

BAB IV HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari ulasan aplikasi seabank di google play store menggunakan bahasa pemrograman python yang dijalankan di google *collaboratory*. Berikut merupakan *Script scrapping* ulasan pada pemrograman Python:

1. Tahap pertama dalam pengambilan data adalah menginstall library Google Play Scrapper.

```
!pip install google-play-scraper
```

```
from google_play_scraper import app, reviews  
import pandas as pd  
import datetime
```

2. Melakukan scrapping dengan URL id seabank di google play store

```
app_id = 'id.co.bankbkemobile.digitalbank'  
start_date = datetime.date(2022, 1, 1)  
end_date = datetime.date(2023, 12, 20)
```

3. Proses scrapping ulasan pengguna seabank di google play store

```
result, continuation_token = reviews(  
    app_id,  
    lang='id',  
    count=3804,  
    sort='newest'  
)
```

Berikut merupakan hasil scapping ulasan aplikasi seabank menggunakan bahasa pemrograman python dapat dilihat pada gambar 4.1 dibawah ini.

	B	C	D	E	F	G	H	I	J	K	L
1	userName	userImage	content	score	thumbsUp	thumbsDown	reviewCreat	replyContent	repliedAt	appVersion	
2	Fikri Fahmi Udin	https://play-lh.google	Jangan sampai naruh duit banyak, bahaya	1	2645	2.60.0	10/12/2023 05:22	Hai Sobat SeaBank, n	10/12/2023 06:17	2.60.0	
3	Dian Mutamimah	https://play-lh.google	Aplikasi sampah ,saya telpon beberapa l	1	1500	2.59.0	28/11/2023 12:14	Hai Sobat SeaBank, n	28/11/2023 14:10	2.59.0	
4	dita heydiar	https://play-lh.google	Aplikasi gak ada GUNA. Kalo terjadi kenc	1	385	2.59.0	01/12/2023 15:31	Hai Sobat SeaBank, n	02/12/2023 02:04	2.59.0	
5	desi anandias	https://play-lh.google	bad service handle satu kendala hampir	1	255	2.58.0	10/11/2023 06:31	Sebelumnya mohon	10/11/2023 07:09	2.58.0	
6	dunia jeje	https://play-lh.google	Aplikasi yang tidak bisa kasih solusi, kon	1	530	2.60.0	08/12/2023 05:48	Hai Sobat SeaBank, n	08/12/2023 06:28	2.60.0	
7	TAZALQY DUARIBUEN	https://play-lh.google	Akun Seabank saya di blokir sedangkan :	1	326	2.59.0	04/12/2023 03:12	Hai Sobat SeaBank, n	02/10/2023 15:11	2.59.0	
8	Azmi Maliki	https://play-lh.google	tolong jika memang ada gangguan dalam	2	156	2.59.0	20/11/2023 06:30	Hi Sobat SeaBank, m	20/11/2023 07:28	2.59.0	
9	vir go	https://play-lh.google	apk ter bagus menurut saya karna buat t	5	93	2.60.0	12/12/2023 07:43	Halo Sobat SeaBank,	12/12/2023 08:39	2.60.0	
10	Primarini Riati	https://play-lh.google	Setiap menghubungi CS baru berapa me	1	136	2.60.0	11/12/2023 08:09	Hai Sobat SeaBank, v	11/12/2023 08:06	2.60.0	
11	Maruli Ruli	https://play-lh.google	Hari ini saya sangat kecewa lagi di mana	1	75	2.58.0	09/11/2023 04:10	Hai Sobat SeaBank, n	09/11/2023 05:05	2.58.0	
12	HadeJaya Channel	https://play-lh.google	Tadinya mau menjadikan seabank sebag	3	37	2.60.0	18/12/2023 02:59	Hi Sobat SeaBank, m	18/12/2023 03:07	2.60.0	
13	Hafizh Dannur Hafshz	https://play-lh.google	Overall UI aplikasi sudah bagus dan mud	3	100	2.58.0	12/11/2023 02:01	Hi Sobat SeaBank, tei	12/11/2023 06:26	2.58.0	
14	Alya Vinoata	https://play-lh.google	sy sudah isi saldo seabank, mau chekout	3	0	2.59.0	19/12/2023 08:08	Hai Sobat SeaBank, n	19/12/2023 09:13	2.59.0	
15	Lija Ovia Sinaga	https://play-lh.google	first time to use Seabank as saving, sem	5	2	2.60.0	18/12/2023 00:55			2.60.0	
16	Rahmat	https://play-lh.google	Apikasi SIAL memang, tidak pernah sala	1	5	2.60.0	18/12/2023 05:11	Hi Sobat SeaBank, m	18/12/2023 06:10	2.60.0	
17	Della Amanda	https://play-lh.google	Disuruh hubungan rekening sea bank d	1	3	2.60.0	17/12/2023 19:24	Hai Sobat SeaBank, n	17/12/2023 20:39	2.60.0	
18	Semangka L	https://play-lh.google	Ga percaya lagi sama Seabank. Bisa-bisai	1	39	2.58.0	16/11/2023 12:27	Hai Sobat SeaBank, n	16/11/2023 15:52	2.58.0	
19	Miftha B	https://play-lh.google	Mau login disuruh verifikasi wajah lagi, s	1	50	2.59.0	26/11/2023 09:15	Hai Sobat SeaBank, n	26/11/2023 09:12	2.59.0	
20	Sevi Haryanti	https://play-lh.google	Sejauh ini bagus, transaksi dll tanpa ken	5	5	2.61.0	20/12/2023 12:20	Hai Sobat SeaBank, t	24/04/2023 02:28	2.61.0	
21	ubur ubur	https://play-lh.google	Awal2 launching, aman bagus2 aja app n	1	66	2.60.0	13/12/2023 21:15	Hi Sobat SeaBank, m	13/12/2023 22:34	2.60.0	

Gambar 4. 1 Hasil Scapping Ulasan Aplikasi Seabank 3804 ulasan

Hasil scapping ulasan aplikasi seabank digoogle play store menggunakan bahasa pemrograman pyhton berjumlah 3804 ulasan. Ulasan yang diambil kemudian disimpan dalam file CSV yang nantinya diolah menggunakan Google Colab dengan pendekatan bahasa pemrograman Python.

4.2 Text Preprocessing

Pada tahap *preprocessing* data dibagi menjadi beberapa langkah di antaranya, Sebelum memasuki tahap *preprocessing* data, langkah pertama adalah tahap penghapusan duplikat, diikuti dengan beberapa langkah *preprocessing* data lainnya seperti *cleaning*, *case folding*, *tokenizing*, *stopwords removal*, dan *stemming*. Berikut merupakan Script *Preprocessing* data pada pemrograman Python:

1. melakukan import padas dan upload file

```
import pandas as pd

data = pd.read_csv("hasil_scraper_ulasan_app_seabank.csv")
data.info()
```

2. Proses hapus data duplikat

```
df.drop_duplicates(subset ="content", keep = 'first', inplace = True)
```

```
df.info()
```

```
df = pd.DataFrame(df[['at', 'userName', 'score', 'content']])
df.head(10)
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3789 entries, 0 to 3804
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  -
0   at           3789 non-null  object
1   userName     3789 non-null  object
2   score        3789 non-null  int64
3   content      3789 non-null  object
dtypes: int64(1), object(3)
memory usage: 148.0+ KB
```

Hasil setelah menghapus data duplikat ulasan aplikasi seabank yang ada di google play store menggunakan bahasa pemrograman python berjumlah 3789 dari 3804 ulasan. Kemudian data yang berjumlah 3789 akan masuk ke tahap *preprocessing data*.

3.1.1 Cleaning

Dalam melakukan proses cleaning dilakukan 6 tahapan untuk mendapatkan hasil yang maksimal diantaranya menghapus URL, menghapus HTML, menghapus emoji, menghapus simbol, menghapus angka. Berikut merupakan Script cleaning pada pemrograman Python.

1. Tahap cleaning adalah Import re, string, dan nltk

```
import re
import string
import nltk
```

2. Membaca hasil cleaning

```
df['cleaning'] = df['content'].apply(lambda x: remove_URL(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_html(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_emoji(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_symbols(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_numbers(x))
```

Berikut merupakan hasil ulasan yang telah selesai tahap cleaning dapat dilihat pada tabel 4.1 dibawah ini.

Tabel 4.1 Hasil Cleaning

Index	Content	cleaning
0	Jangan sampai naruh duit banyak, bahaya. Bisa tiba tiba ga bisa log ini, dibilang salah pin. telepon cs dibilang data ga sesuai padahal pin yang digunakan biasanya juga itu, data yg dimasukan juga itu. Bahaya banget kalau naruh banyak tiba tiba ga jelas ga bisa masuk kayak gini dan tidak ada solusi untuk masuk kembali. Hanya disuruh inget inget pertanyaan yang jelas jelas jawabannya udah sesuai dengan data ktp	Jangan sampai naruh duit banyak bahaya Bisa tiba tiba ga bisa log ini dibilang salah pin telepon cs dibilang data ga sesuai padahal pin yang digunakan biasanya juga itu data yg dimasukan juga itu Bahaya banget kalau naruh banyak tiba tiba ga jelas ga bisa masuk kayak gini dan tidak ada solusi untuk masuk kembali Hanya disuruh inget inget pertanyaan yang jelas jelas jawabannya udah sesuai dengan data ktp
1	Aplikasi sampah ,saya telpon beberapa kali sampe abis pulsa banyak engga bisa di bantu untuk masalah log in. Keterangan log.in	Aplikasi sampah saya telpon beberapa kali sampe abis pulsa banyak engga bisa di bantu untuk masalah log in Keterangan login

	keblokir sementara, udah di kasih data lengkap tapi bilanganya tidak sesuai ! Masalahnya itu saya ada saldo.nya !!	keblokir sementara udah di kasih data lengkap tapi bilanganya tidak sesuai Masalahnya itu saya ada saldonya
2	Aplikasi gak ada GUNA. Kalo terjadi kendala susah komplainnya. Udah habis pulsa saya 30ribu lebih gak menemukan solusi. CS ngomong lambat banget, bertele tele, masalah tidak terselesaikan !!! Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya !!!! Tidak recommended 🙄	Aplikasi gak ada GUNA Kalo terjadi kendala susah komplainnya Udah habis pulsa saya ribu lebih gak menemukan solusi CS ngomong lambat banget bertele tele masalah tidak terselesaikan Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya Tidak recommended
3	bad service handle satu kendala hampir 1 minggu tidak tuntas stuck disatu prosedur prosedur yg diberikan cs berbelit belit sistem telfon sangat buruk selalu terputus dan harus diamuk dulu baru prosesnya dipercepat! Uang dan waktu terbuang sia sia!	bad service handle satu kendala hampir minggu tidak tuntas stuck disatu prosedur prosedur yg diberikan cs berbelit belit sistem telfon sangat buruk selalu terputus dan harus diamuk dulu baru prosesnya dipercepat Uang dan waktu terbuang sia sia
4	Aplikasi yang tidak bisa kasih solusi, komplant via email bertanya hal berbeda juga jawabannya sama suruh chat langsung di aplikasi, sedangkan chat aplikasi tidak ada pilihan ke cs (cuman di suruh pilih pertanyaan), via telepon juga sama jawabannya tidak memuaskan.	Aplikasi yang tidak bisa kasih solusi komplant via email bertanya hal berbeda juga jawabannya sama suruh chat langsung di aplikasi sedangkan chat aplikasi tidak ada pilihan ke cs cuman di suruh pilih pertanyaan via telepon juga sama

Seharusnya kasih tau bukan tanya balik. Mohon di perbaiki lah biar kita tidak kecewa	jawabannya tidak memuaskan Seharusnya kasih tau bukan tanya balik Mohon di perbaiki lah biar kita tidak kecewa
--	---

3.1.2 Case Folding

Data yang digunakan adalah ulasan aplikasi seabank yang sudah selesai tahap cleaning selanjutnya masuk ke tahap case folding. Berikut merupakan Script *Case Folding* pada pemrograman Python.

```
def case_folding(text):
    if isinstance(text, str):
        lowercase_text = text.lower()
        return lowercase_text
    else:
        return text

df['case_folding'] = df['cleaning'].apply(case_folding)
df.head(5)
```

Berikut merupakan hasil ulasan yang telah selesai tahap case folding dapat dilihat pada tabel 4.2 dibawah ini.

Tabel 4.2 Hasil Case Folding

Index	Content	Case Folding
0	Jangan sampai naruh duit banyak, bahaya. Bisa tiba tiba ga bisa log ini, dibilang salah pin. telepon cs dibilang data ga sesuai padahal pin yang digunakan biasanya juga itu, data yg dimasukan juga itu. Bahaya banget kalau naruh banyak tiba tiba ga jelas ga bisa masuk kayak gini	jangan sampai naruh duit banyak bahaya bisa tiba tiba ga bisa log ini dibilang salah pin telepon cs dibilang data ga sesuai padahal pin yang digunakan biasanya juga itu data yg dimasukan juga itu bahaya banget kalau naruh banyak tiba tiba ga jelas ga bisa masuk kayak

	dan tidak ada solusi untuk masuk kembali. Hanya disuruh inget inget pertanyaan yang jelas jelas jawabannya udah sesuai dengan data ktp	gini dan tidak ada solusi untuk masuk kembali hanya disuruh inget inget pertanyaan yang jelas jelas jawabannya udah sesuai dengan data ktp
1	Aplikasi sampah ,saya telpon beberapa kali sampe abis pulsa banyak engga bisa di bantu untuk masalah log in. Keterangan log.in keblokir sementara,udah di kasih data lengkap tapi bilangnyanya tidak sesuai ! Masalahnya itu saya ada saldo.nya !!	aplikasi sampah saya telpon beberapa kali sampe abis pulsa banyak engga bisa di bantu untuk masalah log in keterangan login keblokir sementaraudah di kasih data lengkap tapi bilangnyanya tidak sesuai masalahnya itu saya ada saldonya
2	Aplikasi gak ada GUNA. Kalo terjadi kendala susah komplainnya. Udah habis pulsa saya 30ribu lebih gak menemukan solusi. CS ngomong lambat banget, bertele tele, masalah tidak terselesaikan !!! Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya !!!! Tidak recommended 🙄	aplikasi gak ada guna kalo terjadi kendala susah komplainnya udah habis pulsa saya ribu lebih gak menemukan solusi cs ngomong lambat banget bertele tele masalah tidak terselesaikan buang buang waktu dan uang aja buat beli pulsa ngehubungin cs nya tidak recommended
3	bad service handle satu kendala hampir 1 minggu tidak tuntas stuck disatu prosedur prosedur yg diberikan cs berbelit belit sistem telfon sangat buruk selalu terputus dan harus diamuk dulu baru prosesnya dipercepat! Uang dan waktu terbuang sia sia!	bad service handle satu kendala hampir minggu tidak tuntas stuck disatu prosedur prosedur yg diberikan cs berbelit belit sistem telfon sangat buruk selalu terputus dan harus diamuk dulu baru prosesnya dipercepat uang dan waktu terbuang sia sia

4	Aplikasi yang tidak bisa kasih solusi, komplant via email bertanya hal berbeda juga jawabannya sama suruh chat langsung di aplikasi, sedangkan chat aplikasi tidak ada pilihan ke cs (cuman di suruh pilih pertanyaan), via telepon juga sama jawabannya tidak memuaskan. Seharusnya kasih tau bukan tanya balik. Mohon di perbaiki lah biar kita tidak kecewa	aplikasi yang tidak bisa kasih solusi komplant via email bertanya hal berbeda juga jawabannya sama suruh chat langsung di aplikasi sedangkan chat aplikasi tidak ada pilihan ke cs cuman di suruh pilih pertanyaan via telepon juga sama jawabannya tidak memuaskan seharusnya kasih tau bukan tanya balik mohon di perbaiki lah biar kita tidak kecewa
---	---	---

3.1.3 Tokenizing

Tokenizing bertujuan agar data dapat diproses pada tahapan selanjutnya yaitu stopwords removal. Berikut merupakan Script tokenizing pada pemrograman Python:

```
def tokenize(text):
    tokens = text.split()
    return tokens

df['tokenize'] = df['case_folding'].apply(tokenize)

df.head(15)
```

Berikut merupakan hasil ulasan yang telah selesai tahap tokenizing dapat dilihat pada tabel 4.3 dibawah ini.

Tabel 4.3 Hasil tokenizing

Index	Content	Tokenizing
0	Jangan sampai naruh duit banyak, bahaya. Bisa tiba tiba ga bisa log ini, dibilang salah pin. telepon cs dibilang data ga sesuai padahal pin	[jangan,sampai,naruh,duit,banyak,bahaya,bisa,tiba,tiba,ga,bisa,log,ini,dibilang,salah,pin,telepon,cs,dibilang,data,ga,sesuai,padahal

	yang digunakan biasanya juga itu, data yg dimasukan juga itu. Bahaya banget kalau naruh banyak tiba tiba ga jelas ga bisa masuk kayak gini dan tidak ada solusi untuk masuk kembali. Hanya disuruh inget inget pertanyaan yang jelas jelas jawabannya udah sesuai dengan data ktp	,pin,yang,digunakan,biasanya,ju ga,itu,data,yg,dimasukan,juga,itu,bahaya,banget,kalau,naruh,banyak,tiba,tiba,ga,jelas,ga,bisa,masuk,kayak,gini,dan,tidak,ada,solusi,untuk,masuk,kembali,hanya,disuruh,inget,inget,pertanyaan,yang,jelas,jelas,jawabannya,udah,sesuai,dengan,data,ktp]
1	Aplikasi sampah ,saya telpon beberapa kali sampe abis pulsa banyak engga bisa di bantu untuk masalah log in. Keterangan log.in keblokir sementara,udah di kasih data lengkap tapi bilangnyanya tidak sesuai ! Masalahnya itu saya ada saldo.nya !!	[aplikasi,sampah,saya,telpon,beberapa,kali,sampe,abis,pulsa,banyak,engga,bisa,di,bantu,untuk,masalah,log,in,keterangan,login,keblokir,sementeraudah,di,kasih,data,lengkap,tapi,bilangnyanya,tidak,sesuai,masalahnya,itu,saya,ada,saldonya]
2	Aplikasi gak ada GUNA. Kalo terjadi kendala susah komplainnya. Udah habis pulsa saya 30ribu lebih gak menemukan solusi. CS ngomong lambat banget, bertele tele, masalah tidak terselesaikan !!! Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya !!!! Tidak recommended 🙄	[aplikasi,gak,ada,guna,kalo,terjadi,kendala,susah,komplainnya,udah,habis,pulsa,saya,ribu,lebih,gak,menemukan,solusi,cs,ngomong,lambat,banget,berteletele,masalah,tidak,terselesaikan, buang, buang,waktu,dan,uang,aja,buat,beli,pulsa,ngehubungin,cs,nya,tidak,recommended]
3	bad service handle satu kendala hampir 1 minggu tidak tuntas stuck disatu prosedur prosedur yg diberikan cs berbelit belit sistem	[bad,service,handle,satu,kendala,hampir,minggu,tidak,tuntas, stuck,disatu,prosedur,prosedur,yg,diberikan,cs,berbelit,belit,sistem,te

	telfon sangat buruk selalu terputus dan harus diamuk dulu baru prosesnya dipercepat! Uang dan waktu terbuang sia sia!	lfon,sangat,buruk,selalu,terputus ,dan,harus,diamuk,dulu,baru,prosesnya,dipercepat,uang,dan,waktu,terbuang,sia,sia]
4	Aplikasi yang tidak bisa kasih solusi, komplant via email bertanya hal berbeda juga jawabannya sama suruh chat langsung di aplikasi, sedangkan chat aplikasi tidak ada pilihan ke cs (cuman di suruh pilih pertanyaan), via telepon juga sama jawabannya tidak memuaskan. Seharusnya kasih tau bukan tanya balik. Mohon di perbaiki lah biar kita tidak kecewa	[aplikasi,yang,tidak,bisa,kasih,solusi,komplant,via,email,bertanya ,hal,berbeda,juga,jawabannya,sama,suruh,chat,langsung,di,aplikasi,se sedangkan,chat,aplikasi,tidak,ada,pilihan,ke,cs,cuman,di,suruh ,pilih,pertanyaan,via,telepon,juga,sama,jawabannya,tidak,memuaskan,seharusnya,kasih,tau,bukan ,tanya,balik,mohon,di,perbaiki,lah,biar,kita,tidak,kecewa]

3.1.4 Stopwords removal

Pada tahap stopwords removal perlu disiapkan library yang akan digunakan yaitu `nltk_data`. Berikut merupakan Script Stopwords removal pada pemrograman Python:

1. Downloading package stopwords to `/root/nltk_data`

```
from nltk.corpus import stopwords
nltk.download('stopwords')
stop_words = stopwords.words('indonesian')
```

2. Menambah stopwords dari file csv dan membaca data

```
def remove_stopwords(text):
    return [word for word in text if word not in stop_words]

df['stopword removal'] = df['tokenize'].apply(lambda x: remove_stopwords(x))
df.head(5)
```

Berikut merupakan hasil ulasan yang telah selesai tahap stopwords removal dapat dilihat pada tabel 4.4 dibawah ini.

Tabel 4.4 Hasil Stopwords Removal

Index	Content	Stopwords removal
0	<p>Jangan sampai naruh duit banyak, bahaya. Bisa tiba tiba ga bisa log ini, dibbilang salah pin. telepon cs dibbilang data ga sesuai padahal pin yang digunakan biasanya juga itu, data yg dimasukan juga itu. Bahaya banget kalau naruh banyak tiba tiba ga jelas ga bisa masuk kayak gini dan tidak ada solusi untuk masuk kembali. Hanya disuruh inget inget pertanyaan yang jelas jelas jawabannya udah sesuai dengan data ktp</p>	<p>[naruh,duit,bahaya,ga,log,dibilan g,salah,pin,telepon,cs,dibilang,d ata,ga,sesuai,pin,data,yg,dimasu kan,bahaya,banget,naruh,ga,ga, masuk,kayak,gini,solusi,masuk,d isuruh,inget,inget,jawabannya,ud ah,sesuai,data,ktp]</p>
1	<p>Aplikasi sampah ,saya telpon beberapa kali sampe abis pulsa banyak engga bisa di bantu untuk masalah log in. Keterangan log.in keblokir sementara,udah di kasih data lengkap tapi bilanganya tidak sesuai ! Masalahnya itu saya ada saldo.nya !!</p>	<p>[aplikasi,sampah,telpon,kali,sam pe,abis,pulsa,engga,bantu,log,in, keterangan,login,keblokir,sement araudah,kasih,data,lengkap,bilan gnya,sesuai,saldonya]</p>
2	<p>Aplikasi gak ada GUNA. Kalo terjadi kendala susah komplainnya. Udah habis pulsa saya 30ribu lebih gak menemukan solusi. CS ngomong lambat banget, bertele</p>	<p>[aplikasi,gak,kalo,kendala,susah, komplainnya,udah,habis,pulsa,ri bu,gak,menemukan,solusi,cs,ngo mong,lambat,banget,bertele,tele, terselesaikan,buang,buang,uang,</p>

	tele, masalah tidak terselesaikan !!! Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya !!!! Tidak recommended 🙄	aja, beli, pulsa, ngehubungin, cs, nya, a, recommended]
3	bad service handle satu kendala hampir 1 minggu tidak tuntas stuck disatu prosedur prosedur yg diberikan cs berbelit belit sistem telfon sangat buruk selalu terputus dan harus diamuk dulu baru prosesnya dipercepat! Uang dan waktu terbang sia sia!	[bad, service, handle, kendala, min gg, tuntas, stuck, disatu, prosedur, prosedur, yg, cs, berbelit, belit, siste m, telfon, buruk, terputus, diamuk, prosesnya, dipercepat, uang, terbu ang, sia, sia]
4	Aplikasi yang tidak bisa kasih solusi, komplant via email bertanya hal berbeda juga jawabannya sama suruh chat langsung di aplikasi, sedangkan chat aplikasi tidak ada pilihan ke cs (cuman di suruh pilih pertanyaan), via telepon juga sama jawabannya tidak memuaskan. Seharusnya kasih tau bukan tanya balik. Mohon di perbaiki lah biar kita tidak kecewa	[aplikasi, kasih, solusi, komplant, v ia, email, berbeda, jawabannya, sur uh, chat, langsung, aplikasi, chat, ap likasi, pilihan, cs, cuman, suruh, pili h, via, telepon, jawabannya, memu askan, kasih, tau, mohon, perbaiki, biar, kecewa]

3.1.5 Stemming

Pada tahap ini perlu disiapkan library yang akan digunakan yaitu library Sastrawi (pustaka Python yang menyediakan algoritma stemming) Berikut merupakan Script Stemming pada pemrograman Python.

1. Menginstall library Sastrawi

```
!pip install Sastrawi
```

```
!pip install Sastrawi

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from nltk.stem import PorterStemmer
from nltk.stem.snowball import SnowballStemmer
```

2. Mengimport StemmerFactory yang diperlukan untuk proses Stemming

```
factory = StemmerFactory()
stemmer = factory.create_stemmer()

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

df['stemming_data'] = df['stopword removal'].apply(lambda x: ' '.join(stem_text(x)))
df.head(5)
```

Berikut merupakan hasil ulasan yang telah selesai tahap stemming dapat dilihat pada tabel 4.5 dibawah ini.

Tabel 4.5 Hasil Stemming

Index	Content	Stemming
0	Jangan sampai naruh duit banyak, bahaya. Bisa tiba tiba ga bisa log ini, dibilang salah pin. telepon cs dibilang data ga sesuai padahal pin yang digunakan biasanya juga itu, data yg dimasukan juga itu. Bahaya banget kalau naruh banyak tiba tiba ga jelas ga bisa masuk kayak gini dan tidak ada solusi untuk masuk kembali. Hanya disuruh inget inget pertanyaan yang jelas jelas jawabannya udah sesuai dengan data ktp	naruh duit bahaya ga log bilang salah pin telepon cs bilang data ga sesuai pin data yg masuk bahaya banget naruh ga ga masuk kayak gin solusi masuk suruh inget inget jawab udah sesuai data ktp
1	Aplikasi sampah ,saya telpon beberapa kali sampe abis pulsa banyak engga	aplikasi sampah telpon kali sampe abis pulsa engga bantu

	bisa di bantu untuk masalah log in. Keterangan log.in keblokir sementara,udah di kasih data lengkap tapi bilanganya tidak sesuai ! Masalahnya itu saya ada saldo.nya !!	log in terang login blokir sementaraudah kasih data lengkap bilang sesuai saldo
2	Aplikasi gak ada GUNA. Kalo terjadi kendala susah komplainnya. Udah habis pulsa saya 30ribu lebih gak menemukan solusi. CS ngomong lambat banget, bertele tele, masalah tidak terselesaikan !!! Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya !!!! Tidak recommended 🙄	aplikasi gak kalo kendala susah komplain udah habis pulsa ribu gak temu solusi cs ngomong lambat banget tele tele selesai buang buang uang aja beli pulsa ngehubungin cs nya recommended
3	bad service handle satu kendala hampir 1 minggu tidak tuntas stuck disatu prosedur prosedur yg diberikan cs berbelit belit sistem telfon sangat buruk selalu terputus dan harus diamuk dulu baru prosesnya dipercepat! Uang dan waktu terbang sia sia!	bad service handle kendala minggu tuntas stuck satu prosedur prosedur yg cs belit belit sistem telfon buruk putus amuk proses cepat uang buang sia sia
4	Aplikasi yang tidak bisa kasih solusi, komplant via email bertanya hal berbeda juga jawabannya sama suruh chat langsung di aplikasi, sedangkan chat aplikasi tidak ada pilihan ke cs (cuman di suruh pilih pertanyaan), via telepon juga sama jawabannya tidak memuaskan. Seharusnya kasih tau	aplikasi kasih solusi komplant via email beda jawab suruh chat langsung aplikasi chat aplikasi pilih cs cuman suruh pilih via telepon jawab muas kasih tau mohon baik biar kecewa

	bukan tanya balik. Mohon di perbaiki lah biar kita tidak kecewa	
--	--	--

4.3 Labeling Data

Pada tahap labeling data menggunakan google colab dan ahli bahasa. 1895 Data yang akan diberikan label pada google colab dan 1894 data yang digunakan pada ahli bahasa, dari Total keseluruhan data sebanyak 3.789 ulasan. sentimen dikategorikan menjadi positif, netral dan negatif. Berikut merupakan Script Labeling data pada pemrograman Python:

4.3.1 Labeling Google Colab

1. import pandas untuk membaca data

```
import pandas as pd

data = pd.read_csv("Hasil_Labelling_Data_Colab.csv", sep=';')
data.head()
```

2. perintah untuk menampilkan hasil labeling google colab

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

sentiment_count = data['sentiment'].value_counts()

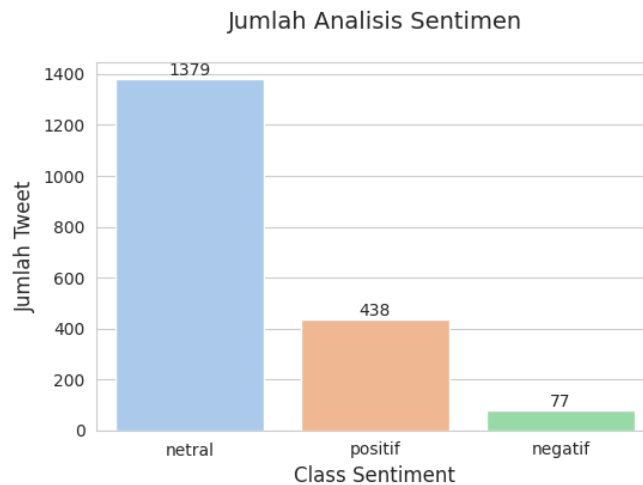
sns.set_style('whitegrid')

fig, ax = plt.subplots(figsize=(6, 4))
ax = sns.barplot(x=sentiment_count.index, y=sentiment_count.values, palette='pastel')
plt.title('Jumlah Analisis Sentimen', fontsize=14, pad=20)
plt.xlabel('Class Sentimen', fontsize=12)
plt.ylabel('Jumlah Tweet', fontsize=12)

for i, count in enumerate(sentiment_count.values):
    ax.text(i, count+0.10, str(count), ha='center', va='bottom')

plt.show()
```

Berikut merupakan hasil labeling google colab kelas sentimen positif, netral dan negatif dapat dilihat pada gambar 4.2 dibawah ini.



Gambar 4. 2 Hasil Labeling Google colab

Hasil pelabelan sentimen google colab menunjukkan bahwa 438 data memiliki sentimen positif, 1379 data memiliki sentimen netral, dan 77 data memiliki sentimen negatif. Berdasarkan hasil pelabelan menggunakan ahli bahasa, dapat disimpulkan bahwa sentimen pengguna aplikasi Seabank pada rentang waktu 01 Januari 2022 hingga 20 Desember 2023 didominasi oleh sentimen netral dari keseluruhan data sebanyak 1894 ulasan.

4.3.2 Labeling Ahli Bahasa

1. Impor padas dan membaca hasil preprocessing

```
import pandas as pd

data = pd.read_csv("Hasil_Labelling_Data ahli bahasa.csv", sep=';')
data.head()
```

2. Perintah untuk menampilkan hasil labeling ahli bahasa


```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

sentiment_count = data['sentiment'].value_counts()

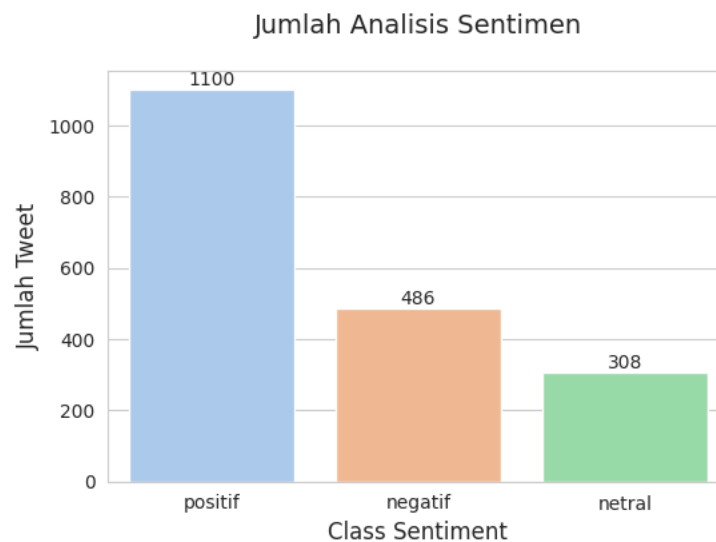
sns.set_style('whitegrid')

fig, ax = plt.subplots(figsize=(6, 4))
ax = sns.barplot(x=sentiment_count.index, y=sentiment_count.values, palette='pastel')
plt.title('Jumlah Analisis Sentimen', fontsize=14, pad=20)
plt.xlabel('Class Sentiment', fontsize=12)
plt.ylabel('Jumlah Tweet', fontsize=12)

for i, count in enumerate(sentiment_count.values):
    ax.text(i, count+0.10, str(count), ha='center', va='bottom')

```

Berikut merupakan hasil labeling ahli bahasa kelas sentimen positif, netral dan negatif dapat dilihat pada gambar 4.3 dibawah ini.



Gambar 4. 3 Hasil Labeling Ahli Bahasa

Hasil pelabelan sentimen ahli bahasa menunjukkan bahwa 1100 data memiliki sentimen positif, 308 data memiliki sentimen netral, dan 486 data memiliki sentimen negatif. Berdasarkan hasil pelabelan menggunakan ahli bahasa, dapat disimpulkan bahwa sentimen pengguna aplikasi Seabank pada rentang waktu 01 Januari 2022 hingga 20 Desember 2023 didominasi oleh sentimen positif.

4.4 Spitting Data

Dalam penelitian ini menggunakan rasio pembagian 80 data training dan 20 data testing (80:20).

1. Mengimport padas dan membaca hasil labelling data csv

```
import pandas as pd

data = pd.read_excel("labelling ahli bahasa dan colab(2).xlsx")
data.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3788 entries, 0 to 3787
Data columns (total 2 columns):
#   Column          Non-Null Count  Dtype
---  -
0   stemming_data   3786 non-null    object
1   sentiment       3788 non-null    object
dtypes: object(2)
memory usage: 59.3+ KB
```

2. Impor train_test_split dan bagi data menjadi data pelatihan dan pengujian

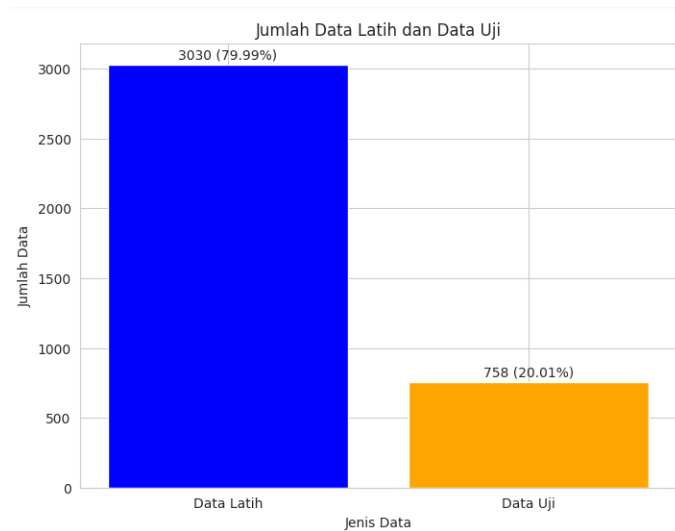
```
from sklearn.model_selection import train_test_split

# Bagi data menjadi data pelatihan dan pengujian
X_train, X_test, y_train, y_test = train_test_split(data['stemming_data'], data['sentiment'], test_size=0.2, random_state=42)
```

3. Menampilkan informasi jumlah data

```
# Menampilkan informasi jumlah data
print(f'Jumlah Data Latih: {len(X_train)}')
print(f'Jumlah Data Uji: {len(X_test)}')
```

Berikut hasil pembagian data yang menampilkan jumlah data dengan rasio 80 data training dan 20 data testing (80:20) dapat dilihat pada gambar 4.4 dibawah ini.



Gambar 4. 4 Pembagian data latih dan data uji

Hasil pembagian data training dan data testing dengan rasio (80:20) menunjukkan bahwa sebanyak 3031 ulasan (atau 79.99% dari total data) ditetapkan sebagai data latih, sementara 758 ulasan (atau 20.01% dari total data) ditetapkan sebagai data uji dari keseluruhan 3788 ulasan. Dengan pembagian yang proporsional antara data latih dan data uji, diharapkan dapat memastikan evaluasi yang akurat terhadap kinerja model dalam menganalisis sentimen ulasan pengguna aplikasi Seabank.

4.5 Models

Tahap pemodelan pada penelitian ini menggunakan dua model yaitu metode Naïve Bayes dan Support Vector Machine untuk klasifikasi.

4.5.1 Naïve Bayes Clasdifikasi

Dalam penelitian ini, digunakan tiga jenis model *Naïve Bayes*, yaitu model *Gaussian Naïve Bayes*, *Multinomial Naïve Bayes*, dan *Bernoulli Naïve Bayes*.

1. tahap pertama import library numpy dan padas

```
import numpy as np
import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer
from sklearn.naive_bayes import GaussianNB, MultinomialNB, BernoulliNB
from sklearn.metrics import accuracy_score, classification_report
from sklearn.preprocessing import LabelEncoder
from sklearn.pipeline import make_pipeline
from sklearn import metrics
```

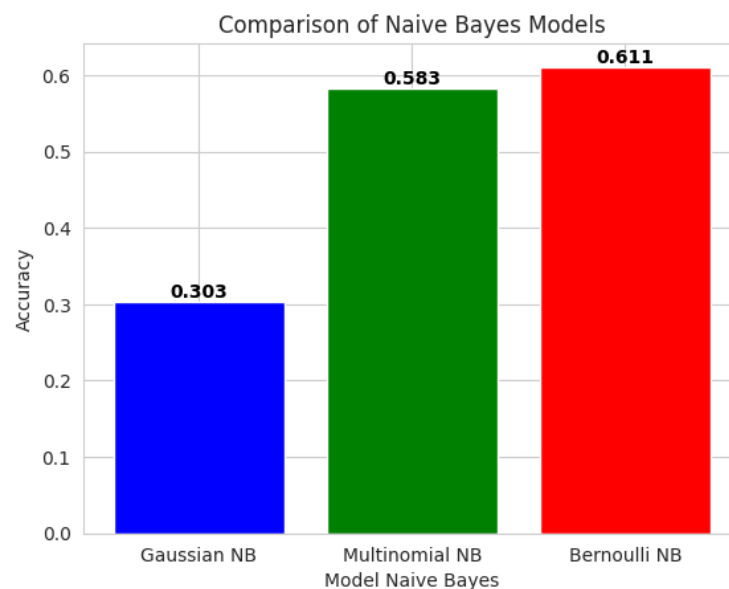
```
# Evaluasi model
y_pred_gnb = gnb.predict(X_test_vectorized.toarray())
y_pred_mnb = mnb.predict(X_test_vectorized)
y_pred_bnb = bnb.predict(X_test_vectorized)

# Print hasil evaluasi jika diperlukan
print("Gaussian Naive Bayes Accuracy:", metrics.accuracy_score(y_test, y_pred_gnb))
print("Multinomial Naive Bayes Accuracy:", metrics.accuracy_score(y_test, y_pred_mnb))
print("Bernoulli Naive Bayes Accuracy:", metrics.accuracy_score(y_test, y_pred_bnb))
```

Tabel 4.6 Hasil Akurasi

No	Model	Akurasi
1	Gaussian naïve bayes	30.3%
2	Multional naïve bayes	58.3%
3	Bernoulli naïve bayes	61.1%

Berikut merupakan hasil perbandingan akurasi tipe model naïve bayes dapat dilihat pada gambar 4.5 dibawah ini.



Gambar 4. 5 Perbandingan Akurasi Model Naïve Bayes

Dalam analisis ini, dilakukan perbandingan kinerja tiga tipe model Naïve Bayes, yaitu Multinomial Naïve Bayes, Bernoulli Naïve Bayes, dan Gaussian Naïve Bayes, dalam melakukan klasifikasi sentimen ulasan pengguna aplikasi Seabank.

Hasil perbandingan menunjukkan bahwa perolehan nilai akurasi tertinggi ditemukan pada model Bernoulli Naïve Bayes dengan akurasi mencapai 0.611. Sementara itu, Multinomial Naïve Bayes mencapai akurasi sebesar 0.583, sedangkan model Gaussian Naïve Bayes memperoleh akurasi terendah dengan nilai 0.303.

Dari hasil akurasi, dapat disimpulkan bahwa dalam konteks analisis sentimen ulasan pengguna aplikasi Seabank, model Bernoulli Naïve Bayes memberikan kinerja yang lebih baik dalam memprediksi sentimen pengguna dibandingkan dengan model Multinomial Naïve Bayes dan Gaussian Naïve Bayes.

2. Untuk menghasilkan laporan klasifikasi yang menyediakan matrix evaluasi berdasarkan hasil prediksi dari suatu model klasifikasi.

```
from sklearn.metrics import classification_report

# Menampilkan classification_report untuk setiap model
print("Gaussian Naive Bayes Classification Report:")
print(classification_report(y_test, y_pred_gnb))

print("\nMultinomial Naive Bayes Classification Report:")
print(classification_report(y_test, y_pred_mnb))

print("\nBernoulli Naive Bayes Classification Report:")
print(classification_report(y_test, y_pred_bnb))
```

Berikut ini merupakan hasil pengukuran evaluasi dari model gaussian naïve bayes, dapat dilihat pada gambar 4.6 dibawah ini.

Gaussian Naive Bayes Classification Report:				
	precision	recall	f1-score	support
negatif	0.16	0.64	0.26	112
netral	0.44	0.13	0.20	348
positif	0.55	0.38	0.45	298
accuracy			0.30	758
macro avg	0.38	0.38	0.30	758
weighted avg	0.44	0.30	0.31	758

Gambar 4. 6 Akurasi model GNB

Berikut hasil perhitungan nilai precision, recall, f1-score pada gaussian naïve bayes yaitu:

1. $Precision = \frac{TP}{TP+FP}$ = Presisi untuk kelas negatif adalah 0.16, untuk kelas netral adalah 0.44, dan untuk kelas positif adalah 0.55.
2. $Recall = \frac{TP}{TP+FN}$ = Recall untuk kelas negatif adalah 0.64, untuk kelas netral adalah 0.13, dan untuk kelas positif adalah 0.38.
3. $F1 - Score = 2 \frac{Precision \times Recall}{Precision+Recall}$ = F1-score untuk kelas negatif adalah 0.26, untuk kelas netral adalah 0.20, dan untuk kelas positif adalah 0.45.
4. Support mencerminkan jumlah sampel dalam setiap kelas. Ada 112 sampel dalam kelas negatif, 348 sampel dalam kelas netral, dan 298 sampel dalam kelas positif.
5. $Accuracy = \frac{TP+TN}{TP+FN+FP+TN}$ = Akurasi sebesar 0.30 menunjukkan bahwa model secara keseluruhan benar dalam 30% prediksinya.

Berikut ini merupakan hasil pengukuran evaluasi dari model multinomial naïve bayes, dapat dilihat pada gambar 4.7 dibawah ini

Multinomial Naive Bayes Classification Report:				
	precision	recall	f1-score	support
negatif	0.51	0.63	0.57	112
netral	0.63	0.35	0.45	348
positif	0.58	0.84	0.69	298
accuracy			0.58	758
macro avg	0.58	0.61	0.57	758
weighted avg	0.60	0.58	0.56	758

Gambar 4. 7 Akurasi model MNB

Berikut hasil perhitungan nilai precision, recall, f1-score pada multinomial naïve bayes yaitu:

1. $Precision = \frac{TP}{TP+FP}$ = Presisi untuk kelas negatif adalah 0.51, untuk kelas netral adalah 0.63, dan untuk kelas positif adalah 0.58.
2. $Recall = \frac{TP}{TP+FN}$ = Recall untuk kelas negatif adalah 0.63, untuk kelas netral adalah 0.35, dan untuk kelas positif adalah 0.84.
3. $F1 - Score = 2 \frac{Precision \times Recall}{Precision+Recall}$ = F1-score untuk kelas negatif adalah 0.57, untuk kelas netral adalah 0.45, dan untuk kelas positif adalah 0.69.
4. $Accuracy = \frac{TP+TN}{TP+FN+FP+TN}$ = Akurasi sebesar 0.58 menunjukkan bahwa model secara keseluruhan benar dalam 58% prediksinya.

Berikut ini merupakan hasil pengukuran evaluasi dari model bernoulli naïve bayes, dapat dilihat pada gambar 4.8 dibawah ini

Bernoulli Naive Bayes Classification Report:				
	precision	recall	f1-score	support
negatif	0.59	0.30	0.40	112
netral	0.58	0.62	0.60	348
positif	0.65	0.71	0.68	298
accuracy			0.61	758
macro avg	0.61	0.55	0.56	758
weighted avg	0.61	0.61	0.60	758

Gambar 4. 8 Akurasi model BNB

Berikut hasil perhitungan nilai precision, recall, f1-score pada bernoulli naïve bayes yaitu:

1. $Precision = \frac{TP}{TP+FP}$ = Presisi untuk kelas negatif adalah 0.59, untuk kelas netral adalah 0.58, dan untuk kelas positif adalah 0.65.
2. $Recall = \frac{TP}{TP+FN}$ = Recall untuk kelas negatif adalah 0.30, untuk kelas netral adalah 0.62, dan untuk kelas positif adalah 0.71.
3. $F1 - Score = 2 \frac{Precision \times Recall}{Precision+Recall}$ = F1-score untuk kelas negatif adalah 0.40, untuk kelas netral adalah 0.60, dan untuk kelas positif adalah 0.68.
4. $Accuracy = \frac{TP+TN}{TP+FN+FP+TN}$ = Akurasi sebesar 0.61 menunjukkan bahwa model secara keseluruhan benar dalam 61% prediksinya.
3. Menghasilkan confusion matrix dari hasil prediksi suatu model klasifikasi


```

from sklearn.metrics import confusion_matrix

# Menampilkan confusion matrix untuk setiap model
print("Gaussian Naive Bayes Confusion Matrix:")
print(confusion_matrix(y_test, y_pred_gnb))

print("\nMultinomial Naive Bayes Confusion Matrix:")
print(confusion_matrix(y_test, y_pred_mnb))

print("\nBernoulli Naive Bayes Confusion Matrix:")
print(confusion_matrix(y_test, y_pred_bnb))

```

4. Fungsi untuk plot confusion matrik

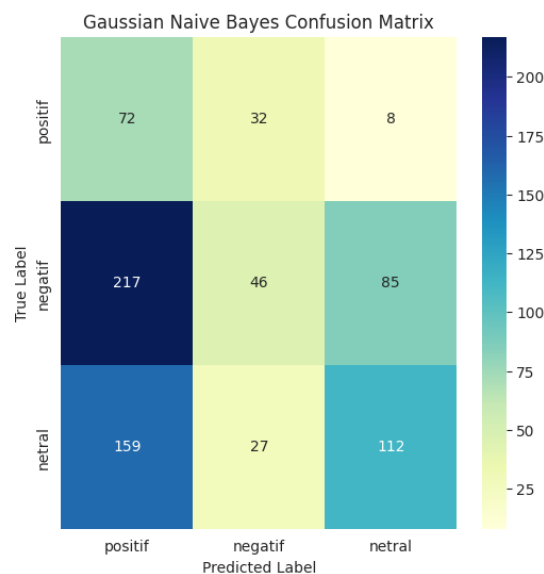
```

# Membuat dan melatih model Gaussian Naive Bayes
gnb = GaussianNB()
gnb.fit(X_train_vectorized.toarray(), y_train)

# Memprediksi label pada data uji
y_pred_gnb = gnb.predict(X_test_vectorized.toarray())

# Memanggil fungsi plot_confusion_matrix untuk menampilkan confusion matrix
plot_confusion_matrix("Gaussian Naive Bayes", y_test, y_pred_gnb)

```



Gambar 4. 9 Confusion matrix GNB

1. Kelas positif (baris 1)

- True positives (TP) : 72 jumlah sampel yang benar termasuk dalam kelas positif dan telah diprediksi dengan benar oleh model sebagai kelas positif.
- False negatives (FN) : 32 jumlah sampel yang sebenarnya termasuk dalam kelas positif, tetapi model memprediksi sebagai kelas negatif.
- False positives (FP) : 8 jumlah sampel yang sebenarnya termasuk dalam kelas netral, tetapi model salah memprediksi sebagai kelas positif.

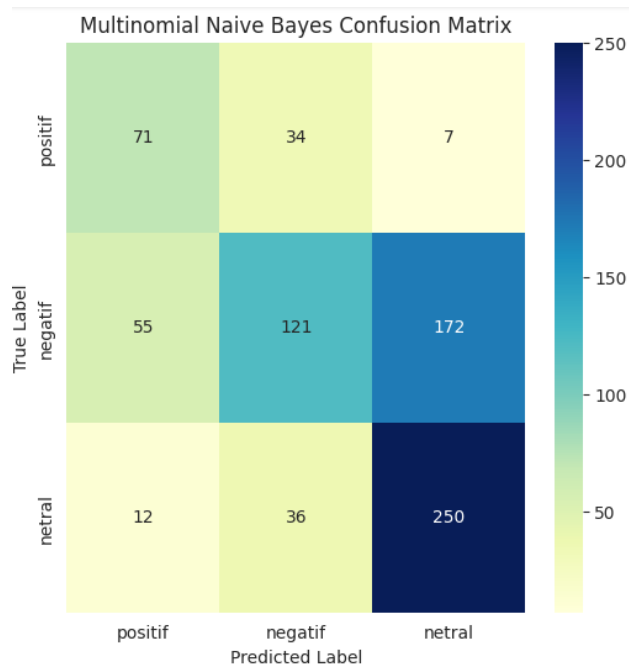
2. Kelas negatif (baris 2)

- False positives (FP) : 217 jumlah sampel yang sebenarnya positif tetapi diprediksi negatif
- True negatives (TN) : 46 jumlah sampel data yang benar termasuk kelas negatif dan telah diprediksi dengan benar oleh model sebagai kelas negatif.
- False negatives (FN) : 85 jumlah sampel yang sebenarnya termasuk netral, tetapi model salah memprediksi sebagai kelas negatif.

3. Kelas netral (baris 3)

- False positives (FP) : 159 jumlah sampel yang sebenarnya positif, tetapi model salah memprediksi sebagai kelas netral.
- False negatives (FN) : 27 jumlah sampel yang sebenarnya negatif tetapi model salah memprediksi sebagai kelas netral.
- True neutral (TN) : 112 jumlah sampel yang benar termasuk dalam kelas netral dan telah diprediksi dengan benar oleh model sebagai kelas netral.

```
# Membuat dan melatih model Multinomial Naive Bayes
mnb = MultinomialNB()
mnb.fit(X_train_vectorized, y_train)
y_pred_mnb = mnb.predict(X_test_vectorized)
plot_confusion_matrix("Multinomial Naive Bayes", y_test, y_pred_mnb)
```



Gambar 4. 10 Confusion matrix MNB

1. Kelas positif (baris 1)

- True positives (TP) : 71 jumlah sampel yang benar termasuk dalam kelas positif dan telah diprediksi dengan benar oleh model sebagai kelas positif.
- False negatives (FN) : 34 jumlah sampel yang sebenarnya kelas positif, tetapi model memprediksi sebagai kelas negatif.
- False positives (FP) : 7 jumlah sampel yang sebenarnya kelas netral, tetapi model salah memprediksi sebagai kelas positif.

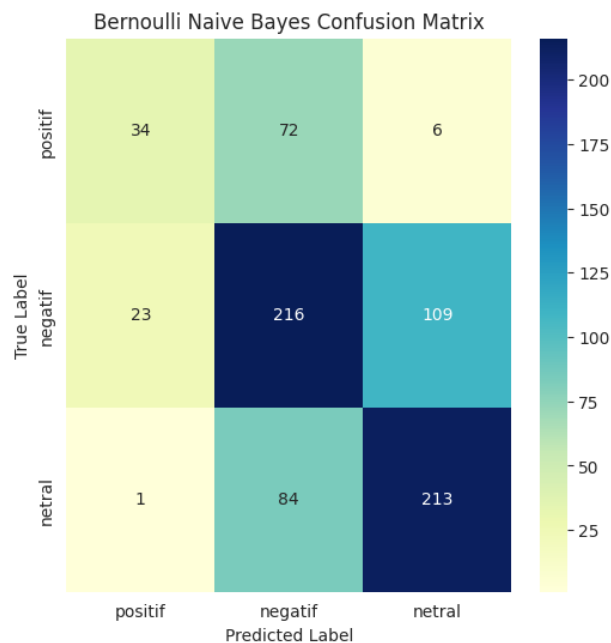
2. Kelas negatif (baris 2)

- False positives (FP) : 55 jumlah sampel yang sebenarnya positif tetapi diprediksi negatif
- True negatives (TN) : 121 jumlah sampel data yang benar termasuk kelas negatif dan telah diprediksi dengan benar oleh model sebagai kelas negatif.
- False negatives (FN) : 172 jumlah sampel yang sebenarnya termasuk netral, tetapi model salah memprediksi sebagai kelas negatif.

3. Kelas netral (baris 3)

- False positives (FP) : 29 jumlah sampel yang sebenarnya positif, tetapi model salah memprediksi sebagai kelas netral.
- False negatives (FN) : 36 jumlah sampel yang sebenarnya negatif tetapi model salah memprediksi sebagai kelas netral.
- True neutral (TN) : 250 jumlah sampel yang benar termasuk dalam kelas netral dan telah diprediksi dengan benar oleh model sebagai kelas netral.

```
# Membuat dan melatih model Bernoulli Naive Bayes
bnb = BernoulliNB()
bnb.fit(X_train_vectorized, y_train)
y_pred_bnb = bnb.predict(X_test_vectorized)
plot_confusion_matrix("Bernoulli Naive Bayes", y_test, y_pred_bnb)
```



Gambar 4. 11 Confusion matrix BNB

1. Kelas positif (baris 1)

- True positives (TP) : 34 jumlah sampel yang benar termasuk dalam kelas positif dan telah diprediksi dengan benar oleh model sebagai kelas positif.

- False negatives (FN) : 72 jumlah sampel yang sebenarnya termasuk dalam kelas positif, tetapi model memprediksi sebagai kelas negatif.
 - False positives (FP) : 6 jumlah sampel yang sebenarnya termasuk dalam kelas netral, tetapi model salah memprediksi sebagai kelas positif.
2. Kelas negatif (baris 2)
- False positives (FP) : 23 jumlah sampel yang sebenarnya positif tetapi diprediksi negatif
 - True negatives (TN) : 216 jumlah sampel data yang benar termasuk kelas negatif dan telah diprediksi dengan benar oleh model sebagai kelas negatif.
 - False negatives (FN) : 109 jumlah sampel yang sebenarnya termasuk netral, tetapi model salah memprediksi sebagai kelas negatif.
3. Kelas netral (baris 3)
- False positives (FP) : 1 jumlah sampel yang sebenarnya positif, tetapi model salah memprediksi sebagai kelas netral.
 - False negatives (FN) : 84 jumlah sampel yang sebenarnya negatif tetapi model salah memprediksi sebagai kelas netral.
 - True neutral (TN) : 213 jumlah sampel yang benar termasuk dalam kelas netral dan telah diprediksi dengan benar oleh model sebagai kelas netral.

4.5.2 Model Support Vector Machines (SVM)

1. Tahap pertama import library seaborn untuk menghasilkan visualisasi data

```
from sklearn.svm import SVC
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.feature_extraction.text import CountVectorizer

# Melakukan vektorisasi pada data latih dan uji
X_train_vectorized = vectorizer.fit_transform(X_train)
X_test_vectorized = vectorizer.transform(X_test)
```

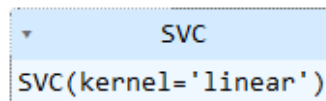
```

# Menampilkan hasil vektorisasi
print("Matriks Vektorisasi untuk Data Latih:")
print(X_train_vectorized.toarray())

# menampilkan sebagian kecil matriks
print("\nSebagian kecil Matriks Vektorisasi untuk Data Latih:")
print(X_train_vectorized[:5, :].toarray())

# Membuat dan melatih model SVM
svm = SVC(kernel='linear')
svm.fit(X_train_vectorized, y_train)

```



Berikut merupakan hasil evaluasi model SVM dapat dilihat pada gambar 4.12 dibawah ini.

SVM Accuracy: 0.6319261213720316

SVM Classification Report:

	precision	recall	f1-score	support
negatif	0.51	0.53	0.52	112
netral	0.61	0.69	0.65	348
positif	0.71	0.61	0.66	298
accuracy			0.63	758
macro avg	0.61	0.61	0.61	758
weighted avg	0.64	0.63	0.63	758

Gambar 4. 12 Akurasi model svm

Berikut hasil perhitungan nilai precision, recall, f1-score pada support vectort machines (SVM) yaitu:

1. $Precision = \frac{TP}{TP+FP}$ = Presisi untuk kelas negatif adalah 0.51, untuk kelas netral adalah 0.61, dan untuk kelas positif adalah 0.71.
2. $Recall = \frac{TP}{TP+FN}$ = Recall untuk kelas negatif adalah 0.53, untuk kelas netral adalah 0.69, dan untuk kelas positif adalah 0.61.

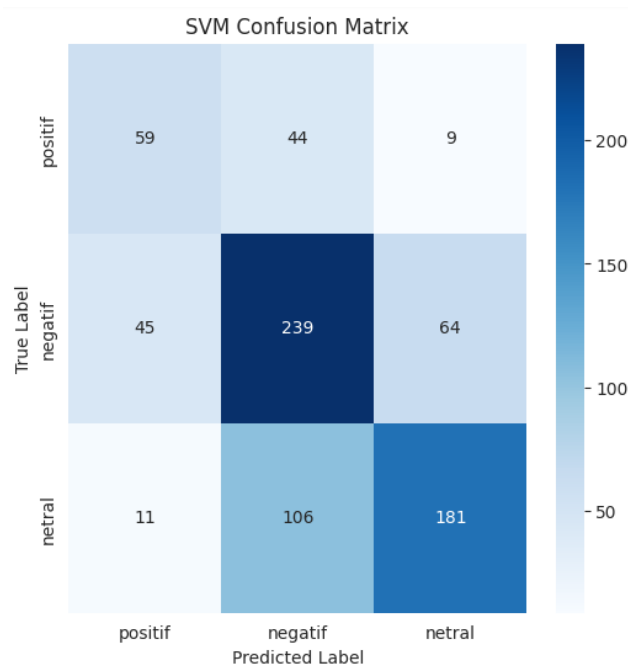
3. $F1 - Score = 2 \frac{Precision \times Recall}{Precision + Recall} =$ F1-score untuk kelas negatif adalah 0.52, untuk kelas netral adalah 0.65, dan untuk kelas positif adalah 0.66.

4. $Accuracy = \frac{TP+TN}{TP+FN+FP+TN} =$ Akurasi sebesar 631 menunjukkan bahwa model secara keseluruhan benar dalam 63% prediksinya.

2. Menampilkan confusion matrix untuk SVM

```
# Memanggil fungsi plot_confusion_matrix untuk menampilkan confusion matrix SVM
plot_confusion_matrix("SVM", y_test, y_pred_svm)
```

Berikut merupakan hasil confusion matrix model support vector machines (SVM) dapat dilihat pada gambar 4.13 dibawah ini.



Gambar 4. 13 Confusion matrix SVM

1. Kelas positif (baris 1)

- True positives (TP) : 59 jumlah sampel yang benar termasuk dalam kelas positif dan telah diprediksi dengan benar oleh model sebagai kelas positif.

- False negatives (FN) : 44 jumlah sampel yang sebenarnya termasuk dalam kelas positif, tetapi model memprediksi sebagai kelas negatif.
 - False positives (FP) : 9 jumlah sampel yang sebenarnya termasuk dalam kelas netral, tetapi model salah memprediksi sebagai kelas positif.
2. Kelas negatif (baris 2)
- False positives (FP) : 45 jumlah sampel yang sebenarnya positif tetapi diprediksi negatif.
 - True positives (TP) : 239 jumlah sampel data yang benar termasuk kelas negatif dan telah diprediksi dengan benar oleh model sebagai kelas negatif.
 - False negatives (FN) : 64 jumlah sampel yang sebenarnya netral, tetapi model salah memprediksi sebagai kelas negatif.
3. Kelas netral (baris 3)
- False positives (FP) : 11 jumlah sampel yang sebenarnya positif, tetapi model salah memprediksi sebagai kelas netral.
 - False negatives (FN) : 106 jumlah sampel yang sebenarnya negatif tetapi model salah memprediksi sebagai kelas netral.
 - True positives (TP) : 181 jumlah sampel yang benar termasuk dalam kelas netral dan telah diprediksi dengan benar oleh model sebagai kelas netral.

4.5.3 Perbandingan Model Naïve Bayes dan SVM

```
import matplotlib.pyplot as plt

# Data akurasi Naive Bayes
naive_bayes_labels = ['GNB', 'MNB', 'BNB']
naive_bayes_accuracies = [metrics.accuracy_score(y_test, y_pred_gnb),
                           metrics.accuracy_score(y_test, y_pred_mnb),
                           metrics.accuracy_score(y_test, y_pred_bnb)]
```



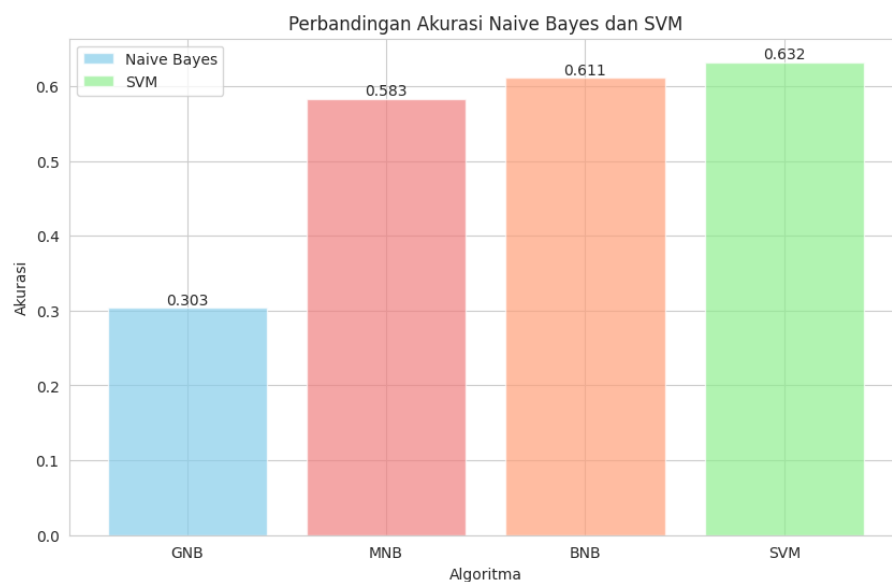
```

# Data akurasi SVM
svm_label = 'SVM'
svm_accuracy = accuracy_score(y_test, y_pred_svm)

# Plotting
plt.figure(figsize=(10, 6))

```

Berikut merupakan hasil perbandingan akurasi tiga model naïve bayes dan model SVM dapat dilihat pada gambar 4.14 dibawah ini.



Gambar 4. 14 Perbandingan akurasi naïve bayes dan SVM

Dalam analisis ini, dilakukan perbandingan kinerja antara dua jenis algoritma klasifikasi, yaitu *Naïve Bayes (Multinomial, Bernoulli, dan Gaussian)* dan *Support Vector Machines (SVM)*. Hasil perbandingan menghasilkan bahwa nilai akurasi tertinggi ditemukan pada model *Support Vector Machines (SVM)*, dengan nilai akurasi mencapai 0.632. Sementara itu, nilai akurasi tertinggi dari *Naïve Bayes* terdapat pada model *Bernoulli Naïve Bayes*, dengan nilai akurasi 0.611. Namun, perlu dicatat bahwa nilai akurasi terendah ditemukan pada model *Gaussian Naïve Bayes*, dengan nilai akurasi sebesar 0.303, sedangkan akurasi *Multinomial Naïve Bayes* adalah 0.583.

Dari hasil analisis ini, dapat disimpulkan bahwa dalam konteks analisis sentimen ulasan pengguna aplikasi Seabank, model *Support Vector Machines (SVM)* memberikan kinerja yang lebih baik dalam memprediksi sentimen pengguna dibandingkan dengan model *Naïve Bayes*. Namun, dari tiga tipe *Naïve Bayes* yang diuji, *Bernoulli Naïve Bayes* menunjukkan kinerja tertinggi, sedangkan *Gaussian Naïve Bayes* menunjukkan kinerja terendah.

4.6 WordCloud

1. Impor pandas as pd

```
import pandas as pd
from wordcloud import WordCloud
import matplotlib.pyplot as plt

# Pisahkan dataset berdasarkan sentimen
sentimen_negatif = data[data['sentiment'] == 'negatif']['stemming_data'].str.cat(sep=' ')
sentimen_positif = data[data['sentiment'] == 'positif']['stemming_data'].str.cat(sep=' ')
sentimen_netral = data[data['sentiment'] == 'netral']['stemming_data'].str.cat(sep=' ')

```

```
# Fungsi untuk membuat dan menampilkan WordCloud
def create_wordcloud(text, title):
    wordcloud = WordCloud(width=800, height=400, random_state=42,
                           max_font_size=100, background_color='black').generate(text)

    plt.figure(figsize=(10, 5))
    plt.imshow(wordcloud, interpolation='bilinear')
    plt.axis('off')
    plt.title(title)
    plt.show()

```

```
# Membuat WordCloud untuk sentimen negatif
create_wordcloud(sentimen_negatif, 'WordCloud Sentimen Negatif')

```

Berikut ini merupakan gambar wordcloud sentimen negatif dapat dilihat pada gambar 4.15 dibawah ini:

Informasi ini memberikan wawasan yang berharga bagi aplikasi Seabank dalam memahami pandangan, kebutuhan, dan pengalaman pengguna terhadap aplikasi mereka. Dengan memperhatikan kata-kata yang sering muncul dalam ulasan pengguna, Seabank dapat mengidentifikasi kekuatan dan kelemahan aplikasi, serta membuat perbaikan atau peningkatan yang sesuai.