

**LAPORAN**  
**RENCANA TUGAS MANDIRI (RTM) Ke-2**  
**MATA KULIAH BIG DATA KELAS B**  
**“MENGIMPLEMENTASIKAN WEB SCRAPING”**



**DISUSUN OLEH:**

Fadlila Agustina ( 21083010050 )

**DOSEN PENGAMPU:**

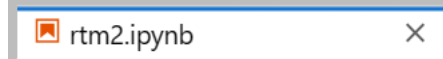
Tresna Maulana Fahrudin S.ST., M.T. (NIP. 199305012022031007)

**PROGRAM STUDI SAINS DATA**  
**FAKULTAS ILMU KOMPUTER**  
**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAWA TIMUR**  
**2023**

## TUGAS 2

Mengimplementasikan *web scraping* dari suatu situs tertentu dengan studi kasus menggunakan sumber data dari halaman google dengan tokoh Ria Ricis. Berikut langkah demi langkah prosesnya :

1. Langkah pertama yaitu membuat file bernama rtm2.ipynb



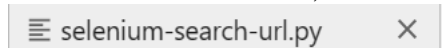
2. Lalu install packages **BeautifulSoup4**, **selenium**, dan **webdriver-manager** dengan menggunakan syntax di bawah ini.

```
pip install BeautifulSoup4
Requirement already satisfied: BeautifulSoup4 in c:\users\fadli\anaconda3\lib\site-packages (4.11.1)
Requirement already satisfied: soupsieve>1.2 in c:\users\fadli\anaconda3\lib\site-packages (from BeautifulSoup4) (2.3.1)
Note: you may need to restart the kernel to use updated packages.

pip install selenium
Requirement already satisfied: selenium in c:\users\fadli\anaconda3\lib\site-packages (4.8.2)
Requirement already satisfied: urllib3[socks]>=1.26 in c:\users\fadli\anaconda3\lib\site-packages (from selenium) (1.26.9)
Requirement already satisfied: certifi>=2021.10.8 in c:\users\fadli\anaconda3\lib\site-packages (from selenium) (2021.10.8)
Requirement already satisfied: trio-websocket>=0.9 in c:\users\fadli\anaconda3\lib\site-packages (from selenium) (0.9.2)
Requirement already satisfied: trio>=0.17 in c:\users\fadli\anaconda3\lib\site-packages (from selenium) (0.22.0)
Requirement already satisfied: idna in c:\users\fadli\anaconda3\lib\site-packages (from trio>=0.17->selenium) (3.3)
Requirement already satisfied: exceptiongroup>=1.0.0rc9 in c:\users\fadli\anaconda3\lib\site-packages (from trio>=0.17->selenium) (1.1.0)
Requirement already satisfied: sortedcontainers in c:\users\fadli\anaconda3\lib\site-packages (from trio>=0.17->selenium) (2.4.0)
Requirement already satisfied: attrs>=19.2.0 in c:\users\fadli\anaconda3\lib\site-packages (from trio>=0.17->selenium) (21.4.0)
Requirement already satisfied: outcome in c:\users\fadli\anaconda3\lib\site-packages (from trio>=0.17->selenium) (1.2.0)
Requirement already satisfied: cffi>=1.14 in c:\users\fadli\anaconda3\lib\site-packages (from trio>=0.17->selenium) (1.15.0)
Requirement already satisfied: async-generator>=1.9 in c:\users\fadli\anaconda3\lib\site-packages (from trio>=0.17->selenium) (1.10)
Requirement already satisfied: sniffio in c:\users\fadli\anaconda3\lib\site-packages (from trio>=0.17->selenium) (1.2.0)
Requirement already satisfied: pycparser in c:\users\fadli\anaconda3\lib\site-packages (from cffi>=1.14->trio>=0.17->selenium) (2.21)
Requirement already satisfied: wsproto>=0.14 in c:\users\fadli\anaconda3\lib\site-packages (from trio-websocket>=0.9->selenium) (1.2.0)
Requirement already satisfied: PySocks!=1.5.7,<2.0,>=1.5.6 in c:\users\fadli\anaconda3\lib\site-packages (from urllib3[socks]>=1.26->selenium) (1.7.1)
Requirement already satisfied: h11<1,>=0.9.0 in c:\users\fadli\anaconda3\lib\site-packages (from wsproto>=0.14->trio-websocket>=0.9->selenium) (0.14.0)
Note: you may need to restart the kernel to use updated packages.

pip install webdriver-manager
Requirement already satisfied: webdriver-manager in c:\users\fadli\anaconda3\lib\site-packages (3.8.5)
Requirement already satisfied: packaging in c:\users\fadli\anaconda3\lib\site-packages (from webdriver-manager) (21.3)
Note: you may need to restart the kernel to use updated packages. Requirement already satisfied: python-dotenv in c:\users\fadli\anaconda3\lib\site-packages (from webdriver-manager) (0.21.1)
Requirement already satisfied: requests in c:\users\fadli\anaconda3\lib\site-packages (from webdriver-manager) (2.27.1)
Requirement already satisfied: tqdm in c:\users\fadli\anaconda3\lib\site-packages (from webdriver-manager) (4.64.0)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\users\fadli\anaconda3\lib\site-packages (from packaging->webdriver-manager) (3.0.4)
Requirement already satisfied: charset-normalizer>=2.0.0 in c:\users\fadli\anaconda3\lib\site-packages (from requests->webdriver-manager) (2.0.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\fadli\anaconda3\lib\site-packages (from requests->webdriver-manager) (1.26.9)
Requirement already satisfied: idna<4,>=2.5 in c:\users\fadli\anaconda3\lib\site-packages (from requests->webdriver-manager) (3.3)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\fadli\anaconda3\lib\site-packages (from requests->webdriver-manager) (2021.10.8)
Requirement already satisfied: colorama in c:\users\fadli\anaconda3\lib\site-packages (from tqdm->webdriver-manager) (0.4.4)
```

3. Setelah terinstall semua, buat file bernama selenium-search-url.py



4. Di dalam file selenium-search-ur.py berisi script seperti di bawah ini dan save.

```
1 from bs4 import BeautifulSoup
2 from selenium import webdriver
3 from webdriver_manager.chrome import ChromeDriverManager
4 chrome_options = webdriver.ChromeOptions()
5 chrome_options.add_argument("--headless")
6 driver = webdriver.Chrome(ChromeDriverManager().install(), chrome_options=chrome_options)
7 # Query to obtain links
8 query = 'Ria Ricis'
9 #Links = [] # Initiate empty List to capture final results
10 # Specify number of pages on google search, each page contains 10 #Links
11 n_pages = 10
12
13 for page in range(1, n_pages):
14     url = "http://www.google.com/search?q=" + query + "&start=" + str((page - 1) * 10)
15     driver.get(url)
16     soup = BeautifulSoup(driver.page_source, 'html.parser')
17     # soup = BeautifulSoup(r.text, 'html.parser')
18
19     search = soup.find_all('div', class_="yuRuf")
20     for h in search:
21         #Links.append(h.a.get('href'))
22         print(h.a.text)
23         print(h.a.get('href'))
```

5. Jalankan syntax di bawah ini pada file rtm2.ipynb untuk memunculkan output ke dalam file bernama hasil-search-url.txt

```
!python selenium-search-ur.py > hasil-search-url.txt
```

C:\Users\fadli\Documents\SMT 4\Big Data\Web scraping\selenium-search-ur.py:6: DeprecationWarning: executable\_path has been deprecated, please pass in a Service object  
driver = webdriver.Chrome(ChromeDriverManager().install(), chrome\_options=chrome\_options)  
C:\Users\fadli\Documents\SMT 4\Big Data\Web scraping\selenium-search-ur.py:6: DeprecationWarning: use options instead of chrome\_options  
driver = webdriver.Chrome(ChromeDriverManager().install(), chrome\_options=chrome\_options)

6. Jika berhasil, akan muncul file bernama hasil-search-url.txt pada folder yang sama dengan file-file di atas.

hasil-search-url 2/24/2023 11:19 PM Text Document 12 KB

7. Isi dari file hasil-search-url.txt adalah link dari google yang mencakup tokoh 'Ria Ricis'.

hasil-search-url - Notepad

File Edit View

Ricis Official - YouTubeyoutube.comhttps://www.youtube.com > channel  
https://www.youtube.com/channel/UC4tS4Q\_Cno5JvCIUXXQ00PA  
5 Artis TikTok Indonesia dengan Followers Terbanyak, Ada Content ...intipseleb.comhttps://www.intipseleb.com > lokal > 60550-5-artis-tiktok...  
https://www.intipseleb.com/lokal/60550-5-artis-tiktok-indonesia-dengan-followers-terbanyak-ada-content-creator-favoritmu#:~:text=Tidak%20hanya%20melalui%20Y...  
RICIS (@riaricis1795) • Instagram photos and videosinstagram.comhttps://www.instagram.com > riaricis1795 > feed  
https://www.instagram.com/riaricis1795/feed/#:~:text=RICIS%20(%40riaricis1795)%20%E2%80%A2%20Instagram%20photos%20and%20videos  
RICIS (@riaricis1795) • Instagram photos and videosinstagram.comhttps://www.instagram.com > ...  
https://www.instagram.com/riaricis1795/?hl=id  
Ria Ricis - Wikipedia bahasa Indonesia, ensiklopedia bebaswikipedia.orghttps://id.wikipedia.org > wiki > Ria\_Ricis  
https://id.wikipedia.org/wiki/Ria\_Ricis  
Ria Ricis (@riaricis) Official - TikTiktiktok.comhttps://www.tiktok.com > @riaricis  
https://www.tiktok.com/@riaricis?lang=id-ID  
Berita dan Informasi Ria ricis Terkini dan Terbaru Hari inidetik.comhttps://www.detik.com > tag > ria-ricis  
https://www.detik.com/tag/ria-ricis  
Profil Ria Ricis - VIVAviva.co.idhttps://www.viva.co.id > siapa > read > 701-ria-ricis  
https://www.viva.co.id/siapa/read/701-ria-ricis  
Berita ria ricis terkini dan terbaru hari ini - Okezone.com -okezone.comhttps://www.okezone.com > tag > ria-ricis  
https://www.okezone.com/tag/ria-ricis  
Minta Dihargai, Ini 8 Potret Ria Ricis yang Ramai Menuai ...kapanlagi.comhttps://www.kapanlagi.com > ... > R > RIA RICIS  
https://www.kapanlagi.com/foto/berita-foto/indonesia/minta-dihargai-ini-8-potret-ria-ricis-yang-ramai-menuai-cibiran-dari-netizen-hargai-orang-lain-dulu.htm  
Berita Ria Ricis Terkini dan Terbaru Hari Ini - SINDONEWSsindonews.comhttps://www.sindonews.com > topic > ria-ricis  
https://www.sindonews.com/topic/48476/ria-ricis  
Berita Ria Ricis Hari Ini - Kabar Terbaru Terkini - Liputan6.comliputan6.comhttps://www.liputan6.com > tag > ria-ricis  
https://www.liputan6.com/tag/ria-ricis  
Ria Ricis Ingin Baby Moana Gak Takut dengan Orang Asing ...suara.comhttps://www.suara.com > lifestyle > komunitas  
https://www.suara.com/lifestyle/2023/02/24/090541/ria-ricis-ingin-baby-moana-gak-takut-dengan-orang-asing-normal-gak-sih  
Youtuber RR dan TR Diprediksi Bakal Bercerai, Nama Ria ...https://www.suara.com > entertainment > gosip  
https://www.suara.com/entertainment/2023/02/24/010000/youtuber-rr-dan-tr-diprediksi-bakal-bercerai-nama-ria-ricis-dan-teuku-ryan-terseret  
Berita Ria Ricis Hari Ini - Kabar Terbaru Terkini - Liputan6.comliputan6.comhttps://www.liputan6.com > tag > ria-ricis  
https://www.liputan6.com/tag/ria-ricis  
6 Potret Ria Ricis dan Baby Moana, Dari Pose Pakai ...https://www.liputan6.com > ShowBiz > Blink-Blink  
https://www.liputan6.com/showbiz/read/5216980/6-potret-ria-ricis-dan-baby-moana-dari-pose-pakai-kacamata-hitam-di-bali-hingga-kompak-berseragam-sd

Ln 1, Col 1 100% Windows (CRLF) ANSI



8. Langkah selanjutnya untuk melakukan scraping pada link google yang dituju adalah dengan cara menginstall packages **newspaper3k**.

```
!pip install newspaper3k
Collecting newspaper3k
  Downloading newspaper3k-0.2.8-py3-none-any.whl (211 kB)
Requirement already satisfied: lxml>=3.6.0 in c:\users\fadli\anaconda3\lib\site-packages (from newspaper3k) (4.8.0)
Requirement already satisfied: PyYAML>=3.11 in c:\users\fadli\anaconda3\lib\site-packages (from newspaper3k) (6.0)
Collecting feedfinder2>=0.0.4
  Downloading feedfinder2-0.0.4.tar.gz (3.3 kB)
Collecting jieba3k>=0.35.1
  Downloading jieba3k-0.35.1.zip (7.4 MB)
Requirement already satisfied: nltk>=3.2.1 in c:\users\fadli\anaconda3\lib\site-packages (from newspaper3k) (3.7)
Requirement already satisfied: python-dateutil>=2.5.3 in c:\users\fadli\anaconda3\lib\site-packages (from newspaper3k) (2.8.2)
Requirement already satisfied: requests>=2.10.0 in c:\users\fadli\anaconda3\lib\site-packages (from newspaper3k) (2.27.1)
Collecting feedparser>=5.2.1
  Downloading feedparser-6.0.10-py3-none-any.whl (81 kB)
Requirement already satisfied: BeautifulSoup4>=4.4.1 in c:\users\fadli\anaconda3\lib\site-packages (from newspaper3k) (4.11.1)
Requirement already satisfied: cssselect>=0.9.2 in c:\users\fadli\anaconda3\lib\site-packages (from newspaper3k) (1.1.0)
Requirement already satisfied: Pillow>=3.3.0 in c:\users\fadli\anaconda3\lib\site-packages (from newspaper3k) (9.0.1)
Collecting tinyssegmenter==0.3
  Downloading tinyssegmenter-0.3.tar.gz (16 kB)
Requirement already satisfied: tldextract>=2.0.1 in c:\users\fadli\anaconda3\lib\site-packages (from newspaper3k) (3.2.0)
Requirement already satisfied: soupsieve>1.2 in c:\users\fadli\anaconda3\lib\site-packages (from BeautifulSoup4>=4.4.1->newspaper3k) (2.3.1)
Requirement already satisfied: six in c:\users\fadli\anaconda3\lib\site-packages (from feedfinder2>=0.0.4->newspaper3k) (1.16.0)
Collecting sgmlib3k
  Downloading sgmlib3k-1.0.0.tar.gz (5.8 kB)
Requirement already satisfied: regex>=2021.8.3 in c:\users\fadli\anaconda3\lib\site-packages (from nltk>=3.2.1->newspaper3k) (2022.3.15)
Requirement already satisfied: tqdm in c:\users\fadli\anaconda3\lib\site-packages (from nltk>=3.2.1->newspaper3k) (4.64.0)
Requirement already satisfied: click in c:\users\fadli\anaconda3\lib\site-packages (from nltk>=3.2.1->newspaper3k) (8.0.4)
```

9. Setelah berhasil terinstall, jalankan script di bawah ini untuk scraping artikel dari situs jawapos.

➤ Artikel 1

```
from newspaper import Article # digunakan untuk mengimpor class Article dari library newspaper

article = Article('https://www.jawapos.com/entertainment/infotainment/19/11/2022/setahun-pernikahan-ria-ricis-teuku-ryan-bersama-menghadapi-cobaan')
# mendefinisikan article dengan class Article dengan parameter pertama adalah Link yang discraping dan parameter yang kedua adalah kode bahasa ya
article.download()
# digunakan untuk download web yang berisi artikel
article.parse()
# untuk mengurai halaman web yang sudah didownload

print("Judul Artikel : ", article.title) # memunculkan output judul artikel
print("Author : ", article.authors) # memunculkan output penulis artikel
print("ISI ARTIKEL : \n", article.text) # memunculkan isi artikel
```

Judul Artikel : Setahun Pernikahan, Ria Ricis-Teuku Ryan Bersama Menghadapi Cobaan  
Author : []  
ISI ARTIKEL :  
JawaPos.com- Sejak 12 November 2022 lalu, usia pernikahan Ria Ricis dan Teuku Ryan genap mencapai 1 tahun. Pasangan yang sudah dianugerahi satu orang anak ini kemudian menggelar perayaan wedding anniversary di bilangan Kemang Jakarta Selatan, Sabtu (19/11), dengan dihadiri keluarga dan sejumlah sahabat.

Teuku Ryan dan Ria Ricis mengakui bahwa dalam satu tahun pernikahan, mereka tidak luput dari badai cobaan. Misalnya saja dimunculkan gosip tak sedap tentang video syur yang dikaitkan dengan Teuku Ryan setelah muncul nama inisial R. Juga dimunculkan rumor tak sedap soal TR numpang hidup ke Ria Ricis padahal faktanya dia juga bekerja di dunia hiburan dan memiliki penghasilan sendiri.

Di tengah badai cobaan yang menimpa rumah tangganya, Ria Ricis dan Teuku Ryan menghadapi bersama. Mereka tidak terlalu memercayai gosip yang beredar di luar sana. Aebab mereka menerapkan rasa saling percaya pada pasangan dengan menjalin komunikasi yang baik dan intensif.

## ➤ Artikel 2

```
from newspaper import Article # digunakan untuk mengimpor class Article dari library newspaper

article = Article('https://www.cnnindonesia.com/hiburan/20230113124035-234-899976/ria-ricis-minta-maaf-soal-kisruh-moana-naik-jetski-jangan-ditiru')
# mendefinisikan article dengan class Article dengan parameter pertama adalah Link yang discraping dan parameter yang kedua adalah kode bahasa yang digunakan
article.download()
# digunakan untuk download web yang berisi artikel
article.parse()
# untuk mengurai halaman web yang sudah didownload

print("Judul Artikel : ", article.title) # memunculkan output judul artikel
print("Author : ", article.authors) # memunculkan output penulis artikel
print("ISI ARTIKEL : \n", article.text) # memunculkan isi artikel
```

```
Judul Artikel : Ria Ricis Minta Maaf Soal Kisruh Moana Naik Jetski: Jangan Ditiru
Author : ['Https', 'Www.Facebook.Com Cnnindonesia', 'Tim']
ISI ARTIKEL :
--
```

Ria Ricis meminta maaf usai membuat vlog dirinya dan Teuku Ryan membawa anak mereka yang berusia lima bulan, Moana, berbuah keributan di media sosial.

Dalam tayangan Rumpi No Secret Trans TV yang tayang pada Senin (9/1), Ria Ricis juga berterima kasih kepada seluruh pihak yang mengingatkan dirinya.

ADVERTISEMENT SCROLL TO RESUME CONTENT

"Kami terima kasih banget buat teman-teman yang sudah ingetin. maaf kalau ada salah. salah sih" katanya sembari tertawa. "Jadi. jangan ditiru."

## 10. Kemudian lakukan proses pembersihan tanda baca agar menjadi sebuah list dan tersimpan dalam file bernama ricis.txt

### ➤ Artikel 1

```
from collections import Counter # untuk menghitung jumlah kemunculan setiap elemen pada sebuah iterable (seperti List, tuple, atau string)
import re # berfungsi untuk memanipulasi string menggunakan ekspresi reguler
# regular expressions

article = article.text
article = article.lower() # digunakan agar tidak ada huruf kapital

# menggunakan fungsi re.sub() untuk substitusi var string
article2 = re.sub(r"[^a-zA-Z\s]", "", article)
bagi = article.split() # membuatnya menjadi list
frekuensi = Counter(bagi) # memberikan seberapa sering kata itu muncul

print('kata yang muncul lebih dari 5 kali akan disimpan ke dalam file ricis.txt')

with open("ricis.txt", "w") as file: # menyimpan hasil output dlm format .txt
    for kalimat in article2: # menggunakan looping for
        file.write(f"{kalimat}") # menyimpan hasil cleaning
        file.write(f"{frekuensi}\n") # menyimpan hasil dari pembagian per kata dalam list
        break
    for kata, frekuensi in frekuensi.items():
        if frekuensi >= 5 : # apabila kata dalam list muncul lebih dari sama dengan 5 kali maka
            file.write(f"{kata}:{frekuensi} kali \n") # akan tersimpan dalam file .txt dan akan memunculkan berapa kali kata tersebut muncul
```

kata yang muncul lebih dari 5 kali akan disimpan ke dalam file ricis.txt

### ➤ Artikel 2

```
from collections import Counter # untuk menghitung jumlah kemunculan setiap elemen pada sebuah iterable (seperti List, tuple, atau string)
import re # berfungsi untuk memanipulasi string menggunakan ekspresi reguler
# regular expressions

article = article.text
article = article.lower() # digunakan agar tidak ada huruf kapital

# menggunakan fungsi re.sub() untuk substitusi var string
article2 = re.sub(r"[^a-zA-Z\s]", "", article)
bagi = article.split() # membuatnya menjadi list
frekuensi = Counter(bagi) # memberikan seberapa sering kata itu muncul

print('kata yang muncul lebih dari 5 kali akan disimpan ke dalam file ricis1.txt')

with open("ricis1.txt", "w") as file: # menyimpan hasil output dlm format .txt
    for kalimat in article2: # menggunakan looping for
        file.write(f"{kalimat}\n") # menyimpan hasil cleaning
        file.write(f"{frekuensi}\n") # menyimpan hasil dari pembagian per kata dalam list
        break
    for kata, frekuensi in frekuensi.items():
        if frekuensi >= 5 : # apabila kata dalam list muncul lebih dari sama dengan 5 kali maka
            file.write(f"{kata}:{frekuensi} kali \n") # akan tersimpan dalam file .txt dan akan memunculkan berapa kali kata tersebut muncul
```

kata yang muncul lebih dari 5 kali akan disimpan ke dalam file ricis1.txt

11. Cek pada file manager dan buka file bernama ricis.txt dan ricis1.txt di dalam file tersebut terdapat isi artikel tanpa tanda baca, banyaknya kata yang muncul di artikel, dan kata yang muncul lebih dari 5 kali.

➤ Artikel 1

```
ricis - Notepad
File Edit View

jawaposcom sejak november lalu usia pernikahan ria ricis dan teuku ryan genap mencapai tahun pasangan yang sudah dianugerahi satu orang anak ini kemudian
teuku ryan dan ria ricis mengakui bahwa dalam satu tahun pernikahan mereka tidak luput dari badai cobaan misalnya saja dimunculkan gosip tak sedap tentang v
di tengah badai cobaan yang menimpa rumah tangganya ria ricis dan teuku ryan menghadapi bersama mereka tidak terlalu memercayai gosip yang beredar di luar s
aku kalau ada apaapa komunikasi dia juga begitu kita lebih percaya sama pasangan sendiri dari pada omongan orang kita menerapkan dalam rumah tangga saling p
kita pribadi nomor komunikasi mudahmudahan allah menjaga rumah tangga kita semua imbuah adik oki setiana dewi itu

teuku ryan memiliki kunci dalam membina rumah tangga supaya tidak mudah bertengkar dengan ria ricis cara yang dilakukannya dengan berusaha menghilangkan ego
alhamdulillah nggak terlalu ada ego apa yang istri mau aku ikut saja asal baik aku nggak merasakan ego aku karena aku sudah punya istri dan anak katanya
cara lain yang mereka lakukan untuk menjaga keharmonisan keluarga adalah dengan memberikan kejutan terhadap pasangan misalnya dengan mengajak pasangan jalan
kadang suami suka ngajak makan aku kan suka makan daging surpirsenya dari suami kalau aku lagi di rumah biasanya yuk kita makan daging sudah aku pesenin pad
dan resep terakhir teuku ryan dan ria ricis selalu memberi tahu pasangan sebelum melakukan sesuatu baik dalam hal urusan pekerjaan ataupun aktivitas lain

Counter({'dan': 9, 'yang': 9, 'ria': 8, 'di': 7, 'dengan': 7, 'aku': 7, 'ricis': 6, 'teuku': 6, 'ryan': 6, 'pasangan': 5, 'rumah': 5, 'dalam': 4, 'mereka':
bat.': 1, 'mengakui': 1, 'bahwa': 1, 'tahun': 1, 'pernikahan': 1, 'luput': 1, 'cobaan.': 1, 'dimunculkan': 1, 'tentang': 1, 'video': 1, 'syun': 1, 'dikaitk
1, 'berusaha': 1, 'menghilangkan': 1, 'diri': 1, 'demi': 1, 'kepentingan': 1, 'besar.': 1, 'alhamdulillah': 1, 'ego.': 1, 'apa': 1, 'mau': 1, 'ikut': 1,

ria:8 kali
ricis:6 kali
dan:9 kali
teuku:6 kali

Ln 1, Col 1 100% Windows (CRLF) ANSI
```

➤ Artikel 2

```
ricis1 - Notepad
File Edit View

mungkin moana belum diajak dulu tapi liat kondisi lanjutnya sembari tertawa dan menegaskan ia juga teuku ryan memang gemar liburan dan mendokumentasikan keg
ria ricis sebelumnya buka suara usai konten naik jetski bareng anaknya moana viral di media sosial

konten itu menuai kontroversi lantaran ria ricis dan suaminya teuku ryan membawa moana yang berusia lima bulan naik jetski tanpa mengenakan pelampung
gambasyoutube

ria ricis kemudian mengatakan melalui instagram story bahwa putrinya baikbaik saja usai kegiatan tersebut ia pun menyertakan tangkapan layar soal perbincang
moms yang masih mengkhawatirkan moana sejak main di laut yang kemarin menunggu kabar ketika pulang tulis ricis dalam unggahan instagram story Kamis
alhamdulillah masyaallah tabarakallah berkat doa suncle moana justru semakin aktif dan malah merangkak ambil mainan alhamdulillah moana sehat lanjutnya
dsa dokter spesialis anak senang sama perkembangan moana tulis ricis

Counter({'ria': 11, 'yang': 11, 'ricis': 10, 'dan': 7, 'moana': 7, 'ia': 5, 'membawa': 4, 'juga': 4, 'naik': 4, 'memang': 4, 'dengan': 4, 'usai': 3, 'diriny
l': 1, 'to': 1, 'resume': 1, 'content': 1, 'kami': 1, 'banget': 1, 'teman-teman': 1, 'ingetin': 1, 'kalau': 1, 'salah.': 1, 'salah': 1, 'sih': 1, 'jadi
': 1, 'kapok': 1, 'kesempatan': 1, 'main.': 1, 'mungkin': 1, '[moana]': 1, 'diajak': 1, 'dulu.': 1, 'liat': 1, 'kondisi.': 1, 'lanjutnya': 1, 'tertawa
'aktif': 1, 'malah': 1, 'merangkak': 1, 'ambil': 1, 'mainan.': 1, 'alhamdulillah': 1, 'sehat.': 1, 'lanjutnya.': 1, 'dsa': 1, '[dokter': 1, 'spesialis':

ria:11 kali
ricis:10 kali
dan:7 kali
yang:11 kali
moana:7 kali
ia:5 kali

Ln 1, Col 1 100% Windows (CRLF) UTF-8
```



12. Untuk membuat 2d, install packages **wordcloud** dan **matplotlib**.

```
!pip install wordcloud
```

```
Requirement already satisfied: wordcloud in c:\users\fadli\anaconda3\lib\site-packages (1.8.2.2)
Requirement already satisfied: pillow in c:\users\fadli\anaconda3\lib\site-packages (from wordcloud) (9.0.1)
Requirement already satisfied: numpy>=1.6.1 in c:\users\fadli\anaconda3\lib\site-packages (from wordcloud) (1.21.5)
Requirement already satisfied: matplotlib in c:\users\fadli\anaconda3\lib\site-packages (from wordcloud) (3.5.1)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\fadli\anaconda3\lib\site-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\fadli\anaconda3\lib\site-packages (from matplotlib->wordcloud) (1.3.2)
Requirement already satisfied: cycler>=0.10 in c:\users\fadli\anaconda3\lib\site-packages (from matplotlib->wordcloud) (0.11.0)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\fadli\anaconda3\lib\site-packages (from matplotlib->wordcloud) (3.0.4)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\fadli\anaconda3\lib\site-packages (from matplotlib->wordcloud) (4.25.0)
Requirement already satisfied: packaging>=20.0 in c:\users\fadli\anaconda3\lib\site-packages (from matplotlib->wordcloud) (21.3)
Requirement already satisfied: six>=1.5 in c:\users\fadli\anaconda3\lib\site-packages (from python-dateutil>=2.7->matplotlib->wordcloud) (1.16.0)
```

```
!pip install matplotlib
```

```
Requirement already satisfied: matplotlib in c:\users\fadi\anaconda3\lib\site-packages (3.5.1)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\fadi\anaconda3\lib\site-packages (from matplotlib) (1.3.2)
Requirement already satisfied: numpy>=1.17 in c:\users\fadi\anaconda3\lib\site-packages (from matplotlib) (1.21.5)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\fadi\anaconda3\lib\site-packages (from matplotlib) (3.0.4)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\fadi\anaconda3\lib\site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: cycler>=0.10 in c:\users\fadi\anaconda3\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: pillow>=6.2.0 in c:\users\fadi\anaconda3\lib\site-packages (from matplotlib) (9.0.1)
Requirement already satisfied: packaging>=20.0 in c:\users\fadi\anaconda3\lib\site-packages (from matplotlib) (21.3)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\fadi\anaconda3\lib\site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\users\fadi\anaconda3\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)
```

13. Untuk membuat wordcloud, dapat menggunakan script dan akan muncul output seperti di bawah ini.

➤ Artikel 1

```
from wordcloud import WordCloud # Library wordcloud
import matplotlib.pyplot as plt
import numpy as np
from PIL import Image # untuk memanipulasi gambar

teks = ''.join(bagi)
mask = np.array(Image.open("joj0.jpeg")) # memanipulasi gambar dgn array
def warna_frekuensi(word, font_size, position, orientation, random_state=None, **kwargs): # settings tampilan
    return tuple((np.random.randint(0, 255, 3))
wordcloud = WordCloud(width = 800, height = 800, background_color = 'white', min_font_size = 10).generate(teks)

plt.figure(figsize = (8, 8), facecolor = None)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off") # menghilangkan bingkai pada library matplotlib
plt.tight_layout(pad = 0) # menyesuaikan layout agar tidak tumpang tindih sehingga dapat dibaca

plt.show()
```



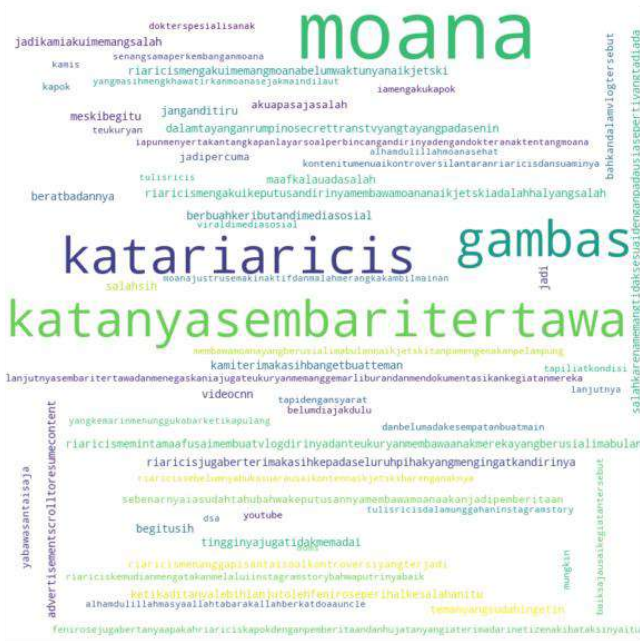
➤ Artikel 2

```
from wordcloud import WordCloud # Library wordCloud
import matplotlib.pyplot as plt
import numpy as np
from PIL import Image # untuk memanipulasi gambar

teks = ''.join(bagi)
mask = np.array(Image.open("joj0.jpeg")) # memanipulasi gambar dgn array
def warna_frekuensi(word, font_size, position, orientation, random_state=None, **kwargs): # settings tampilan
    return tuple(np.random.randint(0, 255, 3))
wordcloud = WordCloud(width = 800, height = 800, background_color = 'white', min_font_size = 10).generate(teks)

plt.figure(figsize = (8, 8), facecolor = None)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off") # menghilangkan bingkai pada library matplotlib
plt.tight_layout(pad = 0) # menyesuaikan Layout agar tidak tumpang tindih sehingga dapat dibaca

plt.show()
```



14. Untuk membuat diagram lingkaran, menggunakan script dan akan muncul output seperti di bawah ini.

➤ Artikel 1

```
daftar_data = bag1

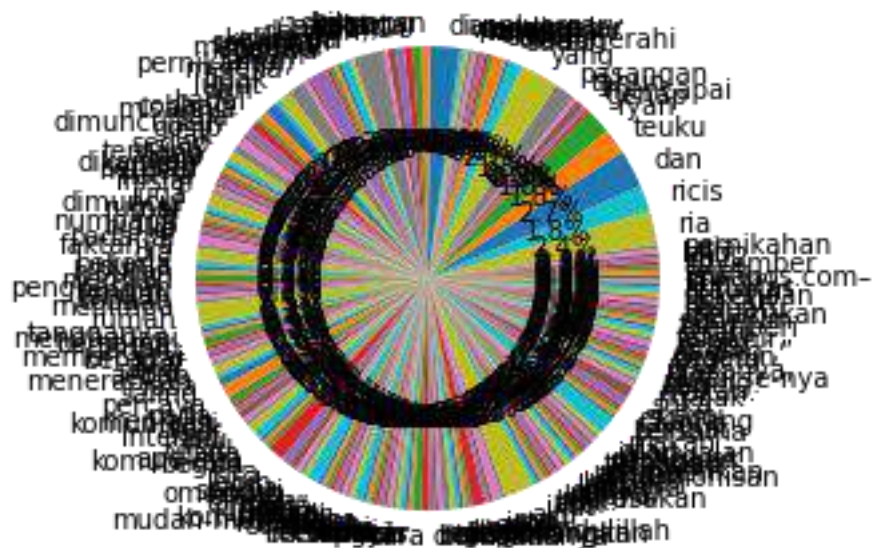
import matplotlib.pyplot as plt
count_data = Counter(daftar_data)

# Mengambil nilai dan label dari setiap item pada dictionary count_data
values = count_data.values()
labels = count_data.keys()

# Membuat diagram lingkaran
plt.pie(values, labels=labels, autopct='%1.1f%%')

# Menampilkan diagram
plt.show()
```





➤ Artikel 2

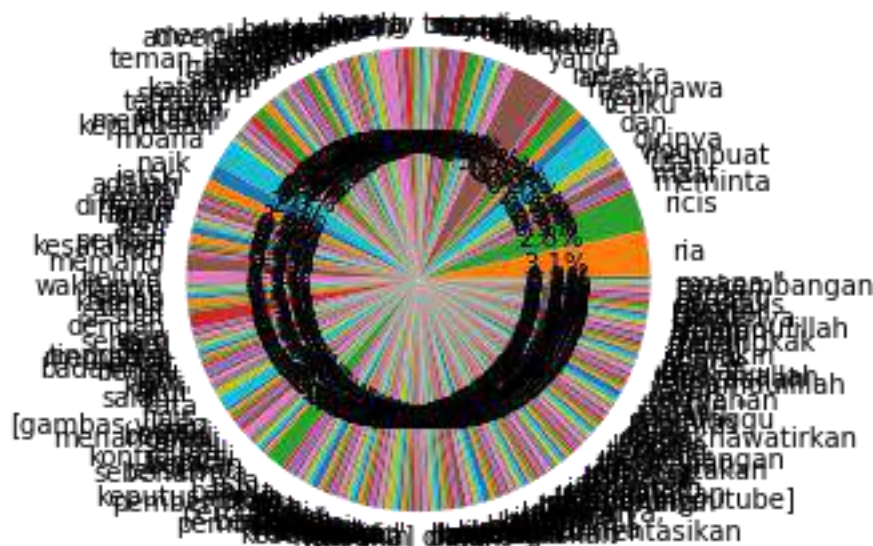
```
daftar_data1 = bagi

import matplotlib.pyplot as plt
count_data = Counter(daftar_data1)

# Mengambil nilai dan label dari setiap item pada dictionary count_data
values = count_data.values()
labels = count_data.keys()

# Membuat diagram lingkaran
plt.pie(values, labels=labels, autopct='%1.1f%%')

# Menampilkan diagram
plt.show()
```



15. Yang terakhir, untuk memunculkan diagram batang dapat menggunakan script dan akan muncul output seperti di bawah ini

➤ Artikel 1

```
plt.figure(figsize=(25, 10)) # mengatur ukuran gambar
plt.rcParams.update({'font.size': 10}) # mengatur ukuran font
plt.bar(labels, values)
plt.xticks(rotation=90) # mengatur agar font label x berotasi 90 derajat
plt.show()
```

[illegible]