Twitter Climate Change Analysis

Scrapping Data

```
#Twitter Scrapping

twitter_auth_token = '71166115989888b6257c18370ee74abdcde5327c' #
change this auth token

# Import required Python package
!pip install pandas

# Install Node.js (because tweet-harvest built using Node.js)
!sudo apt-get update
!sudo apt-get install -y ca-certificates curl gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-repo.gpg.key
| sudo gpg --dearmor -o /etc/apt/keyrings/nodesource.gpg

!NODE_MAJOR=20 && echo "deb
[signed-by=/etc/apt/keyrings/nodesource.gpg]
https://deb.nodesource.com/node_$NODE_MAJOR.x nodistro main" | sudo
tee /etc/apt/sources.list.d/nodesource.list

!sudo apt-get update
!sudo apt-get install nodejs -y

!node -v
```

```
Requirement already satisfied: pandas in
/usr/local/lib/python3.10/dist-packages (2.1.4)
Requirement already satisfied: numpy<2,>=1.22.4 in
/usr/local/lib/python3.10/dist-packages (from pandas) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.10/dist-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.10/dist-packages (from pandas) (2024.2)
Requirement already satisfied: tzdata>=2022.1 in
/usr/local/lib/python3.10/dist-packages (from pandas) (2024.1)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2-
>pandas) (1.16.0)
Get:1 http://security.ubuntu.com/ubuntu jammy-security InRelease [129
kB]
Get:2 https://cloud.r-project.org/bin/linux/ubuntu jammy-cran40/
InRelease [3,626 B]
Hit:3
https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/
```

```
x86_64  InRelease
Hit:4 http://archive.ubuntu.com/ubuntu jammy InRelease
Ign:5 https://r2u.stat.illinois.edu/ubuntu jammy InRelease
Get:6 http://archive.ubuntu.com/ubuntu jammy-updates InRelease [128
kB]
Get:7 https://r2u.stat.illinois.edu/ubuntu jammy Release [5,713 B]
Get:8 https://r2u.stat.illinois.edu/ubuntu jammy Release.gpg [793 B]
Get:9 http://archive.ubuntu.com/ubuntu jammy-backports InRelease [127
kB]
Hit:10 https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu jammy
InRelease
Get:11 http://security.ubuntu.com/ubuntu jammy-security/universe amd64
Packages [1,156 kB]
Hit:12 https://ppa.launchpadcontent.net/graphics-drivers/ppa/ubuntu
jammy InRelease
Get:13 https://r2u.stat.illinois.edu/ubuntu jammy/main amd64 Packages
[2,584 kB]
Hit:14 https://ppa.launchpadcontent.net/ubuntugis/ppa/ubuntu jammy
InRelease
Get:15 http://security.ubuntu.com/ubuntu jammy-security/main amd64
Packages [2,318 kB]
Get:16 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64
Packages [1,445 kB]
Get:17 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64
Packages [2,595 kB]
Get:18 http://archive.ubuntu.com/ubuntu jammy-backports/universe amd64
Packages [33.7 kB]
Get:19 http://archive.ubuntu.com/ubuntu jammy-backports/main amd64
Packages [81.4 kB]
Get:20 https://r2u.stat.illinois.edu/ubuntu jammy/main all Packages
[8,352 kB]
Fetched 19.0 MB in 2s (7,942 kB/s)
Reading package lists... Done
W: Skipping acquire of configured file 'main/source/Sources' as
repository 'https://r2u.stat.illinois.edu/ubuntu jammy InRelease' does
not seem to provide it (sources.list entry misspelt?)
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
ca-certificates is already the newest version (20240203~22.04.1).
curl is already the newest version (7.81.0-1ubuntu1.18).
gnupg is already the newest version (2.2.27-3ubuntu2.1).
gnupg set to manually installed.
0 upgraded, 0 newly installed, 0 to remove and 54 not upgraded.
deb [signed-by=/etc/apt/keyrings/nodesource.gpg]
https://deb.nodesource.com/node_20.x nodistro main
Hit:1 http://archive.ubuntu.com/ubuntu jammy InRelease
Hit:2 http://security.ubuntu.com/ubuntu jammy-security InRelease
Hit:3 https://cloud.r-project.org/bin/linux/ubuntu jammy-cran40/
```

```
InRelease
Hit:4 http://archive.ubuntu.com/ubuntu jammy-updates InRelease
Hit:5 http://archive.ubuntu.com/ubuntu jammy-backports InRelease
Hit:6
https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/
x86_64  InRelease
Get:7 https://deb.nodesource.com/node_20.x nodistro InRelease [12.1
kB]
Ign:8 https://r2u.stat.illinois.edu/ubuntu jammy InRelease
Hit:9 https://r2u.stat.illinois.edu/ubuntu jammy Release
Hit:10 https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu jammy
InRelease
Hit:11 https://ppa.launchpadcontent.net/graphics-drivers/ppa/ubuntu
jammy InRelease
Hit:12 https://ppa.launchpadcontent.net/ubuntugis/ppa/ubuntu jammy
InRelease
Get:13 https://deb.nodesource.com/node_20.x nodistro/main amd64
Packages [8,901 B]
Fetched 21.0 kB in 1s (16.8 kB/s)
Reading package lists... Done
W: Skipping acquire of configured file 'main/source/Sources' as
repository 'https://r2u.stat.illinois.edu/ubuntu jammy InRelease' does
not seem to provide it (sources.list entry misspelt?)
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following NEW packages will be installed:
  nodejs
0 upgraded, 1 newly installed, 0 to remove and 54 not upgraded.
Need to get 31.7 MB of archives.
After this operation, 197 MB of additional disk space will be used.
Get:1 https://deb.nodesource.com/node_20.x nodistro/main amd64 nodejs
amd64 20.17.0-1nodesource1 [31.7 MB]
Fetched 31.7 MB in 1s (56.1 MB/s)
debconf: unable to initialize frontend: Dialog
debconf: (No usable dialog-like program is installed, so the dialog
based frontend cannot be used. at
/usr/share/perl5/Debconf/FrontEnd/Dialog.pm line 78, <> line 1.)
debconf: falling back to frontend: Readline
debconf: unable to initialize frontend: Readline
debconf: (This frontend requires a controlling tty.)
debconf: falling back to frontend: Teletype
dpkg-preconfigure: unable to re-open stdin:
Selecting previously unselected package nodejs.
(Reading database ... 123614 files and directories currently
installed.)
Preparing to unpack .../nodejs_20.17.0-1nodesource1_amd64.deb ...
Unpacking nodejs (20.17.0-1nodesource1) ...
Setting up nodejs (20.17.0-1nodesource1) ...
```

```
Processing triggers for man-db (2.10.2-1) ...
v20.17.0

# Crawl Data

filename = 'twitter_data.csv'
search_keyword = 'climate change since:2024-01-01 until:2024-09-20
lang:en'
limit = 1000

!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --
tab "LATEST" -l {limit} --token {twitter_auth_token}

npm warn deprecated rimraf@3.0.2: Rimraf versions prior to v4 are no
longer supported
npm warn deprecated inflight@1.0.6: This module is not supported, and
leaks memory. Do not use it. Check out lru-cache if you want a good
and tested way to coalesce async requests by a key value, which is
much more comprehensive and powerful.
npm warn deprecated glob@7.2.3: Glob versions prior to v9 are no
longer supported
Tweet Harvest [v2.6.1]

Research by Helmi Satria
Use it for Educational Purposes only!

This script uses Chromium Browser to crawl data from Twitter with your
Twitter auth token.
Please enter your Twitter auth token when prompted.

Note: Keep your access token secret! Don't share it with anyone else.
Note: This script only runs on your local device.


Opening twitter search page...


-- Scrolling... (1) (2) (3)
Filling in keywords: climate change since:2024-01-01 until:2024-09-20
lang:en

 (4)Created new directory: /content/tweets-data


Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 19

-- Scrolling... (1) (2) (3)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 38
```

```
-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 58

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 78

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 98

-- Scrolling... (1) (2) (3)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 117

--Taking a break, waiting for 10 seconds...

-- Scrolling... (1) (2) (3)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 135

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 155

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 173

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 191

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 209

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
```

```
Total tweets saved: 228

--Taking a break, waiting for 10 seconds...

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 248

-- Scrolling... (1) (2) (3)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 267

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 286

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 305

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 324

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 343

--Taking a break, waiting for 10 seconds...

-- Scrolling... (1) (2) (3)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 362

-- Scrolling... (1) (2) (3)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 382

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 402
```

```
-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 421

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 440

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 459

--Taking a break, waiting for 10 seconds...

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 477

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 495

-- Scrolling... (1) (2) (3) (4) (5) (6) (7) (8) (9) (10) (11) (12)
(13) (14) (15) (16) (17) (18) (19) (20) (21)No more tweets found,
please check your search criteria and csv file result
Timeout reached 1 times, making sure again...


Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 514

-- Scrolling... (1) (2) (3)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 533
[v2.6.1] Error parsing response json:
{"_type":"Response","_guid":"response@4edf5a80a720e36b3bd1ac073a8d1533
"}
[v2.6.1] Most likely, you have already exceeded the Twitter rate
limit. Read more on https://x.com/elonmusk/status/1675187969420828672.
[v2.6.1] Error parsing response json:
{"_type":"Response","_guid":"response@55e89b95e81f1c4ca0007cc500d19cdb
"}
[v2.6.1] Most likely, you have already exceeded the Twitter rate
limit. Read more on https://x.com/elonmusk/status/1675187969420828672.
[v2.6.1] Error parsing response json:
```

{"_type":"Response","_guid":"response@54cfb9c28f7623f210a0464bccaaeaf8
"}
[v2.6.1] Most likely, you have already exceeded the Twitter rate limit. Read more on https://x.com/elonmusk/status/1675187969420828672.
[v2.6.1] Error parsing response json:
{"_type":"Response","_guid":"response@77995c38ab8faf0e5d63dfea018b825b
"}
[v2.6.1] Most likely, you have already exceeded the Twitter rate limit. Read more on https://x.com/elonmusk/status/1675187969420828672.
[v2.6.1] Error parsing response json:
{"_type":"Response","_guid":"response@38f39669d5d00b79247ac5b969f77196
"}
[v2.6.1] Most likely, you have already exceeded the Twitter rate limit. Read more on https://x.com/elonmusk/status/1675187969420828672.
[v2.6.1] Error parsing response json:
{"_type":"Response","_guid":"response@b06f16acd110e0c9c0a67be80bc04740
"}
[v2.6.1] Most likely, you have already exceeded the Twitter rate limit. Read more on https://x.com/elonmusk/status/1675187969420828672.
[v2.6.1] Error parsing response json:
{"_type":"Response","_guid":"response@926f4444577656e7ae5b875e28c45694
"}
[v2.6.1] Most likely, you have already exceeded the Twitter rate limit. Read more on https://x.com/elonmusk/status/1675187969420828672.
[v2.6.1] Error parsing response json:
{"_type":"Response","_guid":"response@9e30bf366749f523e44478d9f7c2c371
"}
[v2.6.1] Most likely, you have already exceeded the Twitter rate limit. Read more on https://x.com/elonmusk/status/1675187969420828672.
[v2.6.1] Error parsing response json:
{"_type":"Response","_guid":"response@d0c73df767a3580cb8473ab4dcee73b2
"}
[v2.6.1] Most likely, you have already exceeded the Twitter rate limit. Read more on https://x.com/elonmusk/status/1675187969420828672.
[v2.6.1] Error parsing response json:
{"_type":"Response","_guid":"response@a085be4552e22f60e0491c4ef5d7aa3d
"}
[v2.6.1] Most likely, you have already exceeded the Twitter rate limit. Read more on https://x.com/elonmusk/status/1675187969420828672.
[v2.6.1] Error parsing response json:
{"_type":"Response","_guid":"response@7b7e5a65a2cf74628002db919698ac71
"}
[v2.6.1] Most likely, you have already exceeded the Twitter rate limit. Read more on https://x.com/elonmusk/status/1675187969420828672.

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 550

-- Scrolling... (1)

```
Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 568

--Taking a break, waiting for 10 seconds...

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 586

-- Scrolling... (1) (2) (3)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 603

-- Scrolling... (1) (2) (3)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 623

-- Scrolling... (1) (2) (3)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 643

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 661

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 681

--Taking a break, waiting for 10 seconds...

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 701

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 721

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
```

```
Total tweets saved: 741

-- Scrolling... (1) (2) (3)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 758

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 778

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 797

--Taking a break, waiting for 10 seconds...

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 817

-- Scrolling... (1) (2) (3)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 836

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 854

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 874

-- Scrolling... (1) (2)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 893

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 912

--Taking a break, waiting for 10 seconds...
```

```
-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 930

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 950

-- Scrolling... (1) (2)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 970

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 988

-- Scrolling... (1) (2)

Your tweets saved to: /content/tweets-data/twitter_data.csv
Total tweets saved: 1007
Got 1007 tweets, done scrolling...
```

Import Scrapping Data From CSV

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

# Load the data
df = pd.read_csv('/content/climate_data_with_sentiment.csv')
df.head()
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 1007,\n  \"fields\":
[\n    {\n      \"column\": \"conversation_id_str\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
19233118420585200,\n        \"min\": 1419684030279299075,\n
\"max\": 1836918011514560715,\n        \"num_unique_values\": 835,\n
\"samples\": [\n          1836498727474991152,\n
1836811179802395052,\n          1836890934186758491\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"created_at\",\n
\"properties\": {\n        \"dtype\": \"object\",\n
\"num_unique_values\": 953,\n        \"samples\": [\n          \"Thu
Sep 19 22:40:31 +0000 2024\",\n          \"Thu Sep 19 21:39:04 +0000
2024\",\n          \"Thu Sep 19 20:38:07 +0000 2024\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
```

```
n    },\n    {\n      \"column\": \"favorite_count\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
21,\n        \"min\": 0,\n        \"max\": 478,\n
\"num_unique_values\": 44,\n        \"samples\": [\n          78,\n
163,\n          16\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"full_text\",\n      \"properties\": {\n        \"dtype\":
\"string\",\n        \"num_unique_values\": 996,\n        \"samples\":
[\n          \"@thejournal_ie Another propaganda piece designed to
scare the public. That is domestic terrorism - a crime which the fake
journalists working for this woke rag will not escape the consequences
of. Man-made CO2-driven climate change is a HOAX Apologise now then
stop publishing these lies\",\n          \"The main two scientific
references stated @guardian &amp;amp; @LievenAnatol that #AMOC is
currently weakening are in strong question due to new Nature
publ: ...revealing the #Florida Current has remained remarkably
stable... #ClimateCrisis #climatechange https://t.co/o76oTQV9n9
https://t.co/0RYPTv1XVS\",\n          \"@SalmaKe_mohamed @abierkhatib
No it isn t dying cause of Israel Where you pulled that from ? The
world is a mess cause cultures are clashing do you not agree? Most
wars are started because of religion And to blame one minority for it
is absolutely absurd We are vs all to blame for climate change\"\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"id_str\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": 18489385896311,\n
\"min\": 1836853313238630655,\n        \"max\": 1836918137469472957,\n
\"num_unique_values\": 1007,\n        \"samples\": [\n
1836857237664317940,\n          1836870675581063669,\n
1836868721945538775\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"image_url\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 147,\n
\"samples\": [\n
\"https://pbs.twimg.com/media/GX3U2CsWIAAGgdm.jpg\",\n
\"https://pbs.twimg.com/media/GX3uHi_WUAAAVK9.jpg\",\n
\"https://pbs.twimg.com/media/GX3SGHTWgAAmKdx.jpg\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"in_reply_to_screen_name\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 474,\n        \"samples\": [\n
\"knows_jill\",\n          \"CP24\",\n          \"susanmcgraw88\"\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"lang\",\n      \"properties\":
{\n        \"dtype\": \"category\",\n        \"num_unique_values\":
1,\n        \"samples\": [\n          \"en\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"location\",\n      \"properties\":
{\n        \"dtype\": \"category\",\n        \"num_unique_values\":
450,\n        \"samples\": [\n          \"Michigan, USA\"\n        ],\
```

```
n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"quote_count\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
0,\n        \"min\": 0,\n        \"max\": 9,\n
\"num_unique_values\": 6,\n        \"samples\": [\n          0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"reply_count\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
1,\n        \"min\": 0,\n        \"max\": 38,\n
\"num_unique_values\": 15,\n        \"samples\": [\n          12\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"retweet_count\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
8,\n        \"min\": 0,\n        \"max\": 245,\n
\"num_unique_values\": 23,\n        \"samples\": [\n          17\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"tweet_url\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 1007,\n        \"samples\": [\n
\"https://x.com/leehayward1970/status/1836857237664317940\"\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"user_id_str\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 742144484590418176,\n        \"min\":
6644462,\n        \"max\": 1836873818884788224,\n
\"num_unique_values\": 883,\n        \"samples\": [\n
1833126745643302914\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"username\",\n      \"properties\": {\n        \"dtype\":
\"string\",\n        \"num_unique_values\": 883,\n        \"samples\":
[\n          \"kv0th3_kk\"\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"sentiment_label\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 3,\n
\"samples\": [\n          \"positive\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"df"}
```


```
df.isnull().sum()
```

```
conversation_id_str          0
created_at                   0
favorite_count               0
full_text                    0
id_str                       0
image_url                  859
in_reply_to_screen_name    353
lang                         0
location                   408
quote_count                  0
```

```
reply_count                 0
retweet_count               0
tweet_url                   0
user_id_str                 0
username                    0
sentiment_label             0
dtype: int64
```

Data Preprocessing

```
import re
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
import nltk

# Download necessary NLTK resources
nltk.download('stopwords')
nltk.download('wordnet')

# Initialize preprocessing tools
lemmatizer = WordNetLemmatizer()
stop_words = set(stopwords.words('english'))

# Function for preprocessing text
def preprocess_text(text):
    # Remove special characters and convert to lowercase
    text = re.sub(r'[^a-zA-Z\s]', '', text, re.I | re.A)
    text = text.lower()

    # Tokenization and removal of stopwords
    tokens = text.split()
    tokens = [lemmatizer.lemmatize(word) for word in tokens if word
not in stop_words]

    return ' '.join(tokens)

# Apply preprocessing to the 'full_text' column
df['cleaned_text'] = df['full_text'].apply(preprocess_text)

# Display the first few rows to confirm the preprocessing
df[['full_text', 'cleaned_text']].head()
```

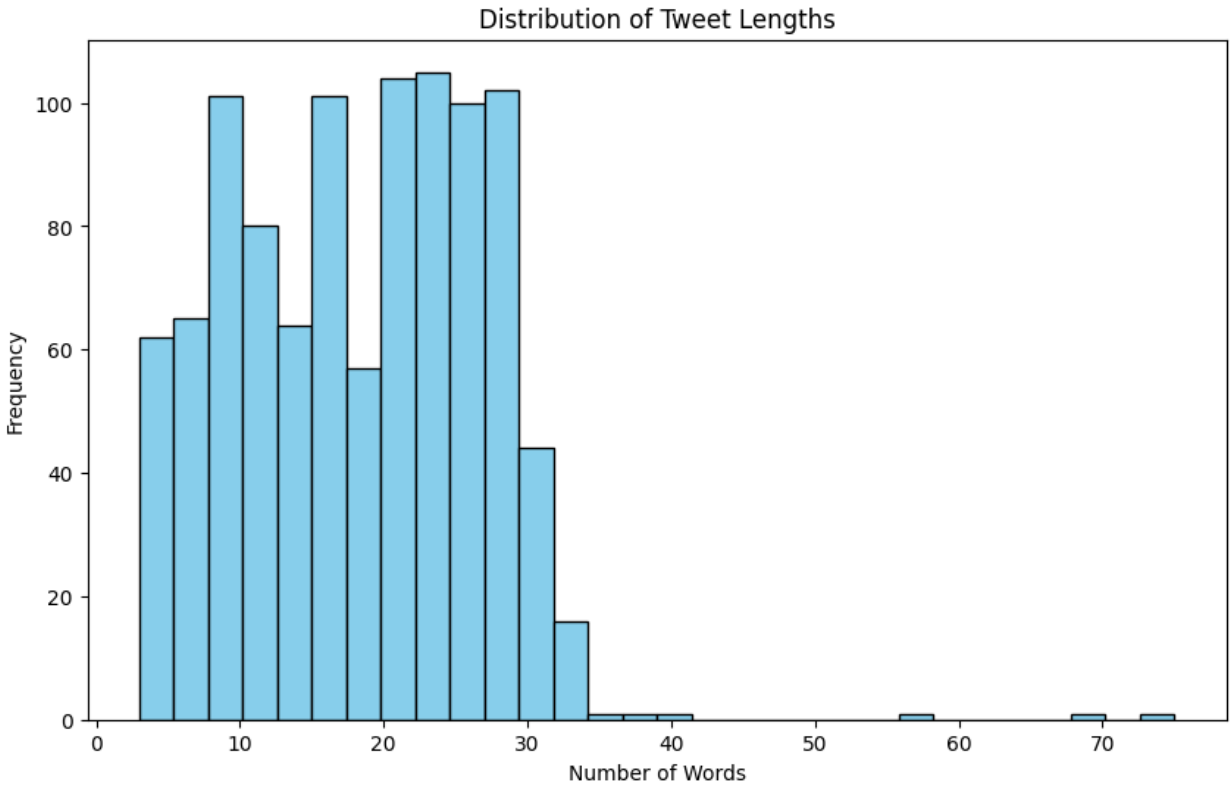
```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
```


```
{"summary":"{\n  \"name\": \"df[['full_text', 'cleaned_text']]\",\n
\"rows\": 5,\n  \"fields\": [\n    {\n      \"column\":
\"full_text\",\n      \"properties\": {\n        \"dtype\":
```

```
\"string\",\n        \"num_unique_values\": 5,\n        \"samples\":
[\n          \"@elonmusk Trump promised to increase extraction use
&amp;amp; sale of fossil fuels calling climate change a hoax. Bye bye
Shanghai\",\n          \"It's past time to #MakePollutersPay. For
years Big Polluters have covered up the truth that fossil fuels are
causing climate change but now we have the chance to hold them
accountable. I signed @EvergreenAction's petition and hope you will
too! https://t.co/dEorCAsix2\",\n          \"@Tim_Walz Gas? What about
climate change?\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"cleaned_text\",\n      \"properties\": {\n        \"dtype\":
\"string\",\n        \"num_unique_values\": 5,\n        \"samples\":
[\n          \"elonmusk trump promised increase extraction use amp
sale fossil fuel calling climate change hoax bye bye shanghai\",\n
\"past time makepolluterspay year big polluter covered truth fossil
fuel causing climate change chance hold accountable signed
evergreenactions petition hope httpstcodeorcasix\",\n
\"timwalz gas climate change\"\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe"}
```


EDA

```
# Add a column for the length of each tweet (number of words)
df['text_length'] = df['cleaned_text'].apply(lambda x: len(x.split()))

# Plot the distribution of tweet lengths
plt.figure(figsize=(10, 6))
plt.hist(df['text_length'], bins=30, color='skyblue',
edgecolor='black')
plt.title('Distribution of Tweet Lengths')
plt.xlabel('Number of Words')
plt.ylabel('Frequency')
plt.show()
# Display some summary statistics of tweet lengths
print("\nSummary Statistics of Tweet Lengths:")
print(df['text_length'].describe())
```



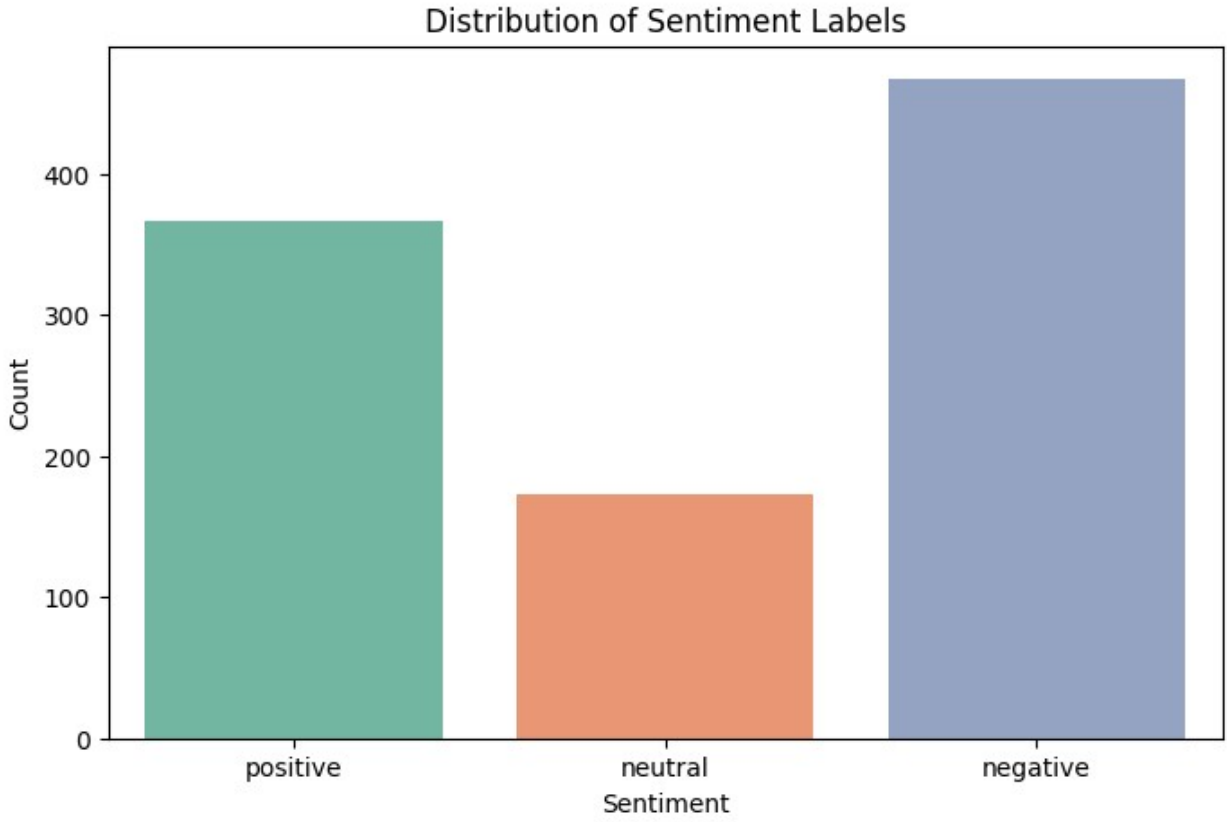

```
Summary Statistics of Tweet Lengths:
count    1007.000000
mean       18.149950
std         8.465833
min         3.000000
25%        11.000000
50%        19.000000
75%        25.000000
max        75.000000
Name: text_length, dtype: float64
```

```
import seaborn as sns

# Visualize the count of each sentiment label
plt.figure(figsize=(8, 5))
sns.countplot(x='sentiment_label', data=df, palette='Set2')
plt.title('Distribution of Sentiment Labels')
plt.xlabel('Sentiment')
plt.ylabel('Count')
plt.show()
```

```
<ipython-input-56-c1bcc2e5b915>:5: 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.

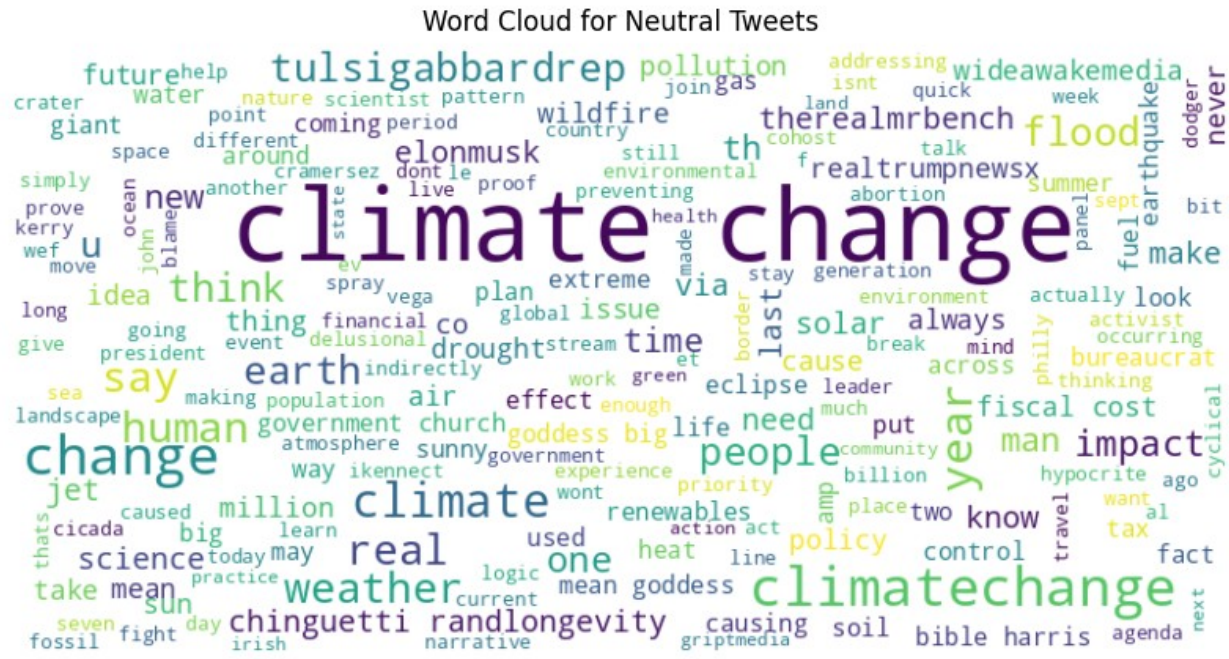
```
  sns.countplot(x='sentiment_label', data=df, palette='Set2')
```




```
from wordcloud import WordCloud

# Function to plot a word cloud for a given sentiment
def plot_wordcloud(sentiment):
    text = " ".join(df[df['sentiment_label'] == sentiment]
['cleaned_text'].tolist())
    wordcloud = WordCloud(width=800, height=400,
background_color='white').generate(text)

    plt.figure(figsize=(10, 6))
    plt.imshow(wordcloud, interpolation='bilinear')
    plt.title(f'Word Cloud for {sentiment.capitalize()} Tweets')
    plt.axis('off')
    plt.show()

# Generate word clouds for positive, negative, and neutral sentiments
plot_wordcloud('positive')
plot_wordcloud('negative')
plot_wordcloud('neutral')
```

Word Cloud for Positive Tweets

Word Cloud for Negative Tweets



TF-IDF Feature Extraction

```
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer

# Split data into training and test sets (80% train, 20% test)
X_train, X_test, y_train, y_test =
train_test_split(df['cleaned_text'], df['sentiment_label'],
test_size=0.2, random_state=42)

# Initialize the TF-IDF Vectorizer
tfidf_vectorizer = TfidfVectorizer(max_features=5000, ngram_range=(1,
2))

# Fit and transform the training data; transform the test data
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)
X_test_tfidf = tfidf_vectorizer.transform(X_test)

# Display the shape of the resulting feature matrices
print(f'Training data shape: {X_train_tfidf.shape}')
print(f'Test data shape: {X_test_tfidf.shape}')
```

```
Training data shape: (805, 5000)
Test data shape: (202, 5000)
```

Logistic Regression

```
from sklearn.linear_model import LogisticRegression

# Initialize Logistic Regression model
```

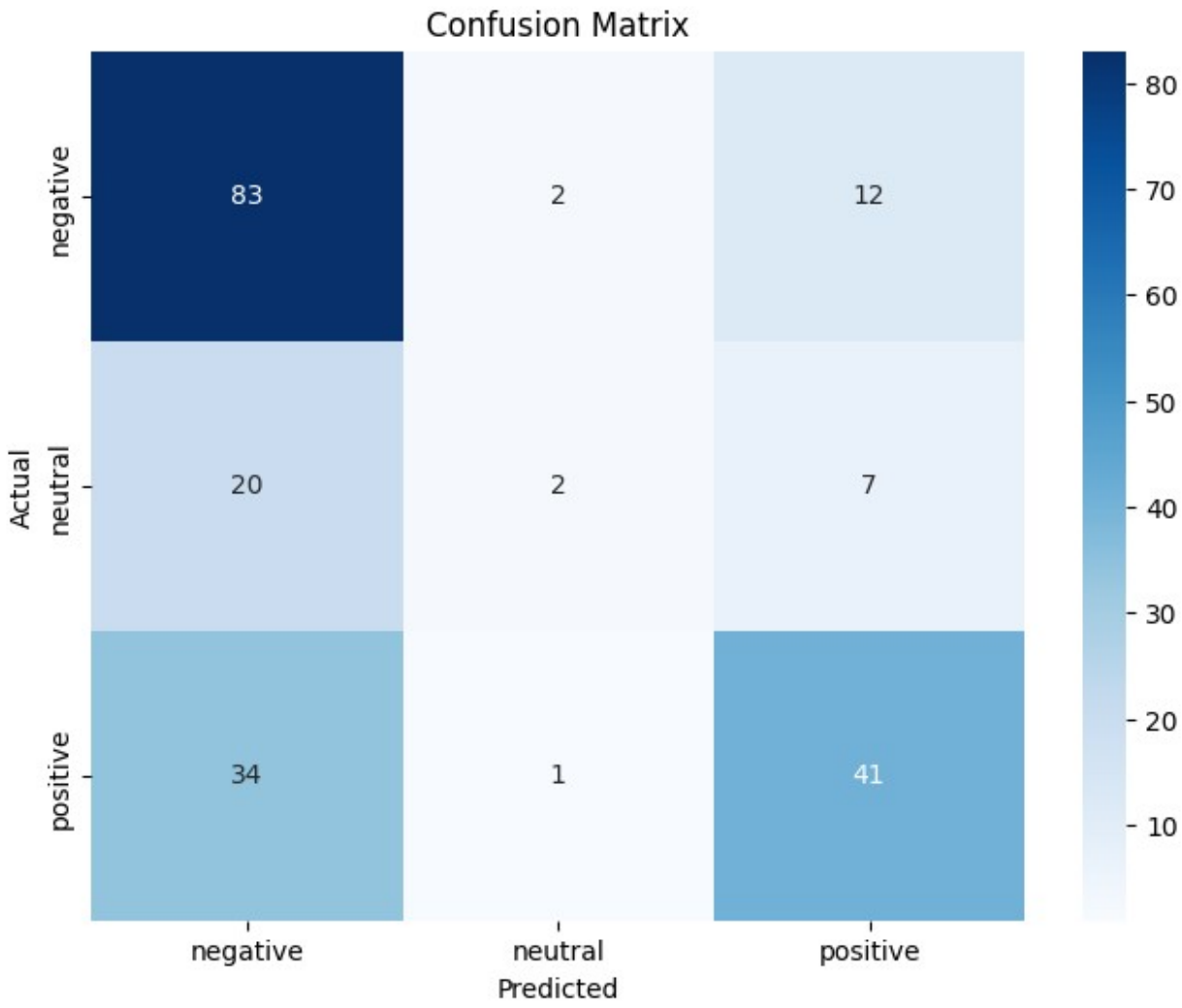
```
model = LogisticRegression(max_iter=200)

# Train the model using the training data
model.fit(X_train_tfidf, y_train)

# Print a message to indicate training completion
print("Model training completed.")
```

```
Model training completed.
```

```
from sklearn.metrics import classification_report, confusion_matrix,
accuracy_score

# Predict sentiment labels for the test data
y_pred = model.predict(X_test_tfidf)

# Evaluate model performance: Accuracy
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy * 100:.2f}%\n')

# Detailed classification report
print("Classification Report:")
print(classification_report(y_test, y_pred))
```

```
Accuracy: 62.38%

Classification Report:
              precision    recall  f1-score   support

    negative       0.61      0.86      0.71        97
     neutral       0.40      0.07      0.12        29
    positive       0.68      0.54      0.60        76

    accuracy                           0.62       202
   macro avg       0.56      0.49      0.48       202
weighted avg       0.61      0.62      0.58       202
```

```
# Confusion Matrix
conf_matrix = confusion_matrix(y_test, y_pred)

# Plot Confusion Matrix
plt.figure(figsize=(8, 6))
sns.heatmap(conf_matrix, annot=True, fmt="d", cmap="Blues",
xticklabels=model.classes_, yticklabels=model.classes_)
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()
```




```
# Calculate overall accuracy
accuracy_logistic = accuracy_score(y_test, y_pred)
print(f"Logistic Regression Accuracy: {accuracy_logistic * 100:.2f}%")

# Analisis kesalahan
error_rate = 1 - accuracy_logistic
print(f"Error Rate: {error_rate * 100:.2f}%")

print("\nPenjelasan Akurasi dan Kesalahan:")
if error_rate > 0.2:
    print("Model memiliki kesalahan tinggi yang mungkin disebabkan
oleh imbalance data atau fitur yang tidak cukup representatif.")
else:
    print("Kesalahan model rendah, tetapi masih bisa diperbaiki dengan
tuning parameter atau metode lain.")
```

```
Logistic Regression Accuracy: 62.38%
Error Rate: 37.62%
```

Penjelasan Akurasi dan Kesalahan:
Model memiliki kesalahan tinggi yang mungkin disebabkan oleh imbalance data atau fitur yang tidak cukup representatif.

```
# Find indices where the predictions are incorrect
incorrect_indices = np.where(y_test != y_pred)[0]

# Loop over some of the incorrectly predicted samples
print("\nBeberapa Contoh Prediksi Salah:")
for index in incorrect_indices[:10]:  # show only first 10 errors
    print(f"Teks: {X_test.iloc[index]}")
    print(f"Label Asli: {y_test.iloc[index]}, Prediksi:
{y_pred[index]}")
    print("-" * 60)
```

```
Beberapa Contoh Prediksi Salah:
Teks: wideawakemedia oops perhaps global cooling ah switched global
warming climate change clever bastard
Label Asli: negative, Prediksi: positive
------------------------------------------------------------
Teks: tulsigabbardrep think right world tell u climate change also
need clarify subscribe preaches climate change goal good people
private jet rich
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: george marshall spent year talking und understanding human acted
climate change also feel lucky count man one friend httpstconclmqjfwbr
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: volcaholic climate change fun p let jail climate protestors
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: kmbinch cant deny climate change agreed paris real needed dealt
whats changed since
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: ad gifted really thought provoking evening science speakeasy
scienceatlife climate change definitely taken lot away evening
heardatspeakeasy httpstcoqjkzcbue
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: wideawakemedia master universe bureaucrat minion say warrior
climate change really global tyranny
Label Asli: neutral, Prediksi: negative
------------------------------------------------------------
Teks: earth may ring system million yr ago httpstcomhbfxrbgoj gi
spatial mapping model modeling geology structuralgeology climatechange
paleoclimate ordovician asteroid rochelimit crater impactcrater crater
```

```
icehouse meteorite platetectonics httpstcotahguzyx
Label Asli: neutral, Prediksi: positive
------------------------------------------------------------
Teks: markevans realtrumpnewsx climate change threaten human health
increasing frequency intensity extreme weather event like storm flood
drought wildfire heatwaves
Label Asli: neutral, Prediksi: negative
------------------------------------------------------------
Teks: climate never got climate change narrative memo
Label Asli: neutral, Prediksi: negative
------------------------------------------------------------
```

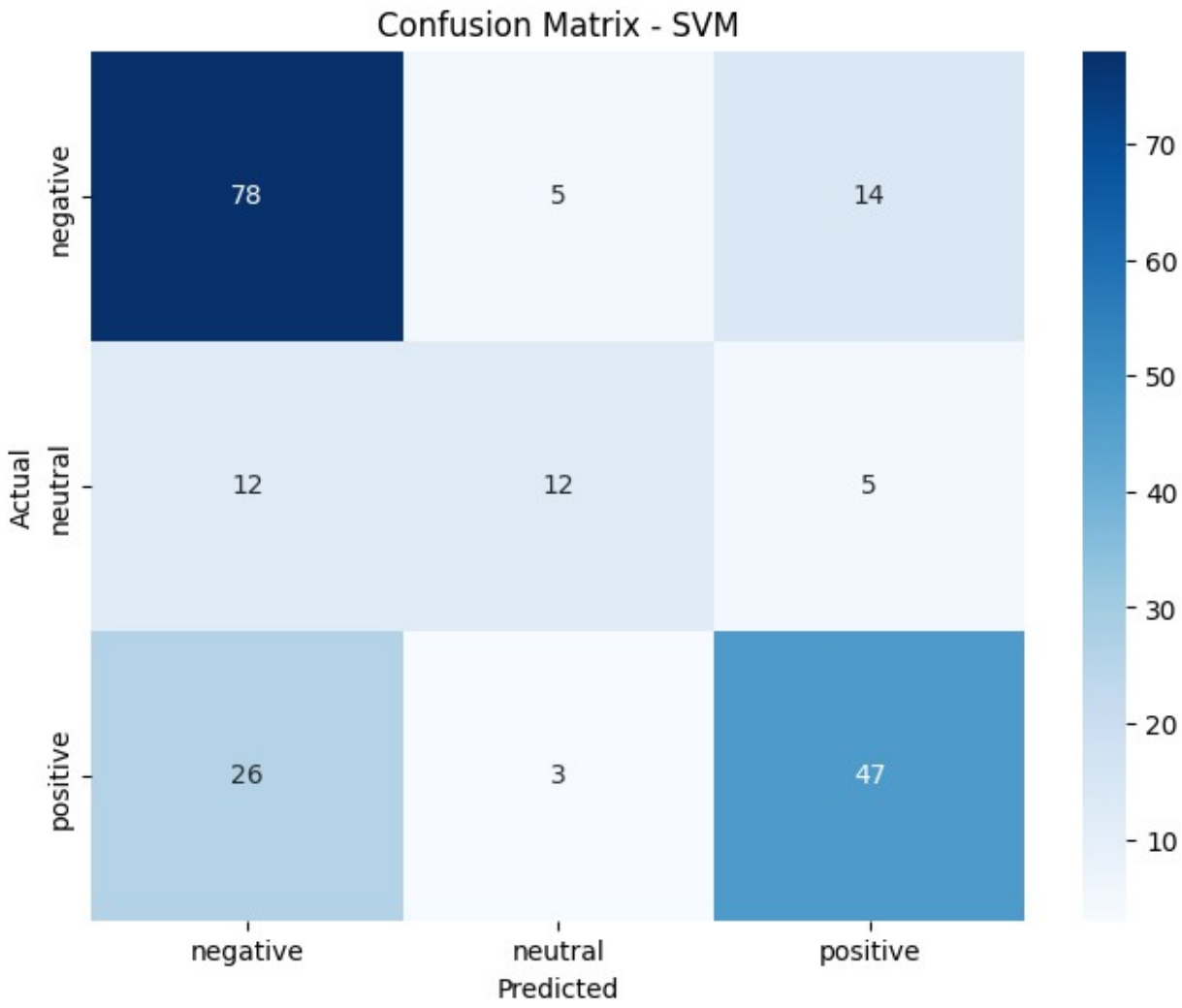
SVM

```
from sklearn.model_selection import GridSearchCV
from sklearn.svm import SVC

# Define parameter grid for SVM
param_grid = {
    'C': [0.1, 1, 10],  # Regularization parameter
    'kernel': ['linear', 'rbf'],  # Kernel type
    'gamma': ['scale', 'auto']  # Kernel coefficient
}

# Initialize SVM model
svm_model = SVC()

# Use GridSearchCV to find the best hyperparameters
grid_search = GridSearchCV(svm_model, param_grid, cv=3,
scoring='accuracy', n_jobs=-1)
grid_search.fit(X_train_tfidf, y_train)

# Get the best parameters and best score
best_params = grid_search.best_params_
best_score = grid_search.best_score_
print(f"Best Parameters: {best_params}")
print(f"Best Cross-Validation Accuracy: {best_score:.2f}")

Best Parameters: {'C': 1, 'gamma': 'scale', 'kernel': 'linear'}
Best Cross-Validation Accuracy: 0.57

# Initialize SVM model with best parameters
best_svm_model = SVC(C=best_params['C'], kernel=best_params['kernel'],
gamma=best_params['gamma'])

# Train the SVM model on the training data
best_svm_model.fit(X_train_tfidf, y_train)

SVC(C=1, kernel='linear')
```

```python
# Predict sentiment labels for the test data using the best SVM model
y_pred_svm = best_svm_model.predict(X_test_tfidf)

# Evaluate model performance: Accuracy
accuracy_svm = accuracy_score(y_test, y_pred_svm)
print(f'SVM Accuracy: {accuracy_svm * 100:.2f}%\n')

# Detailed classification report
print("SVM Classification Report:")
print(classification_report(y_test, y_pred_svm))
```

```
SVM Accuracy: 67.82%

SVM Classification Report:
              precision    recall  f1-score   support

    negative       0.67      0.80      0.73        97
     neutral       0.60      0.41      0.49        29
    positive       0.71      0.62      0.66        76

    accuracy                           0.68       202
   macro avg       0.66      0.61      0.63       202
weighted avg       0.68      0.68      0.67       202
```

```python
# Confusion Matrix for SVM
conf_matrix_svm = confusion_matrix(y_test, y_pred_svm)

# Plot Confusion Matrix for SVM
plt.figure(figsize=(8, 6))
sns.heatmap(conf_matrix_svm, annot=True, fmt="d", cmap="Blues",
xticklabels=best_svm_model.classes_,
yticklabels=best_svm_model.classes_)
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix - SVM')
plt.show()
```




```
# Calculate overall accuracy
accuracy_svm = accuracy_score(y_test, y_pred_svm)
print(f"SVM Accuracy: {accuracy_svm * 100:.2f}%")

# Analisis kesalahan
error_rate_svm = 1 - accuracy_svm
print(f"Error Rate: {error_rate_svm * 100:.2f}%")

print("\nPenjelasan Akurasi dan Kesalahan:")
if error_rate_svm > 0.2:
    print("Kesalahan yang tinggi bisa disebabkan oleh pemilihan
parameter kernel yang kurang tepat atau data yang tidak seimbang.")
else:
    print("Model cukup baik, namun masih bisa ditingkatkan dengan
hyperparameter tuning atau metode lain.")
```

```
SVM Accuracy: 67.82%
Error Rate: 32.18%
```

Penjelasan Akurasi dan Kesalahan:
Kesalahan yang tinggi bisa disebabkan oleh pemilihan parameter kernel yang kurang tepat atau data yang tidak seimbang.

```
# Find indices where the predictions are incorrect
incorrect_indices_svm = np.where(y_test != y_pred_svm)[0]

# Loop over some of the incorrectly predicted samples
print("\nBeberapa Contoh Prediksi Salah (SVM):")
for index in incorrect_indices_svm[:10]:
    print(f"Teks: {X_test.iloc[index]}")
    print(f"Label Asli: {y_test.iloc[index]}, Prediksi:
{y_pred_svm[index]}")
    print("-" * 60)
```

```
Beberapa Contoh Prediksi Salah (SVM):
Teks: wideawakemedia oops perhaps global cooling ah switched global
warming climate change clever bastard
Label Asli: negative, Prediksi: positive
------------------------------------------------------------
Teks: collapse ok climate change something people never control way
rob people blind
Label Asli: negative, Prediksi: positive
------------------------------------------------------------
Teks: tulsigabbardrep think right world tell u climate change also
need clarify subscribe preaches climate change goal good people
private jet rich
Label Asli: positive, Prediksi: neutral
------------------------------------------------------------
Teks: george marshall spent year talking und understanding human acted
climate change also feel lucky count man one friend httpstconclmqjfwbr
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: kmbinch cant deny climate change agreed paris real needed dealt
whats changed since
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: ad gifted really thought provoking evening science speakeasy
scienceatlife climate change definitely taken lot away evening
heardatspeakeasy httpstcoqjkzcbue
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: earth may ring system million yr ago httpstcomhbfxrbgoj gi
spatial mapping model modeling geology structuralgeology climatechange
paleoclimate ordovician asteroid rochelimit crater impactcrater crater
icehouse meteorite platetectonics httpstcotahguzyx
Label Asli: neutral, Prediksi: positive
------------------------------------------------------------
```

```
Teks: markevans realtrumpnewsx climate change threaten human health
increasing frequency intensity extreme weather event like storm flood
drought wildfire heatwaves
Label Asli: neutral, Prediksi: negative
------------------------------------------------------------
Teks: climate never got climate change narrative memo
Label Asli: neutral, Prediksi: negative
------------------------------------------------------------
Teks: vp inflation reduction act biden say named actually climate
change
Label Asli: neutral, Prediksi: negative
------------------------------------------------------------
```

Random Forest

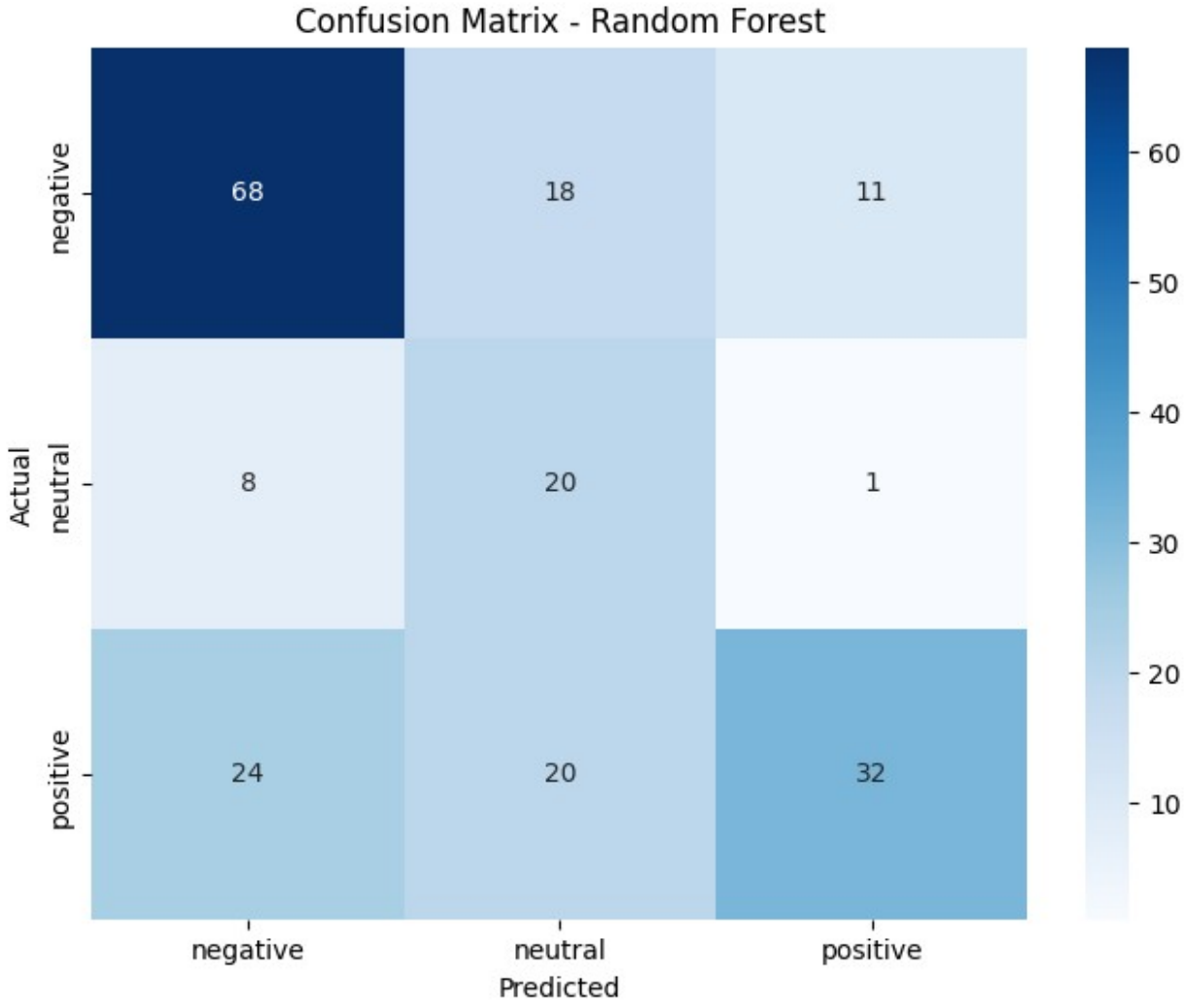
```
from sklearn.ensemble import RandomForestClassifier

# Initialize the Random Forest model
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)

# Train the model on the training data
rf_model.fit(X_train_tfidf, y_train)
```

```
RandomForestClassifier(random_state=42)
```

```
# Predict sentiment labels for the test data using the Random Forest
model
y_pred_rf = rf_model.predict(X_test_tfidf)

# Evaluate model performance: Accuracy
accuracy_rf = accuracy_score(y_test, y_pred_rf)
print(f'Random Forest Accuracy: {accuracy_rf * 100:.2f}%\n')

# Detailed classification report
print("Random Forest Classification Report:")
print(classification_report(y_test, y_pred_rf))

# Confusion Matrix for Random Forest
conf_matrix_rf = confusion_matrix(y_test, y_pred_rf)

# Plot Confusion Matrix for Random Forest
plt.figure(figsize=(8, 6))
sns.heatmap(conf_matrix_rf, annot=True, fmt="d", cmap="Blues",
xticklabels=rf_model.classes_, yticklabels=rf_model.classes_)
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix - Random Forest')
plt.show()
```

```
Random Forest Accuracy: 59.41%
```

```
Random Forest Classification Report:
              precision    recall  f1-score   support

    negative       0.68      0.70      0.69        97
     neutral       0.34      0.69      0.46        29
    positive       0.73      0.42      0.53        76

    accuracy                           0.59       202
   macro avg       0.58      0.60      0.56       202
weighted avg       0.65      0.59      0.60       202
```




```python
# Calculate overall accuracy
accuracy_rf = accuracy_score(y_test, y_pred_rf)
print(f"Random Forest Accuracy: {accuracy_rf * 100:.2f}%")

# Analisis kesalahan
```

```
error_rate_rf = 1 - accuracy_rf
print(f"Error Rate: {error_rate_rf * 100:.2f}%")

print("\nPenjelasan Akurasi dan Kesalahan:")
if error_rate_rf > 0.2:
    print("Model mungkin overfitting karena jumlah estimator yang
terlalu tinggi atau depth tree yang terlalu dalam.")
else:
    print("Model performanya baik, namun bisa dicoba untuk tuning
parameter lebih lanjut.")
```

```
Random Forest Accuracy: 59.41%
Error Rate: 40.59%

Penjelasan Akurasi dan Kesalahan:
Model mungkin overfitting karena jumlah estimator yang terlalu tinggi
atau depth tree yang terlalu dalam.
```

```
# Find indices where the predictions are incorrect
incorrect_indices_rf = np.where(y_test != y_pred_rf)[0]

# Loop over some of the incorrectly predicted samples
print("\nBeberapa Contoh Prediksi Salah (Random Forest):")
for index in incorrect_indices_rf[:10]:
    print(f"Teks: {X_test.iloc[index]}")
    print(f"Label Asli: {y_test.iloc[index]}, Prediksi:
{y_pred_rf[index]}")
    print("-" * 60)
```

```
Beberapa Contoh Prediksi Salah (Random Forest):
Teks: cbcnews trudeau idiot us climate change push agenda give damn
canadian believe clown power
Label Asli: negative, Prediksi: positive
------------------------------------------------------------
Teks: tulsigabbardrep think right world tell u climate change also
need clarify subscribe preaches climate change goal good people
private jet rich
Label Asli: positive, Prediksi: neutral
------------------------------------------------------------
Teks: george marshall spent year talking und understanding human acted
climate change also feel lucky count man one friend httpstconclmqjfwbr
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: volcaholic climate change fun p let jail climate protestors
Label Asli: positive, Prediksi: neutral
------------------------------------------------------------
Teks: kmbinch cant deny climate change agreed paris real needed dealt
whats changed since
Label Asli: positive, Prediksi: negative
```

```
------------------------------------------------------------
Teks: ad gifted really thought provoking evening science speakeasy
scienceatlife climate change definitely taken lot away evening
heardatspeakeasy httpstcoqjkzcbue
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: special edition nj spotlight news take look two way climate
change affecting garden state valuable agriculture industry watch
mynjpbs online httpstcouwsaxdk
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: main two scientific reference stated guardian amp lievenanatol
amoc currently weakening strong question due new nature publ revealing
florida current remained remarkably stable climatecrisis climatechange
httpstcoootqvn httpstcoryptvxvs
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: ctvnews relentless circle jerk vaccine climate change woke
liberalism
Label Asli: negative, Prediksi: neutral
------------------------------------------------------------
Teks: markevans realtrumpnewsx climate change threaten human health
increasing frequency intensity extreme weather event like storm flood
drought wildfire heatwaves
Label Asli: neutral, Prediksi: negative
------------------------------------------------------------
```

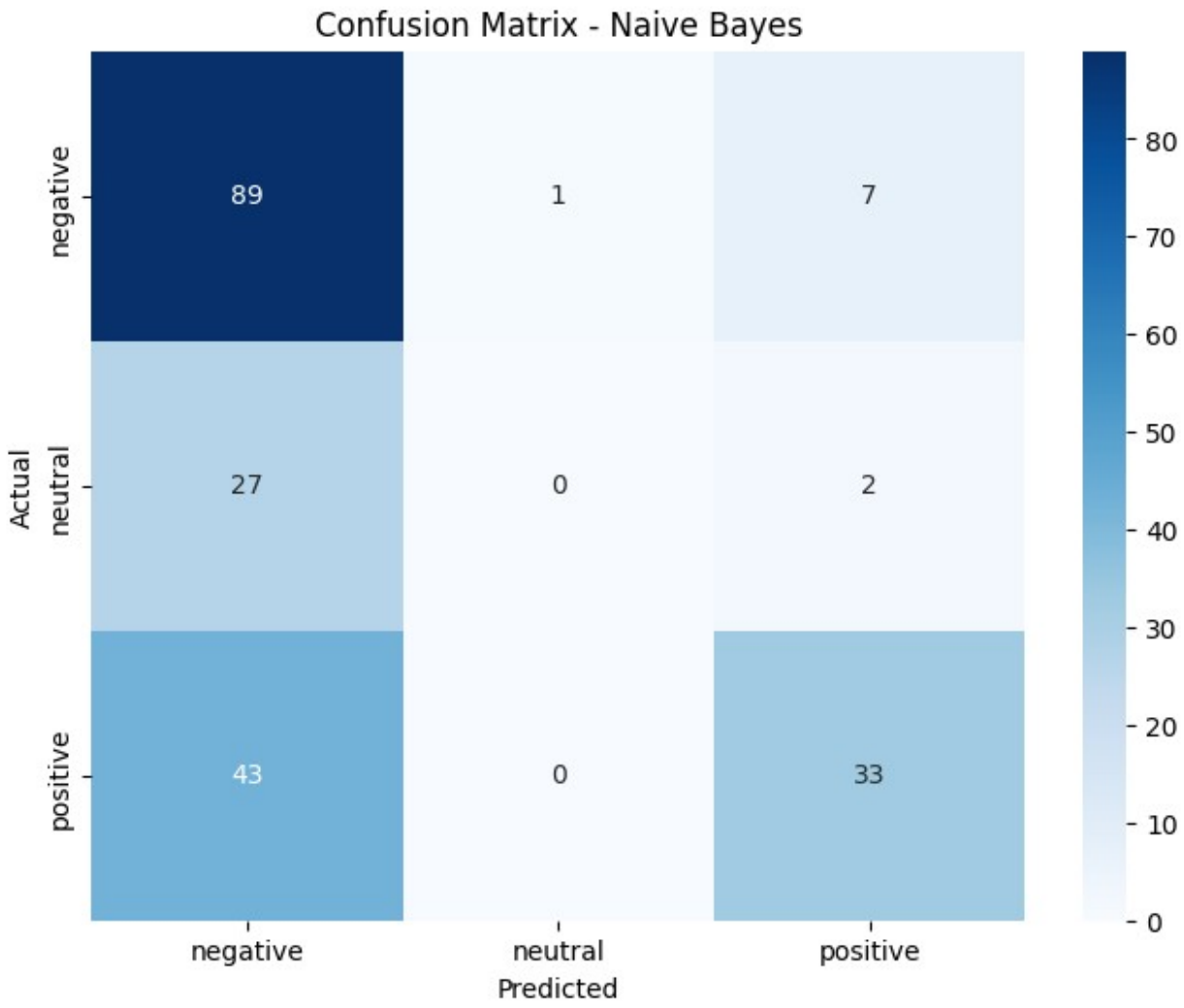
Naive Bayes

```python
from sklearn.naive_bayes import MultinomialNB

# Initialize the Multinomial Naive Bayes model
nb_model = MultinomialNB()

# Train the model on the training data
nb_model.fit(X_train_tfidf, y_train)
```

```
MultinomialNB()
```

```python
# Predict sentiment labels for the test data using the Naive Bayes
model
y_pred_nb = nb_model.predict(X_test_tfidf)

# Evaluate model performance: Accuracy
accuracy_nb = accuracy_score(y_test, y_pred_nb)
print(f'Naive Bayes Accuracy: {accuracy_nb * 100:.2f}%\n')

# Detailed classification report
print("Naive Bayes Classification Report:")
print(classification_report(y_test, y_pred_nb))
```

```
Naive Bayes Accuracy: 60.40%

Naive Bayes Classification Report:
              precision    recall  f1-score   support

    negative       0.56      0.92      0.70        97
     neutral       0.00      0.00      0.00        29
    positive       0.79      0.43      0.56        76

    accuracy                           0.60       202
   macro avg       0.45      0.45      0.42       202
weighted avg       0.56      0.60      0.54       202


# Confusion Matrix for Naive Bayes
conf_matrix_nb = confusion_matrix(y_test, y_pred_nb)

# Plot Confusion Matrix for Naive Bayes
plt.figure(figsize=(8, 6))
sns.heatmap(conf_matrix_nb, annot=True, fmt="d", cmap="Blues",
xticklabels=nb_model.classes_, yticklabels=nb_model.classes_)
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix - Naive Bayes')
plt.show()
```




```
# Calculate overall accuracy
accuracy_nb = accuracy_score(y_test, y_pred_nb)
print(f"Naive Bayes Accuracy: {accuracy_nb * 100:.2f}%")

# Analisis kesalahan
error_rate_nb = 1 - accuracy_nb
print(f"Error Rate: {error_rate_nb * 100:.2f}%")

print("\nPenjelasan Akurasi dan Kesalahan:")
if error_rate_nb > 0.2:
    print("Naive Bayes mungkin kesulitan karena asumsi distribusi data
yang tidak sesuai atau fitur yang kurang informatif.")
else:
    print("Model performanya cukup baik, tetapi bisa ditingkatkan
dengan pemrosesan fitur yang lebih baik.")
```

```
Naive Bayes Accuracy: 60.40%
Error Rate: 39.60%
```

Penjelasan Akurasi dan Kesalahan:
Naive Bayes mungkin kesulitan karena asumsi distribusi data yang tidak sesuai atau fitur yang kurang informatif.

```
# Find indices where the predictions are incorrect
incorrect_indices_nb = np.where(y_test != y_pred_nb)[0]

# Loop over some of the incorrectly predicted samples
print("\nBeberapa Contoh Prediksi Salah (Naive Bayes):")
for index in incorrect_indices_nb[:10]:
    print(f"Teks: {X_test.iloc[index]}")
    print(f"Label Asli: {y_test.iloc[index]}, Prediksi:
{y_pred_nb[index]}")
    print("-" * 60)
```

```
Beberapa Contoh Prediksi Salah (Naive Bayes):
Teks: wideawakemedia oops perhaps global cooling ah switched global
warming climate change clever bastard
Label Asli: negative, Prediksi: positive
------------------------------------------------------------
Teks: tulsigabbardrep think right world tell u climate change also
need clarify subscribe preaches climate change goal good people
private jet rich
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: george marshall spent year talking und understanding human acted
climate change also feel lucky count man one friend httpstconclmqjfwbr
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: volcaholic climate change fun p let jail climate protestors
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: kmbinch cant deny climate change agreed paris real needed dealt
whats changed since
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: got reach heart said talking climatechange way reach heart story
treat see jane goodall wry sharp funny ever speak th birthday
janegoodallinst cityartssf sf httpstcoagpkogvg
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: ad gifted really thought provoking evening science speakeasy
scienceatlife climate change definitely taken lot away evening
heardatspeakeasy httpstcoqjkzcbue
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: wideawakemedia master universe bureaucrat minion say warrior
climate change really global tyranny
```

```
Label Asli: neutral, Prediksi: negative
------------------------------------------------------------
Teks: earth may ring system million yr ago httpstcomhbfxrbgoj gi
spatial mapping model modeling geology structuralgeology climatechange
paleoclimate ordovician asteroid rochelimit crater impactcrater crater
icehouse meteorite platetectonics httpstcotahguzyx
Label Asli: neutral, Prediksi: negative
------------------------------------------------------------
Teks: markevans realtrumpnewsx climate change threaten human health
increasing frequency intensity extreme weather event like storm flood
drought wildfire heatwaves
Label Asli: neutral, Prediksi: negative
------------------------------------------------------------
```

Deep Learning

```
import numpy as np
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout

# Convert TF-IDF sparse matrices to dense format for neural network
input
X_train_dense = X_train_tfidf.toarray()
X_test_dense = X_test_tfidf.toarray()

# Encode the sentiment labels as numeric values
from sklearn.preprocessing import LabelEncoder

label_encoder = LabelEncoder()
y_train_encoded = label_encoder.fit_transform(y_train)
y_test_encoded = label_encoder.transform(y_test)

# Check the encoded classes
label_encoder.classes_

array(['negative', 'neutral', 'positive'], dtype=object)

from tensorflow.keras.optimizers import Adam

# Define the neural network model
model = Sequential()

# Input layer and first hidden layer
model.add(Dense(128, input_dim=X_train_dense.shape[1],
activation='relu'))
model.add(Dropout(0.5))

# Second hidden layer
model.add(Dense(64, activation='relu'))
model.add(Dropout(0.3))
```

```python
# Output layer - using softmax for multiclass classification
model.add(Dense(len(label_encoder.classes_), activation='softmax'))

# Compile the model
model.compile(optimizer=Adam(learning_rate=0.001),
loss='sparse_categorical_crossentropy', metrics=['accuracy'])

# Model summary
model.summary()
```

```
/usr/local/lib/python3.10/dist-packages/keras/src/layers/core/
dense.py:87: UserWarning: Do not pass an `input_shape`/`input_dim`
argument to a layer. When using Sequential models, prefer using an
`Input(shape)` object as the first layer in the model instead.
  super().__init__(activity_regularizer=activity_regularizer,
**kwargs)
```

```
Model: "sequential_3"
```

```
┏━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┳━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┳━━━━━━━━━━━━━━━━━┓
┃ Layer (type)                         ┃ Output Shape                ┃         Param # ┃
┡━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━╇━━━━━━━━━━━━━━━━━━━━━━━━━━━━━╇━━━━━━━━━━━━━━━━━┩
│ dense_11 (Dense)                     │ (None, 128)                 │         640,128 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ dropout_8 (Dropout)                  │ (None, 128)                 │               0 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ dense_12 (Dense)                     │ (None, 64)                  │           8,256 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ dropout_9 (Dropout)                  │ (None, 64)                  │               0 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ dense_13 (Dense)                     │ (None, 3)                   │             195 │
└──────────────────────────────────────┴─────────────────────────────┴─────────────────┘
```

```
 Total params: 648,579 (2.47 MB)

 Trainable params: 648,579 (2.47 MB)
```

```
Non-trainable params: 0 (0.00 B)
```

```python
from tensorflow.keras.callbacks import EarlyStopping,
ReduceLROnPlateau
from tensorflow.keras.layers import BatchNormalization

# Define an optimized neural network model
optimized_model = Sequential()

# Input layer and first hidden layer
optimized_model.add(Dense(256, input_dim=X_train_dense.shape[1],
activation='relu'))
optimized_model.add(BatchNormalization())  # Adding batch
normalization for better training stability
optimized_model.add(Dropout(0.5))

# Second hidden layer
optimized_model.add(Dense(128, activation='relu'))
optimized_model.add(Dropout(0.4))

# Third hidden layer
optimized_model.add(Dense(64, activation='relu'))
optimized_model.add(Dropout(0.3))

# Output layer - using softmax for multiclass classification
optimized_model.add(Dense(len(label_encoder.classes_),
activation='softmax'))

# Compile the model with a lower learning rate for more precise
convergence
optimized_model.compile(optimizer=Adam(learning_rate=0.0005),
loss='sparse_categorical_crossentropy', metrics=['accuracy'])

# Model summary
optimized_model.summary()
```

```
Model: "sequential_4"
```

```
┏━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┳━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┓
━━━━━━━━━━━━━━━━━┓
┃ Layer (type)                          ┃ Output Shape                ┃
Param # ┃
┡━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━╇━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┩
━━━━━━━━━━━━━━━━━┩
│ dense_14 (Dense)                      │ (None, 256)                 │
1,280,256 │
├───────────────────────────────────────┼─────────────────────────────┤
─────────────────┤
│ batch_normalization_2                 │ (None, 256)                 │
1,024 │
│ (BatchNormalization)                  │                             │
```

```
│
├─────────────────────────────────────┼─────────────────────────────┼──
──────────────┤
│ dropout_10 (Dropout)                │ (None, 256)                 │
0 │
├─────────────────────────────────────┼─────────────────────────────┼──
──────────────┤
│ dense_15 (Dense)                    │ (None, 128)                 │
32,896 │
├─────────────────────────────────────┼─────────────────────────────┼──
──────────────┤
│ dropout_11 (Dropout)                │ (None, 128)                 │
0 │
├─────────────────────────────────────┼─────────────────────────────┼──
──────────────┤
│ dense_16 (Dense)                    │ (None, 64)                  │
8,256 │
├─────────────────────────────────────┼─────────────────────────────┼──
──────────────┤
│ dropout_12 (Dropout)                │ (None, 64)                  │
0 │
├─────────────────────────────────────┼─────────────────────────────┼──
──────────────┤
│ dense_17 (Dense)                    │ (None, 3)                   │
195 │
└─────────────────────────────────────┴─────────────────────────────┴──
──────────────┘

 Total params: 1,322,627 (5.05 MB)

 Trainable params: 1,322,115 (5.04 MB)

 Non-trainable params: 512 (2.00 KB)
```

```python
# Early stopping to prevent overfitting and reduce training time
early_stopping = EarlyStopping(monitor='val_loss', patience=3,
restore_best_weights=True)

# Reduce learning rate when a metric has stopped improving
reduce_lr = ReduceLROnPlateau(monitor='val_loss', factor=0.2,
patience=2, min_lr=1e-6)

# Callbacks list
callbacks = [early_stopping, reduce_lr]

# Train the model using the optimized architecture and callbacks
optimized_history = optimized_model.fit(
    X_train_dense, y_train_encoded,
    epochs=20,  # Increase epochs to allow callbacks to take effect
    batch_size=64,  # Larger batch size for more stable gradient
updates
```

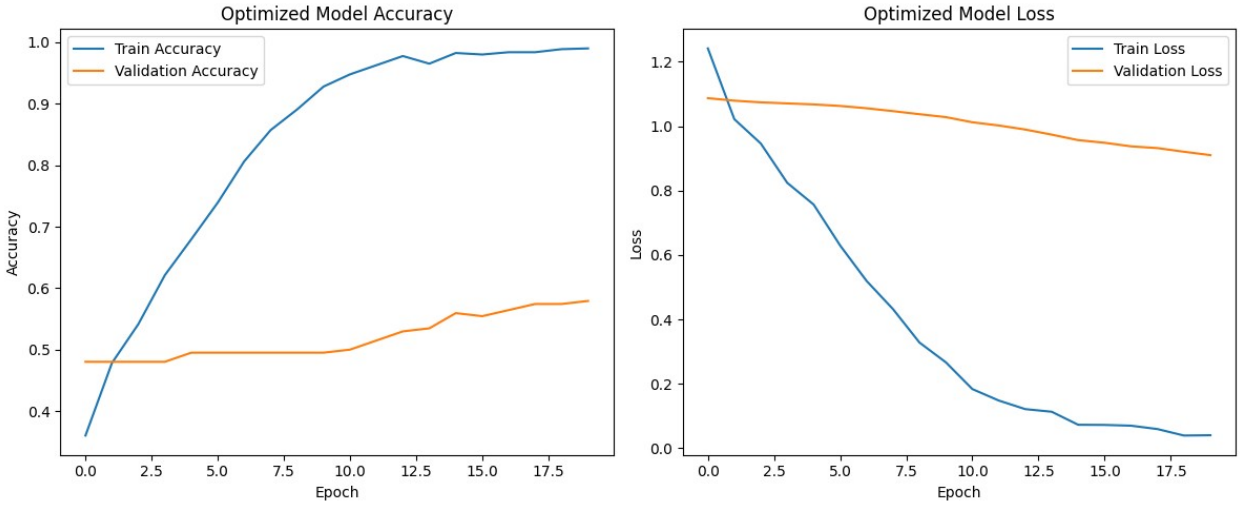
```
    validation_data=(X_test_dense, y_test_encoded),
    callbacks=callbacks,
    verbose=1
)

Epoch 1/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 2s 42ms/step - accuracy: 0.3601 - loss:
1.2724 - val_accuracy: 0.4802 - val_loss: 1.0873 - learning_rate:
5.0000e-04
Epoch 2/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 0s 29ms/step - accuracy: 0.4667 - loss:
1.0402 - val_accuracy: 0.4802 - val_loss: 1.0795 - learning_rate:
5.0000e-04
Epoch 3/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 0s 29ms/step - accuracy: 0.5267 - loss:
0.9665 - val_accuracy: 0.4802 - val_loss: 1.0743 - learning_rate:
5.0000e-04
Epoch 4/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 0s 29ms/step - accuracy: 0.6326 - loss:
0.8323 - val_accuracy: 0.4802 - val_loss: 1.0710 - learning_rate:
5.0000e-04
Epoch 5/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 0s 28ms/step - accuracy: 0.6685 - loss:
0.7635 - val_accuracy: 0.4950 - val_loss: 1.0678 - learning_rate:
5.0000e-04
Epoch 6/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 1s 27ms/step - accuracy: 0.7314 - loss:
0.6496 - val_accuracy: 0.4950 - val_loss: 1.0631 - learning_rate:
5.0000e-04
Epoch 7/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 0s 26ms/step - accuracy: 0.8078 - loss:
0.5377 - val_accuracy: 0.4950 - val_loss: 1.0559 - learning_rate:
5.0000e-04
Epoch 8/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 1s 27ms/step - accuracy: 0.8626 - loss:
0.4282 - val_accuracy: 0.4950 - val_loss: 1.0470 - learning_rate:
5.0000e-04
Epoch 9/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 1s 25ms/step - accuracy: 0.8872 - loss:
0.3395 - val_accuracy: 0.4950 - val_loss: 1.0374 - learning_rate:
5.0000e-04
Epoch 10/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 1s 25ms/step - accuracy: 0.9273 - loss:
0.2721 - val_accuracy: 0.4950 - val_loss: 1.0283 - learning_rate:
5.0000e-04
Epoch 11/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 1s 26ms/step - accuracy: 0.9463 - loss:
0.2021 - val_accuracy: 0.5000 - val_loss: 1.0124 - learning_rate:
5.0000e-04
Epoch 12/20
```

```
13/13 ━━━━━━━━━━━━━━━━━━━━ 0s 25ms/step - accuracy: 0.9645 - loss:
0.1501 - val_accuracy: 0.5149 - val_loss: 1.0024 - learning_rate:
5.0000e-04
Epoch 13/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 1s 39ms/step - accuracy: 0.9714 - loss:
0.1293 - val_accuracy: 0.5297 - val_loss: 0.9898 - learning_rate:
5.0000e-04
Epoch 14/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 1s 41ms/step - accuracy: 0.9554 - loss:
0.1396 - val_accuracy: 0.5347 - val_loss: 0.9739 - learning_rate:
5.0000e-04
Epoch 15/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 1s 43ms/step - accuracy: 0.9880 - loss:
0.0677 - val_accuracy: 0.5594 - val_loss: 0.9570 - learning_rate:
5.0000e-04
Epoch 16/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 1s 41ms/step - accuracy: 0.9851 - loss:
0.0640 - val_accuracy: 0.5545 - val_loss: 0.9487 - learning_rate:
5.0000e-04
Epoch 17/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 1s 44ms/step - accuracy: 0.9825 - loss:
0.0703 - val_accuracy: 0.5644 - val_loss: 0.9376 - learning_rate:
5.0000e-04
Epoch 18/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 1s 38ms/step - accuracy: 0.9758 - loss:
0.0640 - val_accuracy: 0.5743 - val_loss: 0.9321 - learning_rate:
5.0000e-04
Epoch 19/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 1s 41ms/step - accuracy: 0.9873 - loss:
0.0431 - val_accuracy: 0.5743 - val_loss: 0.9208 - learning_rate:
5.0000e-04
Epoch 20/20
13/13 ━━━━━━━━━━━━━━━━━━━━ 1s 42ms/step - accuracy: 0.9867 - loss:
0.0418 - val_accuracy: 0.5792 - val_loss: 0.9103 - learning_rate:
5.0000e-04
```

```
# Evaluate the model on the test set
test_loss_opt, test_accuracy_opt =
optimized_model.evaluate(X_test_dense, y_test_encoded, verbose=0)
print(f"Optimized Neural Network Test Accuracy: {test_accuracy_opt *
100:.2f}%")

# Classification report for optimized model
y_pred_opt_nn = optimized_model.predict(X_test_dense)
y_pred_opt_classes = np.argmax(y_pred_opt_nn, axis=1)

print("\nOptimized Neural Network Classification Report:")
print(classification_report(y_test_encoded, y_pred_opt_classes,
target_names=label_encoder.classes_))
```

```
# Plot training & validation accuracy and loss values
plt.figure(figsize=(12, 5))
plt.subplot(1, 2, 1)
plt.plot(optimized_history.history['accuracy'], label='Train
Accuracy')
plt.plot(optimized_history.history['val_accuracy'], label='Validation
Accuracy')
plt.title('Optimized Model Accuracy')
plt.xlabel('Epoch')
plt.ylabel('Accuracy')
plt.legend()

plt.subplot(1, 2, 2)
plt.plot(optimized_history.history['loss'], label='Train Loss')
plt.plot(optimized_history.history['val_loss'], label='Validation
Loss')
plt.title('Optimized Model Loss')
plt.xlabel('Epoch')
plt.ylabel('Loss')
plt.legend()

plt.tight_layout()
plt.show()
```

```
Optimized Neural Network Test Accuracy: 57.92%
7/7 ━━━━━━━━━━━━━━━━━━━━ 0s 14ms/step

Optimized Neural Network Classification Report:
              precision    recall  f1-score   support

    negative       0.54      0.96      0.69        97
     neutral       0.75      0.10      0.18        29
    positive       0.84      0.28      0.42        76

    accuracy                           0.58       202
   macro avg       0.71      0.45      0.43       202
weighted avg       0.68      0.58      0.51       202
```




```
# Calculate overall accuracy
accuracy_nn = accuracy_score(y_test_encoded, y_pred_opt_classes)
print(f"Neural Network Accuracy: {accuracy_nn * 100:.2f}%")

# Analisis kesalahan
error_rate_nn = 1 - accuracy_nn
print(f"Error Rate: {error_rate_nn * 100:.2f}%")

print("\nPenjelasan Akurasi dan Kesalahan:")
if error_rate_nn > 0.2:
    print("Mungkin arsitektur jaringan belum optimal atau data
training yang kurang seimbang.")
else:
    print("Model Neural Network cukup baik, namun bisa ditingkatkan
dengan menambah epoch atau tuning parameter.")
```

```
Neural Network Accuracy: 57.92%
Error Rate: 42.08%

Penjelasan Akurasi dan Kesalahan:
Mungkin arsitektur jaringan belum optimal atau data training yang
kurang seimbang.
```

```
# Find indices where the predictions are incorrect
incorrect_indices_nn = np.where(y_test_encoded != y_pred_opt_classes)
[0]

# Loop over some of the incorrectly predicted samples
print("\nBeberapa Contoh Prediksi Salah (Neural Network):")
for index in incorrect_indices_nn[:10]:
    print(f"Teks: {X_test.iloc[index]}")
    print(f"Label Asli:
{label_encoder.inverse_transform([y_test_encoded[index]])[0]},
Prediksi:
```

```
{label_encoder.inverse_transform([y_pred_opt_classes[index]])[0]}")
    print("-" * 60)


Beberapa Contoh Prediksi Salah (Neural Network):
Teks: tulsigabbardrep think right world tell u climate change also
need clarify subscribe preaches climate change goal good people
private jet rich
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: george marshall spent year talking und understanding human acted
climate change also feel lucky count man one friend httpstconclmqjfwbr
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: volcaholic climate change fun p let jail climate protestors
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: kmbinch cant deny climate change agreed paris real needed dealt
whats changed since
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: got reach heart said talking climatechange way reach heart story
treat see jane goodall wry sharp funny ever speak th birthday
janegoodallinst cityartssf sf httpstcoagpkogvg
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: ad gifted really thought provoking evening science speakeasy
scienceatlife climate change definitely taken lot away evening
heardatspeakeasy httpstcoqjkzcbue
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: special edition nj spotlight news take look two way climate
change affecting garden state valuable agriculture industry watch
mynjpbs online httpstcouwsaxdk
Label Asli: positive, Prediksi: negative
------------------------------------------------------------
Teks: wideawakemedia master universe bureaucrat minion say warrior
climate change really global tyranny
Label Asli: neutral, Prediksi: negative
------------------------------------------------------------
Teks: earth may ring system million yr ago httpstcomhbfxrbgoj gi
spatial mapping model modeling geology structuralgeology climatechange
paleoclimate ordovician asteroid rochelimit crater impactcrater crater
icehouse meteorite platetectonics httpstcotahguzyx
Label Asli: neutral, Prediksi: negative
------------------------------------------------------------
Teks: markevans realtrumpnewsx climate change threaten human health
increasing frequency intensity extreme weather event like storm flood
drought wildfire heatwaves
```

```
Label Asli: neutral, Prediksi: negative
------------------------------------------------------------
```