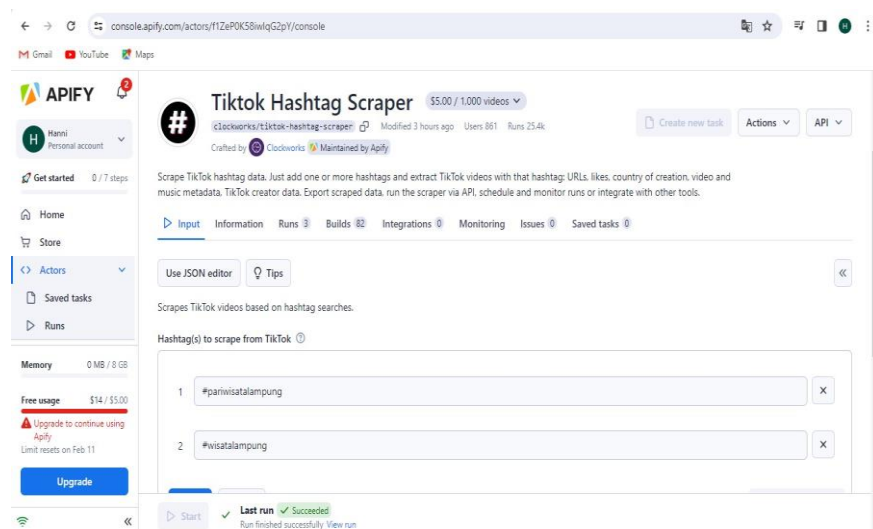


## BAB IV

### HASIL DAN PEMBAHASAN

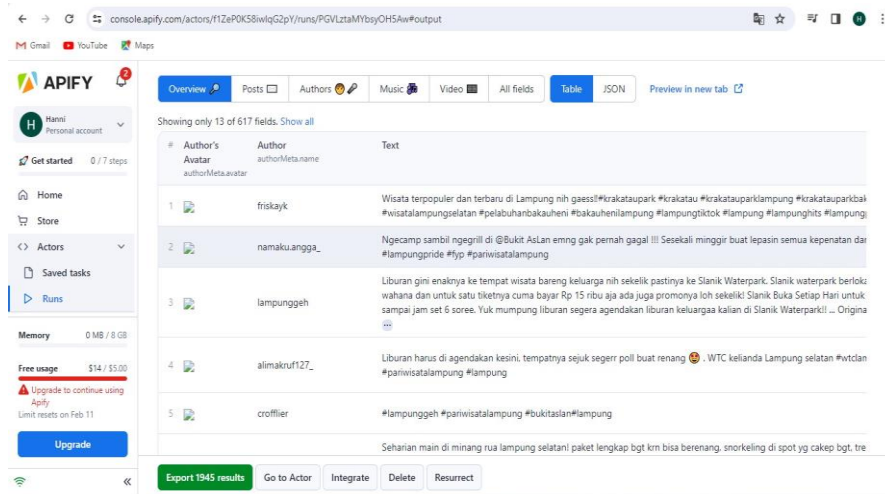
#### 4.1 Hasil Crawling Data

Pada tahapan ini, pengumpulan data pengguna Tiktok terhadap Pariwisata Lampung mengacu pada Hastag dan mengumpulkan data sebanyak 1000 komentar. Pengumpulan Data ini menggunakan bantuan ekstensi Google Chrome yang berfungsi sebagai *web scraper* untuk mengambil data atau komentar yang terdapat pada situs online secara gratis yaitu TikTok Hastag Scrapper dan mengimpornya ke *Microsoft Excel*.



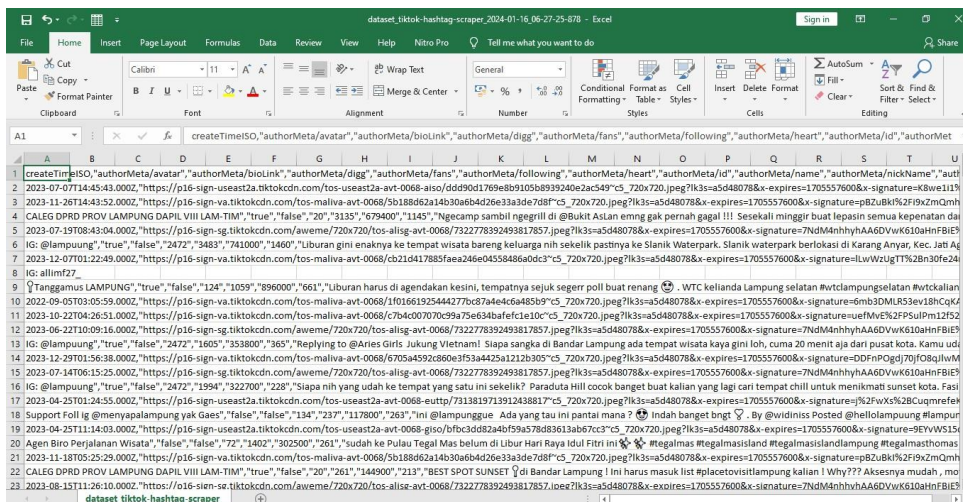
Gambar 4.1 Tiktok Hastag Scrapper

Pada tahapan ini, langkah pertama yang dilakukan adalah login web console.apify.com kemudian klik *store* dan ketik Tiktok, maka akan banyak pilihan seperti Tiktok Scrapper, Tiktok Sound Scrapper dan lain-lain. Karena penelitian ini mengacu pada Hastag maka menggunakan Tiktok Hastag Scrapper. Langkah selanjutnya yaitu meng input kan Hastag yang kita butuhkan seperti #pariwisatalampung #wisatalampung.



Gambar 4.2 Scraping Komentar Tiktok

Tahapan selanjutnya yaitu dapat memulai scraping data komentar dengan meng *exportdata* yang telah terkumpul dan menyimpan hasil ke dalam bentuk Microsoft excel.



Gambar 4.3 Hasil Scraping Data

Selanjutnya, Data yang telah tersimpan dalam bentuk Microsoft Excel, belum dapat dibaca dengan jelas karena data ini yang nanti nya dibaca menggunakan *Tools GoogleCollab* dan akan dilakukan tahapan *Preprocessing*.

```
import pandas as pd\n\ndf = pd.read_csv("dataset_tiktok-hashtag-scraper_2024-01-16_06-27-25-878.csv")\ndf.info()
```

Gambar 4.4 Kode Program untuk membaca data set

Kode program yang terdapat pada gambar 4.4 berfungsi untuk membaca data set yang di Scrapping sebelumnya serta membaca informasi yang terdapat didalam data set. Selanjutnya menampilkan komentar yang akan dibaca terdapat pada gambar 4.5

authorMeta/ttSeller	authorMeta/verified	authorMeta/video	shareCount	playCount	commentCount	text
True	False	277	8098	1100000	2247	Wisata terpopuler dan terbaru di Lampung nih g...
True	False	20	3135	679400	1145	Ngecamp sambil ngegrill di @Bukit AsLan emng g...
True	False	2472	3483	741000	1460	Liburan gini enaknya ke tempat wisata bareng k...
True	False	124	1059	896000	661	Liburan harus di agendakan kesini, tempatnya s...
False	False	9	914	168300	359	#lampunggeh #pariwisatalampung #bukitastan#lam...
...	...	...	...	...	...	...
False	False	286	1663	688400	611	Tempat di Lampung keren-keren kan? . Video by ...

Gambar 4.5 Hasil Tampilan Data Set

Data set yang ditampilkan pada tahapan ini adalah data keseluruhan yang dihasilkan saat Scrapping data, masih terdapat field yang dibutuhkan dan tidak dibutuhkan, yang nanti nya akandi sesuaikan dengan field yang kita butuhkan, terdapat pada gambar 4.6

```
data = pd.DataFrame(df[['createTimeISO', 'authorMeta/name', 'shareCount', 'playCount', 'commentCount', 'text']])
data.head(1000)
```

	createTimeISO	authorMeta/name	shareCount	playCount	commentCount	text
0	2023-07-07T14:45:43.000Z	friskayk	8098	1100000	2247	Wisata terpopuler dan terbaru di Lampung nih g...
1	2023-11-26T14:43:52.000Z	namaku.angga_	3135	679400	1145	Ngecamp sambil ngegrill di @Bukit AsLan emng g...
2	2023-07-19T08:43:04.000Z	lampunggeh	3483	741000	1460	Liburan gini enaknya ke tempat wisata bareng k...
3	2023-12-07T01:22:49.000Z	alimakruf127_	1059	896000	661	Liburan harus di agendakan kesini, tempatnya s...
4	2022-09-05T03:05:59.000Z	crofflier	914	168300	359	#lampunggeh #pariwisatalampung #bukitastan#lam...
...	...	...	...	...	...	...
995	2023-12-20T11:39:49.000Z	pikniklampung	1663	688400	611	Tempat di Lampung keren-keren kan? . Video by ...
996	2022-11-25T01:09:34.000Z	chatbuatkamu	894	435000	446	Pantai emang nggak pernah salah. Lu yang salah...
997	2023-11-16T14:13:46.000Z	awnoona	930	558800	290	marina emg seindah itu☐☐ #pantamarinalampung ...
998	2023-08-13T11:27:42.000Z	riowicak_	1965	442700	811	Kmne lagi yee?#rekomendasihealinglampung #fyp ...
999	2024-01-06T17:10:28.000Z	jayawisatapahawang	1365	867400	1110	Kalo Belum Bisa Ke Banda Neira Minimal Ke Lamp...

Gambar 4.6 Kode Program Data Frame

Pada tahapan ini, Data Frame berfungsi untuk memisahkan field-field yang kita butuhkan, field yang kita butuhkan pada penelitian ini yaitu *Create Time, Name, Share, Coment dan Text*. Data ini lah yang nanti nya akan di proses pada Preprocessing.

```
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1414 entries, 0 to 1413
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  ---                -
0   createTimeISO         1414 non-null   object
1   authorMeta/name      1414 non-null   object
2   shareCount           1414 non-null   int64
3   playCount            1414 non-null   int64
4   commentCount         1414 non-null   int64
5   text                 1413 non-null   object
dtypes: int64(3), object(3)
memory usage: 66.4+ KB
```

Gambar 4.7 Kode Program untuk membaca jumlah data set

Gambar 4.7 terdapat kode program untuk membaca jumlah data set yang kita miliki, data set yang kita miliki sesuai dengan variabel yang kita butuhann yaitu sebanyak 1413 data yang akan masuk ke tahapan Preprocessing. Sebelum masuk ke dalam tahapan Preprocessing, langkah pertama yaitu Menghapus data Duplikat terdapat pada Gambar 4.8

#### HAPUS DATA DUPLIKAT

```
[ ] data.drop_duplicates(subset ="text", keep = 'first', inplace = True)

data.info()
```

Gambar 4.8 Kode Program Menghapus Data Duplikat

Menghapus Data Duplikat ini berfungsi untuk menghapus data yang double, dengan tujuan jika ada baris komentar yang sama terhadap data tersebut maka yang akan di ambil adalah baris pertama pada data set tersebut.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1367 entries, 0 to 1413
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   createTimeISO         1367 non-null   object
1   authorMeta/name      1367 non-null   object
2   shareCount           1367 non-null   int64
3   playCount            1367 non-null   int64
4   commentCount         1367 non-null   int64
5   text                  1366 non-null   object
dtypes: int64(3), object(3)
memory usage: 74.8+ KB
```

Gambar 4.9 Hasil Hapus Data Duplikat

Gambar 4.9 adalah Hasil data duplikat yang sebelumnya pada Data Frame terdapat 1413 data dan setelah melakukan proses Hapus data duplikat terdapat 1367 Data.

## 4.2 Pre Processing

Tahapan ini terdiri dari beberapa proses karena data komentar yang di hasilkan tidak sepenuhnya menggunakan kata baku dan akan di persiapkan melalui proses Preprocessing. Tahap Preprocessing dilakukan dengan menggunakan bantuan *library* pada Bahasa Pemrograman Python. Penerapan tahap Preprocessing data pada penelitian ini dilakukandengan beberapa tahapan.

### 4.2.1 Cleaning Data

Proses Cleaning data pada tahapan ini bertujuan untuk membersihkan data dari hal yang tidak perlu seperti tanda baca, menghapus angka dan lain sebagainya. Adapun kode program Cleaning data terdapat pada gambar 4.10

**CLEANING DATA**

```
[ ] import re
import string
import nltk

# Fungsi untuk menghapus URL
def remove_URL(tweet):
    if tweet is not None and isinstance(tweet, str):
        url = re.compile(r'https?://\S+|www\.\S+')
        return url.sub(r'', tweet)
    else:
        return tweet

# Fungsi untuk menghapus HTML
def remove_html(tweet):
    if tweet is not None and isinstance(tweet, str):
        html = re.compile(r'<.*?>')
        return html.sub(r'', tweet)
    else:
        return tweet
```

Gambar 4.10 Kode Program menghapus URL dan HTML

```
# Fungsi untuk menghapus simbol
def remove_symbols(tweet):
    if tweet is not None and isinstance(tweet, str):
        tweet = re.sub(r'^a-zA-Z0-9\s', '', tweet) # Menghapus semua simbol
    return tweet

# Fungsi untuk menghapus angka
def remove_numbers(tweet):
    if tweet is not None and isinstance(tweet, str):
        tweet = re.sub(r'\d', '', tweet) # Menghapus semua angka
    return tweet
```

Gambar 4.11 Kode Program menghapus Emoji

```
# Fungsi untuk menghapus emoji
def remove_emoji(tweet):
    if tweet is not None and isinstance(tweet, str):
        emoji_pattern = re.compile("[
            u"\U0001F600-\U0001F64F" # emoticons
            u"\U0001F300-\U0001F5FF" # symbols & pictographs
            u"\U0001F680-\U0001F6FF" # transport & map symbols
            u"\U0001F700-\U0001F77F" # alchemical symbols
            u"\U0001F780-\U0001F7FF" # Geometric Shapes Extended
            u"\U0001F800-\U0001F8FF" # Supplemental Arrows-C
            u"\U0001F900-\U0001F9FF" # Supplemental Symbols and Pictographs
            u"\U0001FA00-\U0001FA6F" # Chess Symbols
            u"\U0001FA70-\U0001FAFF" # Symbols and Pictographs Extended-A
            u"\U0001F004-\U0001F0CF" # Additional emoticons
            u"\U0001F1E0-\U0001F1FF" # flags
        ]+", flags=re.UNICODE)
        return emoji_pattern.sub(r'', tweet)
    else:
        return tweet
```

Gambar 4.12 Kode Program menghapus Simbol dan Angka

```
df['cleaning'] = df['text'].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))

df.head(100)
```

Gambar 4.13 Kode Program Menampung hasil Cleaning

Kode Program yang terdapat pada gambar 4.13 berfungsi untuk menampung Hasil *Cleaning* yang sudah dilakukan, dengan cara memanggil masing-masing fungsi seperti *Remove\_URL*, *Remove\_HTML*, *Remove\_Emoji*, *Remove\_symbols*, *Remove\_Numbers* dan akan ditampung dalam bentuk *table*

1	Cleaning
2	Wisata terpopuler dan terbaru di Lampung nih gaesskrakataupark krakatau krakatau parklampung krakatau parkbakauhenlampungselatan wisataviral tempatwisatahits wisatalampung v
3	Ngecamp sambil ngegrill di Bukit AsLan emng gak pernah gagal Sesekali minggir buat lepasin semua kepenatan dari aktifitas keseharian bukitaslan wisatalampung placetovisit wisita
4	Liburan gini enaknya ke tempat wisata bareng keluarga nih sekelik pastinya ke Slank Waterpark Slank waterpark berlokasi di Karang Anyar Kec Jati Agung Kabupaten Lampung Sel
5	Liburan harus di agendakan kesini tempatnya sejuk segerr poll buat renang WTC kelanda Lampung selatan wtclampungselatan wtckalandlampung kaland lampungselatan pariwis
6	lampunggeh pariwisatalampung bukitaslanlampung
7	Seharian main di minang rus lampung selatan paket lengkap bgt krn bisa berenang snorkeling di spot yg cakep bgt trekking ke green canyon trekking ke goa dan danau seru lampung la
8	Replying to Aries Girls Jukung Vietnam Siapa sangka di Bandar Lampung ada tempat wisata kaya gini loh cuma menit aja dari pusat kota Kamu udah pernah ke Jukung Vietnam ba
9	Lampung Punya Potensi Wisata Kelas DUNIA pariwisatalampung lampunggeh krui petrustjandra dpdrlampung
10	Siapa nih yang udah ke tempat yang satu ini sekelik Paraduta Hill cocok banget buat kalian yang lagi cari tempat chill untuk menikmati sunset kota Fasilitasnya juga lengkap dan ada i
11	ini lampungue Ada yang tau ini pantai mana Indah banget bngt By widiniss Posted hellolampung lampungue infoselampung lamlungmenyapa HelloLampung MenyapaLampun
12	sudah ke Pulau Tegal Mas belum di Libur Hari Raya Idul Fitri ini tegalmas tegalmasisland tegalmasislandlampung tegalmasthomas tegalmasthomasbeach tegalmasthomaslampung tega
13	BEST SPOT SUNSET di Bandar Lampung Ini harus masuk list placetovisitlampung kalian Why Aksesnya mudah motor mobil gas aja sampe lokasi jalannya cakep pas naik disug
14	SIMPAN UNTUK NANTI Halo sekelik Lampung Geh yuk ikut kita city tour Bandar Lampung naik helikopter yang ada di Puncak Mas Dari atas ketinggian kita tuh bisa ngeliat in
15	lampungselatan lampunggeh bandarlampung preweddingshoot lampungpride lampunggeh pariwisatalampung preweddingstyle fypviral foryou fyp
16	fyp fyp fypdonggggggg pengantin pantai lampung pariwisatalampung minangruabeach praweddingcasual ombak laksminua jangan Lupa follow Instagram Laksminua
17	ATV FOREST ADVENTURE SUDAH DI BUKA NIH Wahhh tambah penasaran gak si guys Hihhi Yuk jadwalkan liburan kalian ke Bukit AsLan Jangan sampai terlewatkan guys bu
18	Siapa nih yang udh punya niatan mau camping tapi takut ribet karena gapunya perlengkapannya Tenang aja sekelik ini dia nih tempat yang punya solusinya Ngecamp di Wira Garden
19	Avoo merapat Sandiaga Uno karvanakmur labuhanmaringgai lampungtimur dinaspariwisata pariwisataindonesia tiktok pariwisatalampung anakmotor wisatamalam

Gambar 4.14 Hasil Cleaning

## 4.2.2 Case Folding

Tahapan Preprocessing yang kedua yaitu *Case Folding* dengan mengubah semua huruf menjadi huruf kecil dengan tujuan untuk menyeragamkan karakter pada data.

### CASE FOLDING

```
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(1000)
```

Gambar 4.15 Kode Program Case Folding

Kode Program yang terdapat pada gambar 4.15 adalah kode case folding yang berfungsi untuk lowercase atau mengubah huruf menjadi huruf kecil.

`df['case_folding'] = df['cleaning'].apply(case_folding)` ialah fungsi untuk membaca data hasil cleaning yang sebelumnya sudah dilakukan dan selanjutnya akan dilakukan *case folding*.

shareCount	playCount	commentCount	text	cleaning	case_folding
8098	1100000	2247	Wisata terpopuler dan terbaru di Lampung nih g...	Wisata terpopuler dan terbaru di Lampung nih g...	wisata terpopuler dan terbaru di lampung nih g...
3135	679400	1145	Ngecamp sambil ngegrill di @BukitAsLan emng g...	Ngecamp sambil ngegrill di BukitAsLan emng ga...	ngecamp sambil ngegrill di bukit aslan emng ga...
3483	741000	1460	Liburan gini enaknya ke tempat wisata bareng k...	Liburan gini enaknya ke tempat wisata bareng k...	liburan gini enaknya ke tempat wisata bareng k...
1059	896000	661	Liburan harus di agendakan kesini, tempatnya s...	Liburan harus di agendakan kesini tempatnya se...	liburan harus di agendakan kesini tempatnya se...
914	168300	359	#lampunggeh #pariwisatalampung #bukitaslan#iam...	lampunggeh pariwisatalampung bukitaslanlampung	lampunggeh pariwisatalampung bukitaslanlampung

Gambar 4.16 Hasil Case Folding

### 4.2.3 Tokenization

Tahapan *Preprocessing* yang selanjutnya ialah *Tokenization* yaitu proses untuk membagi teks menjadi token atau memisahkan setiap kata.

```

TOKENIZATION
def tokenize(text):
    if isinstance(text, str):
        tokens = text.split()
        return tokens
    else:
        return []

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

df.head(1000)

```

Gambar 4.17 Kode Program Tokenization

Kode Program 4.17 ialah proses Tokenization yang akan memisahkan setiap kata dalam bentuk koma

***df['tokenize'] = df ['case\_folding'].apply(tokenize)***

***df['tokenize']*** berfungsi untuk menampung hasil tokenisasi ***df ['case\_folding'].apply(tokenize)*** berfungsi untuk membaca data hasil casefolding yang akan dilakukan tokenisasi



text	cleaning	case_folding	tokenize
Wisata terpopuler dan terbaru di Lampung nih g...	Wisata terpopuler dan terbaru di Lampung nih g...	wisata terpopuler dan terbaru di lampung nih g...	[wisata, terpopuler, dan, terbaru, di, lampung...
Ngecamp sambil ngegrill di @Bukit AsLan emng g...	Ngecamp sambil ngegrill di Bukit AsLan emng ga...	ngecamp sambil ngegrill di bukit aslan emng ga...	[ngecamp, sambil, ngegrill, di, bukit, aslan, ...
Liburan gini enaknya ke tempat wisata bareng k...	Liburan gini enaknya ke tempat wisata bareng k...	liburan gini enaknya ke tempat wisata bareng k...	[liburan, gini, enaknya, ke, tempat, wisata, b...
Liburan harus di agendakan kesini, tempatnya s...	Liburan harus di agendakan kesini tempatnya se...	liburan harus di agendakan kesini tempatnya se...	[liburan, harus, di, agendakan, kesini, tempat...
#lampunggeh #pariwisatalampung #bukitaslan#iam...	lampunggeh pariwisatalampung bukitaslanlampung	lampunggeh pariwisatalampung bukitaslanlampung	[lampunggeh, pariwisatalampung, bukitaslanlamp...
...	...	...	...
Tempat di Lampung kerenkeren kan? . Video by ...	Tempat di Lampung kerenkeren kan Video by Inf...	tempat di lampung kerenkeren kan video by inf...	[tempat, di, lampung, kerenkeren, kan, video, ...
Pantai emang nggak pernah salah. Lu yang salah...	Pantai emang nggak pernah salah Lu yang salah ...	pantai emang nggak pernah salah lu yang salah ...	[pantai, emang, nggak, pernah, salah, lu, yang...

Gambar 4.18 Hasil Tokenization

#### 4.2.4 Stopword Removal

Tahapan Preprocessing yang selanjutnya ialah Stopword Removal yaitubertujuan untuk menghapus kata yang tidak relevan didalam suatu kalimat atau tidak memiliki makna berdasarkan kamusnya.

```

STOPWORD REMOVAL

[6] from nltk.corpus import stopwords
    nltk.download('stopwords')
    stop_words = stopwords.words('indonesian')

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Unzipping corpora/stopwords.zip.

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(1000)

```

Gambar 4.19 Kode Program Stopword Removal

Kode program yang terdapat pada gambar 4.19 adalah tahapan *Stopword Removal*

*df['stopword removal'] berfungsi sebagai penampung hasil Stopword Removal = df['tokenize'].apply(lambda x: remove\_stopword(x))*

berfungsi untuk membaca hasil tokenize dan perintah menghapus kata yang tidak memiliki makna

text	cleaning	case_folding	tokenize	stopword removal
Wisata terpopuler dan terbaru di Lampung nih g...	Wisata terpopuler dan terbaru di Lampung nih g...	wisata terpopuler dan terbaru di lampung nih g...	[wisata, terpopuler, dan, terbaru, di, lampung...	[wisata, terpopuler, terbaru, lampung, nih, ga...
Ngecamp sambil ngegrill di @Bukit AsLan emng g...	Ngecamp sambil ngegrill di Bukit AsLan emng ga...	ngecamp sambil ngegrill di bukit aslan emng ga...	[ngecamp, sambil, ngegrill, di, bukit, aslan, ...	[ngecamp, ngegrill, bukit, aslan, emng, gak, g...
Liburan gini enaknya ke tempat wisata bareng k...	Liburan gini enaknya ke tempat wisata bareng k...	liburan gini enaknya ke tempat wisata bareng k...	[liburan, gini, enaknya, ke, tempat, wisata, b...	[liburan, gini, enaknya, wisata, bareng, kelua...
Liburan harus di agendakan kesini, tempatnya s...	Liburan harus di agendakan kesini tempatnya se...	liburan harus di agendakan kesini tempatnya se...	[liburan, harus, di, agendakan, kesini, tempat...	[liburan, agendakan, kesini, tempatnya, sejuk...
#lampunggeh #pariwisatalampung #bukitaslanlamp...	lampunggeh pariwisatalampung bukitaslanlampung	lampunggeh pariwisatalampung bukitaslanlampung	[lampunggeh, pariwisatalampung, bukitaslanlamp...	[lampunggeh, pariwisatalampung, bukitaslanlamp...
...	...	...	...	...
Tempat di Lampung kerenkeren kan? Video by ...	Tempat di Lampung kerenkeren kan Video by Inf...	tempat di lampung kerenkeren kan video by inf...	[tempat, di, lampung, kerenkeren, kan, video, ...	[lampung, kerenkeren, video, by, info, hilingh...
Pantai emang nggak pernah salah. Lu yang salah...	Pantai emang nggak pernah salah Lu yang salah ...	pantai emang nggak pernah salah lu yang salah ...	[pantai, emang, nggak, pernah, salah, lu, yang...	[pantai, emang, nggak, salah, lu, salah, wisat...
marina emg seindah itu□□ #pantaimarinalampung ...	marina emg seindah itu pantaimarinalampung for...	marina emg seindah itu pantaimarinalampung for...	[marina, emg, seindah, itu, pantaimarinalampung...	[marina, emg, seindah, pantaimarinalampung, fo...
Kmne lagi yee? rekomendasihealinlampu...	Kmne lagi rekomendasihealinlampu...	kmne lagi rekomendasihealinlampu...	[kmne, lagi, rekomendasihealinlampu...	[kmne, rekomendasihealinlampu...

Gambar 4.20 Hasil Stopword Removal

#### 4.2.5 Stemming Data

Tahapan terakhir yaitu Stemming data berfungsi untuk mereduksi kata yang di infleksikan menjadi kata dasar, misalnya “bermain-main” maka yang di ambil hanya kata “main”.

```
!pip install Sastrawi

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

Collecting Sastrawi
  Downloading Sastrawi-1.0.1-py2.py3-none-any.whl (209 kB)
    ----- 209.7/209.7 kB 3.6 MB/s eta 0:00:00
Installing collected packages: Sastrawi
Successfully installed Sastrawi-1.0.1

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)
```

Gambar 4.21 Kode Program Stemming Data

Kode Program pada gambar 4.21 ialah proses terakhir pada Preprocessing yaitu *Stemming Data*.

`df['stemming_data']` berfungsi sebagai penampung hasil *stemming*

`df['stopword removal']` berfungsi sebagai pembaca data hasil stopword removal yang akan masuk ke tahapan *stemming*

text	cleaning	case_folding	tokenize	stopword removal	stemming_data
Wisata terpopuler dan terbaru di Lampung nih g...	Wisata terpopuler dan terbaru di Lampung nih g...	wisata terpopuler dan terbaru di lampung nih g...	[wisata, terpopuler, dan, terbaru, di, lampung...	[wisata, terpopuler, terbaru, lampung, nih, ga...	wisata populer baru lampung nih gaesskrakataup...
Ngecamp sambil ngegrill di @Bukit AsLan emng g...	Ngecamp sambil ngegrill di Bukit AsLan emng ga...	ngecamp sambil ngegrill di bukit aslan emng ga...	[ngecamp, sambil, ngegrill, di, bukit, aslan, ...	[ngecamp, ngegrill, bukit, aslan, emng, gak, g...	ngecamp ngegrill bukit aslan emng gak gagal mi...
Liburan gini enakny ke tempat wisata bareng k...	Liburan gini enakny ke tempat wisata bareng k...	liburan gini enakny ke tempat wisata bareng k...	[liburan, gini, enakny, ke, tempat, wisata, b...	[liburan, gini, enakny, wisata, bareng, kelua...	libur gin enak wisata bareng keluarga nih keli...
Liburan harus di agendakan kesini, tempatnya s...	Liburan harus di agendakan kesini tempatnya se...	liburan harus di agendakan kesini tempatnya se...	[liburan, harus, di, agendakan, kesini, tempat...	[liburan, harus, di, agendakan, kesini, tempatnya, sejuk...	libur agenda kesini tempat sejuk segerr poll r...
#lampunggeh #pariwisatalampung #bukitaslan#lam...	lampunggeh pariwisatalampung bukitaslanlampung	lampunggeh pariwisatalampung bukitaslanlampung	[lampunggeh, pariwisatalampung, bukitaslanlamp...	[lampunggeh, pariwisatalampung, bukitaslanlamp...	lampunggeh pariwisatalampung bukitaslanlampung

Gambar 4.22 Hasil Stemming Data

```
[ ] df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1367 entries, 0 to 1413
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  ---                -
0   createTimeISO         1367 non-null   object
1   authorMeta/name      1367 non-null   object
2   shareCount           1367 non-null   int64
3   playCount            1367 non-null   int64
4   commentCount         1367 non-null   int64
5   text                  1366 non-null   object
6   cleaning              1366 non-null   object
7   case_folding          1366 non-null   object
8   tokenize              1367 non-null   object
9   stopword removal     1367 non-null   object
10  stemming_data        1367 non-null   object
dtypes: int64(3), object(8)
memory usage: 128.2+ KB
```

Gambar 4.23 Kode Program baca data Hasil Preprocessing

Kode Program yang terdapat pada gambar 4.23 ialah berfungsi sebagai pembaca data hasil tahapan preprocessing. Data yang telah melewati tahapan

preprocessing seperti cleaning dan lain-lain sebanyak 1367 yang siap untuk masuk ketahapan berikutnya.

### 4.3 Menghapus Data Kosong

Data set hasil Preprocessing akan masuk ke tahapan Penghapusan Data kosong apabila data yang telah memasuki tahapan preprocessing terdeteksi terdapat baris yang kosong maka akan di hapus.

#### *HAPUS DATA BERNILAI KOSONG (NAN)*

```
[ ] data = df.dropna()

[ ] data.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1366 entries, 0 to 1413
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   createTimeISO         1366 non-null   object
1   authorMeta/name      1366 non-null   object
2   shareCount           1366 non-null   int64
3   playCount            1366 non-null   int64
4   commentCount         1366 non-null   int64
5   text                  1366 non-null   object
6   cleaning              1366 non-null   object
7   case_folding         1366 non-null   object
8   tokenize              1366 non-null   object
9   stopword removal     1366 non-null   object
10  stemming_data         1366 non-null   object
dtypes: int64(3), object(8)
memory usage: 128.1+ KB
```

Gambar 4.24 Kode Program dan Hasil Penghapusan Data Kosong

Kode program yang terdapat pada gambar 4.24 adalah penghapusan data kosong yang sesuai dengan gambar 4.23 terdapat 1367 data hasil preprocessing dan setelah dilakukan tahapan penghapusan data kosong terdapat 1366 data. Maka, terdapat 1 baris kolom yang kosong dan telah dihapus.

### 4.4 Hasil Preprocessing

```
[ ] data.to_csv('Hasil_Preprocessing_Data.csv', encoding='utf8', index=False)
```

Gambar 4.25 Kode Program download hasil Preprocessing

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1 createTim	author	Me	share	Cou	play	Count	comment	text	cleaning	case	fold	tokenize	stopword	stemming	data					
2 2023-07-0	friskayk	8098	1100000	2247	Wisata	ter	Wisata	ter	Wisata	,	Wisata	,	Wisata	populer	baru	lampung	nih	gaes	krakatau	park
3 2023-11-2	namaku.a	3135	679400	1145	Ngecamp	Ngecamp	Ngecamp	Ngecamp	Ngecamp	Ngecamp	Ngecamp	Ngecamp	Ngecamp	Ngecamp	Ngecamp	Ngecamp	Ngecamp	Ngecamp	Ngecamp	Ngecamp
4 2023-07-1	lampung	3483	741000	1460	Liburan	gi	Liburan	gi	Liburan	,	Liburan	,	Liburan	,	Liburan	,	Liburan	,	Liburan	,
5 2023-12-0	alimakruf	1059	896000	661	Liburan	hi	Liburan	hi	Liburan	hi	Liburan	hi	Liburan	hi	Liburan	hi	Liburan	hi	Liburan	hi
6 2023-09-0	crofflier	914	168300	359	#lampung	lampung	lampung	lampung	lampung	lampung	lampung	lampung	lampung	lampung	lampung	lampung	lampung	lampung	lampung	lampung
7 2023-10-2	10cinam	1191	327300	514	Seharian	i	Seharian	i	Seharian	i	Seharian	i	Seharian	i	Seharian	i	Seharian	i	Seharian	i
8 2023-06-2	lampung	1905	353800	305	Replying	t	Replying	t	Replying	t	Replying	t	Replying	t	Replying	t	Replying	t	Replying	t
9 2023-12-2	petrustjar	334	211200	317	Lampung	Lampung	Lampung	Lampung	Lampung	Lampung	Lampung	Lampung	Lampung	Lampung	Lampung	Lampung	Lampung	Lampung	Lampung	Lampung
10 2023-07-1	lampung	1994	322700	228	Siapa	nih	Siapa	nih	Siapa	nih	Siapa	nih	Siapa	nih	Siapa	nih	Siapa	nih	Siapa	nih
11 2023-04-2	menyapal	237	117800	263	ini	@lam	ini	lampu	ini	lampu	ini	lampu	ini	lampu	ini	lampu	ini	lampu	ini	lampu
12 2023-04-2	nusantara	1402	302500	261	sudah	ke	sudah	ke	sudah	ke	sudah	ke	sudah	ke	sudah	ke	sudah	ke	sudah	ke
13 2023-11-1	namaku.a	261	144900	213	BEST	SPOT	BEST	SPOT	BEST	SPOT	BEST	SPOT	BEST	SPOT	BEST	SPOT	BEST	SPOT	BEST	SPOT
14 2023-08-1	lampung	343	183900	145	SIMPAN	SIMPAN	L	simpan	u	simpan	,	simpan	,	simpan	,	simpan	,	simpan	,	simpan
15 2023-10-0	hansphoti	545	135000	301	#lampung	lampung	lampung	lampung	lampung	lampung	lampung	lampung	lampung	lampung	lampung	lampung	lampung	lampung	lampung	lampung
16 2022-07-1	laksmi_mi	35	518500	24	#fyp	#fyp	fyp	fyp	fyp	fyp	fyp	fyp	fyp	fyp	fyp	fyp	fyp	fyp	fyp	fyp
17 2023-07-0	buktiaslar	619	149600	346	ATV	FORE	ATV	FORE	atv	forest	atv	forest	atv	forest	atv	forest	atv	forest	atv	forest
18 2023-07-3	lampung	555	160700	168	Siapa	nih	Siapa	nih	Siapa	nih	Siapa	nih	Siapa	nih	Siapa	nih	Siapa	nih	Siapa	nih
19 2022-12-3	pantai_mi	387	127100	258	Ayoo	mer	Ayoo	mer	Ayoo	mer	Ayoo	mer	Ayoo	mer	Ayoo	mer	Ayoo	mer	Ayoo	mer
20 2023-08-1	lampung	283	200200	98	SIMPAN	SIMPAN	L	simpan	u	simpan	,	simpan	,	simpan	,	simpan	,	simpan	,	simpan
21 2023-08-1	lampung	269	114900	81	warga	lam	warga	lam	warga	lam	warga	lam	warga	lam	warga	lam	warga	lam	warga	lam
22 2023-08-1	lampung	605	158000	122	KUNJUNJUN	KUNJUNJUN	Kunjungi	Kunjungi	Kunjungi	Kunjungi	Kunjungi	Kunjungi	Kunjungi	Kunjungi	Kunjungi	Kunjungi	Kunjungi	Kunjungi	Kunjungi	Kunjungi
23 2023-08-1	lampung	69	136400	223	GESER	SA	GESER	SA	geseran	geser	,	geser	,	geser	,	geser	,	geser	,	geser

Gambar 4.26 Hasil Preprocessing Data

## 4.5 Normalisasi Kata

Setelah tahapan Pre Processing selesai tahapan selanjutnya adalah Normalisasi Kata yaitu memperbaiki kata yang tidak baku menjadi kata baku.

### NORMALISASI KATA

```
import pandas as pd

data = pd.read_csv("Hasil_Preprocessing_Data (1).csv")
data.head(5)
```

Gambar 4.27 Kode Program Normalisasi Kata

Tahapan Normalisasi kata yang pertama yaitu membaca hasil PreProcessing data untuk mengetahui kata-kata yang tidak baku yang terdapat pada tahapan Preprocessing Data.

```
import pandas as pd
import numpy as np
from PIL import Image
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
import matplotlib.pyplot as plt

df = '.join(data['stemming_data'].tolist())

stopwords = set(STOPWORDS)
stopwords.update(['https', 'co', 'RT', '...', 'amp'])

wc = WordCloud(stopwords=stopwords, background_color="black", max_words=500, width=800, height=400)

wc.generate(df)

plt.figure(figsize=(10, 5))
plt.imshow(wc, interpolation='bilinear')
plt.axis("off")
plt.show()
```

Gambar 4.28 Kode Program Word Cloud Preprocessing



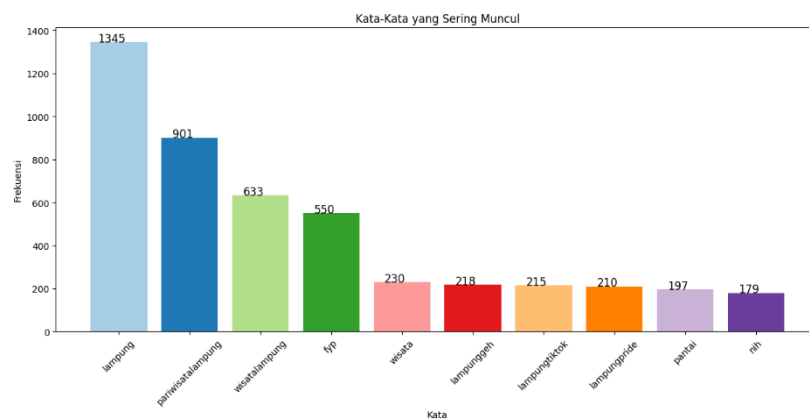
Gambar 4.29 Hasil WordCloud PreProcessing Data

```

▶ top_words = word_counts.most_common(500) # Select top 500 words
word, count = zip(*top_words)

# Export top 500 words to CSV
with open('top_500_words.csv', 'w', newline='') as csvfile:
    writer = csv.writer(csvfile)
    writer.writerow(['Word', 'Frequency'])
    for w, c in top_words:
        writer.writerow([w, c])
    
```

Gambar 4.30 Kode Program Frekuensi Kata (Kata-Kata yang sering muncul)



Gambar 4.31 Hasil Frekuensi Kata

```

import matplotlib.pyplot as plt
from collections import Counter
import csv

text = " ".join(data["stemming_data"])

tokens = text.split()
word_counts = Counter(tokens)

top_words = word_counts.most_common(10)
word, count = zip(*top_words)

colors = plt.cm.Paired(range(len(word)))

plt.figure(figsize=(15, 6))
bars = plt.bar(word, count, color=colors)
plt.xlabel("Kata")
plt.ylabel("Frekuensi")
plt.title("Kata-Kata yang Sering Muncul")
plt.xticks(rotation=45)

for bar, num in zip(bars, count):
    plt.text(bar.get_x() + bar.get_width() / 2 - 0.1, num + 1, str(num), fontsize=12, color='black', ha='center')

plt.show()

```

Gambar 4.32 Kode Program Perintah Ekspor Frekuensi Kata

	A	B
1	Word	Frequency
2	lampung	1345
3	pariwisatalampung	901
4	wisatalampung	633
5	fyp	550
6	wisata	230
7	lampunggeh	218
8	lampungtiktok	215
9	lampungpride	210
10	pantai	197
11	nih	179
12	pariwisata	153
13	bandarlampung	150
14	pulau	134
15	pariwisataindonesia	133
16	pahawang	127
17	ya	119
18	bandar	115
19	lampungselatan	114
20	barat	114
21	kelik	112
22	libur	110
23	kalianda	104

top\_500\_words

Gambar 4.33 Hasil Ekspor Frekuensi Kata

	tidak_baku	kata_baku
1		
2	woww	wow
3	aminn	amin
4	met	selamat
5	netaas	menetas
6	keberpa	keberapa
7	eeeehhhh	eh
8	kata2nyaaa	kata-katanya
9	hallo	halo
10	kaka	kakak
11	ka	kak
12	daah	dah
13	aaaahhhh	ah
14	yaa	ya
15	smga	semoga
16	slalu	selalu
17	amiin	amin
18	kk	kakak
19	trus	terus
20	kk	kakak
21	sii	sih
22	nyenengin	menyenangkan
23	bet	baneet

colloquial-indonesian-lexicon

Gambar 4.34 Kamus Kata Baku

Gambar 4.34 ialah kamus kata baku, kamus yang berisi kata tidak baku yang terdapat pada data set yang kita miliki. Kamus kata baku ini yang akan digunakan untuk di ekspor kedalam Google Collab dan dilakukan Normalisasi Kata.

```

# Baca dataset
data = pd.read_csv("Hasil_Preprocessing_Data (1).csv")
# data.info()
df_dataset1 = pd.DataFrame(data[['createTimeISO', 'authorMeta/name', 'shareCount', 'playCount', 'commentCount', 'steming_data']])

# Baca kamus kata tidak baku
kamus_data = pd.read_excel("kamus_katabaku.xlsx")
kamus_tidak_baku = dict(zip(kamus_data['tidak_baku'], kamus_data['kata_baku']))

```

Gambar 4.35 Kode Program Membaca data set

```

import pandas as pd

# Fungsi penggantian kata tidak baku
def replace_taboo_words(text, kamus_tidak_baku):
    if isinstance(text, str):
        words = text.split()
        replaced_words = []
        kalimat_baku = []
        kata_diganti = []
        kata_tidak_baku_hash = []

        for word in words:
            if word in kamus_tidak_baku:
                baku_word = kamus_tidak_baku[word]
                if isinstance(baku_word, str) and all(char.isalpha() for char in baku_word):
                    replaced_words.append(baku_word)
                    kalimat_baku.append(baku_word)
                    kata_diganti.append(word)
                    kata_tidak_baku_hash.append(hash(word))
                # else:
                #     replaced_words.append('<karakter eksplisit>')
                #     kalimat_baku.append(word)
            else:
                replaced_words.append(word)
        replaced_text = ' '.join(replaced_words)
    else:
        replaced_text = ''
        kalimat_baku = []
        kata_diganti = []
        kata_tidak_baku_hash = []

    return replaced_text, kalimat_baku, kata_diganti, kata_tidak_baku_hash

```

Gambar 4.36 Kode Program Perintah pengganti kata tidak baku

```

# Terapkan fungsi penggantian kata tidak baku
df_dataset1['hasil_normalisasi'], df_dataset1['Kata_Baku'], df_dataset1['Kata_Tidak_Baku'], df_dataset1['Kata_Tidak_Baku_Hash'] = zip(*df_dataset1['steming_data'].apply(lambda x: replace_taboo_words(x, kamus_tidak_baku)))

df_dataset1.head(5)

```

Gambar 4.37 Kode Program Menerapkan Fungsi kata tidak baku



	createTimeISO	author/meta/name	shareCount	playCount	commentCount	stening_data	hasil_normalisasi	Kata_Baku	Kata_Tidak_Baku	Kata_Tidak_Baku_Hash
0	2023-07-07T14:45:43.000Z	friskayk	8098	1100000	2247	wisata populer baru lampung nih gaesskrakataup...	wisata populer baru lampung nih gaesskrakataup...	[ini, informasi]	[nih, fyp]	[-7577283511413283941, 4994126468648512952]
1	2023-11-26T14:43:52.000Z	namaku.angga_	3135	679400	1145	ngecamp ngegrill bukit aslan emng gak gagal m...	ngecamp ngegrill bukit asai memang tidak gagal...	[asal, memang, tidak, melepas, informasi]	[aslan, emng, gak, lepasin, fyc]	[-4987209506859623563, -7939820256425442655, 8...]
2	2023-07-19T08:43:04.000Z	lampungeh	3483	741000	1460	libur gin enak wisata bareng keluarga nih ke...	libur gin enak wisata bareng keluarga ini ke...	[ini, rupiah, saja, ayo, iya, lampung]	[nih, rp, aja, yuk, geh, lampungeh]	[-7577283511413283941, -1282991528816001642, ...]
3	2023-12-07T01:22:49.000Z	alimakru127_	1059	896000	661	libur agenda kesini tempat sejuk seger poll...	libur agenda kesini tempat sejuk seger poll re...	[segar]	[seger]	[-5104688920782368958]
4	2022-09-05T03:05:59.000Z	croffier	914	168300	359	lampungeh pariwisata lampung buktiaslanlampung	lampung buktiaslanlampung	[lampung]	[lampungeh]	[-916804421287130257]

Gambar 4.38 Hasil pengganti kata baku

```
data.to_csv('Hasil_Normalisasi_Data.csv', encoding='utf8', index=False);
```

Gambar 4.39 Kode Program Membuat Wordcloud hasil Normalisasi

Tahapan pada gambar 4.39 ialah membuat wordcloud hasil Normalisasi untuk melihat perbedaan kata dari sebelum dan sesudah dilakukan Splitt 1000 Data

```
import pandas as pd
import numpy as np
from PIL import Image
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
import matplotlib.pyplot as plt

# Mengisi nilai NaN dengan string kosong ('')
df['hasil_normalisasi'] = df['hasil_normalisasi'].fillna('')

# Menggabungkan teks dari kolom 'hasil_normalisasi'
text = ' '.join(data['hasil_normalisasi'].astype(str).tolist())

stopwords = set(STOPWORDS)
stopwords.update(['https', 'co', 'RT', '...', 'amp'])

wc = WordCloud(stopwords=stopwords, background_color="black", max_words=500, width=800, height=400)

wc.generate(text)

plt.figure(figsize=(10, 5))
plt.imshow(wc, interpolation='bilinear')
plt.axis("off")
plt.show()
```

Gambar 4.40 Split 1000 Data

Pada Gambar 4.40 Split 1000 Data pada tahapan ini berfungsi untuk membagi data set menjadi 70:30 untuk masuk ke tahapan Labelling Data.

Pada tahapan Labelling Data yang terbagi menjadi 2 tahapan Labelling, yaitu

Tahap pertama : 300 data di Label kan menggunakan Google Collab

Tahap kedua : 700 data di Label kan menggunakan 3 ahli Bahasa

```

data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1366 entries, 0 to 1365
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   createTimeISO         1366 non-null   object
1   authorMeta/name      1366 non-null   object
2   shareCount           1366 non-null   int64
3   playCount            1366 non-null   int64
4   commentCount         1366 non-null   int64
5   text                 1366 non-null   object
6   cleaning             1366 non-null   object
7   case_folding         1366 non-null   object
8   tokenize             1366 non-null   object
9   stopword removal    1366 non-null   object
10  stemming_data        1366 non-null   object
dtypes: int64(3), object(8)
memory usage: 117.5+ KB

```

Gambar 4.41 Data Info

Pada gambar 4.41 Data Info berfungsi sebagai informasi data yang ada setelah memasuki beberapa proses tahapan. Pada tahapan ini terdapat 1366 data yang nantinya akan di lakukan proses pembagian data dan akan ada 1000 data yang tersisa sesuai data yang akan kita gunakan.

```

df = pd.DataFrame(data[['createTimeISO', 'authorMeta/name', 'shareCount', 'playCount', 'commentCount', 'hasil_normalisasi']].iloc[0:1006])

df.head(5)

```

Gambar 4.42 Perintah untuk mengambil data dari index 0 ke index 1006

```

df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1006 entries, 0 to 1005
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   createTimeISO         1006 non-null   object
1   authorMeta/name      1006 non-null   object
2   shareCount           1006 non-null   int64
3   playCount            1006 non-null   int64
4   commentCount         1006 non-null   int64
5   hasil_normalisasi    1000 non-null   object
dtypes: int64(3), object(3)
memory usage: 47.3+ KB

```

Gambar 4.43 Kode Program data info

Data info pada tahapan ini, berfungsi untuk melihat jumlah data setelah memasuki tahapan pengindeksan label, selanjutnya akan masuk ke tahapan penghapusan data kosong kembali.

#### HAPUS DATA BERNILAI KOSONG (NAN)

```

▶ data = df.dropna()

[ ] data.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1000 entries, 0 to 1005
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   createTimeISO         1000 non-null   object
1   authorMeta/name       1000 non-null   object
2   shareCount            1000 non-null   int64
3   playCount             1000 non-null   int64
4   commentCount          1000 non-null   int64
5   hasil_normalisasi     1000 non-null   object
dtypes: int64(3), object(3)
memory usage: 54.7+ KB

```

Gambar 4.44 Menghapus Data Kosong

Pada tahapan penghapusan data kosong ini berfungsi untuk menghapus data yang kosong setelah memasuki tahapan sebelumnya. Sebelum masuk ke tahapan penghapusan data kosong terdapat data 1006 hasil dari pengindeksan data, selanjutnya setelah dilakukan tahapan penghapusan data kosong tersisa 1000 data yang artinya terdapat 6 data kosong yang telah di hapus, 1000 data ini yang nantinya akan digunakan dalam pembagian data yang akan masuk ke tahapan Labelling Data set

```

▶ # Bagi dataset menjadi dua bagian
proses_labelling_colab = df[:300]
proses_labelling_ahli_bahasa = df[300:1000]

# Simpan setiap bagian ke dalam file CSV terpisah
proses_labelling_colab.to_csv('proses_labelling_colab.csv', index=False)
proses_labelling_ahli_bahasa.to_csv('proses_labelling_ahli_bahasa.csv', index=False)

```

Gambar 4.45 Membagi data set menjadi 2 bagian

createTimeISO	authorMeta/name	shareCount	playCount	commentCount	hasil_normalisasi
2023-07-07T14:45:43.000Z	friskayk	8098	1100000	2247	wisata populer baru lampung ini gaesskrakataupark krakatau parklampung krakatauparkbakauhenlampung
2023-11-26T14:43:52.000Z	namaku.angga_	3135	679400	1145	ngecamp ngegrill bukit asal memang tidak gagal minggu melepas penat aktifitas hari buktiaslan wisatalampung plac
2023-07-19T08:43:04.000Z	lampunggeh	3483	741000	1460	libur gin enak wisata bareng keluarga ini kelik pasti slank waterpark slank waterpark lokasi karang anyar kec jati ags
2023-12-07T01:22:49.000Z	allmakruf127_	1059	896900	661	libur agenda kesini tempat sejuk seger poli renang wtc kelilanda lampung selatan wtlampungselatan wtkalilanda
2022-09-05T03:05:59.000Z	croffier	914	168300	359	lampung buktiaslanlampung
2023-10-22T04:26:51.000Z	10cinnamonroll	1191	327900	514	hari main memang rua lampung selatan paket lengkap banget karena renang snorkeling spot yang cakep banget trek
2023-06-22T10:09:16.000Z	lampunggeh	1605	353800	365	replying tapi aries girls jukung vietnam sangka bandar lampung wisata kayak gin loh ment saja pusat kota sudah juk
2023-12-29T01:56:38.000Z	petrustjandra.id	334	211200	317	lampung potensi wisata kelas dunia lampung kurang petrustjandra ddrilampung
2023-07-14T06:15:25.000Z	lampunggeh	1994	322700	228	ini sudah kelik paraduta hill cocok banget cari chill nikmat sunset kota fasilitas lengkap musholla luas loh kelik lokasi
2023-04-25T01:24:55.000Z	menyapalampung	237	117800	263	lampunggue tau pantai indah banget banget wisata posted heliolampung lampunggue infosekolampung lamlungm
2023-04-23T11:14:03.000Z	nusantaratravelindo.id	1402	302500	261	pulau tegal mas libur raya idul Fitri tegalmas tegalmasiland tegalmasilandlampung tegalmaslandthomas tegalmaslthom
2023-11-18T05:25:29.000Z	namaku.angga_	261	144900	213	best spot sunset bandar lampung masuk list placetowisitlampung why akses mudah motor mobil gas saja sampai lok
2023-08-15T11:26:10.000Z	lampunggeh	343	181900	145	simpan halo kelik lampung iya ayo city tour bandar lampung heliopter puncak mas tinggi tuh melihat indah pندان
2022-10-08T23:29:27.000Z	hansphotography_09	545	135000	301	lampung preweddingshoot lampung preweddingstyle informasi foryou informasi
2023-07-16T22:55:56.000Z	laksmi_mua	35	518500	24	informasi informasi fypdonggggggg pengantin pantai lampung minangsuabeach praweddingcasual ombak laksmin
2023-07-07T05:04:56.000Z	buktiaslan	619	149600	346	atv forest adventure buka ini wah penasaran tidak sih guys hih ayo jadwal libur bukit asal lewat guys buktiaslan buk
2023-07-30T02:36:05.000Z	lampunggeh	555	160700	168	ini sudah niat camping takut ribet lengkap tenang saja kelik ini solusi ngecamp wira garden tidak khawatir lengkap c
2022-12-11T05:35:38.000Z	pantai_mutiara_baru	387	127100	258	ayo rapat sandiaga uno karyamakmur labuhamaringgai lampungtimur dinasariwisata tiktok anakmotor wisatamal
2023-12-30T02:35:40.000Z	allmakruf127_	269	114900	81	warga lampung tau pantai cantik pantai marina kalilanda lampung selatan pantaimarinlampung pantaimarina kalil
2023-08-12T08:48:20.000Z	lampunggeh	283	200200	98	simpan halo kelik lampung iya taman wisata lembah hijau ini lokasi jalan radem imba kusuma ratu sukadana ham bar
2023-08-12T08:24:00.000Z	lampunggeh	605	158000	122	unjung halo kelik lampung iya ini tau asik lihat sunset indah lampung lihat sunset indah lampung cocok banget lana
2023-08-13T02:26:09.000Z	lamouneeeh	69	136400	223	eser peristiwa jalan daerah maula perata banoun lamounz provinsi mluk kayak alam budava oetensi sumber dawa m

Gambar 4.46 Data Proses Labelling Collab

Gambar 4.46 ialah hasil pembagian data set yaitu proses labelling collab terdapat 300 data ter ekspor dalam bentuk csv

createTimeISO	authorMeta/name	shareCount	playCount	commentCount	hasil_normalisasi
2022-09-01T04:59:04.000Z	fetralamp	367	1000000	357	acara lempar selendang pantai bahari ketapang muli mekhanai sabu marga dalam muda mud ketpang mulimekhanaisabumargadalom salampangkonsalankemuaikh
2023-04-21T07:09:11.000Z	noviasafe	1104	268200	534	balas kurang pesisir barat sudah placetogolampung tanjungsietia pesisirbarat mandiribeach pulaupisangpesisirbarat labuhanjukung kurang lampunggeh fyp wisata
2022-10-10T10:34:57.000Z	hansphoiti	676	386700	599	lampungbali gekbali preweddingshoot wisatalampung pantallampung teamcentury informasi foryou lampunghiti
2024-01-01T08:04:12.000Z	allmakruf127_	1451	612400	513	dengar yang bilang orang lampung orang pantai gigi hui ya benar pantai cantik aksesnya pantalgigituragamus kelumlayan pantallampur
2022-12-21T02:02:39.000Z	putriaulia	2042	407400	870	wisatalampung wisatandarlampung farmday farmdedicationpark lampung wisata wisatanak wisatalkuliner informasi kulinerlampung wisata wisat
2023-11-21T07:09:11.000Z	noviasafe	1244	374800	443	mageran tidak ajak hehe lampung wisatalampung pesisirbarat cantilgakadabatasan corlaratulive
2023-06-11T02:02:39.000Z	potretlam	878	384100	815	replying tapi ig potretlampung healing view sejuk masuk ribus semuaibaditiktok amanditiktok wisatalampung lampung
2022-09-01T04:59:04.000Z	slank.waf	897	528800	867	ngenalin wahana air slank guys slankwaterpark slankwaterparklampung tiktok informasi wisatalampung tiktokpeltfyp
2023-05-01T08:04:12.000Z	llovelamp	3666	677500	685	pantai lampung ala santorini konsep ya instagramable kini banget pantai marina kalilanda lampung selatan video putrillalawati llovelampung lampung wisatalamp
2023-12-21T02:02:39.000Z	kawa_fa	644	278800	779	capcut libur mana libur lampung wisatalampung selankwaterpark
2023-06-21T07:09:11.000Z	slank.waf	478	1100000	712	sudah coba wahana air guys slankwaterpark slankwaterparklampung wisatalampungselatan wisatalampung informasi informasi
2023-12-01T08:04:12.000Z	lampunggi	748	461200	755	taman nasional woy kambas truk buka desember destinasi wisata sambut wisawawan kilu tulup rencana balai truk masyarakat desa senaga simulasi wisata taman
2023-03-21T07:09:11.000Z	wisatabeh	2490	272500	0	benar muluk guys kelayabubblepop wisatabehalang sebalangbeach wisatalampung jusjithglowing bujanggenlampung temlampung sumaterapride bakauhe
2023-03-21T07:09:11.000Z	risvian08	2764	493100	551	bingung cari inap nyaman healing coba mampir lembahbatuheritage bandar lampung lembahbatulampung lembahbatuheritage penganpanbandarlampung wisata
2023-03-21T07:09:11.000Z	putriaulia	4941	436900	815	balas calon istri mu video komentar yaadistikim sampai selesai enjoy lampung lampung informasi wisatalampung wisatandarlampung stycationidlampung wis
2023-01-21T07:09:11.000Z	sukasuka	2226	424500	951	wisata murah bagus lampung selatan air ya dingin banget woy segar langsung mata air gunung kesini informasi fypdonggggggg lampung lampung wisatalampung
2024-01-01T08:04:12.000Z	allmakruf127_	1002	444400	1045	balas misid gae ini penasaran jalan pantai gigi hui lampung tonton sampai biar tidak nyasar tau medan tiket masuk ke orang murah pantalgigituragamus pantai
2023-06-01T08:04:12.000Z	wisatapah	1182	309600	384	paket trip tanpa minimal serta paket open trip khusus weekend open trip pahamaw korang open trip tegal mas korang paket privat trip every day paket cocok fan
2023-06-21T07:09:11.000Z	ciktrav	869	219400	514	tau penyai bagus bersih mana ciktrav pantalkalanda wisatalkalanda wisatalampung wisatalampungselatan
2021-10-21T07:09:11.000Z	vansi765	2205	432900	262	marak wisatalampung sepakepakejedeerr
2023-07-21T07:09:11.000Z	potretlam	1585	535000	339	save saja rekomendasi nongki suka view kota lampung semuaibadilampung wisatalampung lampung
2023-11-21T07:09:11.000Z	wisatalam	928	315700	346	paraduta hill paradutahillig wisata wisatalampung kumpungvietnam

Gambar 4.47 Data Proses Labelling Ahli Bahasa

Gambar 4.47 ialah hasil pembagian data set yaitu proses labelling ahli bahasa terdapat 700 data ter ekspor dalam bentuk csv yang akan di labelkan oleh 3 ahli bahasa

```

import pandas as pd

# Membaca dataset
dataset1 = pd.read_csv("proses_labelling_ahli_bahasa.csv")
dataset1.info()
print("")

dataset2 = pd.read_csv("proses_labelling_colab.csv")
dataset2.info()

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 700 entries, 0 to 699
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   createTimeISO         700 non-null   object
1   authorMeta/name       700 non-null   object
2   shareCount            700 non-null   int64
3   playCount             700 non-null   int64
4   commentCount          700 non-null   int64
5   hasil_normalisasi     696 non-null   object
dtypes: int64(3), object(3)
memory usage: 32.9+ KB

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 300 entries, 0 to 299
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   createTimeISO         300 non-null   object
1   authorMeta/name       300 non-null   object
2   shareCount            300 non-null   int64
3   playCount             300 non-null   int64

```

Gambar 4.48 Membaca data set hasil pembagian Labelling

Pada gambar 4.48 adalah membaca data info hasil pembagian data set terdapat 700 data set yang siap dilabelkan menggunakan ahli bahasa dan terdapat 300 data set yang akan dilabelkan menggunakan Google Collab

```

data.to_csv('Hasil_Normalisasi_Data_2.csv',encoding='utf8', index=False);

```

Gambar 4.49 Ekspor Hasil Normalisasi Data 2

## 4.6 Labelling Dataset

Tahapan yang selanjutnya adalah Labelling Dataset yang akan melabelkan data dengan kategori Positif, Negative dan Netral.

## 4.6.1 Labelling Dataset Collab

### LABELLING DATASET

```

import pandas as pd

data = pd.read_csv("proses_labelling_collab.csv")
data.info()

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 300 entries, 0 to 299
Data columns (total 6 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   createTimeISO         300 non-null   object
 1   authorMeta/name       300 non-null   object
 2   shareCount            300 non-null   int64
 3   playCount             300 non-null   int64
 4   commentCount          300 non-null   int64
 5   hasil_normalisasi     298 non-null   object
dtypes: int64(3), object(3)
memory usage: 14.2+ KB

```

Gambar 4.50 Perintah membaca data labelling collab

### LABELLING NLTK

```

!pip install nltk
import nltk
from nltk.sentiment import SentimentIntensityAnalyzer
import pandas as pd

nltk.download('punkt')
nltk.download('vader_lexicon')

```

```

Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-packages (3.8.1)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk) (1.3.2)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk) (2023.12.25)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk) (4.66.1)
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package vader_lexicon to /root/nltk_data...
True

```

Gambar 4.51 Labelling NLTK

Labelling NLTK ialah Library yang berfungsi untuk menganalisis data yang dimiliki

```

from nltk.sentiment import SentimentIntensityAnalyzer
import numpy as np

data = SentimentIntensityAnalyzer()

labels = []
scores = []

for text in df['hasil_normalisasi']:
    if isinstance(text, str):
        sentiment_scores = data.polarity_scores(text)
        compound_score = sentiment_scores['compound']

        scores.append(compound_score)

        if compound_score > 0:
            label = 'positif'
        elif compound_score < 0:
            label = 'negatif'
        else:
            label = 'netral'

        labels.append(label)
    else:
        labels.append('netral')
        scores.append(np.nan)

df['sentiment_score'] = scores
df['sentiment'] = labels

data = df[['createTimeISO', 'authorMeta/name', 'shareCount', 'playCount', 'commentCount', 'hasil_normalisasi', 'sentiment_score', 'sentiment']]
data.head(10)

```

Gambar 4.52 Memberikan Label berdasarkan Score

Pada gambar 4.52 ialah memberikan label dan score berdasarkan kamus NLTK, dan berdasarkan hasil normalisasi maka akan dilakukan analisis yang di namakan *compound*

```

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

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

sns.set_style('whitegrid')

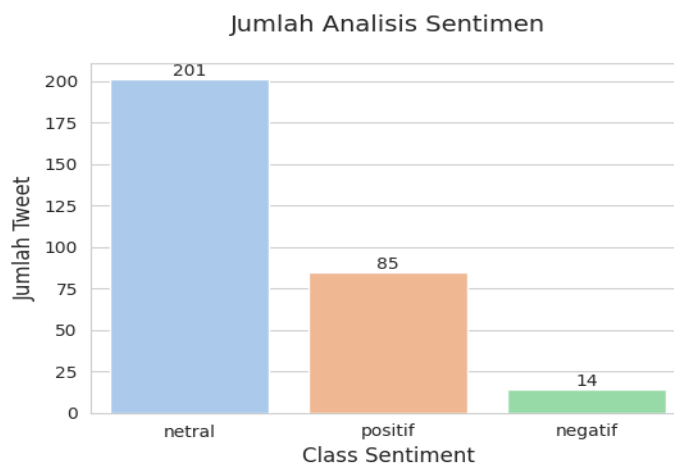
fig, ax = plt.subplots(figsize=(6, 4))
ax = sns.barplot(x=sentiment_count.index, y=sentiment_count.values, palette='pastel')
plt.title('Jumlah Analisis Sentimen', fontsize=14, pad=20)
plt.xlabel('Class Sentiment', fontsize=12)
plt.ylabel('Jumlah Tweet', fontsize=12)

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

plt.show()

```

Gambar 4.53 Perintah untuk menampilkan hasil Labelling Collab



Gambar 4.54 Hasil Analisis Sentiment Google Collab

Hasil Analisis Sentimen berdasarkan Labelling Collab yang terdapat pada gambar 4.54 terdapat 201 data komentar yang bernilai netral, 85 data komentar bernilai positif dan 14 data komentar bernilai Negatif.

```

df.to_csv('Hasil_Labelling_Data_Colab.csv', encoding='utf8', index=False)

```

Gambar 4.55 Ekspor Hasil Labelling Data Collab

#### 4.6.2 Labelling Ahli Bahasa

Pada tahapan Labelling Ahli bahasa ini menggunakan 3 ahli bahasa dengan hasil pelabelan yang berbeda beda dengan kategori pelabelan -1 untuk komentar negatif, 0 untuk netral dan 1 untuk komentar positif. Penghitungan pelabelan ahli bahasa ini menggunakan 700 komentar dan ditambah dengan 300 hasil pelabelan Collab.

1	createTim	authorMe	shareCou	playCount	comment	hasil_normalisasi	Labeling Bahasa 1	Labeling Bahasa 2	Labeling Bahasa 3	Kesimpulan
2	2022-11-1	rajomansi	0	894	2	otw jogjapariwisata	0	0	0	0
3	2022-12-2	pantal_mi	1	2738	5	gass kawan sandiaga	0	0	0	0
4	2023-09-2	sigertour.	10	2624	0	segenap crew sigert	0	0	1	0
5	2024-01-1	petrustjar	5	4356	6	pesona pesisir barat	0	0	0	0
6	2023-05-0	karyawisa	4	1029	1	putri indobesia pula	0	0	1	0
7	2023-03-1	nusantara	91	5140	0	lets gooo pulau pah	0	0	1	0
8	2023-02-2	ahmadjad	0	353	0	buru join trip pahaw	0	0	1	0
9	2022-10-2	minangru	0	608	6	terima kasih tim dok	0	0	0	0
10	2024-01-0	parosilma	7	4981	5	gowes wisata olahra	0	0	1	0
11	2022-11-1	rajomansi	0	659	1	ras trans	0	0	0	0
12	2023-09-0	untir_unti	1	822	0	capcut bismillah sen	0	0	0	0
13	2023-09-2	janwaradv	1	1475	6	capcut explorelampi	0	0	1	0
14	2023-12-0	adriparahi	1	1086	4	lokal desember rastr	0	0	0	0
15	2023-11-1	bukittaslar	0	5123	0	heiiii heiiii buddy as	0	0	1	0
16	2024-01-1	ieanonly	1	874	10	kepaksianpernong sl	0	0	0	0
17	2022-07-1	laksmi_mi	0	1820	0	informasi informasi	0	0	0	0
18	2023-04-2	iniaguy	1	1006	6	halloween arif lokal	0	0	-1	0
19	2023-06-0	ijunaidi29	0	524	5	jalan jalan tidak uan	0	0	0	0
20	2023-08-0	tirtawisat	0	586	1	tirtawisata tirtawisa	0	0	0	0
21	2023-09-1	adriparahi	4	1402	3	brilianshelani terimi	0	0	0	0
22	2022-12-1	amirnoob	0	382	0	cara hibur jadianime	0	0	0	0
23	2022-11-2	rajomansi	1	371	0	kamu	0	0	0	0

Gambar 4.56 Hasil Labelling Ahli Bahasa

Pada pelabelan menggunakan 3 ahli Bahasa ini menghasilkan pelabelan yang sama antara Labelling Ahli Bahasa 1 dan Labelling ahli Bahasa 2, dan hasilnya akan di ambil kesimpulan untuk masuk ke dalam tahap pengujian Algoritma Naïve Bayes



## ▼ KESIMPULAN AHLI BAHASA

```

import pandas as pd

# Baca kamus kata tidak baku
data = pd.read_csv("labelling ahli bahasa kesimpulan.csv", delimiter=';')
data.info()

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  -
0   createTimeISO         1000 non-null   object
1   authorMeta/name       1000 non-null   object
2   shareCount            1000 non-null   int64
3   playCount            1000 non-null   int64
4   commentCount          1000 non-null   int64
5   hasil_normalisasi     994 non-null    object
6   Labeling Bahasa 1     995 non-null    object
7   Labeling Bahasa 2     995 non-null    object
8   Labeling Bahasa 3     995 non-null    object
9   Kesimpulan            993 non-null    float64
dtypes: float64(1), int64(3), object(6)
memory usage: 78.2+ KB

```

Gambar 4.57 Membaca data set Labelling Ahli Bahasa

```

# Fungsi untuk mengklasifikasikan polaritas
def klasifikasi_polaritas(Kesimpulan):
    if Kesimpulan == 0 or Kesimpulan == 0.0:
        return 'Netral'
    elif Kesimpulan < 0:
        return 'Negatif'
    else:
        return 'Positif'

# Menerapkan fungsi klasifikasi_polaritas pada kolom polarity
df['sentiment'] = df['Kesimpulan'].apply(klasifikasi_polaritas)

df.head(500)

```

Gambar 4.58 Klasifikasi Polarity

Pada tahapan Klasifikasi Polarity ini berfungsi untuk membagi kategori negative, positif dan netral berdasarkan nilai polarity yang telah ditentukan.

```

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

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

sns.set_style('whitegrid')

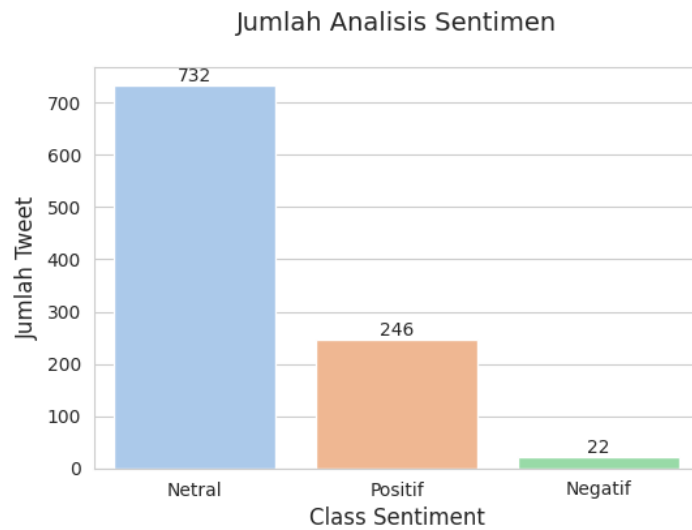
fig, ax = plt.subplots(figsize=(6, 4))
ax = sns.barplot(x=sentiment_count.index, y=sentiment_count.values, palette='pastel')
plt.title('Jumlah Analisis Sentimen', fontsize=14, pad=20)
plt.xlabel('Class Sentiment', fontsize=12)
plt.ylabel('Jumlah Tweet', fontsize=12)

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

plt.show()

```

Gambar 4.59 Perintah menampilkan hasil pelabelan



Gambar 4.60 Hasil Analisis Sentiment berdasarkan Pelabelan

Hasil Analisis Sentiment berdasarkan pelabelan ahli Bahasa terdapat pada gambar 4.60 dengan hasil sentiment kategori positif sebanyak 246 komentar, 22 komentar berkategori negative dan kategori netral sebanyak 732 komentar.

#### SPLIT DATASET (80:20)

```
import pandas as pd

data = pd.read_csv("Hasil_Labelling_Data_Kesimpulan_Ahli_Bahasa.csv")
data.head(10)
```

Gambar 4.61 Split Data set

Tahapan pada gambar 4.61 yaitu split data set berfungsi untuk membagi data training dan data testing menjadi 80:20 yaitu dengan 800 data training dan 200 menjadi data testing

```
import matplotlib.pyplot as plt

# Jumlah data latih dan data uji
train_size = len(X_train)
test_size = len(X_test)

# Membuat plot
plt.figure(figsize=(8, 6))
bars = plt.bar(['Data Latih', 'Data Uji'], [train_size, test_size], color=['blue', 'orange'])

# Menambahkan label untuk setiap bar (dalam kurung persentase)
for bar in bars:
    height = bar.get_height()
    plt.text(bar.get_x() + bar.get_width()/2, height + 20, f'{height} ({height / (train_size + test_size) * 100:.2f}%)',
             ha='center', va='bottom')

plt.title("Jumlah Data Latih dan Data Uji")
plt.xlabel("Jenis Data")
plt.ylabel("Jumlah Data")
plt.show()
```

Gambar 4.62 Proses pembagian data Training dan Data Testing

## 4.7 Pengujian Algoritma Naïve Bayes terhadap Hasil

Tahapan yang selanjutnya adalah penerapan Algoritma Naïve Bayes yang berfungsi sebagai uji validitas dari hasil labelling yang telah di labelkan oleh ketiga Ahli Bahasa

### MODEL NAIVE BAYES CLASDIFIKASI (NBC)

```
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
```

Gambar 4.63 Penerapan Naïve Bayes

```
import matplotlib.pyplot as plt

# Membuat list untuk menyimpan akurasi
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)]

# Membuat label untuk setiap model
labels = ['Gaussian NB', 'Multinomial NB', 'Bernoulli NB']

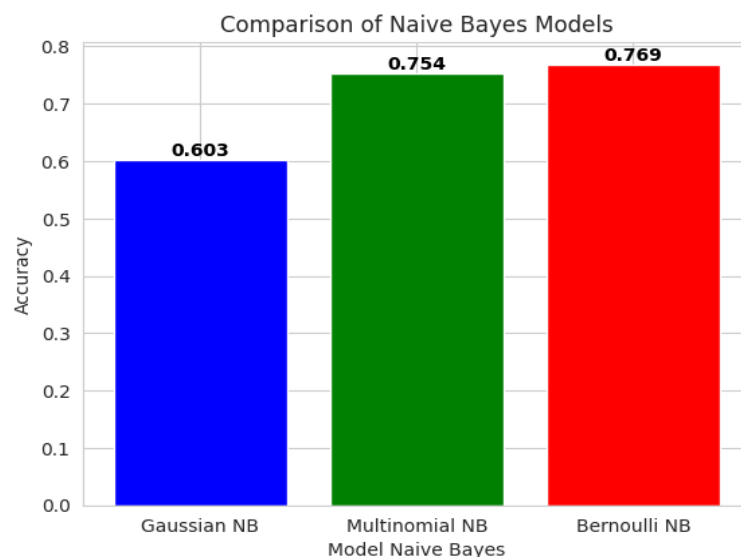
# Membuat bar plot
bars = plt.bar(labels, accuracies, color=['blue', 'green', 'red'])

# Menambahkan label dan judul
plt.xlabel('Model Naive Bayes')
plt.ylabel('Accuracy')
plt.title('Comparison of Naive Bayes Models')

# Menambahkan nilai akurasi di atas setiap bar
for bar, accuracy in zip(bars, accuracies):
    yval = bar.get_height()
    plt.text(bar.get_x() + bar.get_width()/2, yval, round(accuracy, 3), ha='center', va='bottom', color='black', weight='bold')

# Menampilkan plot
plt.show()
```

Gambar 4.64 Perintah untuk menampilkan hasil



Gambar 4.65 Hasil Algoritma Naïve Bayes

```

# Mengganti NaN dengan string kosong pada data latih dan uji
X_train = X_train.fillna('')
X_test = X_test.fillna('')

# Membuat CountVectorizer
vectorizer = CountVectorizer()

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

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

# Membuat dan melatih model Multinomial Naive Bayes
mnb = MultinomialNB()
mnb.fit(X_train_vectorized, y_train)

# Membuat dan melatih model Bernoulli Naive Bayes
bnb = BernoulliNB()
bnb.fit(X_train_vectorized, y_train)

# 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))

```

Gambar 4.66 Perintah untuk menampilkan klasifikasi report

```

Gaussian Naive Bayes Classification Report:
precision    recall  f1-score   support

Negatif      0.00    0.00    0.00         2
Netral       0.84    0.63    0.72        154
Positif      0.30    0.53    0.38         43

accuracy          0.60        199
macro avg         0.38    0.39    0.37        199
weighted avg      0.71    0.60    0.64        199

Multinomial Naive Bayes Classification Report:
precision    recall  f1-score   support

Negatif      0.00    0.00    0.00         2
Netral       0.84    0.86    0.85        154
Positif      0.45    0.42    0.43         43

accuracy          0.75        199
macro avg         0.43    0.43    0.43        199
weighted avg      0.74    0.75    0.75        199

Bernoulli Naive Bayes Classification Report:
precision    recall  f1-score   support

Negatif      0.00    0.00    0.00         2
Netral       0.81    0.92    0.86        154
Positif      0.48    0.28    0.35         43

accuracy          0.77        199
macro avg         0.43    0.40    0.40        199
weighted avg      0.73    0.77    0.74        199

```

Gambar 4.67 Hasil Accuracy

### 4.7.1 Confusion Matrix

Confusion Matrix berfungsi sebagai alat evaluasi kinerja dari sebuah model

```

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))

```

Gambar 4.68 Perintah Confusion Matrix

```

Gaussian Naive Bayes Confusion Matrix:
[[ 0  0  2]
 [ 5 97 52]
 [ 1 19 23]]

Multinomial Naive Bayes Confusion Matrix:
[[ 0  1  1]
 [ 1 132 21]
 [ 0  25 18]]

Bernoulli Naive Bayes Confusion Matrix:
[[ 0  2  0]
 [ 0 141 13]
 [ 0  31 12]]

```

Gambar 4.69 Hasil Confusion Matrix

```

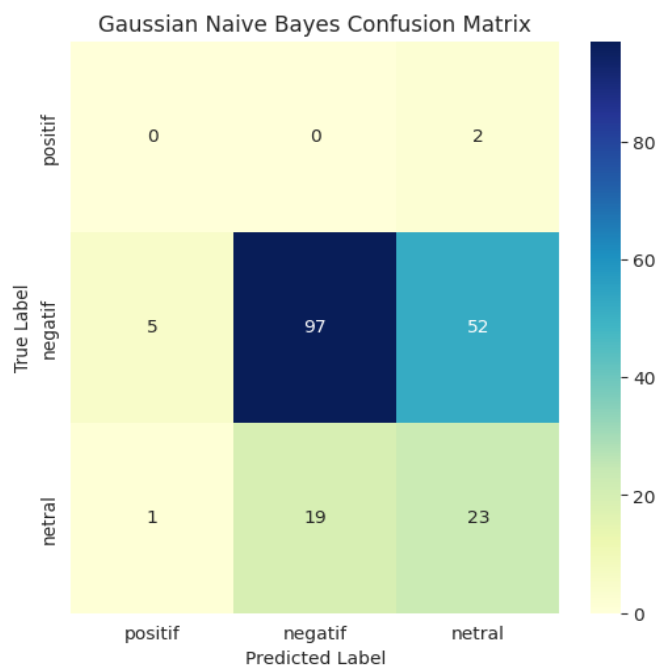
# Membuat dan melatih model Gaussian Naive Bayes
gnb = GaussianNB()
gnb.fit(X_train_vectorized.toarray(), y_train)

# 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 4.70 Melatih Model Gaussian NB CM



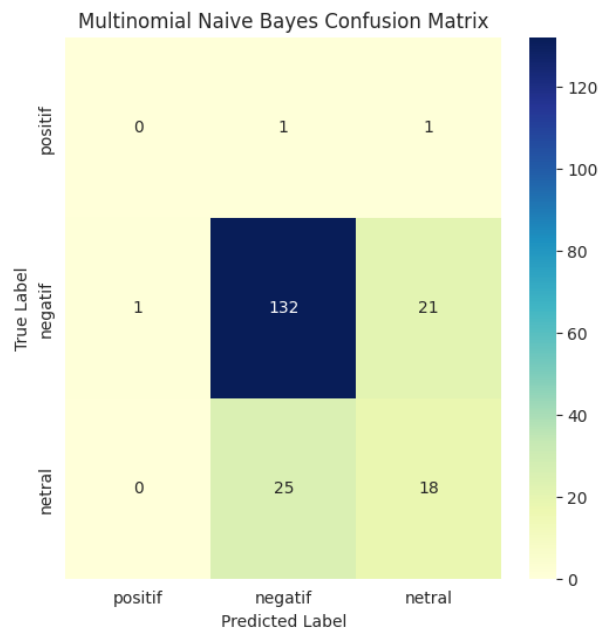
Gambar 4.71 Hasil Model Gaussian NB CM

```

# 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 4.72 Melatih model Multinomial NB CM



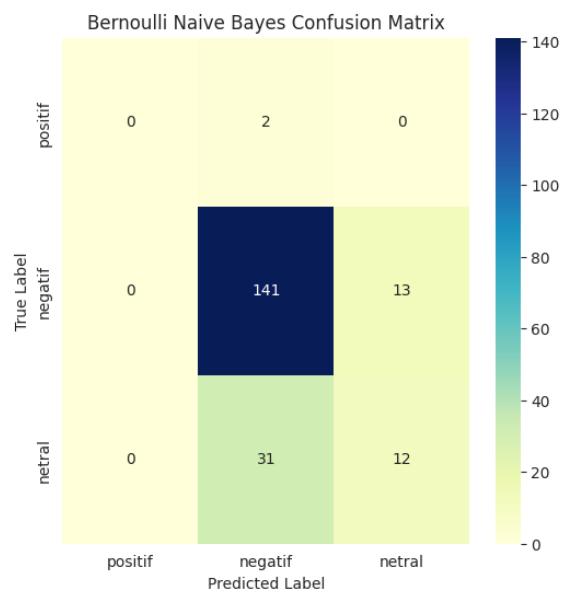
Gambar 4.73 Hasil Model Multinomial NB CM

```

▶ # 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 4.74 Melatih Model Bernoulli NB CM



Gambar 4.75 Hasil Model Multinomial NB CM

## 4.8 Word Cloud

Word Cloud berfungsi untuk memvisualkan data teks

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

# Pisahkan dataset berdasarkan sentimen
sentimen_negatif = data[data['sentiment'] == 'Negatif']['hasil_normalisasi'].str.cat(sep=' ')
sentimen_positif = data[data['sentiment'] == 'Positif']['hasil_normalisasi'].str.cat(sep=' ')
sentimen_netral = data[data['sentiment'] == 'Netral']['hasil_normalisasi'].str.cat(sep=' ')

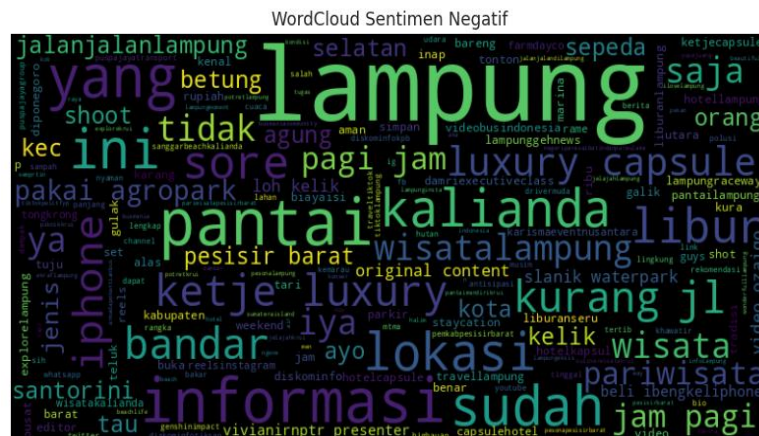
# Fungsi untuk membuat dan menampilkan WordCloud
def create_wordcloud(text, title):
    wordcloud = WordCloud(width=800, height=400, random_state=42, max_font_size=100, background_color='black').generate(text)

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

Gambar 4.76 Fungsi untuk menampilkan Wordcloud

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

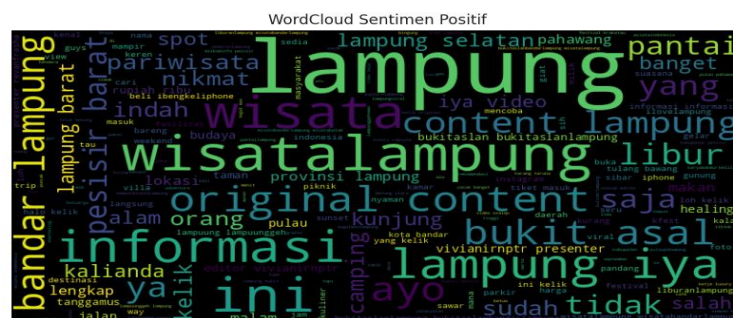
Gambar 4.77 Perintah Wordcloud Sentiment Negatif



Gambar 4.78 Wordcloud Sentiment Negative

```
# Membuat WordCloud untuk sentimen positif
create_wordcloud(sentimen_positif, 'WordCloud Sentimen Positif')
```

Gambar 4.79 Perintah Wordcloud Positif



Gambar 4.80 Wordcloud Sentiment Positif

