

# LAPORAN TUGAS PRAKTIKUM TEXT MINING NAIVE BAYES, LDA DAN MATRIX COHERENCE

Laporan diperuntukan sebagai pemenuh syarat penilaian praktikum mata kuliah Data Mining

Dosen Pengampu : Illham Faishal Mahdy, S.Stat., M.Stat.  
Oleh : Catherine V. Pang 2C2220008 VI/A  
Hari, Tanggal: Senin, 24 April 2025  
Pertemuan Ke: 3  
Obyek: Tuning jumlah topik yang dapat dihasilkan dari LDA dan dievaluasi dengan matriks Coherence

PROGRAM STUDI S1 SAINS DATA FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS KOPERASI INDONESIA 2025

```
import kagglehub

# Download latest version
path = kagglehub.dataset_download("rezkyyayang/reviews-of-indonesian-app-startups-on-playstore")

print("Path to dataset files:", path)
```

Path to dataset files: /kaggle/input/reviews-of-indonesian-app-startups-on-playstore

```
import pandas as pd

# Read the CSV file dengan path yang benar
df_gojek = pd.read_csv('/kaggle/input/reviews-of-indonesian-app-startups-on-playstore/gojek.csv')

# Display the first few rows of the dataframe
df_gojek.head()
```

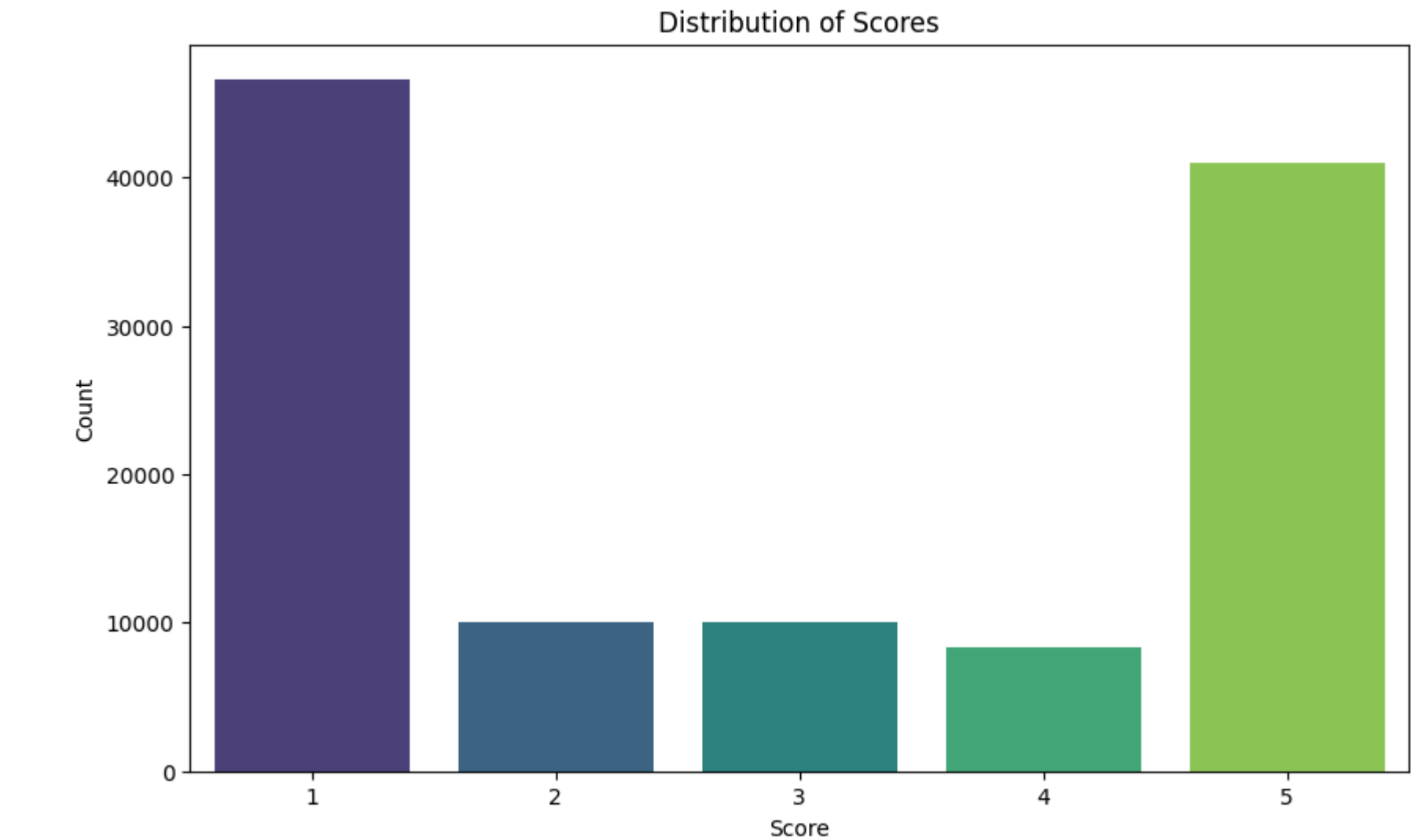
	id	created_at	score	content	thumbsUpCount
0	0	2024-10-17 21:38	1	Gojek udah ga kayak dlu. Skrg ongkirnya mahal....	57
1	1	2024-10-15 20:07	2	Setelah update jadi ngelag dan banyak bug. Say...	104
2	2	2024-10-15 16:24	2	Tolong untuk Go Food, bagian terdekat, enak ya...	48
3	3	2024-10-24 18:02	1	Search alamatnya payah👎. Pas dicari lewat mapn...	2
4	4	2024-10-11 22:36	1	UI doang bagus UX nya jelek parah. aplikasi go...	78

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

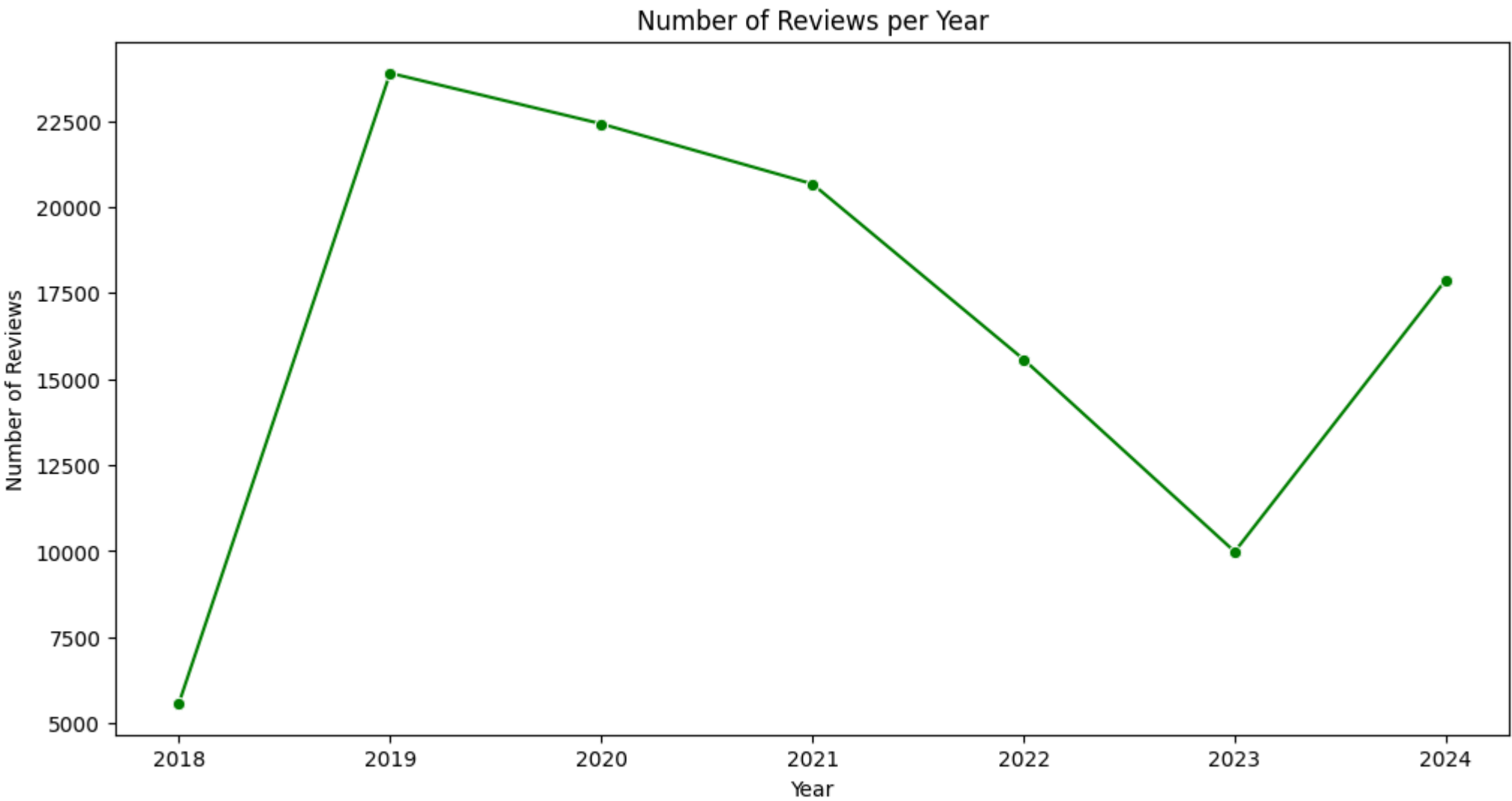
# Analisis deskriptif untuk kolom 'score'
score_counts = df_gojek['score'].value_counts().sort_index()
plt.figure(figsize=(10, 6))
sns.barplot(x=score_counts.index, y=score_counts.values, palette='viridis')
plt.title('Distribution of Scores')
plt.xlabel('Score')
plt.ylabel('Count')
plt.show()
```

```
<ipython-input-9-fd547a61fb78>:7: FutureWarning:
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

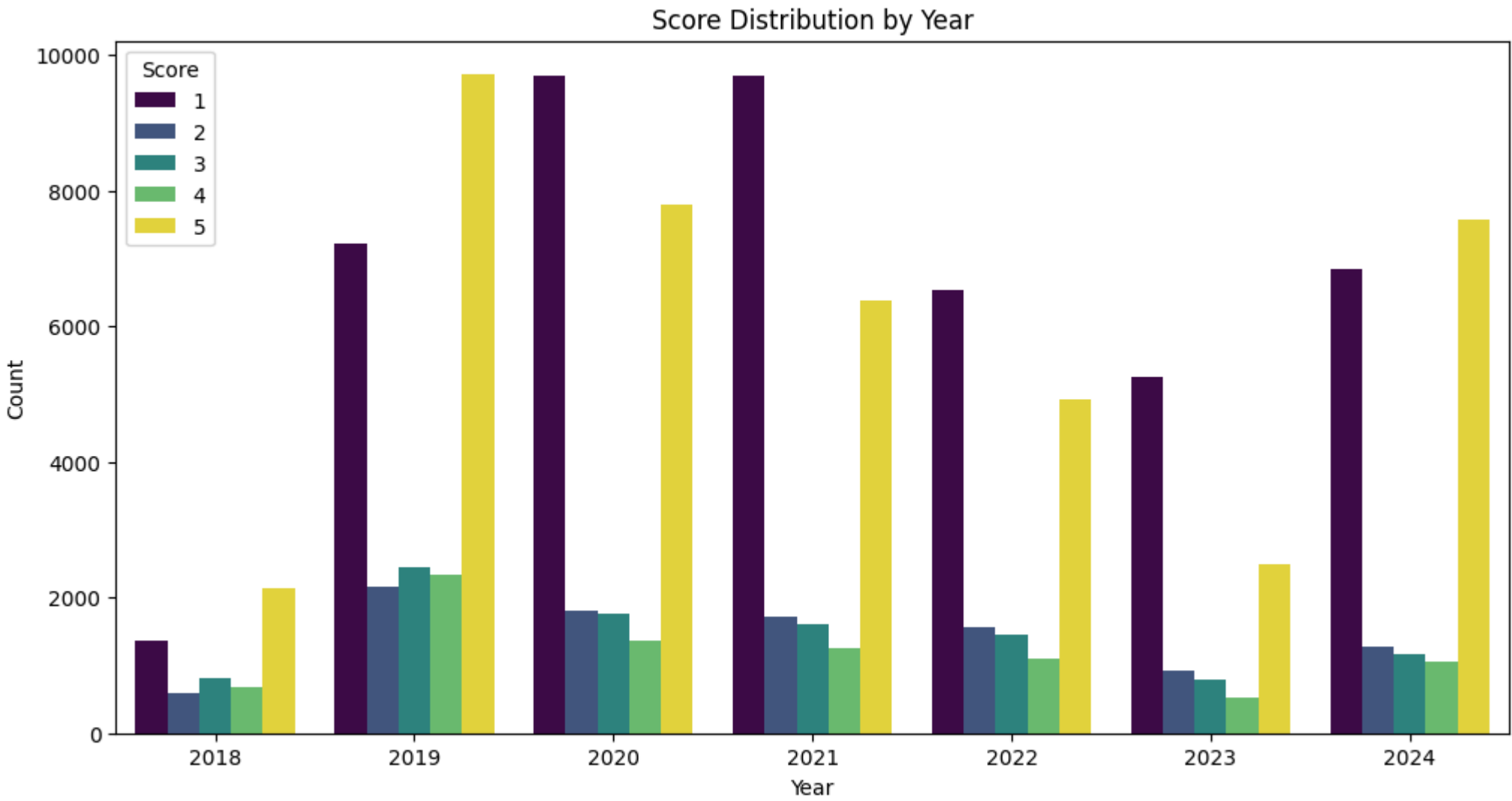
sns.barplot(x=score_counts.index, y=score_counts.values, palette='viridis')
```



```
# Analisis deskriptif untuk kolom 'created_at'
df_gojek['created_at'] = pd.to_datetime(df_gojek['created_at'])
df_gojek['year'] = df_gojek['created_at'].dt.year
year_counts = df_gojek['year'].value_counts().sort_index()
plt.figure(figsize=(12, 6))
sns.lineplot(x=year_counts.index, y=year_counts.values, marker='o', color='green')
plt.title('Number of Reviews per Year')
plt.xlabel('Year')
plt.ylabel('Number of Reviews')
plt.show()
```



```
# Cross tabulasi antara 'year' dan 'score'
plt.figure(figsize=(12, 6))
sns.countplot(x='year', hue='score', data=df_gojek, palette='viridis')
plt.title('Score Distribution by Year')
plt.xlabel('Year')
plt.ylabel('Count')
plt.legend(title='Score')
plt.show()
```



```
df_gojek_2024 = df_gojek[df_gojek['year'] == 2024]
df_gojek_2024
```




	id	created_at	score	content	thumbsUpCount	year
0	0	2024-10-17 21:38:00	1	Gojek udah ga kayak dlu. Skrg ongkirnya mahal....	57	2024
1	1	2024-10-15 20:07:00	2	Setelah update jadi ngelag dan banyak bug. Say...	104	2024
2	2	2024-10-15 16:24:00	2	Tolong untuk Go Food, bagian terdekat, enak ya...	48	2024
3	3	2024-10-24 18:02:00	1	Search alamatnya payah 🙄. Pas dicari lewat mapn...	2	2024
4	4	2024-10-11 22:36:00	1	UI doang bagus UX nya jelek parah. aplikasi go...	78	2024
...	...	...	...	...	...	...
115729	115729	2024-04-06 20:11:00	2	Sekarang susah nyari driver di saat hujan dera...	1	2024
115731	115731	2024-02-12 10:19:00	5	Trimakasih gojek dan aplilasi ini memudah kan ...	0	2024
115733	115733	2024-04-01 14:10:00	5	min mau tnyaa aku mau transfer kok gak bsa ya ...	0	2024
115932	115932	2024-07-23 19:41:00	2	Sekarang ada DC lapangannya ya?	0	2024
115953	115953	2024-02-15 19:51:00	1	Saldo gopay saya menghilang dengan sendirinya....	4	2024

17886 rows × 6 columns

```
# Use the .head() method to get the first 100 rows
df_gojek_2024_head = df_gojek_2024.head(100)

# Now df_gojek_2024_head contains only the first 100 rows
df_gojek_2024_head
```



	id	created_at	score	content	thumbsUpCount	year
	0	0	2024-10-17 21:38:00	1	Gojek udah ga kayak dlu. Skrg ongkirnya mahal....	57 2024
	1	1	2024-10-15 20:07:00	2	Setelah update jadi ngelag dan banyak bug. Say...	104 2024
	2	2	2024-10-15 16:24:00	2	Tolong untuk Go Food, bagian terdekat, enak ya...	48 2024
	3	3	2024-10-24 18:02:00	1	Search alamatnya payah 🤦. Pas dicari lewat mapn...	2 2024
	4	4	2024-10-11 22:36:00	1	UI doang bagus UX nya jelek parah. aplikasi go...	78 2024
...	...		...		...	...
	95	95	2024-10-08 16:41:00	1	Setiap order gofood ,pasti drivernya kirim pes...	4 2024
	96	96	2024-10-01 16:31:00	1	Transaksi pembayaran later selalu gagal 1 . de...	4 2024
	97	97	2024-10-21 20:03:00	1	knp gofood mode hemat banyak driver yg cancel ...	1 2024
	98	98	2024-10-15 23:47:00	1	Clame asuransi sangat susah dan cacat pelayana...	1 2024
	99	99	2024-10-14 13:15:00	1	Nggak ada sesi konfirmasi setelah memesan gofo...	1 2024
100 rows × 6 columns						

TEXT PREPROCESSING

```
!pip install Sastrawi
!pip install deep_translator
!pip install vaderSentiment
```

```
Collecting Sastrawi
  Downloading Sastrawi-1.0.1-py2.py3-none-any.whl.metadata (909 bytes)
  Downloading Sastrawi-1.0.1-py2.py3-none-any.whl (209 kB)
    ----- 209.7/209.7 kB 4.4 MB/s eta 0:00:00

Installing collected packages: Sastrawi
Successfully installed Sastrawi-1.0.1
Collecting deep_translator
  Downloading deep_translator-1.11.4-py3-none-any.whl.metadata (30 kB)
Requirement already satisfied: beautifulsoup4<5.0.0,>=4.9.1 in /usr/local/lib/python3.11/dist-packages (from deep_translator) (4.13.4)
Requirement already satisfied: requests<3.0.0,>=2.23.0 in /usr/local/lib/python3.11/dist-packages (from deep_translator) (2.32.3)
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.11/dist-packages (from beautifulsoup4<5.0.0,>=4.9.1->deep_translator) (2.7)
Requirement already satisfied: typing-extensions>=4.0.0 in /usr/local/lib/python3.11/dist-packages (from beautifulsoup4<5.0.0,>=4.9.1->deep_translator) (4.13.2)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.23.0->deep_translator) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.23.0->deep_translator) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.23.0->deep_translator) (2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.23.0->deep_translator) (2025.4.26)
  Downloading deep_translator-1.11.4-py3-none-any.whl (42 kB)
    ----- 42.3/42.3 kB 1.7 MB/s eta 0:00:00

Installing collected packages: deep_translator
Successfully installed deep_translator-1.11.4
Collecting vaderSentiment
  Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl.metadata (572 bytes)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (from vaderSentiment) (2.32.3)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests->vaderSentiment) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests->vaderSentiment) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests->vaderSentiment) (2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests->vaderSentiment) (2025.4.26)
  Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl (125 kB)
    ----- 126.0/126.0 kB 2.6 MB/s eta 0:00:00

Installing collected packages: vaderSentiment
Successfully installed vaderSentiment-3.3.2
```

```
#Import library yang dibutuhkan
import csv
import nltk
import re
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from nltk.tokenize import word_tokenize
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemover, ArrayDictionary
import string
```

```
def filtering_text(text): #disebut juga trasholding
    # Check if text is a list and join it into a string
    if isinstance(text, list):
        text = ' '.join(text)
    # mengubah tweet menjadi huruf kecil
    text = text.lower()
    # menghilangkan url
    text = re.sub(r'https?:\/\/\/S+', '',text)
    # menghilangkan mention, link, hashtag
    text = ' '.join(re.sub("([@#][A-Za-z0-9]+)|(\w+:\/\/\/S+)", " ", text).split())
    #menghilangkan karakter byte (b')
    text = re.sub(r'(b'\{1,2}\)',"", text)
    # menghilangkan yang bukan huruf
    text = re.sub('[^a-zA-Z]', ' ', text)
    # menghilangkan digit angka
    text = re.sub(r'\d+', '', text)
    #menghilangkan tanda baca
    text = text.translate(str.maketrans("", "",string.punctuation))
    return text
```

```
df_gojek_2024_head['content'] = df_gojek_2024_head['content'].astype(str)
```

```
<ipython-input-19-fa9f116344c9>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
df_gojek_2024_head['content'] = df_gojek_2024_head['content'].astype(str)
```

```
df_gojek_2024_head['filtered'] = df_gojek_2024_head['content'].apply(filtering_text)
print(df_gojek_2024_head['filtered'].head())
```

```
0    gojek udah ga kayak dlu    skrg ongkirnya mahal ...
1    setelah update jadi ngelag dan banyak bug    say...
2    tolong untuk go food    bagian terdekat    enak ya...
3    search alamatnya payah    pas dicari lewat mapn...
4    ui doang bagus ux nya jelek parah    aplikasi go...
Name: filtered, dtype: object
<ipython-input-20-b3f582bd7e42>:1: SettingWithCopyWarning:
```



A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy).  
df\_gojek\_2024\_head['filtered'] = df\_gojek\_2024\_head['content'].apply(filtering\_text)

```
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords

# Indonesian stopwords
indonesian_stopwords = stopwords.words('indonesian')

# Create a custom stopwords remover
factory = StemmerFactory()
stemmer = factory.create_stemmer()
stopword_factory = ArrayDictionary(indonesian_stopwords)
stopword = StopWordRemover(stopword_factory)

def remove_stopwords(text):
    text = stopword.remove(text)
    return text

df_gojek_2024_head['no_stopwords'] = df_gojek_2024_head['filtered'].apply(remove_stopwords)
print(df_gojek_2024_head['no_stopwords'].head())
```



[nltk\_data] Downloading package stopwords to /root/nltk\_data...  
[nltk\_data] Package stopwords is already up-to-date!  
0 gojek udah ga kayak dlu skrg ongkirnya mahal ...  
1 update ngelag banyak bug menambah alamat di ...  
2 tolong go food terdekat enak dulu pakai jar...  
3 search alamatnya payah pas dicari mapnya ket...  
4 ui doang bagus ux nya jelek parah aplikasi go...  
Name: no\_stopwords, dtype: object  
<ipython-input-21-401e661d67d0>:18: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy).  
df\_gojek\_2024\_head['no\_stopwords'] = df\_gojek\_2024\_head['filtered'].apply(remove\_stopwords)

```
import nltk

# Download the 'punkt' and 'punkt_tab' resources
nltk.download('punkt')
nltk.download('punkt_tab')

def word_tokenize_wrapper(text):
    return word_tokenize(text)

df_gojek_2024_head['tokenized'] = df_gojek_2024_head['no_stopwords'].apply(word_tokenize_wrapper)
print(df_gojek_2024_head['tokenized'].head())
```



[nltk\_data] Downloading package punkt to /root/nltk\_data...  
[nltk\_data] Package punkt is already up-to-date!  
[nltk\_data] Downloading package punkt\_tab to /root/nltk\_data...  
[nltk\_data] Unzipping tokenizers/punkt\_tab.zip.  
0 [gojek, udah, ga, kayak, dlu, skrg, ongkirnya,...  
1 [update, ngelag, banyak, bug, menambah, alamat...  
2 [tolong, go, food, terdekat, enak, dulu, pakai...  
3 [search, alamatnya, payah, pas, dicari, mapnya...  
4 [ui, doang, bagus, ux, nya, jelek, parah, apli...  
Name: tokenized, dtype: object  
<ipython-input-22-d54e403b642f>:10: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy).  
df\_gojek\_2024\_head['tokenized'] = df\_gojek\_2024\_head['no\_stopwords'].apply(word\_tokenize\_wrapper)

```
# stemming
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

def stemming(text_cleaning):
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    do = []
    for w in text_cleaning:
        dt = stemmer.stem(w)
        do.append(dt)
    d_clean = []
    d_clean = " ".join(do)
    print(d_clean)
    return d_clean

df_gojek_2024_head['stemming'] = df_gojek_2024_head['tokenized'].apply(stemming)
```



gojek udah ga kayak dlu skrg ongkirnya mahal ongkir gofood jarak km kayak ongkir jarak km trus tambah biaya layanan kalo pake mode hemat proses banget blm nunggu dapet drivernya trus goride udah ngelag banyak bug tambah alamat di set respon aplikasi pesan makan jarak menu sangat jauh buka pilih menu muncul pop up biasa tekan tutup tidak ada respon saat tekan dari menu resto tolong go food dekat enak dulu pakai jarak bisa resto makan dekat jarak nya pakai menit gak urut jarak gak nemu resto dekat sekali saya lumayan kalau tempat beli gak hapal resto dekat yg search alamat payah pas cari map ketemu jalan dll kalo ketik malah gk payah banget nyari alamat yg jarak ya gila aja kali cari scroll map serius fitur navigasi bodoh pernah saya alami ui doang bagus ux nya jelek parah aplikasi gofoodnya bikin bingung ngapus pesan aja ribet bgt pencet pesan ga tau bug gimana minimal kasih pin biar ga langsung order di cancel ga enak dr ada update x hancur order makan resto tp buka menu lemottttt beratttttt gimana pilih menu buka menu aja lot berat kesini makin sono makin makain aja hufffttttt langgan go ride utk jarak pendek akhir ndak nyaman rekomendasi driver yg jarak jemput banget sama kadang lebih parah jarak nganternya agregat rugi bahan bakar n tunggu aplikasi tiada bayar pajak kendaraan status proses komplain ada respon transaksi batal oleh gojek bayar pajak kendaraan telat kena denda rp alam sesak bayar tokopedia instan langsung hasil c benahin deh makin baik dan profesional ngga knapa kalo beli makan sulit bgt dpt driver cuaca ngga ujan rumah tengah kota rugi bgt udah acc resto tpi nunggu driver lama bgt smpe menit bat tolong sampein para driver kalo emang bom buat jalan gausah dinyalain aplikasi customer udah nunggu drivernya diem tempat aja estimasi driver sampe menit driver datang dari menit gilir c tolong batal pesan laku langgan impactnya ke driver riwayat batal order masuk akun susah order kemudian karena driver susah payah menyetabilkan forma jam online rating kerja terima order bulan kali pesan gocar kecewa lokasi jakarta titik gak sesuai hal ulang kali maps yg apps tumpang benar tp beda dg maps di driver otomatis driver sama maps nya beberapa yg gk bebal ikuti aplikasi berat hp ram kasihan konsumen milik hp android seperti coolpad masih setia pakai jasa gojek lebih nyaman hp mereka tak untuk loading sebab versi makin berat mb nya makin tinggi akhir selalu error kali buka app pesan gofood menu toko ga tampil besar layar putih tungguin baru muncul menu pun klik ga muncul muncul pop up ngeliat jelas menu yg saya klik td keranjang aplikasi susah user ganti nomor sudah daftar susah ubah pin login ga menu lupa pin bagaimana masuk tiga kurang bagi go transit suka munculin barcodeny scan barcodenya kecewa sih aplikasi nya error jaring bagus hp bagus tpi pilih layanan gosend gak jalan menit nunggu ngirim chat gabisa telpon gabisa driver yg itu di map stuck jalan alami masalah beberapa yg salah satu saja sore kemarin mes makan ongkir k keti makan nanya sampai ongkir k jadi harga ongkir naik otomatis tiba aplikasi sering alami dalam beberapa minggu boleh tarif jangkau promo voucher go point banyak udah males pakai aplikasi tarik lumayan naik biaya go food naik point pakai untuk aja fitur chat tpi komunikasi arah gmana nih aplikasi t sudah uninstal aplikasi gojek customer service call center nya sekali fungsi sekali bisa hubungi bisa komunikasi langsung useless gambar gojek lagi empati customernya utk gocar driver tel aplikasi bug nya akhir order goride x yg panggil driver tuju yg dari akun saya order x kali cek aplikasi tiba order goridenya saat saldo gopay isi iklan nya kyk shopee anjg asik main game tekan silang masuk aplikasi toll emang hilang uang ku gara gara banyak iklan nih harus grinding in lagi jam bisa SSR mohon baik sistem masak order goride pesan gak sengaja double klik nyantol driver konsumen yg gak enak driver nya aplikasi belah kayak gak double pesan selesai untung drivernya sabar harus hari tiap order gocar gojek astaga banget dapetnya sekali dapat drivernya protes bilang kalo ngambil paket hemat nggak modal BBM kalo dr pov konsumen nyari yg murah ya cuman sangat dari barusan order gocar tampil aplikasi gak muncul layar pantau kemarin bikin salah paham bapak drivernya bisa lokasi maps kesel banget CS nya nggak responsif drivernya nggak yang aneh bin ajaib please depan tolong training driver drivernya untuk driver motor kami jelas aman data pengunna aplikasi duga pihak sengaja manfaat nomor daftar dlm aplikasi saya minggu di hubungi nomor spam tawar produk saya baru mendownload aplikasi tapi banyak nomor beli makan jajan karna nggak enak badan go mart drivernya nggak yang nerima jelek banget kalau pikir selalu ngasi tip mereka tolong deh baik iya puluh ratus driver nggak masuk notifiknya n system juga yg perlu perhati kait drivernya khusus driver motor bersih helm sedia tumpang tolong buat standard tentu gojek biar tumpang nyaman


banget beli makan proses resto tiba cancel jam toko tutup online juga bisa proses selesai offline toko buka asa php sudah adu kait akun yg bisa du akses namun nominal saldo akun rasa tipu ada jamin uang namun tidak akses nomor telepon yg hilang tidak kait gmail respon csnya bgt ngga responsif ada selesai buruk tumpang tunggu menit driver tolak antar tp mau cancel kalau driver gerak arah tumpang waktu menit cari driver ganti mau pick up tumpang bug tdk jls buru koneksi internet tdk mulu data sdh nyata kuota aman buka aplikasi bisa sy cb pakai hp lgi sama koneksi internet tdk knp ya gojek ikhlas promo mode hemat pesan makan pakai antar mode hemat harus nunggu jam dapat driver semntara makan udah disiapin resto tidak batal tolong ya gojek kasih solusi kait sistem cari driver gofood gatau sistem driver tpi banget driver entah pencet gimana chat ga dibales gk kasih info nunggu jam tiba ganti driver makan ser main ongkir deket mahal isi embel embel biaya nang dll trus promo gak yang bagus makan nya murah serasa makan mahal biaya ongkir tangan nya tiap paket hemat grab driver mobil motor nya susah banget dapetnya banyak sekitar gilir ganti ongkos mahal langsung dapet driver apa diadain paket hemat kalu ada yang siap ngantar dan nu punya alam gocar kurang kena pakai promo hemat sudah aplikasi sedia kami di remeh oknum driver gocar sampai hina kami pakai promo hemat tsb pdhl customer tau tau harga customer yang malis lu si aplikasi mana muncul iklan gojek gausah banyak iklan aplikasi mana mengganggu pula orang tau gojek harus muncul iklan aplikasi kecewa banget udah nunggu jam beli gofood gak dapet driver nya kali pilih driver yang bener kerja keras cuma main aja iyaa tau karna ujan gak gitu nya udah dapet driver nya makan udah jd puas krn tepat harga sikap layan yg ramah kurang perlu verifikasi titik titik jemput map akhir dpt geser titik lokasi sy utk jemput geser m harap verifikasi kesini jengkel guna maps gojek relevan jarak tempuh pendek sangat rugi drivernya map gojek singkat timbang jalan sudah tutup portal jam malam buat debat tumpang drivernya pihak it develc bagi aktifitas knapa gabisa refresh ya pesan jaring bagus hp bagus aplikasi lancar jaya gojek nya aja rada eror aplikasi belah ga loh kayak gitu malu produk negri ya sedikit kritik moga sistem gajelas udah pakai paylater tiba blok karna lambat coba cek jatuh tempo tgl nov gue udah bayar tgl okt bayar tgl pas gaji ga masalah ini lg rezeki bayar sblm tgl jatuh tempo di bl ya pesan goride dapat drive posisi menit harus tunggu ke tmpt kerja saya menit saya cancel minta jgn dicancel pesan goride utk cpt aman malah tolong kasi posisi driver yg tolong driver s guna sampai ok nyaman mungkin sayang mau banyak promonya pakai aplikasi dompet digital yg pisah aplikasi ini juga tolong bayar bank jago syariah nya go transit biar gak top up gopay nya driver suka cancel enak nya ya orang nunggu pas driver dekat lokasi tiba di cancel bukan dua kali sering perbaiki masalah pemberitahuan notifikasi diberitahukan kalo gak driver driver sibuk pas mesan goride untuk kali tiba driver pesan goride pertama aman aja jujur aplikasi bagus amat bantu akhir ini knpa gofood pas ga pakai gilir pakai akun suami bisa pakai lokasi nya di akun saya muncul tulis gofood pas bisa di jangkau area sudah kali aplikasi gojek titik map benar klu gofood titik temu alamat kirim titik selalu sama mohon baik rute pilih gps maps jauh rute biasa tidak macet bagaana atas di aplikasi menu adu masalah aplikasi heran bgt driver nih pingin laku ga niat jemput jarak deket ga yg jemput like wtf mah jalan sepi ga macet tolak mulu njg nolak rejeki bgt dah kasi tumpang gofood banget udah dapet driver drivernya gak konfirmasi ambil alias tidak responsif status driver luncur restoran lama sih biar ganti driver otomatis kasi restonya suka nyiapin makan du hari gopay g pakai klik menu pilih bayar cm sedia pilih tunai pdhl msh byk gopaynya trs uninstall diinstal ulang mlh g login masuk pin tulis ada koneksi pdhl koneksi kali blok suruh coba libur bal anjur bayar nontunai gopay krna bnyak driver mau ambil order banyak keluhan driver karna ambil saldo vendor itu potong nya jadi untuk karna taxi online lebih baik pindah grab supay kasih promo gofood kek udah langgan gofood plus bulan makin kesini ngga kasih promo bjir harga resto naikin beda harga beli langsung di tempat yang dateng porsi beda kadang sen ngga leng kecewa banget apk saya guna setia gojek skrg jarang promk sering banget telat kirim klo pas pesan makan apk nya gak kayak dlu skrg

```
#!pip install deep_translator
```

```
from deep_translator import GoogleTranslator

# Menerjemahkan setiap kata dalam setiap entri kolom 'full_text' ke dalam bahasa Inggris
translated_tweets = []
for tweet in df_gojek_2024_head['stemming']:
    translated_words = []
    for word in tweet.split():
        try:
            translated_word = GoogleTranslator(source='id', target='en').translate(word)
        except TranslationNotFound:
            # If translation not found, keep the original word
            translated_word = word
        print(f"Translation not found for: {word}, using original word.")
    translated_words.append(translated_word)
    translated_tweet = ' '.join(translated_words)
    translated_tweets.append(translated_tweet)

# Menambahkan kolom baru 'tweet_english' ke dalam data
df_gojek_2024_head['translated'] = translated_tweets
```

 <ipython-input-24-a36d5a6b9595>:19: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
df_gojek_2024_head['translated'] = translated_tweets
```

```
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

# Initialize VADER sentiment analyzer
analyzer = SentimentIntensityAnalyzer()


def analyze_sentiment(text):
    scores = analyzer.polarity_scores(text)
    return scores['compound']

df_gojek_2024_head['vader_sentiment'] = df_gojek_2024_head['translated'].apply(analyze_sentiment)

# Function to label sentiment based on compound score
def label_sentiment(score):
    if score >= 0.05:
        return 'positive'
    elif score <= -0.05:
        return 'negative'
    else:
        return 'neutral'


df_gojek_2024_head['sentiment_label'] = df_gojek_2024_head['vader_sentiment'].apply(label_sentiment)

df_gojek_2024_head[['translated', 'vader_sentiment', 'sentiment_label']].head()
```

 <ipython-input-25-e958a07cadd1>:10: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
df_gojek_2024_head['vader_sentiment'] = df_gojek_2024_head['translated'].apply(analyze_sentiment)
```

 <ipython-input-25-e958a07cadd1>:21: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
df_gojek_2024_head['sentiment_label'] = df_gojek_2024_head['vader_sentiment'].apply(label_sentiment)
```

	translated	vader_sentiment	sentiment_label
0	Gojek Already Ga Like dlu now postage expensiv...	-0.0516	negative
1	Update Ngelag Lots bug plus address in set res...	-0.5267	negative
2	Help Go Food near nice Formerly use distance C...	-0.4501	negative
3	Search address bad right search folder meet ro...	-0.6187	negative
4	UI doang Good UX his Bad critical application ...	0.2023	positive

```
# Menghitung jumlah masing-masing label
label_counts = df_gojek_2024_head['sentiment_label'].value_counts()

label_counts
```



sentiment_label	count
negative	55
positive	40
neutral	5

dtype: int64

MULTINOMIAL NAIVE BAYES

```
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

# Pisahkan data menjadi fitur (teks) dan target (label sentimen)
X = df_gojek_2024_head['translated']
y = df_gojek_2024_head['sentiment_label']

# Bagi data menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Buat objek TF-IDF vectorizer
tfidf_vectorizer = TfidfVectorizer()

# Fit dan transform data latih
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)

# Transform data uji
X_test_tfidf = tfidf_vectorizer.transform(X_test)

# Buat dan latih model Multinomial Naive Bayes
nb_classifier = MultinomialNB()
nb_classifier.fit(X_train_tfidf, y_train)

# Lakukan prediksi pada data uji
y_pred = nb_classifier.predict(X_test_tfidf)

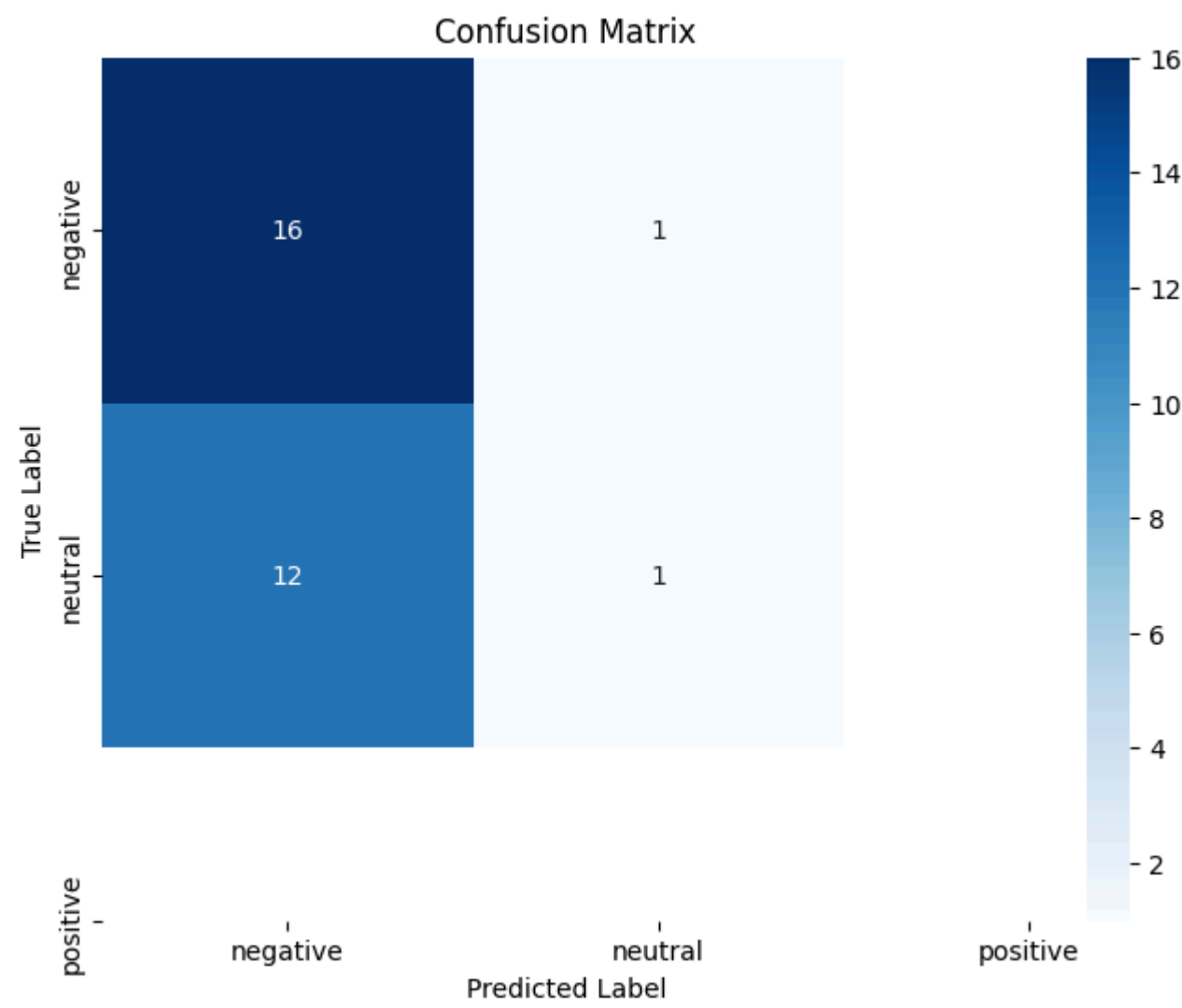
# Evaluasi model
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy}")
print(classification_report(y_test, y_pred))

# Membuat confusion matrix
cm = confusion_matrix(y_test, y_pred)

# Plot confusion matrix
plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues',
            xticklabels=['negative', 'neutral', 'positive'],
            yticklabels=['negative', 'neutral', 'positive'])
plt.xlabel('Predicted Label')
plt.ylabel('True Label')
plt.title('Confusion Matrix')
plt.show()
```



	precision	recall	f1-score	support
negative	0.57	0.94	0.71	17
positive	0.50	0.08	0.13	13
accuracy			0.57	30
macro avg	0.54	0.51	0.42	30
weighted avg	0.54	0.57	0.46	30



LATENT DIRICHLET ALLOCATION

```
from sklearn.decomposition import LatentDirichletAllocation
from sklearn.feature_extraction.text import CountVectorizer
from nltk.corpus import stopwords # Import stopwords from NLTK

# Get Indonesian stopwords from NLTK
indonesian_stopwords = stopwords.words('indonesian')

# Create a CountVectorizer with Indonesian stopwords
```



```
vectorizer = CountVectorizer(max_df=0.95, min_df=2, stop_words=indonesian_stopwords)
dtm = vectorizer.fit_transform(df_gojek_2024_head['stemming'])

# Create a LatentDirichletAllocation model
LDA = LatentDirichletAllocation(n_components=5, random_state=42) # You can adjust n_components

# Fit the model to the document-term matrix
LDA.fit(dtm)

# Print the top words for each topic
for index, topic in enumerate(LDA.components_):
    print(f"Top 15 words for Topic #{index}")
    print([vectorizer.get_feature_names_out()[i] for i in topic.argsort()[-15:]])
    print("\n")

➡ Top 15 words for Topic #0
['order', 'yg', 'gojek', 'gofood', 'aja', 'customer', 'ga', 'udah', 'tolong', 'gak', 'pakai', 'bayan', 'aplikasi', 'nya', 'driver']

Top 15 words for Topic #1
['banget', 'gabisa', 'nomor', 'promo', 'iklan', 'pesan', 'biaya', 'ongkir', 'makan', 'driver', 'nya', 'bagus', 'gojek', 'gak', 'aplikasi']

Top 15 words for Topic #2
['batal', 'sengaja', 'makan', 'tp', 'muncul', 'buka', 'gofood', 'bayan', 'nggak', 'aplikasi', 'driver', 'order', 'menu', 'yg', 'pesan']

Top 15 words for Topic #3
['rugi', 'ga', 'drivernya', 'jemput', 'biaya', 'banget', 'kalo', 'hemat', 'jarak', 'tumpang', 'bgt', 'layan', 'gojek', 'yg', 'driver']

Top 15 words for Topic #4
['tp', 'drivernya', 'gak', 'tolong', 'masuk', 'yg', 'beda', 'maps', 'kali', 'susah', 'langgan', 'aplikasi', 'order', 'titik', 'driver']

/usr/local/lib/python3.11/dist-packages/sklearn/feature_extraction/text.py:402: UserWarning: Your stop_words may be inconsistent with your preprocessing. Tokenizing the stop words generate
warnings.warn(

```

▼ Tugas Topic Modelling Menggunakan LDA

Pada tugas ini, saya diminta untuk menerapkan algoritma **Latent Dirichlet Allocation (LDA)** untuk melakukan **topic modelling** terhadap kumpulan dokumen teks. Secara umum, LDA akan membentuk sejumlah topik berdasarkan distribusi kata dalam dokumen. Namun, selain menentukan topik secara langsung, saya juga diminta untuk melakukan **penentuan jumlah topik (komponen)** yang optimal dengan cara berikut:

1. **Lakukan tuning jumlah topik (n-topik)**, misalnya dari 2 hingga 10 topik.
2. Untuk setiap jumlah topik, latih model LDA dan hitung **coherence score**-nya.
3. Bandingkan nilai coherence score dari setiap model.
4. Pilih model dengan **coherence score tertinggi** sebagai model terbaik.
5. Buat interpretasi dari topik-topik yang dihasilkan oleh model terbaik tersebut.

**Catatan:** Coherence score digunakan sebagai metrik evaluasi kualitas model topic modelling. Semakin tinggi nilainya, semakin baik kualitas pemisahan dan interpretabilitas topik yang dihasilkan.

**Kesimpulan akhir yang harus disampaikan dalam laporan:**

Berdasarkan model LDA dengan jumlah topik terbaik (dengan coherence score tertinggi), topik-topik yang dihasilkan adalah sebagai berikut... *(diikuti interpretasi topik)*

```
...
Disini digunakan
#!pip install gensim==4.3.3
#!pip uninstall numpy==1.26.4
#!pip install numpy==1.24.3
...
```

▼ Pengerjaan Tugas

```
from gensim import corpora
from gensim.models import LdaModel, CoherenceModel
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
import nltk
import matplotlib.pyplot as plt

nltk.download('punkt')
nltk.download('stopwords')

# Preprocessing: Tokenisasi, lowercasing, filtering stopwords
stop_words = set(stopwords.words('indonesian'))
texts = []

for text in df_gojek['content'].dropna():
    tokens = word_tokenize(text.lower())
    filtered = [word for word in tokens if word.isalpha() and word not in stop_words]
    texts.append(filtered)

# Membuat dictionary dan corpus untuk LDA
dictionary = corpora.Dictionary(texts)
corpus = [dictionary.doc2bow(text) for text in texts]

# Fungsi untuk menghitung coherence score dari berbagai jumlah topik
def compute_coherence_values(dictionary, corpus, texts, start, limit, step):
    coherence_values = []
    model_list = []
    for num_topics in range(start, limit, step):
        model = LdaModel(corpus=corpus, num_topics=num_topics, id2word=dictionary, random_state=42)
        model_list.append(model)
        coherencemodel = CoherenceModel(model=model, texts=texts, dictionary=dictionary, coherence='c_v')
        coherence_values.append(coherencemodel.get_coherence())
    return model_list, coherence_values

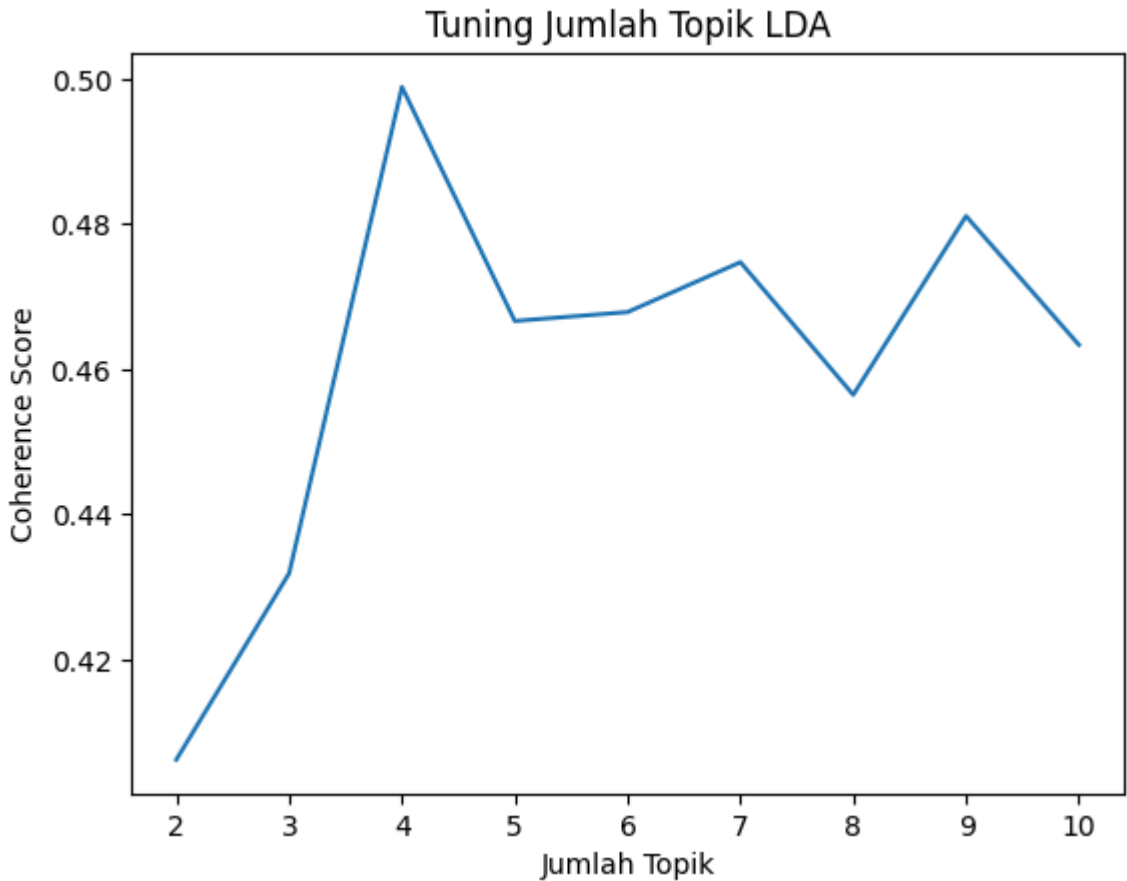
# Menentukan jumlah topik terbaik berdasarkan coherence score
start, limit, step = 2, 11, 1
model_list, coherence_values = compute_coherence_values(dictionary, corpus, texts, start, limit, step)

# Visualisasi coherence score
x = range(start, limit, step)
```

```
plt.plot(x, coherence_values)
plt.xlabel("Jumlah Topik")
plt.ylabel("Coherence Score")
plt.title("Tuning Jumlah Topik LDA")
plt.show()

# Menampilkan coherence tertinggi
best_index = coherence_values.index(max(coherence_values))
best_model = model_list[best_index]
print(f"Model terbaik memiliki {start + best_index} topik dengan coherence score {coherence_values[best_index]:.4f}")
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```



Model terbaik memiliki 4 topik dengan coherence score 0.4989

```
def show_top_words_cumulative_from_lda(lda_model, threshold=0.8):
    num_topics = lda_model.num_topics
    for topic_id in range(num_topics):
        topic_terms = lda_model.get_topic_terms(topicid=topic_id, topn=None) # Ambil semua term dari topik
        # Sortir berdasarkan bobot tertinggi
        topic_terms_sorted = sorted(topic_terms, key=lambda x: -x[1])

        total_weight = sum(weight for _, weight in topic_terms_sorted)
        cumulative_weight = 0.0
        selected_words = []

        for word_id, weight in topic_terms_sorted:
            word = lda_model.id2word[word_id]
            cumulative_weight += weight
            selected_words.append(word)
            if cumulative_weight >= threshold * total_weight:
                break

        print(f"\nTop words (cumulative ≥ {int(threshold*100)}%) for Topic #{topic_id}:")
        print(selected_words)
```

```
show_top_words_cumulative_from_lda(best_model, threshold=0.8)
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Top words (cumulative ≥ 80%) for Topic #0:
['gopay', 'akun', 'gojek', 'gak', 'ya', 'ga', 'aplikasi', 'nya', 'masuk', 'tolong', 'saldo', 'udah', 'uang', 'upgrade', 'blokir', 'mohon', 'hp', 'nomer', 'gk', 'transfer', 'knpa', 'aja', '

Top words (cumulative ≥ 80%) for Topic #1:
['gojek', 'driver', 'membantu', 'aplikasi', 'yg', 'nya', 'bagus', 'cepat', 'mudah', 'kasih', 'semoga', 'pelayanan', 'terima', 'mantap', 'transportasi', 'ramah', 'drivernya', 'terimakasih',

Top words (cumulative ≥ 80%) for Topic #2:
['yg', 'driver', 'smpe', 'sy', 'gojek', 'sya', 'gk', 'sma', 'chat', 'go', 'pesan', 'cancel', 'tdk', 'tidk', 'tmn', 'pesanan', 'ya', 'order', 'tman', 'tpi', 'gofood', 'klo', 'bs', 'msih', '

Top words (cumulative ≥ 80%) for Topic #3:
['nya', 'aja', 'ga', 'gojek', 'gak', 'aplikasi', 'promo', 'ya', 'udah', 'mahal', 'voucher', 'gk', 'pake', 'banget', 'pesen', 'gofood', 'update', 'yg', 'kalo', 'apk', 'go', 'diskon', 'harga
```

### Interpretasi

Analisis **LDA (Latent Dirichlet Allocation)** mengelompokkan kata-kata dalam dokumen menjadi beberapa topik dominan berdasarkan distribusi kata. Berdasarkan kata-kata yang paling sering muncul dalam masing-masing topik, interpretasi topik dapat dirangkum sebagai berikut. *Topik 0* merepresentasikan keluhan pengguna terhadap GoPay dan berbagai masalah teknis aplikasi. Ciri khasnya adalah banyaknya kata bermuatan negatif seperti gagal, blokir, pending, hilang, komplain, dan error, dengan isu utama meliputi saldo yang bermasalah, akun yang diblokir, layanan pelanggan yang kurang responsif, serta gangguan teknis seperti error pada login atau kode OTP. *Topik 1* mencerminkan pengalaman positif pengguna terhadap layanan transportasi Gojek, terutama dalam interaksi dengan driver. Topik ini ditandai dengan kata-kata apresiatif seperti mantap, bagus, nyaman, cepat, dan terbaik, serta berfokus pada pelayanan driver, kendaraan, dan aktivitas harian seperti pergi ke sekolah atau tempat kerja. *Topik 2* berhubungan dengan minat pengguna terhadap promo, diskon, dan harga layanan, dengan kata-kata dominan seperti promo, diskon, harga, voucher, dan kode. Keluhan terhadap harga yang mahal atau tidak wajar juga muncul, disertai ungkapan apresiasi terhadap potongan harga atau keluhan ketika promo berakhir. Terakhir, *Topik 3* menggambarkan pengalaman pengguna dalam menggunakan layanan GoFood, dengan kata-kata yang berkaitan dengan makanan dan pengantaran seperti gofood, resto, pesanan, antar, dan kurir. Masalah yang sering muncul meliputi kualitas makanan yang dingin atau basi, pesanan yang salah, keterlambatan pengantaran, serta kurangnya respons dari restoran. Keempat topik ini menunjukkan spektrum pengalaman pengguna terhadap layanan Gojek, dari aspek teknis dan finansial hingga kepuasan layanan transportasi dan makanan.