengumpulan Data

Dengan pengambilan data melalui metode Studi Pustaka dan *web scraping*, penelitian ini berhasil memperoleh data yang diperlukan dengan cepat. Hasil dari pengumpulan data tersebut telah disajikan dalam gambar 4.1

ULASAN	RATING
mantep banget bahannya bagus dan ukurannya panjang, pengiriman juga cepet	5
mantep banget bahannya bagus dan panjang, pengiriman juga cepet	5
bagus bahannya adem dan premium, pengiriman juga cepet	5
bagus bahannya adem dan ukurannya panjang, pengiriman juga cepet	5
terima kasih kak sdh di terima dgn baik dan sesuai pesanan	4
terima kasih kak sdh di terima dgn baik dan sesuai pesanan	4
bahan nya tipis... model nya gak sama kyk difoto	4
bahan nya tipis... model nya gak sama kyk difoto	5
aneh banget kerudungnya inner cuma dagu, engga nutup leher, tidak sesuai gambar inner juga gedhe banget, padahal kepala ku	1
kainnya kurang ke depan, garis atas sambungan inner juga kelihatan tebal bikin kepala kotak, sudah dikasih ke saudara ga ada yg	2
ini aku ambil yg warna beige untuk daily bener bener recommended, karena bahannya adem ga bikin gerah terus innernya tipika	5
alhamdulillah	5
alhamdulillah bagus, agak tipis, namun tidak nerawang banget...warnanya real pic	5
bahannya adem	5
Ukuran pas. Bahan lembut, bagus, adem	5
Bahannya nyaman, dipakai gampang, warna sesuai	5
Sesuai.. thank you	5
produk sesuai gambar, tinggi 170 masih cocok	4
Aneh bgt innernya cm sampe dagu	4
size kurang lebar dan kurang panjang.	4

Gambar 4.1 Hasil Pengumpulan Data.

2. Web Scraping

Proses pengumpulan data ini menggunakan Jupyter notebook dan Bahasa pemrograman Python untuk melakukan pengumpulan data pada *website* Tokopedia. Proses ini dilakukan dengan cara memasang library Selenium dan fitur *web driver* yang tersedia pada Microsoft Edge. Data yang akan di ambil dalam proses ini yaitu, Ulasan dan Rating dari pelanggan. Berikut penjelasan script yang telah di buat.

```

1 from selenium import webdriver
2 from selenium.webdriver.edge.service import Service
3 from selenium.webdriver.edge.options import Options
4 from selenium.webdriver.common.by import By
5 from selenium.webdriver.support.ui import WebDriverWait
6 from selenium.webdriver.support import expected_conditions as EC
7 from selenium.common.exceptions import TimeoutException, WebDriverException, ElementClickInterceptedException
8 from time import sleep
9 import csv
10 import random
11
12 # Fungsi untuk menginisialisasi WebDriver
13 def init_driver():
14     options = Options()
15     options.use_chromium = True
16     options.add_argument("window-size=1280,800")
17     options.add_argument("disable-gpu")
18     options.add_argument("--disable-extensions")
19     options.add_argument(f"user-agent=Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/{random.randint(70, 89)}")
20
21     service = Service(executable_path='C:\\Program Files\\Microsoft Edge\\MicrosoftEdgeDriver.exe') # Ganti path jika perlu
22     driver = webdriver.Edge(service=service, options=options)
23     return driver
24
25 # Inisialisasi driver
26 driver = init_driver()
27
28 # URL website yang ingin di-scrap
29 url = "https://www.tokopedia.com/hijabmaula/review"
30
31 # CSS selectors untuk ulasan dan rating bintang
32 ulasan_selector = "[data-unify='Typography'][class*='css-ybiclm-unf-heading etqvo2ff8'] span[data-testid='lblitemUlasan']"
33 rating_selector = "[data-testid='icnStarRating']"

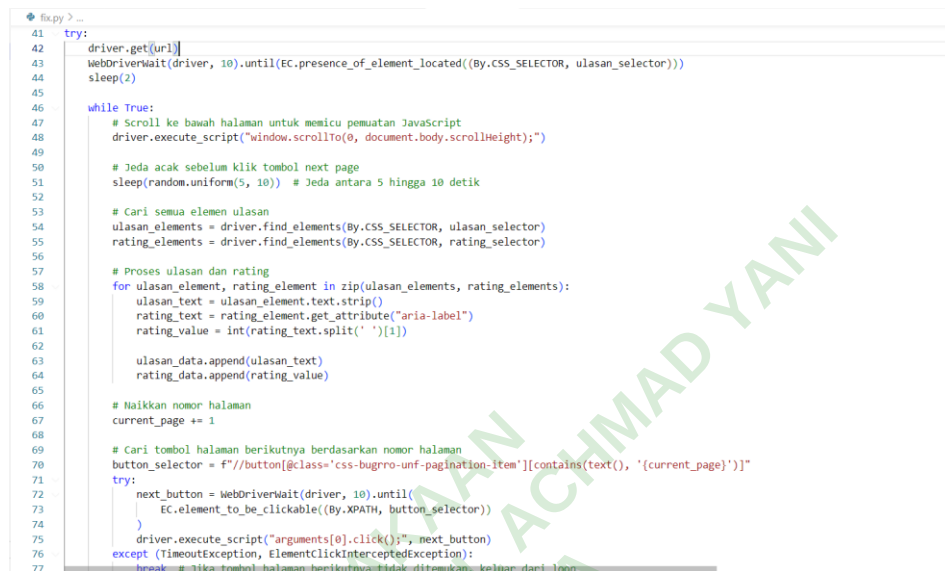
```

Gambar 4.2 Scraping Tokopedia.

Pada Gambar 4.2, fungsi *'from'* dan *'import'* digunakan untuk memanggil perpustakaan yang diperlukan dalam proses pengambilan data (scraping). Fungsi *Selenium* dan *Edge Driver* memungkinkan pengendalian browser secara otomatis—seperti membuka halaman *web*, mengisi formulir, mengklik tombol, dan lain sebagainya. Fungsi *import csv* ialah untuk setelah berhasil melakukan pengambilan data dari Tokopedia Maula Hijab file hasil *scraping* tersebut akan otomatis tersimpan dalam format csv.

Fungsi *'def init_driver()'*: diaktifkan saat program dijalankan dan akan membuka tampilan layar sesuai dengan ukuran yang diinginkan. 'Ulasan

selector’ dan *rating selector*’ merupakan data CSS yang digunakan untuk mengambil data dari halaman *website* Tokopedia Maula Hijab .



```

41 try:
42     driver.get(url)
43     WebDriverWait(driver, 10).until(EC.presence_of_element_located((By.CSS_SELECTOR, ulasan_selector)))
44     sleep(2)
45
46 while True:
47     # Scroll ke bawah halaman untuk memicu pemuatan JavaScript
48     driver.execute_script("window.scrollTo(0, document.body.scrollHeight);")
49
50     # Jeda acak sebelum klik tombol next page
51     sleep(random.uniform(5, 10)) # Jeda antara 5 hingga 10 detik
52
53     # Cari semua elemen ulasan
54     ulasan_elements = driver.find_elements(By.CSS_SELECTOR, ulasan_selector)
55     rating_elements = driver.find_elements(By.CSS_SELECTOR, rating_selector)
56
57     # Proses ulasan dan rating
58     for ulasan_element, rating_element in zip(ulasan_elements, rating_elements):
59         ulasan_text = ulasan_element.text.strip()
60         rating_text = rating_element.get_attribute("aria-label")
61         rating_value = int(rating_text.split(' ')[1])
62
63         ulasan_data.append(ulasan_text)
64         rating_data.append(rating_value)
65
66     # Naikkan nomor halaman
67     current_page += 1
68
69     # Cari tombol halaman berikutnya berdasarkan nomor halaman
70     button_selector = f"//button[@class='css-bugpro-unf-pagination-item'][contains(text(), '{current_page}'])"
71
72     try:
73         next_button = WebDriverWait(driver, 10).until(
74             EC.element_to_be_clickable((By.XPATH, button_selector))
75         )
76         driver.execute_script("arguments[0].click();", next_button)
77     except (TimeoutException, ElementClickInterceptedException):
78         break # Jika tombol halaman berikutnya tidak ditemukan, keluar dari loop

```

Gambar 4.3 *Scraping Tokopedia.*

Pada tangkapan gambar 4.3 ini fungsi penggunaan **try:**, **While True:** untuk saat program di jalankan memberikan jeda random antara 5 sampai 10 detik untuk menampilkan *website* tersebut. Fungsi **try** dan **except** adalah untuk menangani kemungkinan terjadinya *error* saat menjalankan *Selenium*, seperti halaman gagal dimuat atau ulasan yang tidak ditemukan.

3. Hasil Web Scraping

ULASAN	RATING
mantep banget bahannya bagus dan ukurannya pan...	5
mantep banget bahannya bagus dan panjang, peng...	5
bagus bahannya adem dan premium, pengiriman ju...	5
bagus bahannya adem dan ukurannya panjang, pen...	5
terima kasih kak sdh di terima dgn baik dan se...	4
...	...
lumayan oke bahan adem ga nerawanng	5
bahan tipis sekali jahitan muka hanya di keli...	5
tipis sekalii..saran ya sekelas mall harusnya ...	5
lumayanlah ini tebal tp bahan panas	4
recommended	5

Gambar 4.4 Web scraping.

Gambar 4.3 menunjukkan proses *scraping* data ulasan terbaru pada produk Maula Hijab di Tokopedia. Sebanyak 495 data ulasan berhasil dikumpulkan dan akan digunakan sebagai dasar analisis sentimen untuk mengukur tingkat kepuasan pelanggan.

4. Preprocessing Teks

ULASAN	RATING	Clean_Text
0 mantep banget bahannya bagus dan ukurannya pan...	5	mantep banget bahannya bagus dan ukurannya pan...
1 mantep banget bahannya bagus dan panjang, peng...	5	mantep banget bahannya bagus dan panjang pengi...
2 bagus bahannya adem dan premium, pengiriman ju...	5	bagus bahannya adem dan premium pengiriman jug...
3 bagus bahannya adem dan ukurannya panjang, pen...	5	bagus bahannya adem dan ukurannya panjang peng...
4 terima kasih kak sdh di terima dgn baik dan se...	4	terima kasih kak sudah di terima dengan baik d...
5 terima kasih kak sdh di terima dgn baik dan se...	4	terima kasih kak sudah di terima dengan baik d...
6 bahan nya tipis... model nya gak sama kyk difoto	4	bahan nya tipis model nya tidak sama seperti d...
7 bahan nya tipis... model nya gak sama kyk difoto	5	bahan nya tipis model nya tidak sama seperti d...
8 aneh banget kerudungnya inner cuma dagu, engga...	1	aneh banget kerudungnya inner cuma dagu tidak ...
9 kainnya kurang ke depan, qaris atas sambungan ...	2	kainnya kurang ke depan qaris atas sambungan i...

Gambar 4.5 Cleaning teks.

Dalam proses ini sebelum melakukan *preprocessing* teks kita harus melakukan pembersihan data. Pada proses ini ditambahkan fungsi seperti normalisasi kata yang di singkat dan menjadikan teks aslinya menurut kbbs, menghapus tanda baca dan menghapus nomor. Fungsi ini di lakukan agar data mudah untuk di olah. Berikut tahapan proses *Preprocessing* teks:

a. *Case Folding*.



```
def casefolding(text):
    return text.casefold()

data['casefolding'] = data['Clean_Text'].apply(casefolding)
data[['Clean_Text', 'casefolding']].head(17)
```

✓ 0.0s

	Clean_Text	casefolding
0	mantep banget bahannya bagus dan ukurannya pan...	mantep banget bahannya bagus dan ukurannya pan...
1	mantep banget bahannya bagus dan panjang pengi...	mantep banget bahannya bagus dan panjang pengi...
2	bagus bahannya adem dan premium pengiriman jug...	bagus bahannya adem dan premium pengiriman jug...
3	bagus bahannya adem dan ukurannya panjang peng...	bagus bahannya adem dan ukurannya panjang peng...
4	terima kasih kak sudah di terima dengan baik d...	terima kasih kak sudah di terima dengan baik d...
5	terima kasih kak sudah di terima dengan baik d...	terima kasih kak sudah di terima dengan baik d...
6	bahan nya tipis model nya tidak sama seperti d...	bahan nya tipis model nya tidak sama seperti d...
7	bahan nya tipis model nya tidak sama seperti d...	bahan nya tipis model nya tidak sama seperti d...
8	aneh banget kerudungnya inner cuma dagu tidak ...	aneh banget kerudungnya inner cuma dagu tidak ...
9	kainnya kurang ke depan garis atas sambungan i...	kainnya kurang ke depan garis atas sambungan i...
10	ini aku ambil yang warna beige untuk daily ben...	ini aku ambil yang warna beige untuk daily ben...
11	alhamdulillah	alhamdulillah
12	alhamdulillah bagus agak tipis namun tidak tera...	alhamdulillah bagus agak tipis namun tidak tera...
13	bahannya adem	bahannya adem
14	Ukuran pas Bahan lembut bagus adem	ukuran pas bahan lembut bagus adem
15	Bahannya nyaman dipakai qampanq warna sesuai	bahannya nyaman dipakai qampanq warna sesuai

Gambar 4.6 *Case Folding*.

Proses pertama dalam preprocessing teks ialah *case folding*. Fungsi kode di atas untuk menambahkan kolom baru dalam file excel Bernama *case folding* dan mengganti teks huruf besar menjadi huruf kecil semua.

b. Tokenisasi

```
def tokenize(text):
    return word_tokenize(text)

# Terapkan Tokenisasi
data['Tokenized'] = data['casefolding'].apply(tokenize)

# Tampilkan Data yang Sudah Ditokenisasi
data[['casefolding', 'Tokenized']].head()
```

✓ 0.0s × Tokenisasi + Tag

	casefolding	Tokenized
0	mantep banget bahannya bagus dan ukurannya pan...	[mantep, banget, bahannya, bagus, dan, ukurann...
1	mantep banget bahannya bagus dan panjang pengi...	[mantep, banget, bahannya, bagus, dan, panjang...
2	bagus bahannya adem dan premium pengiriman jug...	[bagus, bahannya, adem, dan, premium, pengirim...
3	bagus bahannya adem dan ukurannya panjang peng...	[bagus, bahannya, adem, dan, ukurannya, panjan...
4	terima kasih kak sudah di terima dengan baik d...	[terima, kasih, kak, sudah, di, terima, dengan...

Gambar 4.7 Tokenisasi.

Pada tangkapan gambar 4.7 Fungsi kode ini menambahkan satu kolom baru yaitu Tokenize, selanjutnya di isi data pada *case folding* di setiap kalimat di tambahkan kolom baru yaitu berdasarkan kata pertoken.

c. Stopword

```
# Daftar Stopwords Bahasa Indonesia
stop_words = set(stopwords.words('indonesian'))

# Fungsi Penghapusan Stopwords
def remove_stopwords(tokens):
    return [word for word in tokens if word not in stop_words]

# Terapkan Penghapusan Stopwords
data['No_Stopwords'] = data['Tokenized'].apply(remove_stopwords)

# Tampilkan Data Setelah Stopwords Dihapus
data[['Tokenized', 'No_Stopwords']].head()
```

✓ 0.0s × Stopword Removal + Tag

	Tokenized	No_Stopwords
0	[mantep, banget, bahannya, bagus, dan, ukurann...	[mantep, banget, bahannya, bagus, ukurannya, p...
1	[mantep, banget, bahannya, bagus, dan, panjang...	[mantep, banget, bahannya, bagus, pengiriman, ...
2	[bagus, bahannya, adem, dan, premium, pengirim...	[bagus, bahannya, adem, premium, pengiriman, c...
3	[bagus, bahannya, adem, dan, ukurannya, panjan...	[bagus, bahannya, adem, ukurannya, pengiriman,...
4	[terima, kasih, kak, sudah, di, terima, dengan...	[terima, kasih, kak, terima, sesuai, pesanan]

Gambar 4.8 Stopword.

Selanjutnya, *Stopword*. Tangkapan gambar 4.8 diatas program tersebut berfungsi menambahkan kolom baru dan menghapus kata yang tidak memiliki arti seperti “dan”, “di” dan seterusnya.

d. *Stemming*

```
# Fungsi Stemming
factory = StemmerFactory()
stemmer = factory.create_stemmer()

def stem_text(tokens):
    return [stemmer.stem(word) for word in tokens]

# Terapkan Stemming
data['Stemmed'] = data['No_Stopwords'].apply(stem_text)

# Tampilkan Data Setelah Stemming
data[['No_Stopwords', 'Stemmed']].head()
```

✓ 18.5s × Stemming + Tag

	No_Stopwords	Stemmed
0	[mantep, banget, bahannya, bagus, ukurannya, p...	[mantep, banget, bahan, bagus, ukur, kirim, ce...
1	[mantep, banget, bahannya, bagus, pengiriman, ...	[mantep, banget, bahan, bagus, kirim, cepat]
2	[bagus, bahannya, adem, premium, pengiriman, c...	[bagus, bahan, adem, premium, kirim, cepat]
3	[bagus, bahannya, adem, ukurannya, pengiriman,...	[bagus, bahan, adem, ukur, kirim, cepat]

Gambar 4.9 *Stemming*.

Tangkapan gambar 4.9 diatas program tersebut berfungsi menambahkan kolom baru dan mengganti kata menjadi kata dasarnya. Setelah Pengolahan data tersebut akan di simpan dalam format Microsoft Excel.

5. Hasil *Preprocessing* Teks

Setelah melalui tahapan pengolahan data teks atau *preprocessing* teks, data ulasan akan terlihat seperti pada gambar di bawah ini. Kata-kata seperti tanda baca dan angka pada teks diubah menjadi bentuk dasar kata.

1	ULASAN	RAT	Clean_Text	casefolding	Tokenized	Stopwords
2	mantep banget bahannya bagus dan ukurannya panjang	5	mantep banget bahannya bagus dan uk	mantep banget bahannya bagus dan uku	['mantep', 'banget', 'bahannya', 'bagus', 'dan', 'ukurannya']	['mantep', 'banget',
3	mantep banget bahannya bagus dan panjang, pengirim	5	mantep banget bahannya bagus dan pa	mantep banget bahannya bagus dan pan	['mantep', 'banget', 'bahannya', 'bagus', 'dan', 'panjang']	['mantep', 'banget',
4	bagus bahannya adem dan premium, pengiriman juga	5	bagus bahannya adem dan premium pe	bagus bahannya adem dan premium per	['bagus', 'bahannya', 'adem', 'dan', 'premium', 'pengirim']	['bagus', 'bahannya
5	bagus bahannya adem dan ukurannya panjang, pengi	5	bagus bahannya adem dan ukurannya g	bagus bahannya adem dan ukurannya pa	['bagus', 'bahannya', 'adem', 'dan', 'ukurannya', 'panjang']	['bagus', 'bahannya
6	terima kasih kak sdh di terima dgn baik dan sesuai pes	4	terima kasih kak sudah di terima dengai	terima kasih kak sudah di terima dengan	['terima', 'kasih', 'kak', 'sudah', 'di', 'terima', 'dengan', 'b']	['terima', 'kasih', 'ka
7	terima kasih kak sdh di terima dgn baik dan sesuai pes	4	terima kasih kak sudah di terima dengan	terima kasih kak sudah di terima dengan	['terima', 'kasih', 'kak', 'sudah', 'di', 'terima', 'dengan', 'b']	['terima', 'kasih', 'ka
8	bahan nya tipis... model nya gak sama kyk difoto	4	bahan nya tipis model nya tidak sama s	bahan nya tipis model nya tidak sama se	['bahan', 'nya', 'tipis', 'model', 'nya', 'tidak', 'sama', 'sepe']	['bahan', 'nya', 'tipis
9	bahan nya tipis... model nya gak sama kyk difoto	5	bahan nya tipis model nya tidak sama s	bahan nya tipis model nya tidak sama se	['bahan', 'nya', 'tipis', 'model', 'nya', 'tidak', 'sama', 'sepe']	['bahan', 'nya', 'tipis
10	aneh banget kerudungnya inner cuma dagu, engga nu	1	aneh banget kerudungnya inner cuma c	aneh banget kerudungnya inner cuma da	['aneh', 'banget', 'kerudungnya', 'inner', 'cuma', 'dagu', '']	['aneh', 'banget', 'ki
11	kainnya kurang ke depan, garis atas sambungan inner	2	kainnya kurang ke depan garis atas sam	kainnya kurang ke depan garis atas samb	['kainnya', 'kurang', 'ke', 'depan', 'garis', 'atas', 'sambun']	['kainnya', 'garis', 's
12	ini aku ambil yg warna beige untuk daily bener bener	5	ini aku ambil yang warna beige untuk d	ini aku ambil yang warna beige untuk da	['ini', 'aku', 'ambil', 'yang', 'warna', 'beige', 'untuk', 'daily']	['ambil', 'warna', 'be
13	alhamdulillah	5	alhamdulillah	alhamdulillah	['alhamdulillah']	['alhamdulillah']
14	alhamdulillah bagus, agak tipis, namun tidak nerawang	5	alhamdulillah bagus agak tipis namun ti	alhamdulillah bagus agak tipis namun tid	['alhamdulillah', 'bagus', 'agak', 'tipis', 'namun', 'tidak', '']	['alhamdulillah', 'bag
15	bahannya adem	5	bahannya adem	bahannya adem	['bahannya', 'adem']	['bahannya', 'adem']

Gambar 4.10 Hasil *Preprocessing* teks.

Pada gambar 4.5 setelah tahapan *preprocessing* selesai, output hasil file siap digunakan untuk analisis atau pemodelan lebih lanjut. *Output* ini memiliki format yang lebih terstruktur dan bersih dibandingkan data mentah awal.

6. Pemberian Label pada Data

Pada proses ini, pemberian label pada data dilakukan secara otomatis dengan *machine learning* dan manual. *Machine learning* digunakan untuk menerapkan sentimen pada kolom teks ulasan dan kolom rating yang telah diberikan oleh pelanggan Maula Hijab di Tokopedia. Proses manual dilakukan untuk memeriksa dan memastikan tidak ada kesalahan dalam pelabelan yang dilakukan oleh penerapan *machine learning*.

a. Penjelasan kode

Pada tangkapan Gambar 4.11 di bawah fungsi *analyze_sentiment* ditambahkan untuk menerima dua parameter: text (teks yang akan dianalisis) dan rating (nilai rating). Fungsi ini mengkategorikan sentimen teks berdasarkan kolom "*Case folding*" dan "*Rating*".

```
# Inisialisasi pipeline sentiment-analysis dengan model Bahasa Indonesia
sentiment_pipeline = pipeline("sentiment-analysis", model="w11wo/indonesian-roberta-base-sentiment-classifier")

# Fungsi untuk menganalisis sentimen berdasarkan rating dan teks
def analyze_sentiment_based_on_rating(text, rating):
    if rating > 3:
        return "Positif"
    elif rating < 3:
        return "Negatif"
    else:
        # Jika rating 3, gunakan pipeline untuk menganalisis teks
        result = sentiment_pipeline(text)[0]
        return "Positif" if result['label'] == "positive" else "Negatif"

# Misalkan df adalah DataFrame yang sudah di-load sebelumnya
df['Kategori Sentimen'] = df.apply(lambda row: analyze_sentiment_based_on_rating(row['Case folding'], row['Rating']), axis=1)

output_file_path = "C:/New/hasil_labelling_.xlsx"
df.to_excel(output_file_path, index=False)
```

Gambar 4.11 Pelabelan data.

Jika rating di atas 3, sentimennya akan dianggap "Positif". Jika rating di bawah 3, sentimennya dianggap "Negatif". Namun, jika ratingnya tepat 3, model bahasa akan menganalisis teks ulasan untuk menentukan sentimennya. Hasil analisis kemudian digunakan untuk mengelompokkan ulasan ke dalam kategori "Positif" atau "Negatif", lalu hasil program akan di simpan dalam format Excel bernama hasil_labelling.

b. Hasil labelling menggunakan *machine learning*.

Ulasan	Rating	Clean_Text	Case folding	Tokenized	Stopwords	Stemmed	Hasil preprocessing	Kategori Senti
mante ba	5	mante ba mante ba mante ba	mante ba mante ba mante ba	['mante', 'ba', 'mante', 'ba', 'mante', 'ba']	['mante', 'ba']	['mante', 'ba']	mante ba mante ba mante ba	Positif
mante ba	5	mante ba mante ba mante ba	mante ba mante ba mante ba	['mante', 'ba', 'mante', 'ba', 'mante', 'ba']	['mante', 'ba']	['mante', 'ba']	mante ba mante ba mante ba	Positif
bagus bahu	5	bagus bahu bagus bahu bagus	bagus bahu bagus bahu bagus	['bagus', 'bahu', 'bagus', 'bahu', 'bagus', 'bahu']	['bagus', 'bahu']	['bagus', 'bahu']	bagus bahu bagus bahu bagus	Positif
bagus bahu	5	bagus bahu bagus bahu bagus	bagus bahu bagus bahu bagus	['bagus', 'bahu', 'bagus', 'bahu', 'bagus', 'bahu']	['bagus', 'bahu']	['bagus', 'bahu']	bagus bahu bagus bahu bagus	Positif
terima kas	4	terima kas terima kasih kak sudah di terima dengan k	terima kas terima kasih kak sudah di terima dengan k	['terima', 'kas', 'terima', 'kas', 'sudah', 'di', 'terima', 'dengan', 'k']	['terima', 'kas', 'sudah', 'di', 'terima', 'dengan', 'k']	['terima', 'kas', 'sudah', 'di', 'terima', 'dengan', 'k']	terima kas terima kasih kak sudah di terima dengan k	Positif
terima kas	4	terima kas terima kasih kak sudah di terima dengan k	terima kas terima kasih kak sudah di terima dengan k	['terima', 'kas', 'terima', 'kas', 'sudah', 'di', 'terima', 'dengan', 'k']	['terima', 'kas', 'sudah', 'di', 'terima', 'dengan', 'k']	['terima', 'kas', 'sudah', 'di', 'terima', 'dengan', 'k']	terima kas terima kasih kak sudah di terima dengan k	Positif
bahan nya	4	bahan nya bahan nya tipis model nya tidak sama sep	bahan nya bahan nya tipis model nya tidak sama sep	['bahan', 'nya', 'bahan', 'nya', 'tipis', 'model', 'nya', 'tidak', 'sama', 'sep']	['bahan', 'nya', 'tipis', 'model', 'nya', 'tidak', 'sama', 'sep']	['bahan', 'nya', 'tipis', 'model', 'nya', 'tidak', 'sama', 'sep']	bahan nya bahan nya tipis model nya tidak sama sep	Positif
bahan nya	5	bahan nya bahan nya tipis model nya tidak sama sep	bahan nya bahan nya tipis model nya tidak sama sep	['bahan', 'nya', 'bahan', 'nya', 'tipis', 'model', 'nya', 'tidak', 'sama', 'sep']	['bahan', 'nya', 'tipis', 'model', 'nya', 'tidak', 'sama', 'sep']	['bahan', 'nya', 'tipis', 'model', 'nya', 'tidak', 'sama', 'sep']	bahan nya bahan nya tipis model nya tidak sama sep	Positif
aneh bang	1	aneh bang aneh banget kerudung nya inner cuma dag	aneh bang aneh banget kerudung nya inner cuma dag	['aneh', 'bang', 'aneh', 'banget', 'kerudung', 'nya', 'inner', 'cuma', 'dag']	['aneh', 'bang', 'aneh', 'banget', 'kerudung', 'nya', 'inner', 'cuma', 'dag']	['aneh', 'bang', 'aneh', 'banget', 'kerudung', 'nya', 'inner', 'cuma', 'dag']	aneh bang aneh banget kerudung nya inner cuma dag	Negatif
kainnya ku	2	kainnya ku kainnya kurang ke depan garis atas sambu	kainnya ku kainnya kurang ke depan garis atas sambu	['kainnya', 'ku', 'kainnya', 'kurang', 'ke', 'depan', 'garis', 'atas', 'sambu']	['kainnya', 'ku', 'kainnya', 'kurang', 'ke', 'depan', 'garis', 'atas', 'sambu']	['kainnya', 'ku', 'kainnya', 'kurang', 'ke', 'depan', 'garis', 'atas', 'sambu']	kainnya ku kainnya kurang ke depan garis atas sambu	Negatif
ini aku ami	5	ini aku ami ini aku ambil yang warna beige untuk daili	ini aku ami ini aku ambil yang warna beige untuk daili	['ini', 'aku', 'ami', 'ini', 'aku', 'ambil', 'yang', 'warna', 'beige', 'untuk', 'daili']	['ini', 'aku', 'ami', 'ini', 'aku', 'ambil', 'yang', 'warna', 'beige', 'untuk', 'daili']	['ini', 'aku', 'ami', 'ini', 'aku', 'ambil', 'yang', 'warna', 'beige', 'untuk', 'daili']	ini aku ami ini aku ambil yang warna beige untuk daili	Positif
alhamdulillah	5	alhamdulillah alhamdulillah	alhamdulillah alhamdulillah	['alhamdulillah', 'alhamdulillah']	['alhamdulillah']	['alhamdulillah']	alhamdulillah alhamdulillah	Positif
alhamdulillah	5	alhamdulillah alhamdulillah bagus agak tipis namun tidak	alhamdulillah alhamdulillah bagus agak tipis namun tidak	['alhamdulillah', 'alhamdulillah', 'bagus', 'agak', 'tipis', 'namun', 'tidak']	['alhamdulillah', 'alhamdulillah', 'bagus', 'agak', 'tipis', 'namun', 'tidak']	['alhamdulillah', 'alhamdulillah', 'bagus', 'agak', 'tipis', 'namun', 'tidak']	alhamdulillah alhamdulillah bagus agak tipis namun tidak	Positif
bahannya	5	bahannya bahannya adem	bahannya bahannya adem	['bahannya', 'bahannya', 'adem']	['bahannya', 'bahannya', 'adem']	['bahannya', 'bahannya', 'adem']	bahannya bahannya adem	Positif
Ukuran pa	5	Ukuran pa ukuran pas bahan lembut bagus adem	ukuran pa ukuran pas bahan lembut bagus adem	['ukuran', 'pa', 'ukuran', 'pas', 'bahan', 'lembut', 'bagus', 'adem']	['ukuran', 'pa', 'ukuran', 'pas', 'bahan', 'lembut', 'bagus', 'adem']	['ukuran', 'pa', 'ukuran', 'pas', 'bahan', 'lembut', 'bagus', 'adem']	ukuran pa ukuran pas bahan lembut bagus adem	Positif
Bahannya	5	Bahannya bahannya nyaman dipakai gampang warn	bahannya bahannya nyaman dipakai gampang warn	['bahannya', 'bahannya', 'nyaman', 'dipakai', 'gampang', 'warn']	['bahannya', 'bahannya', 'nyaman', 'dipakai', 'gampang', 'warn']	['bahannya', 'bahannya', 'nyaman', 'dipakai', 'gampang', 'warn']	bahannya bahannya nyaman dipakai gampang warn	Positif
Sesuai...	5	Sesuai thai sesuai thank you	Sesuai thai sesuai thank you	['sesuai', 'thai', 'sesuai', 'thank', 'you']	['sesuai', 'thai', 'sesuai', 'thank', 'you']	['sesuai', 'thai', 'sesuai', 'thank', 'you']	Sesuai thai sesuai thank you	Positif
produk bgt	4	produk ses produk sesuai gambar tinggi masih cocok	produk ses produk sesuai gambar tinggi masih cocok	['produk', 'ses', 'produk', 'sesuai', 'gambar', 'tinggi', 'masih', 'cocok']	['produk', 'ses', 'produk', 'sesuai', 'gambar', 'tinggi', 'masih', 'cocok']	['produk', 'ses', 'produk', 'sesuai', 'gambar', 'tinggi', 'masih', 'cocok']	produk ses produk sesuai gambar tinggi masih cocok	Positif
Aneh bgt ti	4	Aneh bang aneh banget kerudung nya inner cuma dagu	aneh bang aneh banget kerudung nya inner cuma dagu	['aneh', 'bang', 'aneh', 'banget', 'kerudung', 'nya', 'inner', 'cuma', 'dagu']	['aneh', 'bang', 'aneh', 'banget', 'kerudung', 'nya', 'inner', 'cuma', 'dagu']	['aneh', 'bang', 'aneh', 'banget', 'kerudung', 'nya', 'inner', 'cuma', 'dagu']	aneh bang aneh banget kerudung nya inner cuma dagu	Positif
size kurang	4	size kurang size kurang lebar dan kurang panjang	size kurang size kurang lebar dan kurang panjang	['size', 'kurang', 'size', 'kurang', 'lebar', 'dan', 'kurang', 'panjang']	['size', 'kurang', 'size', 'kurang', 'lebar', 'dan', 'kurang', 'panjang']	['size', 'kurang', 'size', 'kurang', 'lebar', 'dan', 'kurang', 'panjang']	size kurang size kurang lebar dan kurang panjang	Positif
seperti bia	5	seperti bia seperti biasa beli di sini karna jahitan rapi	seperti bia seperti biasa beli di sini karna jahitan rapi	['seperti', 'bia', 'seperti', 'biasa', 'beli', 'di', 'sini', 'karna', 'jahitan', 'rapi']	['seperti', 'bia', 'seperti', 'biasa', 'beli', 'di', 'sini', 'karna', 'jahitan', 'rapi']	['seperti', 'bia', 'seperti', 'biasa', 'beli', 'di', 'sini', 'karna', 'jahitan', 'rapi']	seperti bia seperti biasa beli di sini karna jahitan rapi	Positif
bagus ra	5	bagus rapi bagus rapi jahitannya	bagus rapi bagus rapi jahitannya	['bagus', 'rapi', 'bagus', 'rapi', 'jahitannya']	['bagus', 'rapi', 'bagus', 'rapi', 'jahitannya']	['bagus', 'rapi', 'bagus', 'rapi', 'jahitannya']	bagus rapi bagus rapi jahitannya	Positif
kualitas m	4	kualitas m kualitas menurun sekarang bahan lebih tij	kualitas m kualitas menurun sekarang bahan lebih tij	['kualitas', 'm', 'kualitas', 'menurun', 'sekarang', 'bahan', 'lebih', 'tij']	['kualitas', 'm', 'kualitas', 'menurun', 'sekarang', 'bahan', 'lebih', 'tij']	['kualitas', 'm', 'kualitas', 'menurun', 'sekarang', 'bahan', 'lebih', 'tij']	kualitas m kualitas menurun sekarang bahan lebih tij	Positif

Gambar 4.12 Gambar Hasil Pelabelan.

Pada gambar 4.12 setelah menjalankan program hasil pelabelan data akan seperti di atas.

c. Hasil penerapan labelling manual.

Setelah melakukan pelabelan otomatis menggunakan *machine learning* pada Gambar 4.12, dilakukan juga pelabelan manual untuk memastikan akurasi hasil pelabelan. Pelabelan manual ini berfungsi sebagai langkah verifikasi untuk menghindari kesalahan yang mungkin terjadi dalam proses pelabelan otomatis.

Kategori Positif	Kategori Negatif
sesuai pesanan.	pendek banget depan nya tidak sesuai deskripsi.
bahannya nyaman adem nggak menerawang semua jahitannya rapi aku suka banget sama warnanya.	bahan nya tipis model nya tidak sama seperti difoto.
mantep banget bahannya bagus dan ukurannya panjang pengiriman juga cepat.	jelek model nya penjual d chat krn kurang pham pun nggk d bles chat nya
bahannya nyaman dipakai gampang warna sesuai.	bahan nya tipis banget ya ampun panas juga.

Pada proses ini setelah melakukan pelabelan pada data ialah menampilkan teks visualisasi, pada tahapan ini untuk melakukan visualisasi peneliti menggunakan *library* Word Cloud. Fungsi ini

bertujuan untuk melihat teks apa yang sering muncul pada ulasan yang telah di berikan oleh pelanggan.

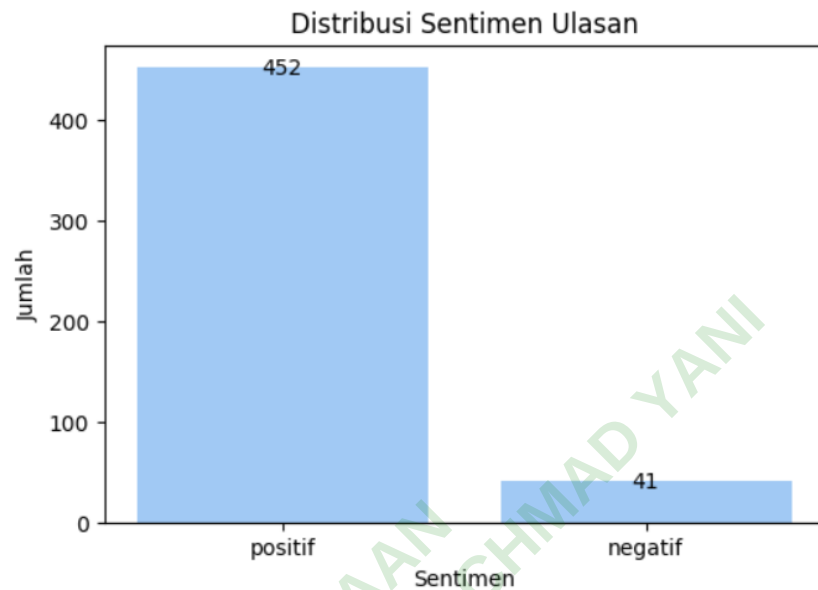
a. Kategori Positif



Gambar 4.14 Hasil Visualisasi teks positif.

Hasil visualisasi dalam gambar 4.14. Visualisasi ini, kata-kata yang berukuran lebih besar seperti "bagus," "bahan," "banget," dan "warna" menunjukkan bahwa kata-kata tersebut paling sering digunakan dalam ulasan atau komentar positif. Kata-kata lain yang berukuran lebih kecil, seperti "nyaman," "adem," "sesuai," dan "kiriman," juga muncul cukup sering, tetapi frekuensinya lebih rendah dibandingkan kata-kata yang lebih besar

c. Penyaluran teks pada kelas kategori



Gambar 4.16 Distribusi setiap kategori teks.

Dari tangkapan gambar 4.14 setelah melalui tahapan pelabelan data ulasan dapat diidentifikasi bahwa kelas positif berjumlah 452 data dan kelas negatif berjumlah 41 data.

8. Pembobotan Data

Setelah melalui tahapan visualisasi pengolahan data, pembobotan kata dilakukan menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*).

term	TF	TF-IDF
0 0 {'mantep': 0.14285714285714285, 'banget', ...	{'mantep': 0.7291318897885446, 'banget', ...	
1 1 {'mantep': 0.16666666666666666, 'banget', ...	{'mantep': 0.8506538714199687, 'banget', ...	
2 2 {'bagus': 0.16666666666666666, 'bahan', ...	{'bagus': 0.3993121712362671, 'bahan', ...	
3 3 {'bagus': 0.16666666666666666, 'bahan', ...	{'bagus': 0.3993121712362671, 'bahan', ...	
4 4 {'terima': 0.16666666666666666, 'kasih', ...	{'terima': 0.6675518233086171, 'kasih', ...	
...	...	...
489 489 {'lumayan': 0.2, 'oke': 0.2, 'bahan', ...	{'lumayan': 0.8246187951016174, 'oke', ...	
490 490 {'bahan': 0.1, 'tipis': 0.1, 'jahit', ...	{'bahan': 0.1520404290063703, 'tipis', ...	
491 491 {'tipis': 0.08333333333333333, 'sekali', ...	{'tipis': 0.36756467066332227, 'sekali', ...	
492 492 {'lumayan': 0.25, 'tebel': 0.25, 'baha...	{'lumayan': 1.0307734938770217, 'tebel', ...	
493 493 {'recommended': 1.0}	{'recommended': 5.5093883366279774}	

494 rows × 3 columns

Gambar 4.17 Pembobotan kata TF-IDF.

Pada Gambar 4.15, menggunakan persamaan (2) dan (3) untuk menghitung bobot masing-masing kata. Hasil dari proses ini adalah identifikasi kata-kata yang paling penting dan relevan dalam dokumen atau kumpulan dokumen.

9. Klasifikasi Naïve Bayes

Pada proses ini untuk melatih model setelah melakukan pembobotan menggunakan TF-IDF yaitu, melakukan klasifikasi menggunakan metode Naïve Bayes. Proses ini menggunakan data latih dan data uji sebagai tertera pada table di bawah ini.

a. Penjelasan kode

```
X = df['Stemmed']
y = df['Kategori Sentimen']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.1, random_state=42)

vectorizer = TfidfVectorizer()
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)

X_train_tfidf = X_train_tfidf.toarray()
X_test_tfidf = X_test_tfidf.toarray()

y_train = y_train.astype(str)
y_train = np.array(y_train)

try:
    smote = SMOTE(random_state=42)
    X_train_res, y_train_res = smote.fit_resample(X_train_tfidf, y_train)
except ValueError as e:
    print("Error saat resampling:", e)
    raise

# Pelatihan Model
model = MultinomialNB()
model.fit(X_train_res, y_train_res)

# Pengujian Model
y_pred = model.predict(X_test_tfidf)

# Evaluasi Model
accuracy = accuracy_score(y_test, y_pred)
report = classification_report(y_test, y_pred, zero_division=1)

# Menampilkan hasil evaluasi
print("Akurasi:", accuracy)
print("Laporan Klasifikasi:\n", report)
```

Gambar 4.18 Klasifikasi Naïve Bayes.

Penjelasan kode dalam gambar 4.18. Tahap pelatihan model, model Naive Bayes tipe Multinomial diinisialisasi dengan menggunakan **MultinomialNB()**. Model ini dipilih karena sangat sesuai untuk

klasifikasi teks yang telah diubah menjadi representasi frekuensi atau TF-IDF.

Model kemudian dilatih menggunakan data latih (*X_train_res*) dan label latih (*y_train_res*) yang telah diimbangi menggunakan metode *smote* untuk mengatasi ketidakseimbangan kelas.

Setelah proses pelatihan selesai, model yang telah dilatih diterapkan untuk memprediksi data uji (*X_test_tfidf*).

Hasil prediksi disimpan dalam variabel *y_pred*. Selanjutnya, kinerja model dievaluasi dengan menghitung akurasi menggunakan fungsi *accuracy_score*. Akurasi ini menunjukkan rasio prediksi yang benar terhadap total prediksi, memberikan indikasi seberapa baik model dalam mengklasifikasikan data uji secara keseluruhan.

Selain itu, laporan klasifikasi dihasilkan menggunakan fungsi *classification_report*. Laporan ini mencakup berbagai metrik evaluasi seperti *precision*, *recall*, *F1-score*, dan *support* untuk setiap kelas yang ada di data uji. Penggunaan parameter *zero_division=1* bertujuan untuk menghindari kesalahan pembagian dengan nol dengan menggantinya dengan nilai 1.

Terakhir, nilai akurasi dan laporan klasifikasi dicetak. Akurasi menunjukkan persentase prediksi model yang benar. Laporan klasifikasi memberikan detail lebih lanjut mengenai performa model untuk setiap kelas, termasuk *precision* (ketepatan), *recall* (kemampuan model untuk mendeteksi semua sampel yang termasuk dalam kelas tersebut), *F1-score* (rata-rata antara *precision* dan *recall*), serta *support* (jumlah kejadian sebenarnya dari setiap kelas dalam data uji).

b. Hasil pengujian

Berikut hasil pengujian metode naïve bayes pada gambar 4.19.

Akurasi: 0.92
Laporan Klasifikasi:

	precision	recall	f1-score	support
negative	0.57	0.80	0.67	5
positive	0.98	0.93	0.95	45
accuracy			0.92	50
macro avg	0.77	0.87	0.81	50
weighted avg	0.94	0.92	0.93	50

Gambar 4.19 Hasil pengujian Klasifikasi Naïve Bayes.

c. Hasil uji akurasi

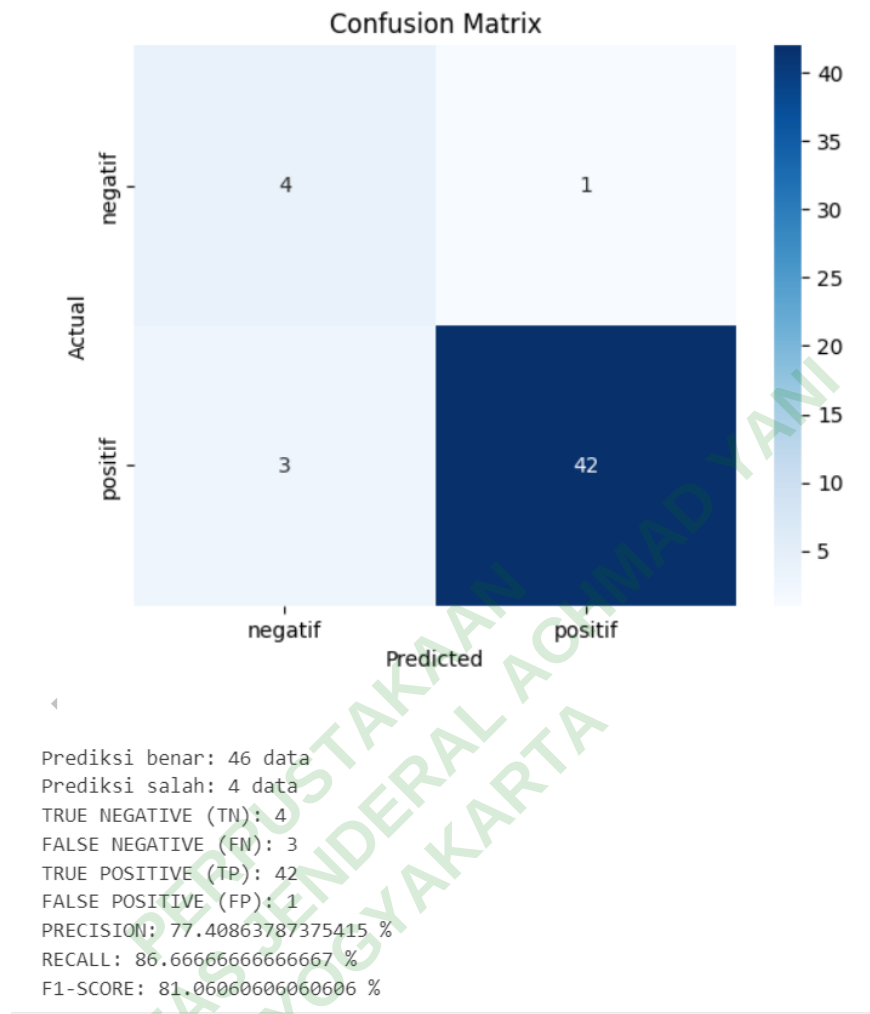
Tabel 4.2 Hasil Akurasi Naive Bayes.

Data Latih	Data Uji	Akurasi Naïve Bayes
Data latih 90%	Data Uji 10%	92%
Data latih 80%	Data Uji 20%	90%
Data latih 70%	Data Uji 30%	90%

Dari paparan pada table 4.2 dapat disimpulkan bahwa penggunaan 90% data latih dan 10% data uji menghasilkan akurasi tertinggi, yaitu 92%.

10. Confusion Matrix

Setelah melakukan pengujian model menggunakan metode Naive Bayes, langkah selanjutnya adalah melakukan evaluasi menggunakan *confusion matrix*. *Confusion matrix* berfungsi untuk melihat informasi secara detail mengenai performa model, tidak hanya akurasi secara keseluruhan, tetapi juga performa pada tiap kelas (positif dan negatif). *Output confusion matrix* akan menghasilkan berbagai metrik evaluasi seperti *precision*, *recall*, dan *F1-score*. Berikut hasil pengujian *Confusion Matrix* pada gambar 4.20 di bawah ini



Gambar 4.20 Hasil Confusion Matrix.

Berdasarkan gambar 4.20 dapat diidentifikasi bahwa model memprediksi data benar senilai 46 data dan model memprediksi salah 4 data. Hasil pengujian *Confusion Matrix* ini menyimpulkan bahwa kelas positif lebih dominan di bandingkan kelas negatif.