

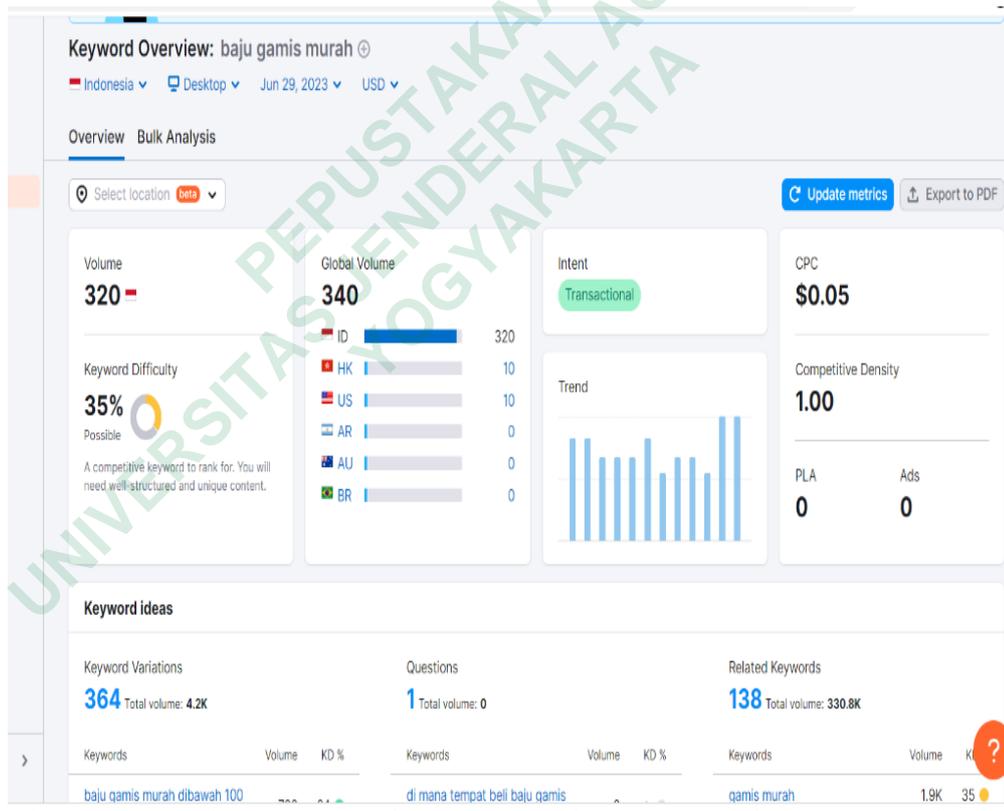
BAB 4

HASIL PENELITIAN

4.1 RINGKASAN HASIL PENELITIAN

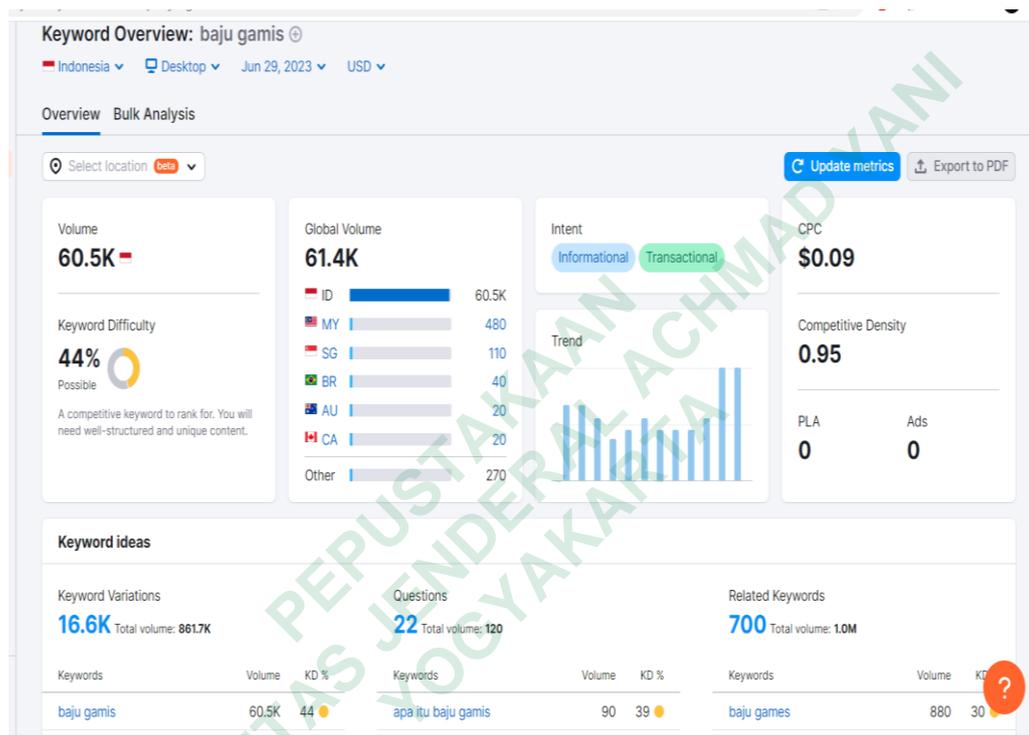
Penulis melakukan beberapa pengujian pada beberapa kata kunci sebagai berikut:

kata kunci yang pertama penulis coba yaitu “**baju gamis murah**” pada percobaan ini penulis mendapat *volume* pencarian tentang baju gamis murah sebanyak 320 dan *keyword ideas* sebanyak 364, berikutnya *related keywords* sebanyak 138 seperti gambar di bawah ini:



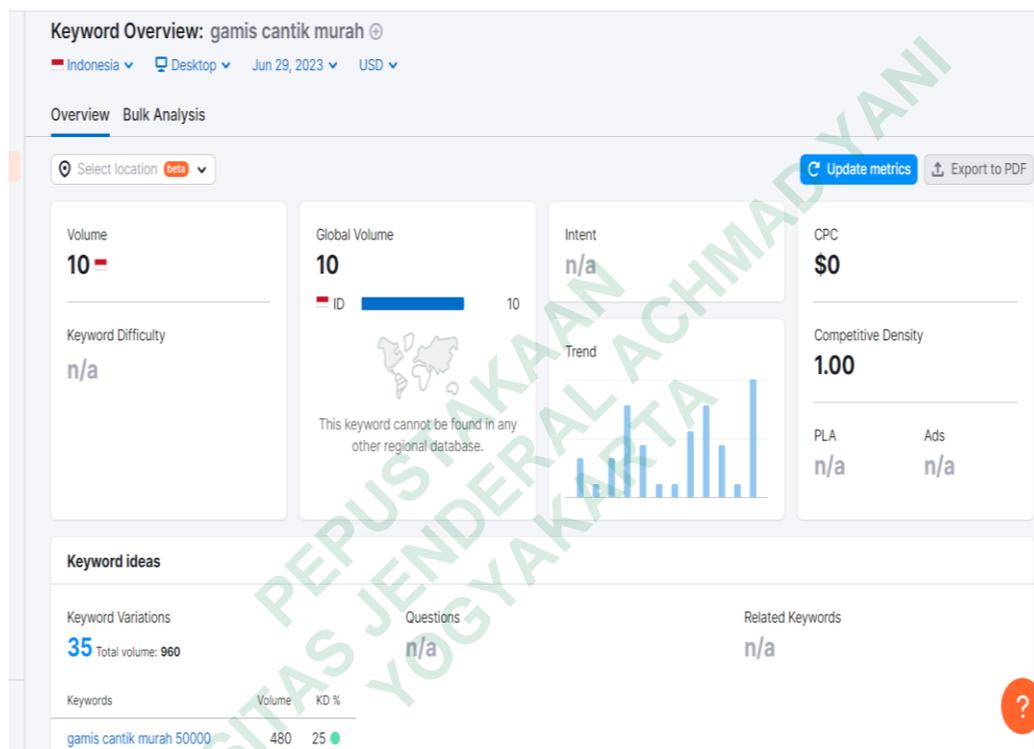
Gambar 4. 1 hasil pencarian kata kunci baju gamis murah

kata kunci kedua yang digunakan yaitu “**baju gamis**” pada percobaan ini mendapat *volume* pencarian tentang baju gamis sebanyak 60.5k dan *keywords ideas* 16.6k, berikutnya *related keywords* sebanyak 700 seperti gambar dibawah:



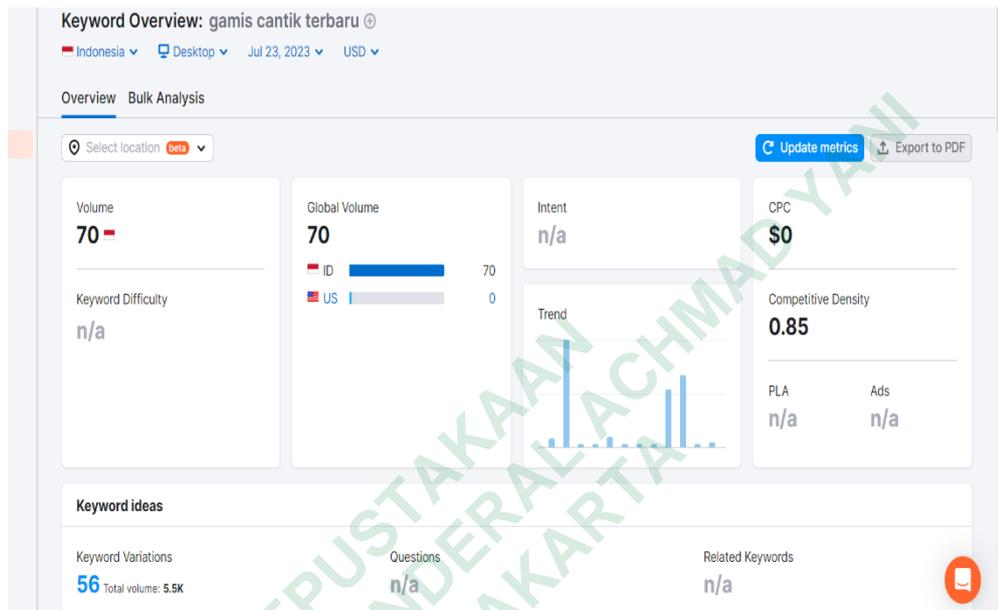
Gambar 4. 2 Hasil Pencarian Kata Kunci Baju Gamis

kata kunci ketiga yaitu “**gamis cantik murah**” pada percobaan ini penulis mendapat *volume* pencarian tentang gamis cantik murah sebanyak 10, dan *keywords ideas* 35, berikutnya *related keywords* 0. Seperti gambar dibawah:



Gambar 4. 3 Hasil Pencarian Kata Kunci Gamis Cantik Murah

Kata kunci terakhir yaitu “**gamis cantik terbaru**” pada percobaan ini penulis mendapatkan *volume* pencarian tentang gamis murah cantik sebanyak 70, dan *keywords ideas* 56, berikutnya *related keywords* 0. Seperti gambar dibawah:



Gambar 4. 4 Hasil Pencarian Kata Kunci Gamis Murah Cantik

4.1.1.1 Twitter

Twitter adalah sebuah *platform* media sosial yang penggunaanya bisa melakukan pengirim pesan singkat atau membaca pesan yang dinamakan dengan “*tweet*”, *twitter* sendiri pada saat sekarang ini sangat populer di Indonesia atau pun dunia yang penggunaanya yang sangat beragam, *twitter* juga sebagai penyalur saluran dalam dunia jurnalistik, komunikasi dan pemasaran untuk berkomunikasi dengan *audiens*, membagikan informasi dan membangun merek.

Hashtag adalah sebuah tanda (#) yang diikuti oleh kata dan frasa yang bertujuan mempermudah pencarian konten yang berkaitan dengan kata atau topik yang sedang dicari lalu *hashtag* dengan mudah menemukan dan melihat konten yang berkaitan.

Berikut beberapa percobaan di *twitter*:

Pertama penulis mencoba dengan kata **“Baju gamis murah”** dalam pencarian data dengan *scraping* langsung, penulis mendapatkan *tweet* di *twitter* sebanyak 399. Seperti gambar dibawah:

The screenshot shows a Jupyter Notebook interface with the following content:

- File Explorer:** Lists files such as `sample_data`, `baju_gamis_lang-id-since-2023-01-...`, `baju_gamis_murah_lang-id-since-20...`, `baju_gamis_murahlang-id-since-20...`, `data_baju_gamis.csv`, `data_baju_gamis_murah.csv`, `data_gamis_cantik_murah.csv`, `data_gamis_murah_terbaru.csv`, `data_gamis_terbaru.csv`, `gamis_cantik_murah_lang-id-since...`, `gamis_murah_terbaru_lang-id-since...`, and `gamis_terbaru_lang-id-since-2023-0...`.
- Data Table:** Displays a DataFrame with 399 rows and 15 columns. The first three rows are:

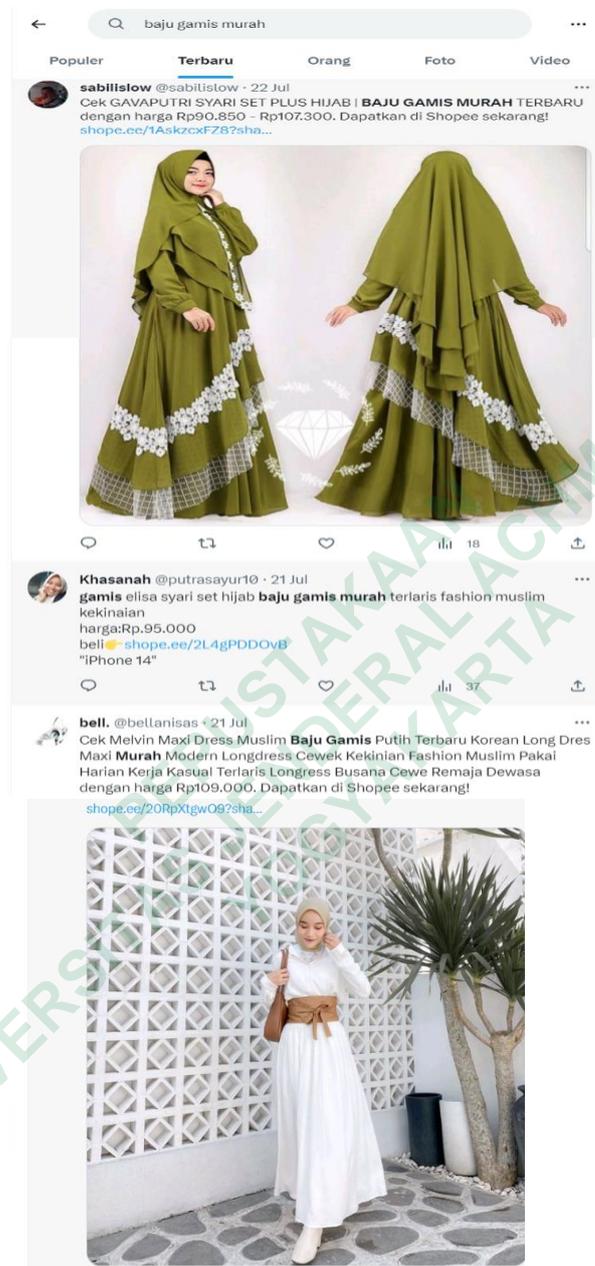
Index	URL	Date	Text	Retweets	Replies	Retweeted By	Retweeted To	Retweeted From	Retweeted To
396	https://twitter.com/HeruPra26690615/status/161...	2023-01-05 07:21:44+00:00	#GAMIS SYARI JILBAB 2 LAYER UKURAN S M L XL X...	0	0	0	0	0	1610899332f
397	https://twitter.com/thebatikCOD/status/160352...	2023-01-01 12:16:02+00:00	#grosirdaster #dastermurah baju gamis batik mu...	0	0	0	0	0	1609523841f
398	https://twitter.com/hanhyun_00/status/16094063...	2023-01-01 04:29:16+00:00	GAMIS SYARI PLUS JILBAB 2 LAYER UKURAN S M L X...	1	0	0	0	0	1609402320f
- Code Cell:**

```
# Cek jumlah data yang didapatkan
num_tweets = len(tweets_df)
print(f"Jumlah tweet dalam dataframe untuk kata [baju gamis murah ] adalah {num_tweets}.")
```

Jumlah tweet dalam dataframe untuk kata [baju gamis murah] adalah 399.

Gambar 4. 5 Hasil *Scraping* Dengan Kata Kunci Baju Gamis Murah

Contoh tampilan hasil pencarian kata kunci “**baju gamis murah**” di *twitter* seperti gambar dibawah:



Gambar 4. 6 Pencarian Di *Twitter* Dengan Kata Kunci Baju Gamis Murah

Kedua penulis mencoba dengan kata “**Baju gamis**” dalam pencarian data dengan *scraping* langsung, penulis mendapatkan *tweet* di *twitter* sebanyak 2500. Seperti gambar dibawah:

The screenshot shows a Jupyter Notebook interface with the following content:

File Edit Lihat Sisipkan Runtime Fitur Bantuan Semua perubahan telah disimpan

File Explorer: sample_data, baju_gamis_lang_id_since-2023-01-..., baju_gamis_murah_lang_id_since-20..., baju_gamis_murahlang_id_since-20..., data_baju_gamis.csv, data_baju_gamis_murah.csv, data_gamis_cantik_murah.csv, data_gamis_murah_terbaru.csv, data_gamis_terbaru.csv, gamis_cantik_murah_lang_id_since..., gamis_murah_terbaru_lang_id_since..., gamis_terbaru_lang_id_since-2023-0...

Code Cell [15]:

ID	URL	Date	Content	Retweets	Replies	Views	Score
2577	https://twitter.com/Hafteh26/status/1642637400...	2023-04-02 21:17:29+00:00	Temukan dan dapatkan Baju gamis wanita terbaru...	0	0	0	0 1642637400
2578	https://twitter.com/rizkikanu/status/164263455...	2023-04-02 21:06:12+00:00	Pasar Tanah Abang Mulai Dipadati Pengunjung, B...	0	0	0	0 1642634550
2579	https://twitter.com/Orchidaceae_Pe/status/1642...	2023-04-02 20:53:07+00:00	Gamis yang w mau dah habis. Yaudahlah ya. Ga s...	0	0	0	0 1642631260

2580 rows x 15 columns

Code Cell [16]:

```
# Cek jumlah data yang didapatkan
num_tweets = len(tweets_df)
print(f"Jumlah tweet dalam dataframe untuk kata [baju gamis ] adalah {num_tweets}.")
```

Output: Jumlah tweet dalam dataframe untuk kata [baju gamis] adalah 2580.

0 d selesai pada 18:55

Gambar 4. 7 Hasil *Scraping* Dengan Kata Kunci Baju Gamis

Contoh tampilan hasil pencarian kata kunci “baju gamis” di *twitter* seperti gambar dibawah:



Gambar 4. 8 Pencarian Di *Twitter* Dengan Kata Kunci Baju Gamis

Ketiga penulis mencoba dengan kata “**Gamis cantik murah**” dalam pencarian data dengan *scraping* langsung, penulis mendapatkan *tweet* di *twitter* sebanyak 78. Seperti gambar dibawah:

The screenshot shows a Jupyter Notebook interface with a file explorer on the left and a code cell with its output in the center. The file explorer shows a directory structure with files like 'sample_data', 'baju_gamis_lang-id-since-2023-01-...', 'baju_gamis_murah_lang-id-since-20...', 'data_baju_gamis.csv', 'data_baju_gamis_murah.csv', 'data_gamis_cantik_murah.csv', 'data_gamis_murah_terbaru.csv', 'data_gamis_terbaru.csv', 'gamis_cantik_murah_lang-id-since-...', 'gamis_murah_terbaru_lang-id-since-...', and 'gamis_terbaru_lang-id-since-2023-0-...'. The code cell contains the following Python code:

```
# cek jumlah data yang didapatkan
num_tweets = len(tweets_df)
print(f"Jumlah tweet dalam dataframe untuk kata [gamis cantik murah ] adalah {num_tweets}.")
```

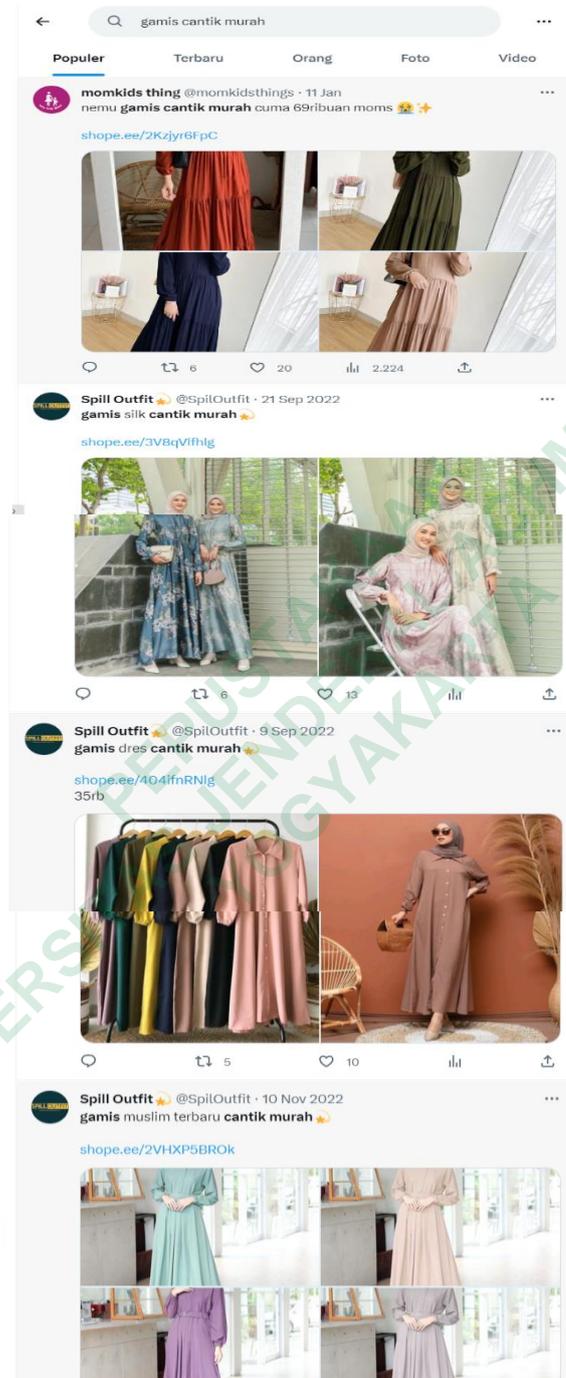
The output of the code cell shows the result of the print statement:

```
Jumlah tweet dalam dataframe untuk kata [gamis cantik murah ] adalah 76.
```

The background of the image contains a large, diagonal watermark: "UNIVERSITAS JENDERAL ACHMAD YANUAR REPUBLIK INDONESIA YOGYAKARTA".

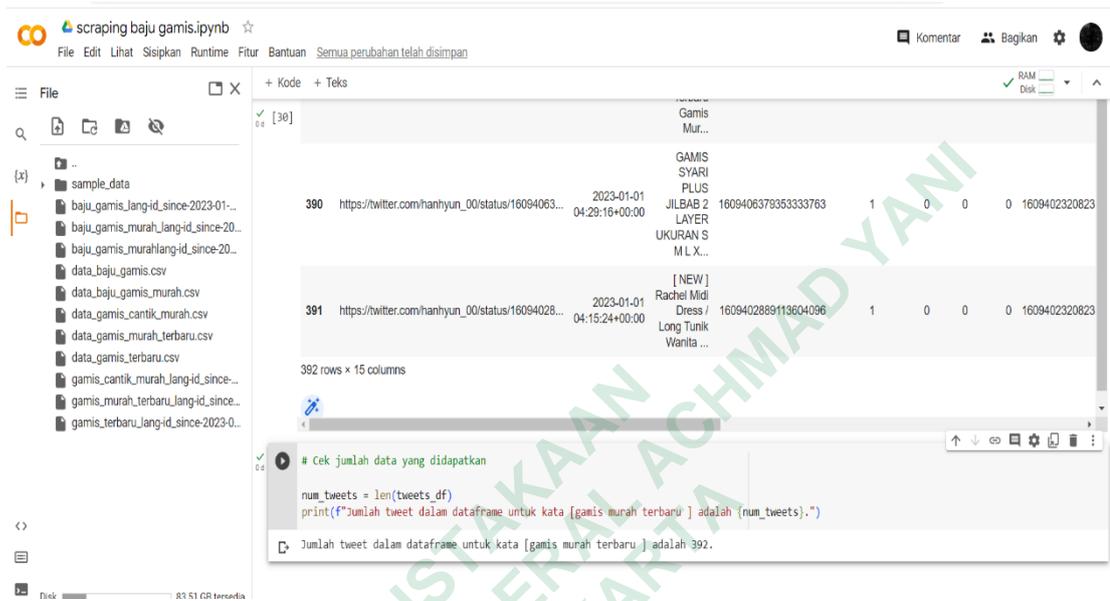
Gambar 4. 9 Hasil *Scraping* Dengan Kata Kunci Gamis Cantik Murah

Contoh tampilan hasil pencarian kata kunci “**gamis cantik murah**” di *twitter* seperti gambar dibawah:



Gambar 4. 10 Pencarian Di *Twitter* Dengan Kata Kunci Gamis Cantik Murah

Terakhir penulis mencoba dengan kata “**Gamis murah terbaru**” dalam pencarian data dengan *scraping* langsung, kata ini penulis mendapatkan *tweet* di *twitter* sebanyak 392. Seperti gambar dibawah:



The screenshot shows a Jupyter Notebook interface with the following content:

File Explorer: Lists various files and folders related to the scraping process, including CSV files and directories.

DataFrame View: Displays a table with 392 rows and 15 columns. Two rows are visible:

Index	URL	Date	Time	Content	Retweets	Replies	Retweeted	Retweeted	Retweeted	Retweeted
390	https://twitter.com/hanhyun_00/status/16094063...	2023-01-01	04:29:16+00:00	JILBAB 2 LAYER UKURAN S M L X...	1	0	0	0	0	1609402320823
391	https://twitter.com/hanhyun_00/status/16094028...	2023-01-01	04:15:24+00:00	[NEW] Rachel Midi Dress / Long Tunik Wanita ...	1	0	0	0	0	1609402320823

Code Cell:

```
# Cek jumlah data yang didapatkan
num_tweets = len(tweets_df)
print(f"Jumlah tweet dalam dataframe untuk kata [gamis murah terbaru ] adalah {num_tweets}.")
```

Output: Jumlah tweet dalam dataframe untuk kata [gamis murah terbaru] adalah 392.

Gambar 4. 11 Hasil *Scraping* Dengan Kata Kunci Gamis Murah Terbaru

Contoh tampilan hasil pencarian kata kunci “gamis murah terbaru” di *twitter* seperti gambar dibawah:



Gambar 4. 12 Pencarian Di *Twitter* Dengan Kata Kunci Gamis Murah Terbaru

4.1.2 Proses Scraping

- 1) Input kata kunci yang akan di *scraping*
- 2) Menggunakan *library snsrape*
- 3) Autentikasi Dalam kode *Python*
- 4) Impor pustaka:

Import `pandas` as `pd` digunakan untuk mengimpor perpustakaan `pandas` dan menentukan alias `pd` untuk kenyamanan.

`import ast` digunakan untuk mengimpor pustaka `ast` yang digunakan dalam operasi pemformatan string ke dalam kamus.

`import json` digunakan untuk mengimpor library `json` untuk memproses data JSON.

`fileJSON: tweets_df=pd.read_json(filename, lines=True)`

digunakan untuk membaca file JSON dengan nama file dan menghasilkan bingkai data `tweets_df` `panda`. `lines=True` berarti setiap baris dalam file JSON adalah objek yang terpisah.

Ganti nama kolom:

`kolom_baru` adalah kamus yang berisi pasangan nilai baru dan nama kolom untuk digunakan dalam bingkai data.

mengganti nama kolom yang dimaksud. Jika Anda memerlukan bantuan untuk mengganti nama kolom di `pandas DataFrame`. Seperti gambar dibawah:

```
[ ] import os
import datetime

# Batasi jumlah hasil yang diambil
max_results =3000

# Gunakan Twitter search untuk mencari tweet yang di-favoritkan minimal 10000 kali dan berbahasa Indonesia
twitter_search = "baju gamis lang:id since:2023-01-01 until:2023-05-1"

# Tentukan nama file dengan format "<kueri pencarian> <tanggal saat ini>.json"
filename = f"{twitter_search.replace(' ', '_').replace(':', '-')}_{datetime.date.today().strftime('%Y-%m-%d')}.json"

# Ambil tweet yang cocok dengan kueri pencarian dan simpan dalam file JSON dengan nama file yang telah ditentukan
os.system(f"snsrape --jsonl --max-results {max_results} twitter-search '{twitter_search}'> {filename}")
```

256

Gambar 4. 13 Kodingan *Google colab / Scraping*

Manipulasi data:

Di bagian ini, pilih kolom tertentu untuk digunakan dalam bingkai data dengan menggunakan `tweets_df.loc[:, [...]]`. Kolom dipilih berdasarkan nama kolom asli. Selanjutnya, lakukan beberapa manipulasi data. Misalnya, gunakan fungsi `terapkan` untuk mengekstrak kolom "Media" dan "Tautan" dari kolom "media" dan "tautan", dan gunakan fungsi `lamda` untuk mengekstrak.

Menampilkan bingkai data:

`display(tweets_df)` digunakan untuk menampilkan kerangka data `tweets_df` dalam output.

Simpan ke file CSV:

`tweets_df.to_csv(NAME_FILE_CSV, index=False)` digunakan untuk menyimpan bingkai data `tweets_df` ke file CSV bernama `NAME_FILE_CSV`. `index=False` berarti indeks tidak akan disimpan dalam file CSV. Seperti gambar dibawah:

```
import pandas as pd
import ast
import json

# Membaca file JSON hasil dari perintah CLI sebelumnya dan membuat dataframe pandas
tweets_df = pd.read_json(filename, lines=True)

NAMA_FILE_CSV = 'data_baju_gamis.csv'

# Membuat kamus untuk mengganti nama kolom
new_columns = {
    'conversationId': 'Conv. ID',
    'url': 'URL',
    'date': 'Date',
    'rawContent': 'Tweet',
    'id': 'ID',
    'replyCount': 'Replies',
    'retweetCount': 'Retweets',
    'likeCount': 'Likes',
    'quoteCount': 'Quotes',
    'bookmarkCount': 'Bookmarks',
    'lang': 'Language',
    'links': 'Links',
    'media': 'Media',
    'retweetedTweet': 'Retweeted Tweet',
    'username': 'Username'
}

if len(tweets_df) == 0:
    print('Pencarian tidak ditemukan coba ganti keyword lain, keywordmu: ', twitter_search)
    exit()
else:
    # Memilih kolom yang akan digunakan dan mengganti nama kolom menggunakan kamus yang telah dibuat
    tweets_df = tweets_df.loc[:, ['url', 'date', 'rawContent', 'id',
                                  'replyCount', 'retweetCount', 'likeCount', 'quoteCount',
                                  'conversationId', 'lang', 'links',
                                  'media', 'retweetedTweet', 'bookmarkCount', 'username']]
    tweets_df = tweets_df.rename(columns=new_columns)

    # Ekstrak fullurl dari kolom media dan url dari kolom links
    tweets_df['Media'] = tweets_df['Media'].apply(lambda x: x[0]['fullurl'] if isinstance(x, list) and x and isinstance(x[0], dict) and 'fullurl' in x[0] else None)
    tweets_df['Links'] = tweets_df['Links'].apply(lambda x: x[0]['url'] if isinstance(x, list) and x and isinstance(x[0], dict) and 'url' in x[0] else None)

    # Menampilkan dataframe tweets_df
    display(tweets_df)

# Simpan ke csv
tweets_df.to_csv(NAMA_FILE_CSV, index=False)
```

Gambar 4. 14 Kodingan *Google colap / Scraping*

4.1.2.1 Scraping Data

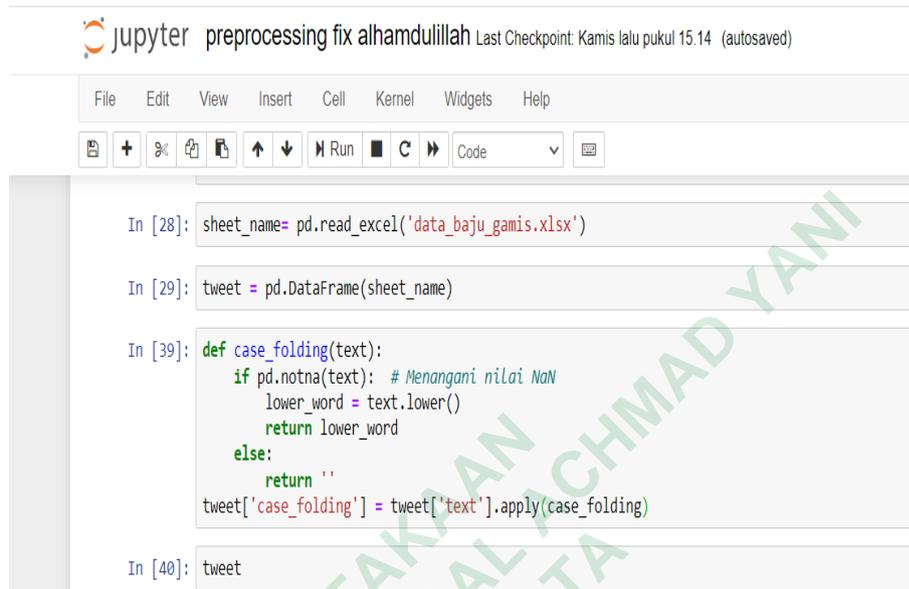
Data dibawah adalah data yang telah di pindahkan ke *microsoft excel* dari hasil *scraping* dan dikelompokkan sesuai keterangannya masing - masing terdiri dari : *url, date* atau tanggal, *tweet, id, reples, retweets, likes, quotes, negara, media, dan userneme.*

A1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	URL	Date	Tweet	ID	Replies	Retweets	Likes	Quotes	Conv. ID	Language	Links	Media	Retweeted Tweet	Bookmarks	Username
2	https://twitter.com	2023-04-30 15:3	sebenarnya eng	1652696920748	0	0	0	0	1652696920748	in					0 kokocrushj
3			👉 Selma Maxi												
3	https://twitter.com	2023-04-30 15:1	Cek Harga : http	1652694023914	1	0	0	0	1652693978704	in	https://shp.ee/sy	https://pbs.twimg.com/media/Fu-l			0 BestRomi
4	https://twitter.com	2023-04-30 15:0	@mingkiw nex	1652689346917	0	0	0	0	1652680536396	in					0 JonghoJoe
5	https://twitter.com	2023-04-30 14:4	@sanniae wkwk	1652684997319	0	0	0	0	1652680536396	in					0 JonghoJoe
6	https://twitter.com	2023-04-30 14:3	@byeolbox rada	1652684105719	0	0	0	0	1652680536396	in					0 JonghoJoe
7	https://twitter.com	2023-04-30 14:3	Aera Baju Rajut	1652683849137	0	0	0	0	1652683849137	in	https://shope.ee	https://pbs.twimg.com/media/Fu-l			0 FitriYa070f
8	https://twitter.com	2023-04-30 14:3	@sparklingnumn	1652683792363	1	0	0	0	1652680536396	in					0 JonghoJoe
9	https://twitter.com	2023-04-30 14:3	sumpah ya adsk	1652682531739	0	0	0	0	1652682531739	in					0 cho_toma
10	https://twitter.com	2023-04-30 14:3	Temukan KAFTA	1652682145750	0	0	1	0	1652682145750	in	https://shope.ee	https://pbs.twimg.com/media/Fu-l			0 Akhysmail
11	https://twitter.com	2023-04-30 13:0	kemaren ngeliat	1652661527185	1	0	0	0	1652661527185	in					0 lknowscut
12			Gamis wanita se Harga Produk: F Harga Diskon: R Belinya disini 👉												
12	https://twitter.com	2023-04-30 12:4	https://t.co/LiW	1652655259402	0	0	0	0	1652655259402	in	https://c.lazada.c	https://pbs.twimg.com/media/Fu9			0 UmmuNau
13	https://twitter.com	2023-04-30 12:3	@tanyakanri Bai	1652653817400	2	0	0	0	1652544572898	in					0 unispacex
14	https://twitter.com	2023-04-30 12:2	Sangat Mantap.	1652651028196	0	0	0	0	1652651028196	in					0 UkhtyZuryi
15	https://twitter.com	2023-04-30 11:5	@kenjagung jua	1652643295947	1	0	0	0	1652641257599	in					0 Dulceternit
16			Mau Gamis Terb JANGAN SAMP Baju edisi terb:												
2582			Hello want 🍀 Outer Gamis Pa 👉 4.7 👉 72.100 kancing depan p												
2582	https://twitter.com	2023-04-02 23:1	https://t.co/AFFH	1642666691594	0	0	0	0	1642666691594	in	https://shope.ee	https://pbs.twimg.com/media/Fsvs			0 yangkuma
2583	https://twitter.com	2023-04-02 22:4	Temukan Damai	1642660372204	0	0	0	0	1642660372204	in	https://shope.ee	https://pbs.twimg.com/media/Fsvr			0 hasanwjy6
2584			Baju lebaran, ga												
2584	https://twitter.com	2023-04-02 22:3	A thread	1642656257646	10	0	4	0	1642656257646	in					1 blurbkchng
2585	https://twitter.com	2023-04-02 22:2	hipergamis tuh k	1642653289320	0	0	0	0	1642653289320	in					0 regytok
2586	https://twitter.com	2023-04-02 22:3	Cek (Dijamin Re	1642658160304	0	0	0	0	1642658160304	in	https://shope.ee	https://pbs.twimg.com/media/Fsvk			0 HosOfStar
2587	https://twitter.com	2023-04-02 22:3	Cek RB SARIME	1642656806978	0	0	0	0	1642656806978	in	https://shope.ee	https://pbs.twimg.com/media/Fsvj			0 syahid_zik
2588	https://twitter.com	2023-04-02 21:5	@never_alonely	1642646488454	0	0	0	0	1642556000514	in					0 Safa_Andr
2589	https://twitter.com	2023-04-02 22:1	/ootd/ saranin be	1642650615171	53	1	15	0	1642650615171	in					16 OOTDFES
2590	https://twitter.com	2023-04-02 22:1	Cek Kids Mayra	1642651473321	0	1	1	1	1642651473321	in	https://shope.ee	https://pbs.twimg.com/media/Fsvy			0 1002indah
2591	https://twitter.com	2023-04-02 21:2	Cek Promo setel	1642640237808	0	0	1	0	1642640237808	in	https://shope.ee	https://pbs.twimg.com/media/Fsvl			0 mustsus
2592	https://twitter.com	2023-04-02 21:1	Temukan dan de	1642637400752	0	0	0	0	1642637400752	in	https://shope.ee	https://pbs.twimg.com/media/Fsvz			0 Hafish26
2593	https://twitter.com	2023-04-02 21:0	Pasar Tanah Abi	1642634558804	0	0	0	0	1642634558804	in	https://banten24.com/pasar-tanah-abang-mulai-dip				0 rizkikaru
2594	https://twitter.com	2023-04-02 20:5	Gamis yang w rr	1642631269250	0	0	0	0	1642631269250	in					0 Orchidacee
2595															
2596															

Gambar 4. 15 Ekspor Data ke Microsoft Excel

4.1.2.2 Preprocessing Data

- 1) *Casefolding* adalah proses mengubah semua huruf dalam teks menjadi huruf kecil atau huruf besar. Seperti gambar dibawah:



```

jupyter preprocessing fix alhamdulillah Last Checkpoint: Kamis lalu pukul 15.14 (autosaved)
File Edit View Insert Cell Kernel Widgets Help
[+] [x] [copy] [paste] [undo] [redo] [run] [stop] [refresh] [code] [help]

In [28]: sheet_name= pd.read_excel('data_baju_gamis.xlsx')

In [29]: tweet = pd.DataFrame(sheet_name)

In [39]: def case_folding(text):
          if pd.notna(text): # Menangani nilai NaN
              lower_word = text.lower()
              return lower_word
          else:
              return ''
          tweet['case_folding'] = tweet['text'].apply(case_folding)

In [40]: tweet
  
```

Gambar 4. 16 Kodingan Dari *Case Folding*

sebelum di <i>preprocessing</i>			setelah di <i>proprocessing</i>		
no		text			case_folding
0	1	sebenarnya enggak, cuma baju ku kebanyakan kao...			sebenarnya enggak, cuma baju ku kebanyakan kao...
1	2	ðYCEY Selma Maxi Dress kondangan-Maxi Dress Bus...			ðyœy selma maxi dress kondangan-maxi dress bus...
2	3	@mingkiww next sg24 baju gamis/Koko ðY□f			@mingkiww next sg24 baju gamis/koko ðy□f
3	4	@sanniue wkwkwk kalo ada dua outfit maka satu...			@sanniue wkwkwk kalo ada dua outfit maka satu...
4	5	@byeolbux rada terlihat vibes lebaran dan Nata...			@byeolbux rada terlihat vibes lebaran dan nata...
5	6	Aera Baju Rajut Wanita Korean Style Blouse Tur...			aera baju rajut wanita korean style blouse tur...
6	7	@sparklingnunna Next bisa si season greeting p...			@sparklingnunna next bisa si season greeting p...
7	8	sumpah ya adek gw ngadi2 nyuruh pake baju gami...			sumpah ya adek gw ngadi2 nyuruh pake baju gami...
8	9	Temukan KAFTAN MADINA JUMBO GAMIS RAYON PREMIU...			temukan kaftan madina jumbo gamis rayon premiu...
9	10	kemaren ngeliat outer brukat cakep banget. bua...			kemaren ngeliat outer brukat cakep banget. bua...
10	11	Gamis wanita set hijab cadar FRENCH REMPEL baj...			gamis wanita set hijab cadar french rempel baj...
11	12	@tanyakanrl Barbie lengkap dengan baju, sepatu...			@tanyakanrl barbie lengkap dengan baju, sepatu...

Gambar 4. 17 Tampilan Hasil *Case Folding* Sebelum Dan Setelah di *preprocessing*

2) *Filtering / cleaning* adalah proses menghapus karakter-karakter yang tidak diperlukan seperti tanda baca dan karakter khusus lainnya, tahap ini dapat dijabarkan sebagai berikut *Cleaning* "Pembersihan" dalam pemrosesan data atau pemrosesan teks mengacu pada serangkaian langkah atau proses yang dilakukan untuk menghilangkan *elemen* yang tidak diinginkan, seperti karakter khusus, pemformatan yang tidak konsisten, atau data yang tidak *relevan*, dari data atau teks. Tujuan dari proses pembersihan adalah menyiapkan data atau teks yang bersih dan siap untuk dianalisis atau diproses lebih lanjut, seperti gambar dibawah:

```
In [52]: def remove_tweet_special(case_folding):
text = case_folding.replace('\t', " ").replace('\n', " ").replace('\', "'")
text = text.encode('ascii', 'replace').decode('ascii')
text = ' '.join(re.sub("([@#][A-Za-z0-9]+)(\w+:\w+\/\w+)", " ", text).split())
return text.replace("http://", " ").replace("https://", " ")
tweet['clean'] = tweet['case_folding'].apply(remove_tweet_special)

def remove_number(clean):
text = re.sub(r"\d+", "", clean)
return text
tweet['clean'] = tweet['clean'].apply(remove_number)

def remove_punctuation(clean):
clean_spcl = re.compile('[/(){}[\]|\@,;]')
clean_symbol = re.compile('[^0-9a-z]')
text = clean_spcl.sub('', clean)
text = clean_symbol.sub(' ', clean)
return text
tweet['clean'] = tweet['clean'].apply(remove_punctuation)

def remove_whitespace(clean):
corrected = str(clean)
corrected = re.sub(r"//t", r"\t", corrected)
corrected = re.sub(r"( )1+", r"\1", corrected)
corrected = re.sub(r"(\n)1+", r"\1", corrected)
corrected = re.sub(r"(\r)1+", r"\1", corrected)
corrected = re.sub(r"(\t)1+", r"\1", corrected)
return corrected.strip(" ")
tweet['clean'] = tweet['clean'].apply(remove_whitespace)
print(tweet['clean'])
```

Gambar 4. 18 Kodongan Dari *Filtering / Cleaning*

Sebelum di *preprocessing*

no	text
0	1 sebenarnya enggak, cuma baju ku kebanyakan kao...
1	2 ðŸŒŸ Selma Maxi Dress kondangan-Maxi Dress Bus...
2	3 @mingkiww next sg24 baju gamis/Koko ðŸŒŸf
3	4 @sanniue wkwkwk kalo ada dua outfit maka satun...
4	5 @byeolbux rada terlihat vibes lebaran dan Nata...
5	6 Aera Baju Rajut Wanita Korean Style Blouse Tur...
6	7 @sparklingnunna Next bisa si season greeting p...
7	8 sumpah ya adek gw ngadi2 nyuruh pake baju gami...
8	9 Temukan KAFTAN MADINA JUMBO GAMIS RAYON PREMIU...
9	10 kemaren ngeliat outer brukat cakep banget. bua...
10	11 Gamis wanita set hijab cadar FRENCH REMPEL baj...
11	12 @tanyakanri Barbie lengkap dengan baju, sepatu...
12	13 Sangat Mantap...Baju Gamis Bermotif Cuantikkk ...
13	14 @kenjagung jual baju jongli pls beli itu, mau ...
14	15 Mau Gamis Terbaru, Setelan Wanita Terbaru, Gam...
15	16 Cek Jubbah Shahama SH-019 Baju Gamis Pria Dewa...
16	17 Aku nemu promo menarik di Lazada nih. Yuk coba...
17	18 Yuk coba lihat!nNama Produk: LOOK Baju Gamis...
18	19 NEW - Putri Dress ARMANI SILK PREMIUM M-XXL / ...
19	20 Cek RB SARIMBIT KELUARGA KIA BAJU GAMIS KOKO A...
20	21 liat baju gamis mama gue lucu warna pinky tbtb...
21	22 Rekomendasi Gamis / Tunik ðŸŒŸnðŸŒŸELEA TUN...
22	23 Baik, ini saatnya kita membeli semua yang kita...
23	24 @sayuurbening Baju hari raya udah model gamis ...

Setelah di *preprocessing*

0 sebenarnya enggak cuma baju ku kebanyakan kaos...

1 selma maxi dress kondangan maxi dress busui fr...

2 next sg baju gamis koko

3 wkwkwk kalo ada dua outfit maka satunya pake b...

4 rada terlihat vibes lebaran dan natal jg next ...

5 aera baju rajut wanita korean style blouse tur...

6 next bisa si season greeting pake baju koko ga...

7 sumpah ya adek gw ngadi nyuruh pake baju gamis...

8 temukan kaftan madina jumbo gamis rayon premiu...

9 kemaren ngeliat outer brukat cakep banget buat...

10 gamis wanita set hijab cadar french rempel baj...

11 barbie lengkap dengan baju sepatu dan accessor...

12 sangat mantap baju gamis bermotif cuantikkk ma...

13 jual baju jongli pls beli itu mau liat jagung ...

14 mau gamis terbaru setelan wanita terbaru gamis...

15 cek jubbah shahama sh baju gamis pria dewasa m...

16 aku nemu promo menarik di lazada nih yuk coba ...

17 yuk coba lihat nama produk look baju gamis say...

18 new putri dress armani silk premium m xxl baju...

19 cek rb sarimbit keluarga kia baju gamis koko a...

20 liat baju gamis mama gue lucu warna pinky tbtb...

21 rekomendasi gamis tunik elea tunic baju kondan...

22 baik ini saatnya kita membeli semua yang kita ...

23 baju hari raya udah model gamis semua biar bis...

24 gamis wanita set hijab cadar clarisha darkgrey...

Gambar 4. 19 Tampilan Hasil *Filtering* Sebelum Dan Sesudah di *preprocessing*

- 3) *stopword remover* proses menghapus kata-kata yang umum dan tidak memiliki makna seperti “dan”, “atau”, dan lainnya. Seperti gambar dibawah:

```
In [51]: from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory

factory = StopWordRemoverFactory()
stopword = factory.create_stop_word_remover()

def stopword_removal(tokens):
    if isinstance(tokens, str):
        stopwords = stopword.remove(tokens)
    else:
        stopwords = [stopword.remove(token) for token in tokens]
    return stopwords
tweet['stopword'] = tweet['clean'].apply(stopword_removal)
print(tweet['stopword'])
```

Gambar 4. 20 Kodingan dari *stopword remover*

Sebelum di *preprocessing*

no	text
0	1 sebenarnya enggak, cuma baju ku kebanyakan kao...
1	2 ðŸŒŸ Selma Maxi Dress kondangan-Maxi Dress Bus...
2	3 @mingkiww next sg24 baju gamis/Koko ðŸŒŸf
3	4 @sanniue wkwkwk kalo ada dua outfit maka satun...
4	5 @byeolbux rada terlihat vibes lebaran dan Nata...
5	6 Aera Baju Rajut Wanita Korean Style Blouse Tur...
6	7 @sparklingnunna Next bisa si season greeting p...
7	8 sumpah ya adek gw ngadi2 nyuruh pake baju gami...
8	9 Temukan KAFTAN MADINA JUMBO GAMIS RAYON PREMIU...
9	10 kemaren ngeliat outer brukat cakep banget. bua...
10	11 Gamis wanita set hijab cadar FRENCH REMPEL baj...
11	12 @tanyakanrl Barbie lengkap dengan baju, sepatu...
12	13 Sangat Mantap...Baju Gamis Bermotif Cuanthkkk ...
13	14 @kenjagung jual baju jongli pls beli itu, mau ...
14	15 Mau Gamis Terbaru, Setelan Wanita Terbaru, Gam...
15	16 Cek Jubbah Shahama SH-019 Baju Gamis Pria Dewa...
16	17 Aku nemu promo menarik di Lazada nih. Yuk coba...
17	18 Yuk coba lihat!nNama Produk: LOOK Baju Gamis...
18	19 NEW - Putri Dress ARMANI SILK PREMIUM M-XXL / ...
19	20 Cek RB SARIMBIT KELUARGA KIA BAJU GAMIS KOKO A...
20	21 liat baju gamis mama gue lucu warna pinky tbtb...
21	22 Rekomendasi Gamis / Tunik ðŸŒŸ, ðŸŒŸ, ðŸŒŸELEA TUN...
22	23 Baik, ini saatnya kita membeli semua yang kita...
23	24 @sayuurbening Baju hari raya udah model gamis ...

Setelah di *preprocessing*

0 sebenarnya enggak cuma baju ku kebanyakan kaos...

1 selma maxi dress kondangan maxi dress busui fr...

2 next sg baju gamis koko

3 wkwkwk kalo dua outfit satunya pake baju gamis...

4 rada terlihat vibes lebaran natal jg next sg k...

5 aera baju rajut wanita korean style blouse tur...

6 next si season greeting pake baju koko gamis g...

7 sumpah adek gw ngadi nyuruh pake baju gamis pu...

8 temukan kaftan madina jumbo gamis rayon premiu...

9 kemaren ngeliat outer brukat cakep banget buat...

10 gamis wanita set hijab cadar french rempel baj...

11 barbie lengkap baju sepatu accessories lainnya...

12 sangat mantap baju gamis bermotif cuantikkk ce...

13 jual baju jongli pls beli mau liat jagung pake...

14 mau gamis terbaru setelan wanita terbaru gamis...

15 cek jubbah shahama sh baju gamis pria dewasa m...

16 aku nemu promo menarik lazada nih yuk coba lih...

17 yuk coba lihat nama produk look baju gamis say...

18 new putri dress armani silk premium m xxl baju...

19 cek rb sarimbit keluarga kia baju gamis koko a...

20 liat baju gamis mama gue lucu warna pinky tbtb...

21 rekomendasi gamis tunik elea tunic baju kondan...

22 baik saatnya membeli semua kita mau tas baju s...

23 baju hari raya udah model gamis semua biar dip...

24 gamis wanita set hijab cadar clarisha darkgrey...

Gambar 4. 21 Tampilan Hasil *stopword remover* Sebelum Dan Sesudah di *preprocessing*

Hasil keseluruhan data yang telah di *preprocessing* mulai dari *case folding*, *filtering/cleaning*, dan *stopword remover* yang telah di rapikan agar lebih mudah di baca dan dipahami, seperti gambar dibawah:

no	text	case_folding	clean	stopword
0	1	sebenemnya enggak cuma baju ku kebanyakan kao...	sebenemnya enggak cuma baju ku kebanyakan kao...	sebenemnya enggak cuma baju ku kebanyakan kaos...
1	2	ðŸŒŸ Selma Maxi Dress kondangan-Maxi Dress Bus...	ðŸŒŸ selma maxi dress kondangan-maxi dress bus...	selma maxi dress kondangan maxi dress busui fr...
2	3	@mingkiww next sg24 baju gamis/Koko ðŸŒŸf	@mingkiww next sg24 baju gamis/koko ðŸŒŸf	next sg baju gamis koko
3	4	@sanniue wkwkwk kalo ada dua outfit maka satun...	@sanniue wkwkwk kalo ada dua outfit maka satun...	wkwkwk kalo ada dua outfit maka satunya pake b... wkwkwk kalo dua outfit satunya pake baju gamis...
4	5	@byeolbux rada terlihat vibes lebaran dan Nata...	@byeolbux rada terlihat vibes lebaran dan nata...	rada terlihat vibes lebaran dan natal jg next ... rada terlihat vibes lebaran natal jg next sg k...
5	6	Aera Baju Rajut Wanita Korean Style Blouse Tur...	aera baju rajut wanita korean style blouse tur...	aera baju rajut wanita korean style blouse tur...
6	7	@sparklingnunna Next bisa si season greeting p...	@sparklingnunna next bisa si season greeting p...	next bisa si season greeting pake baju koko ga... next si season greeting pake baju koko gamis g...
7	8	sumpah ya adek gw ngadi2 nyuruh pake baju gami...	sumpah ya adek gw ngadi2 nyuruh pake baju gami...	sumpah ya adek gw ngadi nyuruh pake baju gamis... sumpah adek gw ngadi nyuruh pake baju gamis pu...

Gambar 4. 22 Hasil Dari Keseluruhan Data Yang Telah di *Preprocessing*

4.1.3 Analisis Hasil

Hasil dari perbandingan kata kunci pada tools *semrush* dan *twitter*, penulis mengambil kata kunci “**baju gamis**” karna dari semua percobaan di atas kata kunci baju gamis mendapatkan *volume* pencarian sebanyak 60.5k, *keywords ideas* 16.6k, *related keywords* 700, sedangkan kata kunci “**baju gamis murah**” mendapat *volume* 320, *keywords ideas* 354, *related keyword* 138, pada kata kunci “**gamis cantik murah**” mendapatkan *volume* pencarian sebanyak 10, *keywords ideas* 35, *related keywords* sebanyak 0, dan terakhir kata kunci pencariannya “**gamis murah cantik**” mendapatkan *volume* pencarian sebanyak 70, *keywords ideas* 56, *related keywords* sebanyak 0.

Pada pencarian kata kunci di *twitter* kata kunci “**gamis terbaru**” mendapatkan 2608 *tweet*. berikutnya percobaan dengan *tweet* “**baju gamis murah**” mendapatkan hasil *tweet twitter* sebanyak 399, selanjutnya percobaan dengan *tweet* “**gamis cantik murah**” mendapatkan hasil *tweet* sebanyak 78, dan terakhir dengan percobaan *tweet* “**gamis murah cantik**” mendapatkan hasil *tweet twitter* sebanyak 392.

Oleh karna itu dari semua data yang di dapat di atas penulis menggunakan kata kunci “**baju gamis**” sebagai kata kunci yang yang paling *populer* dan *tren* selama dalam jangka waktu 5 bulan trahir ini.

Kata kunci	Google search			Twitter
	<i>Volume</i>	<i>Keywords ideas</i>	<i>related keywords</i>	<i>tweet</i>
Baju gamis murah	320	364	138	399
Baju gamis	60.5k	16.6k	700	2608
Gamis cantik murah	10	35	0	78
Gamis murah cantik	20	35	0	392

Table 4. 1 Hasil Dari Penelitian