

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Analisis Kebutuhan Penelitian

Bab ini menjelaskan mengenai hasil implementasi berbagai proses yang telah dirancang pada bab 3. Adapun tahapan-tahapannya adalah sebagai berikut:

##### 1. Perangkat yang di gunakan spesifikasi

Perangkat yang di gunakan spesifikasi yang di butuh dalam penelitian :

**Tabel 4.1.** Perangkat spesifikasi

NO	Perangkat Keras	Spesifikasi
1	Laptop	LENOVO
2.	Processor	Athlon Silver
3	RAM	8 GB
4	VGA	AMD RADEON

##### 2. Pengumpulan data

Pengumpulan data menggunakan library google play scrape yaitu mengkoneksikan ulasan ke scraper dengan menggunakan link pada web com.bukalapak contoh sebagai berikut :

```
from google_play_scraper import Sort, reviews
```

```
app_id = 'com.bukalapak.android'
```

```
lang = 'id'
```

```
country = 'id'
```

```

sort_order = Sort.MOST_RELEVANT
target_year = 2023
end_year = 2023

reviews_from_target_years = []

while True:
    result, continuation_token = reviews(
        app_id,
        lang=lang,
        country=country,
        sort=sort_order,
        count=100,
        filter_score_with=None,
        continuation_token=continuation_token
    )

```

Data ulasan Bukalapak yang telah berhasil di dapat kemudian di simpan ke dataframe lalu di jadikan file csv total data bukalapak dapatkan sebanyak 19999 ulasan yang akan di proses kedalam model praposes dan metode nove bayes classifier.

## 4.2 Implementasi *Naïve Bayes Classifier*.

Implementasi yang telah dijelaskan di bab 3 membantu analisis terhadap ulasan untuk menentukan sentimen pada teks dari 3 kelas yaitu negatif, neutral dan positif untuk mendapatkan rekomendasi di aplikasi bukalapak dan melakukan prediksi pada kelas sentiment apakah akan membuat setiap kelas mengalami kenaikan atau rendah dari tiap-tiap kelas pada proses pengujian. Analisis di lakukan terhadap sebanyak 19999 ulasan 2023. Dalam melakukan ini yang menjadi atribut adalah Ulasan teks. Data tersebut diolah dengan menggunakan *naïve bayes classifier*.

### 4.2.1 Hasil *Preprocessing*

Pada tahapan ini melakukan input pada dokumen ulasan yang akan di bersihkan dengan 6 tahapan :

## 1. Case folding

Pada tahap case folding huruf kapital pada semua dokumen tweet diubah menjadi huruf kecil contoh tahap konversi huruf menjadi kecil dapat di dilihat pada tabel 4.2.

**Tabel 4.2.** *Case folding*

Input	Output
@Yoghaaaaa Hi, pesanan kamu sudah dalam pengiriman. Maksimal kamu menerima produk 16/02/2017. Mohon ditunggu ya. Thanks –Bella	@yoghaaaaa hi, pesanan kamu sudah dalam pengiriman. maksimal kamu menerima produk 16/02/2017. mohon ditunggu ya. thanks –bella

## 2. Cleaning

Tahapan *cleansing* merupakan tahap pembersihan kata yang tidak berpengaruh sama sekali terhadap hasil klasifikasi sentimen.. Contoh dari atribut yang tidak penting tersebut adalah yaitu mention yang diawali dengan atribut ('@'), hastag yang diawali dengan atribut ('#'), link yang diawali dengan atribut ('http', 'bit.ly') dan karakter simbol (~!@#\$\$%^&\* ()\_+?<>.,?:{ }[|]) contoh tahap menghapus simbol pada Tabel 4.3.

**Tabel 4.3.** *Cleansing*

Input	Output
@yoghaaaaa hi, pesanan kamu sudah dalam pengiriman. maksimal kamu menerima produk 16/02/2017. mohon ditunggu ya. thanks –bella	hi, pesanan kamu sudah dalam pengiriman maksimal kamu menerima produk 16/02/2017 mohon ditunggu yathanks –bella

### 3. Normalitation

Pada tahap *normalization* ini dilakukan perubahan kata yang tidak sesuai dengan dokumen kamus bahasa indonesia, sehingga dapat mengurangi hasil sentimen dokumen contoh tahap konversi inggris ke indonesia dapat di lihat pada Tabel 4.4.

**Tabel 4.4.** *Normalization*

<i>Input</i>	<i>Output</i>
hi, bukalapak utk sementara tdk dptdiakses & masih dalam proses utk memulihkannya mohon maaf atas ketidaknyamanannya bl	hi bukalapak untuk sementara tidak dapat diakses amp masih dalam proses untuk memulihkannya mohon maaf atasketidaknyamanannya bl

Proses selanjutnya adalah tahap konversi kata baku, karena standar bahasa yang dipakai pada penelitian ini adalah bahasa Indonesia baku pada Tabel 4.4.

**Tabel 4.5.** Contoh Tahap Konversi Kata Baku (*Normalization*)

<i>Input</i>	<i>Output</i>
hai yunni mohon maaf mohon informasikan terlebih dahulu emailakun tokopedia via dm	hai yunni mohon maaf mohon informasi lebih dahulu email akun via

Tahap yang terakhir dalam proses Normalization adalah mengubah kata inggris ke Indonesia, berikut contoh konversi kata inggris pada Table.

**Tabel 4.6.** Contoh Tahap Konversi Kata Inggris (*Normalization*)

<i>Input</i>	<i>Output</i>
ulasan customer service yang menggunakan gosend hari ini	ulasan customer service yang menggunakan gosend hari ini

thanksbuat bintang nya	terimakasih buat bintang nya
------------------------	------------------------------

#### 4. *Tokenizing*

Tahap *tokenizing* Kata dalam dokumen yang dimaksud adalah kata yang dipisah oleh spasi. Sehingga hasil dari proses ini merupakan kata tunggal yang dimasukkan ke dalam *database* untuk keperluan pembobotan.

**Tabel 4.7. *Tokenizing***

<i>Input</i>	<i>Output</i>
pesanan sudah dalam pengiriman, jadisudah tidak bisa dibatalkan	'pesanan' 'sudah' 'dalam' 'pengiriman' 'jadi' 'sudah' 'tidak' 'bisa' 'dibatalkan'

#### 5. *Stopword*

Tahap *stopword removal* merupakan tahap menghilangkan kata yang tidak sesuai dengan topik dokumen, jika ada kata tersebut tidak mempengaruhi akurasi dalam klasifikasi sentimen dokumen.

**Tabel 4.8. *Stopword***

<i>Input</i>	<i>Output</i>
bukan kah resinya sudah saya kirim dan juga sudah di konfirm oleh Bukalapak klo barangnya sudah sampai di gudang Bukalapak	bukan resinya kirim Konfirm Bukalapak klo barangnya Gudang Bukalapak

## 6. Stemming

Stemming Jika ditemukan maka diasumsikan bahwa kata tersebut adalah *root word*. Maka metode berhenti. *Inflection Suffixes* (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”) dibuang. Jikaberupa *particles* (“-lah”, “-kah”, “-tah” atau “-pun”).

**Tabel 4.9.** *Stemming*

<i>Input</i>	<i>Output</i>
kalau tetap tidak terkirim paketnya dianggap batal gimana kalau batal pesanan min pembayaran via cod	kalau tetap tidak kirim paket anggap batal gimana pesan kalau batal pesan min bayar via cod

### 4.2.2 Hasil Pelabelan *Polaritas Dan Subjektivty*

Pelabelan melakukan pemisahan dengan memakai Polaritas dan subjektivitas adalah dua faktor penting yang dapat digunakan untuk menentukan sentimen sebuah teks.

Polaritas mengacu pada apakah sepotong teks mengekspresikan sentimen positif, negatif, atau netral, menganalisis kata dan frasa yang digunakan dalam teks dan membandingkannya dengan daftar kata positif dan negatif yang telah ditentukan sebelumnya atau dengan menggunakan model pembelajaran mesin yang dilatih pada data berlabel.

Subjektivitas mengacu kepada sejauh mana sepotong teks dalam mengekspresikan pendapat atau sudut pandang pribadi. Teks yang sangat subjektif akan mengungkapkan pendapat pribadi yang kuat, sedangkan teks yang lebih objektif akan lebih netral sifatnya Saat mempertimbangkan polaritas dan subjektivitas, sentimen sepotong teks dapat ditentukan sebagai:

1. Positif: jika teks mengungkapkan sentimen positif dan sangat subyektif
2. Negatif: jika teks mengungkapkan sentimen negatif dan sangat subyektif
3. Netral: jika teks mengungkapkan netral atau lebih bersifat objektif.

Hasil setiap dokumen pada setiap sentiment di peroleh dari setiap kata yang bernilai baik buruk dan netral dapat terlihat yaitu :

**a. Dokument Sentiment Positif**

Document 6892:

Dokument : wow aplikasi keren polarity

**Tabel 4.10.** Nilai kata sentiment positif

Nilai kata Dokument
wow: 4.00
aplikasi: 0.00
keren: 0.00
Polarity: 0.10
Subjectivity: 1.00
Sentiment: Positive
Overall Document Sentiment: 4.00

Berdasarkan dokumen sentiment tidak memiliki nilai buruk atau nilai minus dan memiliki keluaran yaitu kata wow: 4.00 keren:1.00 dan dapat menghasilkan data yaitu 4.00 dari keseluruhan nilai pada 1 dokumen menghasilkan dokumen sentiment positif

**b. Dokumen Sentiment Negatif**

Document 18338:

Dokument : pusing mental ya polarity

**Tabel 4.11.** Nilai kata sentiment negatif

Nilai kata
pusing: -0.10
mental: 0.00
ya: 0.00
Polarity: -0.10
Subjectivity: 0.20
Sentiment: Negative
Overall Document Sentiment: 0.00

Berdasarkan dokumen sentiment tidak memiliki nilai buruk atau nilai minus dan memiliki keluaran yaitu kata wow: 4.00 keren:1.00 dan dapat menghasilkan data yaitu 4.00 dari keseluruhan nilai pada 1 dokumen menghasilkan dokumen sentiment negatif.

**c. Dokumen Sentiment Netral**

Document 15225:

Dokumen : pengen bisnis online tau buka lapak bantu

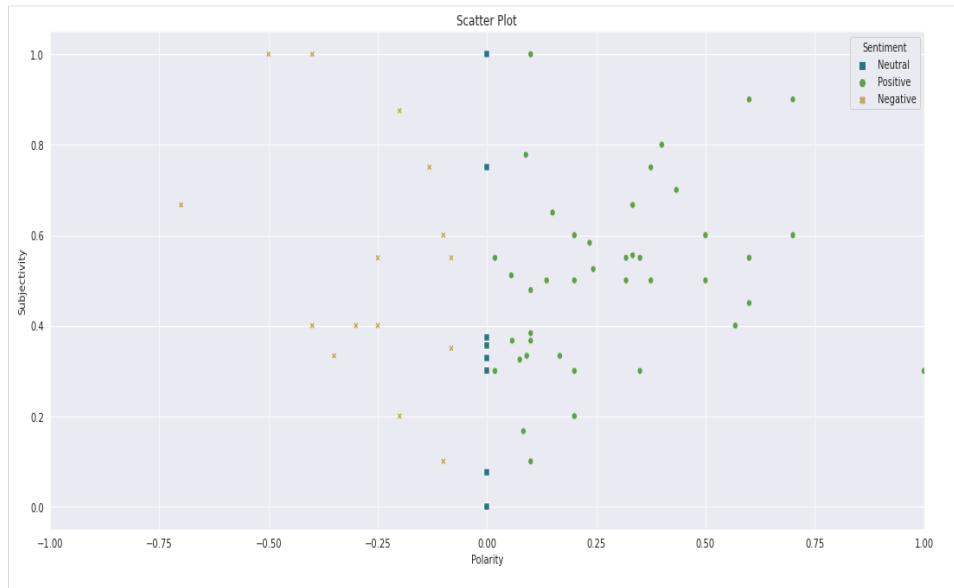
**Tabel 4.12.** Nilai kata sentiment netral Polarity

Nilai kata dokumen netral
pingin: 0.00
bisnis: 0.00
online: 0.00
tau: 0.00
buka: 0.00
lapak: 0.00
Polarity: 0.00
Subjectivity: 0.00
Sentiment: Neutral
Overall Document Sentiment: 0.00

Berdasarkan dokumen sentiment tidak memiliki nilai buruk atau nilai minus dan memiliki keluaran yaitu kata pengen: 0.00, bisnis: 0.00, online: 0.00, tau: 0.00, buka: 0.00, lapak: 0.00, dan dapat menghasilkan data yaitu 4.00 dari keseluruhan nilai pada 1 dokumen menghasilkan dokumen sentiment netral.



Hasil dari melalui metode polarity dan subjectvity yang terbagi 3 sentimen dapat di lihat pada Gambar 4.1.



**Gambar 4.1.** Persebaran *Subjecvity* Dan *Polarity*

Kriteria pelabelan didapat dari nilai setiap kalimat akan di identifikasi merujuk kepada skor polarity dengan ketentuan sebagai berikut:

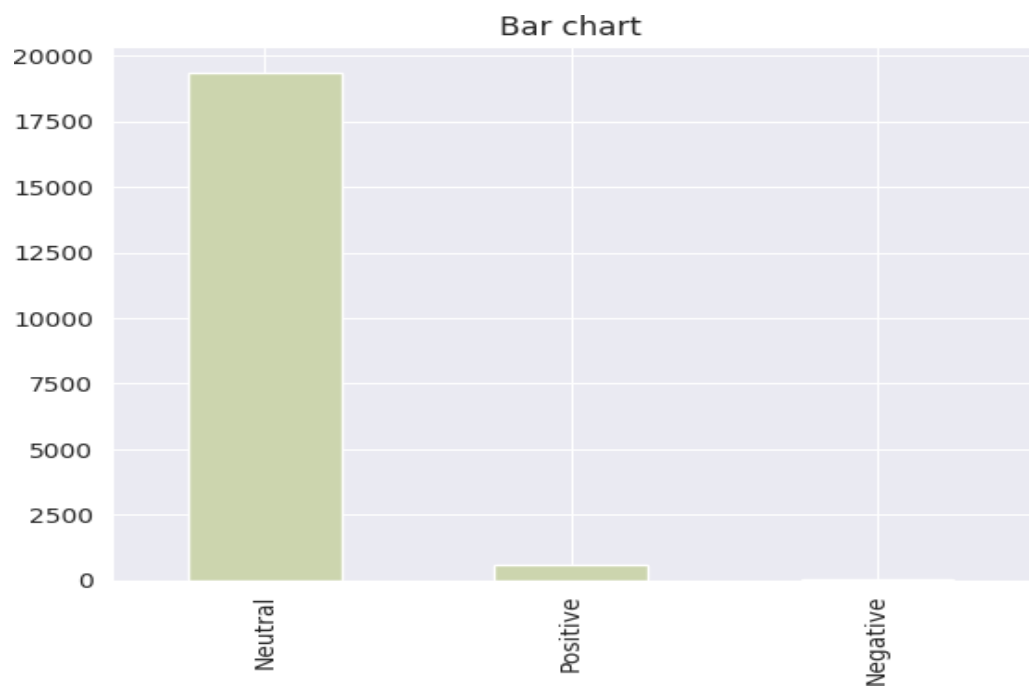
- Skor *polarity* < 0, maka sentimen negatif
- Skor *polarity* = 0, maka sentimen netral
- Skor *polarity* > 0, maka sentimen positif

**Tabel 4.13.** *Sujecvity* Dan *Polarity*

Ulasan siap	<i>Subjecvity</i>	<i>Polarity</i>	Sentiment
topup pulsa telkomsel masuk paylater kuras komplain komplain mohon bantu rugi	0.50	0.5	Positif
bayar kali hari kena biaya layan ribu bayar outlet hadeh habis bayar selesai pakai aplikasi berat biaya	0.0	0.0	Netral

layan terang kena biaya layan mundur			
kecewa ketika voucher voucher balas voucher akun mitra kalo akun mitra voucher kayak kalo adil mitra percaya kalo kayak malas akun mitra nombokin doang untung kagak rugi iya voucher game lingkung kalo kalo kayak pindah	0.10	-0.40	Negatif

Berdasarkan tabel 4.13 hasil melakukan perhitungan visulisasi diagram batang untuk memperjelas hasil dari persebaran pada Gambar 4.2.



**Gambar 4.2.** Visualisasi Hasil angka Pada Setiap Sentimen

Berdasar dari data di dapat pada Gambar 4.2. sentiment netral mendapatkan 1031 data hasil kemudian pada sentiment positive mendapatkan 219 data dan sentiment negatif mendapatkan hasil sangat

kecil yaitu 50 data.

#### 4.2.3 Hasil Term Frequency-Inverse Document Frequency(TF-IDF)

Formula yang di gunakan untuk menghitung bobot Tf masing-masing dokumen terhadap kata kunci pada setiap dokumen.

Rumus

TF IDF :

$$P(\text{label}|\text{fitur}) = \frac{P(\text{label}) * P(\text{fitur}|\text{label})}{P(\text{fitur})}$$

Dimana

setelah bobot (tf) masing masing dokumen diketahui, maka dilakukan proses sortin/pengurutan dimana semakin besar nilai tf, semakin besar tingkat similaritas dokumen tersebut terhadap kata yang di cari demikian sebaliknya. Berikut perhitungannya

#### 1. Sentimen Positif Dokumen

Data yang sudah di siapkan teridentifikasi bahwa dokumen iyalah positif dan akan melakukan perhitungan pada satu dokumen Tabel 4.10.

**Tabel 4.14.** Ulasan Siap Positif

Dokumen	Ulasan preposeccing
Positif	topup pulsa telkomsel masuk paylater kuras komplain komplain mohon bantu rugi

Pengambilan data dari praposesing setelah di pisah melalui tokenized pada Tabel 4.14 dan akan melakukan persamaan kata dari setiap dokumen kata yang paling banyak keluar mempresentasikan bahwa kata tersbut mempunyai nilai probalitas terbesar.

Tabel Perhitungan Bobot Kata Pada Dokumen

**Tabel 4.15.** Perhitungan TF-IDF

No	Term	Frekuensi (nk)	Probalitas P(label fitur)
1	Topup	1	0,095
2	Pulsa	1	0,095
3	Telkomsel	1	0,095
4	Masuk	1	0,095
5	Paylater	1	0,095
6	Kuras	1	0,095
7	Komplain	2	0,142
8	Mohon	1	0,095
9	Bantu	1	0,095
10	Rugi	1	0,095

Pada tabel 4.15 dapat di ketahui:

1. jumlah frekuensi keseluruhan sentimen positif (n) = 10
2. jumlah kosakata keseluruhan sentimen positif (|kosakata) = 11  
probalitas dari setiap term dapat setiap term dapat di cari dengan menggunakan rumus P(label|fitur) pada sub bab 2 adalah:
3.  $P(\text{Komplain} | \text{positif}) = (2+1)/(11+10) = 0,142857143$

## 2. Sentimen Negatif Dokumen

Data yang sudah di siapkan teridentifikasi bahwa dokumen iyalah negatif dan akan melakukan perhitungan pada satu dokumen Tabel 4.16.

**Tabel 4.16.** Ulasan Siap Negatif

Dokumen	Ulasan preposeccing
Negatif	kecewa ketika voucher voucher balas voucher akun mitra kalo akun mitra voucher kayak kalo adil mitra percaya kalo kayak malas akun mitra nombokin

	doang untung kagak rugi iya voucher game lingkung kalo kalo kayak pindah
--	---

Pengambilan data dari praprosesing setelah di pisah melalui tokenized pada Tabel 4.16 dan akan melakukan persamaan kata dari setiap dokumen kata yang paling banyak keluar mempresentasikan bahwa kata tersebut mempunyai nilai probalitas terbesar.

**Tabel 4.17.** Perhitungan Tf-Idf

No	Term	Frekuensi ( $n_k$ )	Probalitas $P(\text{label} \text{fitur})$
1	Kecewa	1	0,048
2	Ketika	1	0,048
3	Voucher	3	0,097
4	Malas	1	0,048
5	Akun	1	0,048
6	Mitra	2	0,073
7	Kayak	2	0,073
8	Kalo	4	0,121
9	Adil	1	0,048
10	Percaya	1	0,048
11	Nombokin	1	0,048
12	Doang	1	0,048
13	Untung	1	0,048
14	Rugi	1	0,048
15	Iya	1	0,048
16	Game	1	0,048
17	Kalo	1	0,048
18	Pindah	1	0,048

Pada tabel 4.17 dapat di ketahui:

1. jumlah frekuensi keseluruhan sentimen negatif ( $n$ ) = 18
2. jumlah kosakata keseluruhan sentimen negatif ( $|\text{kosakata}|$ ) = 25  
probalitas dari setiap term dapat setiap term dapat di cari dengan menggunakan rumus  $P(\text{label}|\text{fitur})$  pada sub bab 2 adalah:
3.  $P(\text{kalo} | \text{positif}) = (4+1)/(25+18) = 0,12195122$

### 3. Sentimen Netral Dokumen

Data yang sudah di siapkan teridentifikasi bahwa dokumen iyalah netral dan akan melakukan perhitungan pada satu dokumen Tabel 4.18.

**Tabel 4.18.** Ulasan Siap Netral

Dokumen	Ulasan preposeccing
Netral	bayar kali hari kena biaya layan ribu bayar outlet hadeh habis bayar selesai pakai aplikasi berat biaya layan terang kena biaya layan mundur

Pengambilan data dari praposesing setelah di pisah melalui tokenized pada Tabel 4.18 dan akan melakukan persamaan kata dari setiap dokumen kata yang paling banyak keluar mempresentasikan bahwa kata tersebut mempunyai nilai probalitas terbesar.

**Tabel 4.19.** Perhitungan Tf-Idf

No	Term	Frekuensi ( $n_k$ )	Probalitas $P(\text{label} \text{fitur})$
1	Bayar	3	0,114
2	Kali	1	0,057
3	Hari	1	0,057
4	Biaya	2	0,085
5	Ribu	1	0,057
6	Outlet	1	0,057
7	Layan	3	0,114
8	HadeH	1	0,057
9	Selesai	1	0,057
10	Pakai	1	0,057
11	Aplikasi	1	0,057
12	Kena	1	0,057
13	Mundur	1	0,057
14	Terang	1	0,057
15	Berat	1	0,057

Pada tabel 4.19 dapat di ketahui:

1. jumlah frekuensi keseluruhan sentimen neutral ( $n$ ) = 15
2. jumlah kosakata keseluruhan sentimen neutral ( $|kosakata|$ ) = 20  
 probalitas dari setiap term dapat setiap term dapat di cari dengan menggunakan rumus  $P(\text{label}|\text{fitur})$  pada sub bab 2 adalah:
3.  $P(\text{bayar} | \text{positif}) = (3+1)/(20+15) = 0,114285714$

#### 4.2.4 Naïve Bayes Classifier Mencarikan Nilai Dari N/Dokumen Frekuensi

Menghitung probalitas term dari setiap sentimen, maka N/df di lakukan perhitungan melanjutkan perhitungan ke 3 dari semua sentimen. Pada tahapan ini di ketahui terdapat 3 data sentimen latih positif negatif dan neutral, sehingga hasilnya:

$$P(\text{positif}) = 1/3 = 0,33 \mid P(\text{negatif}) = 1/3 = 0,33 \mid P(\text{neutral}) = 1/3 = 0,33$$

Dengan menghitung nilai probalitas dari semua kategori sentimen, maka terdapat perubahan nilai probalitas dari setiap term dalam semua kategori sentimen. Perubahan dilakukan dengan menggabungkan total kosakata dari semua sentimen. Total ( $|kosakata|$ ) dari semua sentimen adalah 43 kata sehingga perubahan probalitas sebagai berikut

##### 1. Perubahan *Probalitas Term* Pada Sentimen Positif

Perubahan perhitungan di lakukan pada tabel 4.20 dengan semua fitur dengan membagi bobot yang sama dengan sentiment berbeda di karnakan memilik 3 kelas sentimen dengan bobot kata yang sama maka akan di bagi jumlah term dengan jumlah label.

**Tabel 4.20.** Mencari nilai DF (N) Positif

No	Term	Frekuensi ( $n_k$ )	Probalitas $P(\text{label} \text{fitur})$
1	Topup	1	$(1+1)/(10+43) = 0,037$
2	Pulsa	1	$(1+1)/(10+43) = 0,037$
3	Telkomsel	1	$(1+1)/(10+43) = 0,037$
4	Masuk	1	$(1+1)/(10+43) = 0,037$
5	Paylater	1	$(1+1)/(10+43) = 0,037$

6	Kuras	1	$(1+1)/(10+43) = 0,037$
7	Komplain	2	$(2+1)/(10+43) = 0,056$
8	Mohon	1	$(1+1)/(10+43) = 0,037$
9	Bantu	1	$(1+1)/(10+43) = 0,037$
10	Rugi	1	$(1+1)/(10+43) = 0,037$

## 2. Perubahan *Probabilitas Term* Pada Sentimen Negatif

Selanjutnya perubahan dari kelas negatif perhitungan di lakukan pada tabel 4.21 dengan semua fitur dengan membagi bobot yang sama dengan sentiment berbeda di karnakan memiliki 3 kelas yaitu kelas negatif dengan bobot kata yang sama maka akan di bagi jumlah term dengan jumlah label.

**Tabel 4.21.** Mencari nilai DF (N) Negatif

No	Term	Frekuensi ( $n_k$ )	Probabilitas $P(\text{label} \text{fitur})$
1	Kecewa	1	$(1+1)/(18+43) = 0,032$
2	Ketika	1	$(1+1)/(18+43) = 0,032$
3	Voucher	3	$(3+1)/(18+43) = 0,065$
4	Malas	1	$(1+1)/(18+43) = 0,032$
5	Akun	1	$(1+1)/(18+43) = 0,032$
6	Mitra	2	$(2+1)/(18+43) = 0,049$
7	Kayak	2	$(2+1)/(18+43) = 0,049$
8	Kalo	4	$(4+1)/(18+43) = 0,081$
9	Adil	1	$(1+1)/(18+43) = 0,032$
10	Percaya	1	$(1+1)/(18+43) = 0,032$
11	Nombokin	1	$(1+1)/(18+43) = 0,032$
12	Doang	1	$(1+1)/(18+43) = 0,032$
13	Untung	1	$(1+1)/(18+43) = 0,032$
14	Rugi	1	$(1+1)/(18+43) = 0,032$
15	Iya	1	$(1+1)/(18+43) = 0,032$
16	Game	1	$(1+1)/(18+43) = 0,032$
17	Kalo	1	$(1+1)/(18+43) = 0,032$
18	Pindah	1	$(1+1)/(18+43) = 0,032$



### 3. Perubahan *Probalitas Term* Pada Sentimen Neutral

Perubahan Probalitas Term Pada Sentimen Neutral perhitungan di lakukan pada tabel 4.22 dengan semua fitur dengan membagi bobot yang sama dengan sentiment berbeda di karnakan memiliki 3 kelas sentimen dengan bobot kata yang sama maka akan di bagi jumlah term dengan jumlah label.

**Tabel 4.22.** Mencari nilai DF (N) Netral

No	Term	Frekuensi ( $n_k$ )	Probalitas $P(\text{label} \text{fitur})$
1	Bayar	3	$(3+1)/(15+43) = 0,068$
2	Kali	1	$(1+1)/(15+43) = 0,034$
3	Hari	1	$(1+1)/(15+43) = 0,034$
4	Biaya	2	$(2+1)/(15+43) = 0,051$
5	Ribu	1	$(1+1)/(15+43) = 0,034$
6	Outlet	1	$(1+1)/(15+43) = 0,034$
7	Layan	3	$(3+1)/(15+43) = 0,068$
8	Hadehhh	1	$(1+1)/(15+43) = 0,034$
9	Selesai	1	$(1+1)/(15+43) = 0,034$
10	Pakai	1	$(1+1)/(15+43) = 0,034$
11	Aplikasi	1	$(1+1)/(15+43) = 0,034$
12	Kena	1	$(1+1)/(15+43) = 0,034$
13	Mundur	1	$(1+1)/(15+43) = 0,034$
14	Terang	1	$(1+1)/(15+43) = 0,034$
15	Berat	1	$(1+1)/(15+43) = 0,034$

#### 4.2.5 Cross Validation Data Naïve Bayes Classifier

Dalam tahap Pembagian data menurut data klasifikasi data uji term pada daftar yang akan diklasifikasi di cari nilai probalitas dengan membandingkan term yang ada tabel 4.22 dengan term pada sentimen positif, negatif dan netral. Jika ada term pada tabel 4.22 data latih sentimen positif negatif and netral maka nilai probalitas yang sama tersebut dijadikan nilai probalitas pada term di tabel tetapi jika term pada tabel 4.22. tetapi jika term 4.22 tidak sama, maka fekuensi di hitung nol.

##### 1. Hasil Pehitungan Positif Crossvalidation

Berdasarkan perhitungan pada table 4.23 melakukan seleksi data dengan melihat data terbesar dari data nilai umum yang keluar yang akan menjadi seleksi data yang akan di ikut sertakan dengan data latih dan data uji crossvalidation dan akan menghiung setiap nilai term pada dokomen probalitas peluang.

**Tabel 4.23.** Seleksi Positif

No	Data Uji	Term Latih	Frekuensi ( $n_k$ )	Probalitas $P(\text{label} \text{fitur})$
1	Topup	Tidak ada	0	$(0+1)/(10+43) = 0,018$
2	Pulsa	Tidak ada	0	$(0+1)/(10+43) = 0,018$
3	Telkomsel	Tidak ada	0	$(0+1)/(10+43) = 0,018$
4	Masuk	Tidak ada	0	$(0+1)/(10+43) = 0,018$
5	Paylater	Tidak ada	0	$(0+1)/(10+43) = 0,018$
6	Kuras	Tidak ada	0	$(0+1)/(10+43) = 0,018$
7	Komplain	Ada	1	$(1+1)/(10+43) = 0,037$
8	Mohon	Tidak ada	0	$(0+1)/(10+43) = 0,018$
9	Bantu	Tidak ada	0	$(0+1)/(10+43) = 0,018$
10	Rugi	Tidak ada	0	$(0+1)/(10+43) = 0,018$

Berdasarkan Tabel.23. nilai probalitas data uji pada sentiment netral Tabel

untuk sentimen netral dihitung:

$$P(\text{label}|\text{fitur}) = \frac{P(\text{label}) * P(k_1|\text{label}) * P(k_2|\text{label}) * P(f_1|\text{label}) * \dots * P(f_n|\text{lab})}{\text{SUM}[\text{label}](P(f_1|\text{fitur}) * P(f_2|\text{fitur}) * P(f_3|\text{fitur}) * \dots * P(f_n|\text{fitur}))}$$

$$P(\text{positive} | \text{"Data uji"}) = P(\text{positive}) * P(\text{"Topip"} | \text{positive}) * P(\text{"pulsa"} | \text{positive}) * P(\text{"telkomsel"} | \text{positive}) * P(\text{"masuk"} | \text{positive}) * P(\text{"pulsa"} | \text{positive}) * P(\text{"pulsa"} | \text{positive}) * P(\text{"pulsa"} | \text{positive}) * P(\text{"pulsa"} | \text{positive}) * P(\text{"pulsa"} | \text{positive}) * P(\text{"pulsa"} | \text{positive}) / P(\text{"Tokenizing = 0,33"})$$

$$= (0,037) (0,018) (0,018) (0,018) (0,018) (0,018) (0,018) (0,018) (0,018) (0,018) (0,018) / (0,33) = 2,22402841e-17$$

Dari hasil berikut adalah melihat jarak kata yang unggul di katategorikan 1 dalam perhitungan terhindar dari overviting dan melakukan performa untuk dapat melakukan seleksi data yang biasanya terjadi jika data dari kata yang sama dari sentiment berbeda untuk melakukan pemisahaan di gunakan dan jika nilai nya standar dapat di kategorikan 0 yang mana tidak ada jaminan sebagai kata uji dan untuk 1,38284932e-10e adalah ekponen algoritma pangkat dari penjumlahan nilai yang sama dan nilai paling besar dari setiap kata akan mewakili data latih untuk di uji.

## 2. Hasil Perhitungan Negatif *Crossvalidation*

Selanjutnya berdasarkan perhitungan pada table 4.24 melakukan seleksi data dengan melihat data terbesar dari data nilai umum yang keluar yang akan menjadi seleksi data yang akan di ikut sertakan dengan data latih dan data uji *crossvalidation* dan akan menghiung setiap nilai term pada dokomen probalitas peluang.

**Tabel 4.24.** Seleksi Negatif

No	Data Uji	Term Latih	Frekuensi ( $n_k$ )	Probalitas $P(\text{label} \text{fitur})$
1	Kecewa	Tidak ada	0	$(0+1)/(18+43) = 0,016$
2	Ketika	Tidak ada	0	$(0+1)/(18+43) = 0,016$
3	Voucher	Ada	1	$(1+1)/(18+43) = 0,032$
4	Malas	Tidak ada	0	$(0+1)/(18+43) = 0,016$
5	Akun	Tidak ada	0	$(0+1)/(18+43) = 0,016$
6	Mitra	Ada	1	$(1+1)/(18+43) = 0,032$
7	Kayak	Ada	1	$(1+1)/(18+43) = 0,032$
8	Kalo	Ada	1	$(1+1)/(18+43) = 0,032$
9	Adil	Tidak ada	0	$(0+1)/(18+43) = 0,016$
10	Percaya	Tidak ada	0	$(0+1)/(18+43) = 0,016$
11	Nombokin	Tidak ada	0	$(0+1)/(18+43) = 0,016$
12	Doang	Tidak ada	0	$(0+1)/(18+43) = 0,016$
13	Untung	Tidak ada	0	$(0+1)/(18+43) = 0,016$
14	Rugi	Tidak ada	0	$(0+1)/(18+43) = 0,016$
15	Iya	Tidak ada	0	$(0+1)/(18+43) = 0,016$
16	Game	Tidak ada	0	$(0+1)/(18+43) = 0,016$
17	Kalo	Tidak ada	0	$(0+1)/(18+43) = 0,016$
18	Pindah	Tidak ada	0	$(0+1)/(18+43) = 0,016$

Berdasarkan Tabel.20. nilai probalitas data uji pada sentiment negatif tabel untuk sentimen netral dihitung:

$$\begin{aligned}
 & \left( \quad | \quad \right) \frac{P(\text{label}) * P(k_1|\text{label}) * P(k_2|\text{label}) * P(f_3|\text{label}) * \dots * P(f_n|\text{lab}} \\
 P \text{ label fitur} = & \\
 & \text{SUM}[\text{label}]( P(f_1|\text{fitur}) * P(f_2|\text{fitur}) * P(f_3|\text{fitur}) * \dots * P(f_n|\text{fitur})) \\
 P(\text{negatif} | \text{"Data uji"}) = & P(\text{negatif}) * P(\text{"kecewa"} | \text{negatif}) * P(\text{"ketika"} | \\
 & \text{negatif}) * P(\text{"voucher"} | \text{negatif}) * P(\text{"malas"} | \text{negatif}) * \\
 & P(\text{"akun"} | \text{negatif}) * P(\text{"mitra"} | \text{negatif}) * P(\text{"kayak"} | \text{negatif}) \\
 & * P(\text{"kalo"} | \text{negatif}) * P(\text{"adil"} | \text{negatif}) * P(\text{"percaya"} | \text{negatif}) \\
 & * P(\text{"nombokin"} | \text{negatif}) * P(\text{"doang"} | \text{negatif}) * P(\text{"untung"} | \\
 & \text{negatif}) * P(\text{"rugi"} | \text{negatif}) * P(\text{"iya"} | \text{negatif}) * P(\text{"game"} | \\
 & \text{negatif}) * P(\text{"kalo"} | \text{negatif}) * P(\text{"pindah"} | \text{negatif}) \\
 & / P(\text{"0,33"}) \\
 = & (0,032) (0,032) (0,032) (0,032) (0,016) (0,016) (0,016) \\
 & (0,016)(0,016) (0,016) (0,016) (0,016) (0,016) (0,016) \\
 & (0,016) (0,016) (0,016) (0,016) / (0,33) \\
 = & 2,28963223e-31
 \end{aligned}$$

### 3. Hasil Perhitungan Seleksi Netral *Cross Validation*

Selain itu dilakukan Perhitungan seleksi netral berdasarkan perhitungan pada table 4.25 melakukan seleksi data dengan melihat data terbesar dari data nilai umum yang keluar yang akan menjadi seleksi data yang akan di ikut sertakan dengan data latih dan data uji *crossvalidation* dan akan menghiung setiap nilai term pada dokumen probalitas peluang.

**Tabel 4.25.** Seleksi Netral

No	Data Uji	Term Latih	Frekuensi (n <sub>k</sub> )	Probalitas P(label fitur)
1	Bayar	Ada	1	(1+1)/(15+43) = 0,034
2	Kali	Tidak ada	0	(0+1)/(15+43) = 0,017
3	Hari	Tidak ada	0	(0+1)/(15+43) = 0,017

4	Biaya	Ada	1	$(1+1)/(15+43) = 0,034$
5	Ribu	Tidak ada	0	$(0+1)/(15+43) = 0,017$
6	Outlet	Tidak ada	0	$(0+1)/(15+43) = 0,017$
7	Layan	Ada	1	$(1+1)/(15+43) = 0,034$
8	Hadehhh	Tidak ada	0	$(0+1)/(15+43) = 0,017$
9	Selesai	Tidak ada	0	$(0+1)/(15+43) = 0,017$
10	Pakai	Tidak ada	0	$(0+1)/(15+43) = 0,017$
11	Aplikasi	Tidak ada	0	$(0+1)/(15+43) = 0,017$
12	Kena	Tidak ada	0	$(0+1)/(15+43) = 0,017$
13	Mundur	Tidak ada	0	$(0+1)/(15+43) = 0,017$
14	Terang	Tidak ada	0	$(0+1)/(15+43) = 0,017$
15	Berat	Tidak ada	0	$(0+1)/(15+43) = 0,017$

Berdasarkan Tabel.25. nilai probabilitas data uji pada sentiment netral tabel untuk sentimen netral dihitung:

$$\begin{aligned}
 & \left( \text{Data uji} \mid \text{netral} \right) = \frac{P(\text{netral}) * P(\text{"biaya"} \mid \text{netral}) * P(\text{"ribu"} \mid \text{netral}) * P(\text{"outlet"} \mid \text{netral}) * P(\text{"layan"} \mid \text{netral}) * P(\text{"hadehhh"} \mid \text{netral}) * P(\text{"selesai"} \mid \text{netral}) * P(\text{"pakai"} \mid \text{netral}) * P(\text{"aplikasi"} \mid \text{netral}) * P(\text{"kena"} \mid \text{netral}) * P(\text{"mundur"} \mid \text{netral}) * P(\text{"terang"} \mid \text{netral}) * P(\text{"berat"} \mid \text{netral})}{P(\text{"tokenizing"} = 0,33)} \\
 & P(\text{netral} \mid \text{"Data uji"}) = P(\text{netral}) * P(\text{"biaya"} \mid \text{netral}) * P(\text{"kali"} \mid \text{netral}) * P(\text{"hari"} \mid \text{netral}) * P(\text{"biaya"} \mid \text{netral}) * P(\text{"ribu"} \mid \text{netral}) * P(\text{"outlet"} \mid \text{netral}) * P(\text{"layan"} \mid \text{netral}) * P(\text{"hadehhh"} \mid \text{netral}) * P(\text{"selesai"} \mid \text{netral}) * P(\text{"pakai"} \mid \text{netral}) * P(\text{"aplikasi"} \mid \text{netral}) * P(\text{"kena"} \mid \text{netral}) * P(\text{"mundur"} \mid \text{netral}) * P(\text{"terang"} \mid \text{netral}) * P(\text{"berat"} \mid \text{netral}) / P(\text{"tokenizing"} = 0,33) \\
 & = (0,034) (0,034) (0,034) (0,017) (0,017) (0,017) (0,017) (0,017) (0,017) (0,017) (0,017) (0,017) (0,017) (0,017) (0,017) (0,017) (0,017) / (0,33) \\
 & = 6,9392074e-27
 \end{aligned}$$

Dari ketiga sentiment memiliki hasil perhitungan positif dengan nilai  $2,22402841e-17$ , negatif  $2,8963223e-31$  dan netral  $6,9392074e-27$  dari ketiga sentiment memiliki nilai eksponen di dalam angka di karenakan pada saat proses perhitungan setiap term memiliki nilai yang hampir sama terlihat pada 3 sentiment netral memiliki hasil yang lebih besar dari kedua perhitungan lainnya masing masing data kata yang tidak tereliminasi akan menjadi data uji yang akan di proses pengujian.

#### **4.2.6 Hasil Pembelajaran Dan Uji**

Proses hasil dari pembelajaran akan di hitung dengan confusion matrik dengan data yang sudah di dapat data yang di peroleh bobot nilai yang besar akan di ambil jika sentiment mempunyai rendah dalam perhitungan bobot akan memberikan nilai yang rendah pada pengujian yang akan muncul seberapa besar data yang muncul akan mempengaruhi suatu akurasi seperti berikut:

1. *True Positive* (TP) untuk kelas "Negatif": 1
2. *False Positive* (FP) untuk kelas "Negatif": 7
3. *False Negative* (FN) untuk kelas "Negatif": 4
4. *True Positive* (TP) untuk kelas "Netral": 136
5. *False Positive* (FP) untuk kelas "Netral": 17
6. *False Negative* (FN) untuk kelas "Netral": 62
7. *True Positive* (TP) untuk kelas "Positif": 20
8. *False Positive* (FP) untuk kelas "Positif": 12
9. *False Negative* (FN) untuk kelas "Positif": 1

Terlihat data diatas didapat dari nilai seberapa besar dari yang di peroleh dari masing-masing kelas sentiment kontribusi yang di dapat dari sentiment akan mempengaruhi suatu akurasi yang di dapat pada saat pengujian.





sentiment Positif kata free ongkir dan ongkir termasuk lemah tetapi termasuk kelas positif komentar dan susah tetapi juga terdapat kata mudah aman dan bagus dari kata yang di bobot banyak sekali dari segi positif dan varian saran untuk kedepannya untuk tetap di lakukan dan seoptimal mungkin di lakukan.



**Gambar 4.4.** Sentimen Negatif

Visualisasi Gambar 4.4. menunjukan kata (beli, tipu,minimal, jual, voucher, mending, admin, system , ongkir, sampah, susah, enggak, untung, update ). Komentar pada sentiment Negatif dari sistem pembelian transaksi hubungan banyak sekali yang harus di dilakukan untuk target marketing kedepannya



### 4.3.2 Naïve Bayes Classifier Hasil Dari Prediksi

Selain rekomendasi melakukan prediksi dengan data latih dan uji untuk mendapatkan peluang yang akan datang di masa yang akan datang melakukan prediksi sentimen yang akan muncul berdasarkan pembelajaran naïve bayes gaussian yang di peroleh.

	Nama pengguna	Ulasan_clean	Sentiment	predicted_sentiment_class
0	Nizar Rayyan	paylater batal otomatis kadang nya ganti jatuh...	Neutral	Neutral
1	Fandy Komo	kesini buka lapak bagus layanan komplan susah su...	Neutral	Neutral
2	dedy kelana (CV PRIBUMI BANGKIT)	pasti hp wifi lemoootttt dasar bukalapak kar...	Positive	Positive
3	hidayat mudaris	tolong rumit verifikasi akun saldo bukalapak p...	Negative	Negative
4	ari sandy	aplikasi wadah tipu gampang banget tipu tim sy...	Neutral	Neutral
...	...	...	...	...
1295	Pengguna Google	aplikasi berat bos saran banyak fitur baik per...	Positive	Positive
1296	Pengguna Google	aplikasi bukalapak kecewa order customer kurir...	Neutral	Neutral
1297	Roofie Yk	paylaterpadahal tingkat belanja non cod mohon ...	Neutral	Neutral
1298	Anto Suhendra	tingkat aman nya ya pakai bukalapak betahun pi...	Neutral	Neutral
1299	Erwin Rusdiyanto	haous az tuh paylatermya paylater pakai	Neutral	Neutral

1300 rows × 4 columns

**Gambar 4.6.** Prediksi Hasil *Naïve Bayes Classifier*

Data predict pada gambar 4.6 di hasilkan berdasarkan dari data latih dan data uji melakukan prediksi mendapatkan hasil Neutral 868 Positive 326 dan Negative 106 cukup memiliki perbedaan dengan data di miliki pada saat identifikasi sentimen dengan polarity dan subjecvity dengan hasil Neutral 1056 Positive 198 dan Negative 46 yaitu neutral mengalami pengurangan dan positif mengalami kenaikan dan juga negative mengalami kenaikan.

### 4.3.3 Naïve Bayes Classifier Confusion Matrix

Perhitungan confusion matrix dari gambar 4.7 terdapat beberapa perhitungan yaitu akurasi, recall precision, dan f1-score berdasarkan dari crossvalidation semakin banyak data setiap kelas dari masing-masing kelas akan banyak keluar pada saat proses perhitungan.

```
Accuracy: 0.8051666666666667
Confusion Matrix:
[[ 5  1  1]
 [425 4694 703]
 [ 9  30 132]]
Classification Report:
              precision    recall  f1-score   support

   Negative         0.01         0.71         0.02         7
   Neutral         0.99         0.81         0.89        5822
   Positive         0.16         0.77         0.26         171

 accuracy                   0.81        6000
 macro avg                   0.39        6000
 weighted avg                0.97        6000
```

**Gambar 4.7.** Confusion Matrix

Terlihat data pada gambar 4.7 setiap class memiliki sampling masing masing memberikan nilai terlihat class neutral memiliki hasil tertinggi dan positif memiliki partisipasi tertinggi ke 2 dan negatif memiliki nilai paling kecil nilai yang seperti berikut:

Rumus:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{F1-score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

## 1. Hasil Confusion Matrik

Dari hasil *presisi*, *recall*, *F1* yang telah di hitungan dari rumus di atas sebagai berikut

a. Perhitungan untuk Negative class:

- $Precision$  (Negative) =  $5 / (5 + 425 + 9) \approx 0.0113$  or 1.14%
- $Recall$  (Negative) =  $5 / (5 + 1 + 1) \approx 0.7143$  or 71.43%
- $F1-score$  (Negative) =  $2 * (0.0114 * 0.7143) / (0.0114 + 0.8071) \approx 0.0225$  or 2.25%

b. Perhitungan untuk Neutral class:

- $Precision$  (Neutral) =  $4694 / (1 + 4694 + 30) \approx 0.9935$  or 99.35%
- $Recall$  (Neutral) =  $4694 / (425 + 4694 + 703) \approx 0.8071$  or 80.71%
- $F1-score$  (Neutral) =  $2 * (0.9935 * 0.8071) / (0.9934 + 0.8071) \approx 0.8916$  or 89.16%

c. Perhitungan untuk Positive class:

- $Precision$  (Positive) =  $132 / (1 + 703 + 132) \approx 0.1577$  or 15.77%
- $Recall$  (Positive) =  $312 / (9 + 30) + 132 \approx 0.7719$  or 77.19%
- $F1-score$  (Positive) =  $2 * (0.1577 * 0.8071) / (0.1577 + 0.7719) \approx 0.2612$  or 26.12%

## 2. Hasil Perhitungan Akurasi

Hasil akurasi menghitung keakuratannya, kita dapat menggunakan rumus:

$$\text{Akurasi} = (TP + TN + TN) / (TP + TN + TN + FP + FN + FN)$$

- TP (Positif Benar) = 132
- TN (Negatif Benar) = 5
- TN (Netral Benar) = 4694

$$\text{Akurasi} = (TP + TN + TN) / (TP + TN + TN + FP + FN + FN)$$

$$\text{Akurasi} = (5 + 4694 + 132)$$

$$(5 + 4694 + 132 + 425 + 30 + 703 + 1 + 9 + 1)$$

$$\text{Akurasi} = 4831 / 6000 = 0.8052 \text{ atau } 80,52\%$$

Berdasarkan perhitungan 3x3 class mendapatkan laporan klasifikasi yang diberikan, *classifier* berperforma baik untuk kelas “Neutral” dengan *precision* 99.35%, *recall* 80.71%, dan *f1-score* 89.16%. Untuk kelas “Positive” *classifier* juga berperforma baik dengan *precision* 15.77%, *recall* 77.19% dan *f1-score* 26.12%, namun untuk kelas “Negative” *classifier* kurang berperforma baik dengan *precision* 1.14%, *recall* 71.43%, dan skor *f1* sebesar 2.25%.

Berdasarkan tabel multi class *confusion matrix* 3x3 kelas term frequency didapatkan nilai match sebanyak *True Negatif* 5, *True Netral* 4694, *True Positif* 132 lalu di bagi dengan data total uji sebanyak 4831 karna data keseluruhan yaitu 6000 data menghasilkan akurasi sebesar 80.52%.