

Analisis Sentimen Masyarakat Terhadap Direktorat Jenderal Pajak

Juli Supriyanto Gea^{1*}, Haeni Budiati², Kristian Juri Damai Lase³, Sunneng Sandino Berutu⁴

Program Studi Informatika, Fakultas Sains dan Komputer, Universitas Kristen Immanuel Yogyakarta
Jl. Solo Km. 11, Purwomartani, Kalasan, Sleman, Daerah Istimewa Yogyakarta 55571, Indonesia

E-mail: supriyanto@student.ukrimuniversity.ac.id

Abstrak

Salah satu proses mengenali opini berbentuk teks untuk menentukan emosi secara netral, positif, ataupun negatif disebut sebagai analisis sentimen. Analisis sentimen akan menggali emosi orang-orang dari teks yang disampaikan melalui berbagai sosial media.

Direktorat jenderal pajak sebagai data yang diambil dari twitter untuk dianalisis digunakan penulis untuk mendukung penelitian. Pengumpulan data dilakukan dengan mencari dan mengumpulkan data dengan kata kunci pajak dari twitter, sehingga dapat dianalisis untuk menentukan sentimen masyarakat. Dari data yang dikumpulkan dan telah dilakukan analisis didapatkan sebanyak 40.5% netral, 39.4% positif, dan 20.1% negatif.

Info Naskah:

Naskah masuk: 09 Agustus 2023

Direvisi: 29 November 2023

Diterima: 10 Januari 2024

Abstract

One process of recognizing opinions in the form of text to determine emotions neutrally, positively, or negatively is called sentiment analysis. Sentiment analysis will explore people's emotions from texts conveyed through various social media.

Directorate general of taxes as data taken from twitter for analysis was used by the author to support research. Data collection is done by searching and collecting data with tax keywords from twitter, so that it can be analyzed to determine public sentiment. From the data collected and analyzed, 40.5% were neutral, 39.4% positive, and 20.1% negative.

Keywords:

Dirjen Pajak;

Sentiment Analysis;

Twitter;

Opinion mining;

*Penulis korespondensi:

Juli Supriyanto Gea

E-mail: supriyanto@student.ukrimuniversity.ac.id

1. Pendahuluan

Analisis sentimen dikatakan sebagai proses untuk menganalisis data berbentuk teks guna mengenali opini yang bersifat netral, positif, atau negatif. Proses analisis akan menggali emosi dari masyarakat melalui teks yang diposting melalui sosial media, seperti twitter. Sosial media twitter adalah jejaring media sosial yang digunakan untuk mengekspresikan berbagai macam opini, pengalaman, maupun hal lainnya yang menjadi perhatian banyak pengguna.

Menurut peneliti Rakhman dkk. (2021) menggunakan sentimen analisis dan analisis jaringan sosial pada user twitter guna mengetahui aktor terlibat pada pembentukan opini melalui kampanye Memakai masker, Menjaga jarak, dan Mencuci tangan (3M) di masa covid-19. Hasilnya para user twitter 87% bingung, dan 11% mendukung, sedangkan 2% lainnya sarkasme pada kampanye 3M tersebut.

Menurut Kurniasih dan Suseno (2022), kebijakan pemerintah memberi Bantuan Subsidi Upah (BSU) dalam mengurangi dampak naiknya harga minyak Bahan Bakar Minyak (BBM) menuai tanggapan masyarakat khususnya sosial media twitter. Peneliti menganalisis bahwa tanggapan masyarakat terkait BBM mendapat sentimen positif dengan persentase 28.2% sedangkan sentimen yang negatif dengan persentase 71.8%. Tanggapan masyarakat pada kebijakan BSU dari pemerintah mendapatkan sentimen positif dengan persentase 65.2%, sedangkan sentimen yang negatif dengan persentase 34.8%.

Berdasarkan penelitian Rivanie, dkk (2021) melakukan analisis tentang pendapat masyarakat terkait kinerja dari menteri kesehatan untuk menjelaskan terkait situasi pandemi. Banyak masyarakat yang memberi tanggapan melalui salah satu sosial media, yaitu twitter. Peneliti melakukan pemberian label dari tweet yang digunakan oleh pengguna twitter untuk mengetahui sentimen tanggapan masyarakat dengan menggunakan algoritma SVM dan Navie Bayes. Dari penelitian yang dilakukan didapatkan hasil untuk *Naïve Bayes* 66.45% dan SVM dengan akurasi 72.57%. Algoritma SVM lebih baik dengan nilai akurasi tinggi dalam mengklasifikasikan bagaimana komentar yang negatif dan positif terkait sentimen pada menteri kesehatan.

Peneliti selanjutnya Fahrezi dan Permana (2022) menganalisis opini dari masyarakat dari twitter terhadap suatu organisasi aksi yang cepat dan tanggap. Metode yang digunakan peneliti adalah *Naïve Bayes Classifier* yang memanfaatkan *google colab* berbahasa *python*. Peneliti mencari persentase dan diklasifikasikan dalam opini yang negatif, positif, dan netral dari organisasi aksi yang cepat dan tanggap.

Penelitian Nurmalasari, dkk (2020) BPJS kesehatan mengharapkan para peserta BPJS agar lebih mudah dalam mengakses berbagai layanan pada kesehatan dengan membuat *mobile-JKN* berbentuk aplikasi. Peneliti bertujuan menganalisis respons masyarakat terhadap aplikasi ini yang dibagi dalam tiga klasifikasi, yaitu netral, negatif, atau positif. Peneliti melakukan

scraping data dari twitter dari bulan Januari-November 2019. Dari hasil penelitian didapatkan data sebanyak 528 tweet, dimana 222 data tweet bernilai positif dengan persentase 42%, 153 data tweet bernilai netral dengan persentase 29%, dan 152 data tweet bernilai negatif dengan persentase 29%.

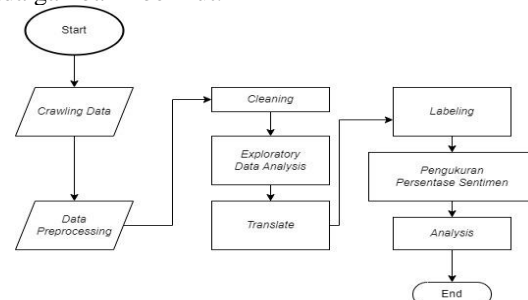
Saragih (2021) melakukan penelitian tentang analisis sentimen mengenai pelayanan pajak pada kendaraan bermotor dengan mengolah data komentar di twitter. Peneliti mendapatkan sebanyak 888 jumlah data untuk negatif dan positif dengan menggunakan mesin pengklasifikasi yaitu *naïve bayes*. Yang digunakan peneliti dalam pengklasifikasian adalah *10-foldCrossValidation* dengan hasil akurasi rata-rata 91.00%.

Direktorat jenderal pajak atau biasa disingkat dirjen pajak sebagai salah satu lembaga pengelola perpajakan pusat di Indonesia menjadi sorotan publik akhir-akhir ini. Penyebabnya dengan munculnya kasus penganiayaan yang diduga dilakukan oleh seorang anak pejabat yang merupakan pegawai pejabat eselon III dirjen pajak wilayah Jakarta Selatan. Dari isu tersebut muncul berbagai opini dari masyarakat yang menduga adanya kasus suap dan korupsi yang dilakukan oleh lembaga dirjen pajak tersebut. Sorotan publik tersebut dituangkan dalam berbagai sosial media, dimana salah satunya adalah twitter.

Dari latar masalah di atas, maka penulis melakukan penelitian terkait sentimen masyarakat yang dibangun dalam sistem terpadu sebagai tugas akhir dalam studi dengan mengangkat judul “Analisis Sentimen Masyarakat Terhadap Direktorat Jenderal Pajak. Peneliti bertujuan untuk menganalisis sentimen masyarakat terhadap dirjen pajak dengan menghitung persentase nilai sentimen yang dikategorikan dalam netral, negatif, dan positif.

2. Metode Penelitian

Dalam melakukan penelitian, metode yang digunakan adalah mempelajari literatur dari berbagai sumber dan mengumpulkan data dari twitter yang kemudian diimplementasikan dengan menggunakan platform *google colab* dengan bahasa *python*. Dengan menggunakan alur *flowchart* dapat lebih mudah dipahami alur proses untuk melakukan analisis seperti pada gambar 1 berikut.



Gambar 1 Flowchart Alur Proses Crawling Data

Pada gambar di atas adalah bentuk flowchart alur proses untuk crawling data yang kemudian diolah agar bisa dianalisis.

3. Hasil dan Pembahasan

Dalam bagian ini membahas tentang program yang akan dibangun dan digunakan dalam melakukan analisis sentimen terhadap dirjen pajak.

3.1 Implementasi Sistem

Implementasi sistem dibangun untuk analisis sentimen masyarakat terhadap dirjen pajak dengan menggunakan platform *google colab* yang menggunakan bahasa *python*. Di dalamnya ada beberapa library yang akan digunakan, yaitu *pandas*, *numpy*, *textblob*, *wordcloud*, *googletranslator*, dan *matplotlib*.

3.2 Implementasi Program

3.2.1 Instalasi

Dalam melakukan *crawling* data, dibutuhkan *tools python 3.9*, *git bash*, dan *windows build tools* untuk mempermudah proses *crawling* data.

3.2.2 Proses Crawling

Dengan sudah dilakukan instalasi tools, maka proses untuk mengumpulkan data twitter atau *crawling* bisa dilakukan. Seperti gambar 2.1 tentang query instalasi di bawah ini.

```
$ pip install pipenv
$ pipenv --python 3.9
$ pipenv shell
$ pipenv install
git+https://github.com/twintproject/twint.git#egg=twint
```

Gambar 2.1 Query untuk instalasi

Gambar di atas menunjukkan query untuk menginstal library dalam membantu proses *crawling* data dari twitter, dengan ditunjukkan pada gambar 2.2 di bawah ini untuk *crawling* data.

MINGW32/c/Users/User/Desktop/Dirjen Pajak

```
User@MINGW32 ~/Desktop/Dirjen Pajak
$ twint-s-dirjenpajak -o dirjenpajak.csv --limit 60 --lang id --since 2023-02-01 --until 2023-03-08 --csv
```

Gambar 2.2 Query untuk *crawling* data twitter

Gambar di atas adalah salah satu bentuk query untuk mengumpulkan data dari twitter yang dibatasi sebanyak 60 data dari kata kunci *dirjenpajak* berbentuk teks bahasa Indonesia dengan selang waktu dari 01-02-2023 hingga 08-03-2023 yang kemudian disimpan sebagai file *.csv*. Dalam mengambil data twitter digunakan *key word* pajak, dirjen pajak, perpajakan, direktorat jenderal pajak, dan lembaga pajak yang semuanya disimpan sebagai file *.csv*.

Data yang dikumpulkan dari waktu bulan Februari hingga Maret berjumlah 6602 data, dengan data terbanyak pada bulan Maret dibandingkan bulan sebelumnya.

3.2.3 Mengimport Library

Untuk mengimport atau memanggil *library* dalam *google colab*, maka dibutuhkan query seperti gambar 2.3 di bawah ini.

```
[1] import pandas as pd #untuk membersihkan data, memanipulasi data, dan analisis data
import numpy as np #untuk melakukan perhitungan saintifik seperti matrix, aljabar, statistik, dsb
from sklearn.feature_extraction.text import CountVectorizer
import nltk #untuk membantu dalam bekerja dengan text/text processing
from textblob import TextBlob
from wordcloud import WordCloud
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
import re
```

Gambar 2.3 Mengimport *library*

Gambar di atas merupakan query untuk memanggil *library* seperti *pandas* untuk menganalisis data, dan *numpy* untuk melakukan komputasi numerik, dan beberapa *library* lainnya yang dapat membantu *google colab* dalam membaca dataset.

3.2.4 Memasukkan Dataset

Setelah memanggil *library*, maka dapat memasukkan dataset untuk membantu *google colab* membaca data tersebut. Berikut query untuk menampilkan penggalan dataset dengan batasan lima data seperti pada gambar 2.4 di bawah ini.

```
def load_data():
    data = pd.read_csv("dataPajak.csv", encoding='iso-8859-1', parse_dates=True)
    return data

db_tabel = load_data()
db_tabel.head(5)
```

Gambar 2.4 Query untuk menampilkan dataset

Hasil dari query pada gambar di atas akan ditunjukkan pada gambar 2.5 di bawah ini.

Index	id	conversation_id	created_at	date	time	timezone	user_id	username	name	place	tweet	language
0	1.636E+23	1.636E+23	2023-03-07 21:17:23	2023-03-07	9:17:23 PM	Asia/Manila	700 0 7.42E+22	lirntge	lirntge	Manila	DirjenPajak paling gask kalo ada yang post hatta tapi orang dalam bisa lolos vektor	in
1	1.636E+23	1.636E+23	2023-03-07 20:17:19	2023-03-07	8:17:19 PM	Asia/Manila	700 0 1.32E+23	mitah7966543	Mitah	Manila	@feridgaban Mitah gimana ini ceritanya @DirjenPajak?	in
2	1.636E+23	1.636E+23	2023-03-07 19:13:59	2023-03-07	7:13:59 PM	Asia/Manila	700 0 2192545394	adra_silent	Damai	Manila	500 M semuanya duit bukan pasir ya? H Kd sdr terbaesa kongkalikong akan selamanya byk jemi2 rafael2 lainnya di @KemenkeuRI terdapat di @dirjenpajak https://t.co/4XPFmCqPB	in
3	1.636E+23	1.636E+23	2023-03-07 18:18:54	2023-03-07	6:18:54 PM	Asia/Manila	700 0 1.12E+23	heruchanulya1	Dolpi	Manila	Diri kepolisiin pindah ke dirjenpajak Dari sambio pindah ke trisambodo	in
4	1.636E+23	1.636E+23	2023-03-07 15:43:57	2023-03-07	3:43:57 PM	Asia/Manila	700 0 89677164	fajaronline	fajar online	Manila	PPATK Bkuk Puluan Rukening terkaal Rafael Akun Trisambodo https://t.co/6912NPO4XU	in

Gambar 2.5 Contoh lima data penggalan dataset

Gambar di atas merupakan contoh penggalan dataset dengan menampilkan lima data yang teratas dari data yang sudah dikumpulkan.

3.2.5 Mengubah Bentuk Kolom

Dalam mengubah bentuk kolom, maka query pada gambar 2.6 berikut digunakan untuk menampilkan data yang terdiri dari kolom *date* dan juga kolom *tweet*.

```
df=pd.DataFrame(db_tabel[['date','tweet']])
df
```

	date	tweet
0	2023-07-03	DirjenPajak paling gasik kalo ada yang post ha...
1	2023-07-03	@faridgaban Halo gimana ini certanya @Dirjenp...
2	2023-07-03	500 M semuanya duit bukan pasir ya H Ki sdh t...
3	2023-07-03	Kepolisian pindah ke dirjenpajak Dari sam...
4	2023-07-03	PPATK Blokir Puluhan Rekening terkait Rafael A...
...
6597	2023-07-03	@Michkp71 Hai, Kak. Layanan DJP yang menelepe...
6598	2023-07-03	@rins3265 Jadi, secara ketentuan, meskipun Bad...
6599	2023-07-03	@sino_k_Veei Siang, Kak. DM hanya diperuntukan ...
6600	2023-07-03	Dapat tips hemat dari founder Smart Drive I...
6601	NaN	NaN

6602 rows x 2 columns

Gambar 2.6 Kolom data

Gambar di atas merupakan query serta hasil dari mengubah kolom dengan menampilkan kolom data dan juga kolom tweet.

3.2.6 Proses Cleaning

Agar data lebih mudah dibaca dan dipahami, maka perlu menghilangkan hyperlink, RT, mentions, dan juga hashtag dengan query yang ditunjukkan pada gambar 2.7 di bawah ini.

```
#function to clean the tweet
def cleanTxt(text):
    text = re.sub('@[A-Za-z0-9]+', '', text) #remove mentions
    text = re.sub('#', '', text) #remove hashtag
    text = re.sub('RT[\s]+', '', text) #remove RT
    text = re.sub('https?:\/\/\S+', '', text) #remove hyperlink
    return text

#clean tweets
df['tweet'] = df['tweet'].apply(cleanTxt)
```

Gambar 2.7 Query untuk cleaning

Gambar di atas merupakan query untuk membersihkan data agar lebih mudah dibaca seperti ditampilkan pada gambar 2.8 di bawah ini.

```
#Show cleaned tweet
df1=df['tweet']
df1

0      DirjenPajak paling gasik kalo ada yang post ha...
1              Halo gimana ini ceritanya ?
2      500 M semuanya duit bukan pasir ya !! Kl sdh t...
3      Dari kepolisian pindah ke dirjenpajak Dari sam...
4      PPATK Blokir Puluhan Rekening terkait Rafael A...
5      Rafael Alun Miliki 40 Rekening Dengan Nila...
6      Kasus Ayah Marlo Dandy Naik Tanpa Penyelidikan...
7      Bagaimana jadinya kalau seperti ini... Rafael...
8      Anggota DPR RI: Renumerasi Pegawai Pajak Harus...
9
10                                     Donee, lirik aku dong min
11                                     YUUUUK
12                                     LIRIK AKU DONG MINNIN
13                                     Done ce
14                                     LIRIK AKUUU MIN
15                                     salkeenn sall
16                                     Halo El salken ya
17      waduuuu, diliat liat makin panas nih kwkwkwk d...
18      otoritasjasakeuangan lhkpn beacukai dirjenpaja...
19                                     Polri Halo RI
20      Harta Rafael Alun Trismadomo Jadi Sorotan, Ter...
21      VIRALKAN !!!!! dan kita tunggu kelanjutan nya...
22
23      Sri Mulyani Sebut Data Ditjen Pajak yang Bocor...
24      Masih adakah instansi pemerintah yg bersih da...
```

Gambar 2.8 Hasil setelah dibersihkan

Gambar di atas menunjukkan hasil data setelah dibersihkan sehingga lebih mudah dibaca dan dipahami.

3.2.7 Bentuk Wordcloud

Bentuk *wordcloud* adalah bentuk gambar yang menyaring data dengan menunjukkan kata yang paling banyak muncul agar lebih mudah untuk ditemukan.

Dapat dilihat pada gambar 2.9 berikut untuk query wordcloud.

```
allWords = ' '.join([twts for twts in df['tweet']])
wordCloud = WordCloud(width=900, height=450, random_state=10, max_font_size=110).generate(allWords)

plt.imshow(wordCloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```

Gambar 2.9 Query untuk wordcloud

Query gambar di atas akan dihasilkan dalam bentuk gambar 2.10 di bawah ini.



Gambar 2.10 Hasil untuk wordcloud

Gambar di atas merupakan hasil untuk wordcloud yang menampilkan kata atau data yang paling sering muncul dalam bentuk kata yang lebih besar.

3.2.8 Grafik Data

Jumlah data perhari dari total jumlah tweet yang sudah dikumpulkan akan ditampilkan pada gambar 2.11 di bawah ini.

df2[['date', 'tweet']]

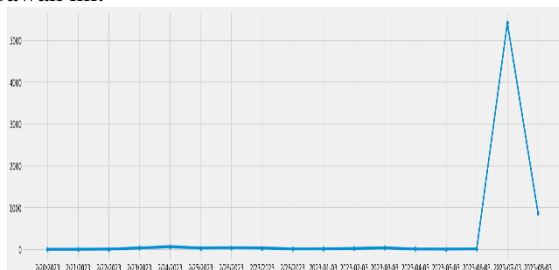
1 to 17 of 17 entries **Filter**

index	date	tweet
0	2/20/2023	1
1	2/21/2023	2
2	2/22/2023	7
3	2/23/2023	33
4	2/24/2023	65
5	2/25/2023	32
6	2/26/2023	41
7	2/27/2023	32
8	2/28/2023	13
9	2023-01-03	15
10	2023-02-03	22
11	2023-03-03	13
12	2023-04-03	9
13	2023-05-03	7
14	2023-06-03	12
15	2023-07-03	5408
16	2023-08-03	866

Show 25 per page

Gambar 2.11 Jumlah tweet perhari

Dari jumlah tweet pada gambar di atas, akan dibuat dalam bentuk grafik seperti pada gambar 2.12 di bawah ini.



Gambar 2.12 Bentuk grafik dari jumlah data perhari

Gambar di atas menampilkan bahwa grafik lebih tinggi ada pada tanggal 07-03-2023 dengan jumlah data tweet terbanyak 5408 jumlah data tweet daripada tanggal sebelum dan sesudahnya.

3.2.9 Translate Data

Dalam penggunaan *textblob* hanya mengenali dalam bentuk bahasa Inggris, maka terlebih dahulu data diterjemahkan menjadi bahasa Inggris karna sebelumnya data masih dalam bentuk bahasa Indonesia. Pada gambar 2.13 di bawah ini menunjukkan query untuk proses melakukan translate data.

```
#translator
!pip install deep-translator

Collecting deep-translator
  Downloading deep-translator-1.11.4-py3-none-any.whl (42 kB)
    42.3/42.3 kB 3.7 MB/s eta 0:00:00
Requirement already satisfied: beautifulsoup4<5.0.0,>=4.9.1 in /usr/local/lib/python3.10/dist-packages (from deep-translator)
Requirement already satisfied: requests<3.0.0,>=2.23.0 in /usr/local/lib/python3.10/dist-packages (from deep-translator)
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-packages (from beautifulsoup4<5.0.0,>=4.9.1)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.23.0)
Requirement already satisfied: certifi>2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.23.0)
Requirement already satisfied: charset-normalizer<2.0.0 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.23.0)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.23.0)
Installing collected packages: deep-translator
Successfully installed deep-translator-1.11.4

from deep_translator import GoogleTranslator

pajakEng=[]
for kalimat in df1:
    translated=GoogleTranslator(source='auto', target='en').translate(kalimat)
    pajakEng.append(translated)
pajakEng

df2=pd.DataFrame(pajakEng)
df2.to_csv('PajakEng1.csv', index=False)
```

Gambar 2.13 Query instal, import, dan translate

Gambar di atas merupakan query untuk menginstal *deep translator* serta import *googletranslator* dan menerjemahkan data ke dalam bahasa Inggris. Setelah melakukan terjemahan, maka hasilnya akan diperlihatkan pada gambar 2.14 di bawah ini.

1	The Director General of Taxes is the craziest ...
2	Hello, what's the story?
3	500 M is all money not sand huh !! If you are ...
4	From the police, they moved to the Director Ge...
...	...
6571	_pajak _1202 If you want to ask for a forgotte...
6572	Hi bro. DGT services that make calls to taxpay...
6573	So, in terms of regulations, even though the A...
6574	_Veei Afternoon, Sis. DM is only intended for ...
6575	Get saving tips from the founder of the Smart ...

Gambar 2.14 Hasil setelah menerjemahkan

Pada gambar di atas terlihat bahwa data sudah diterjemahkan ke dalam bahasa Inggris sehingga dapat dilanjutkan ke proses selanjutnya.

3.2.10 Labeling Data

Setelah proses translate dilakukan, maka proses selanjutnya adalah menentukan label untuk kategori sentimen, yaitu kategori negatif, netral, dan positif. Dalam hal ini, yang dilakukan adalah pertama adalah menentukan nilai dari polaritas dan subjektivitas dari data. Berikut gambar 2.15 query untuk polaritas dan juga subjektivitas di bawah ini.

```
#membuat function untuk menghitung subjectivity
def getSubjectivity(text):
    return TextBlob(text).sentiment.subjectivity

#membuat function untuk menghitung polarity
def getPolarity(text):
    return TextBlob(text).sentiment.polarity

#membuat dua kolom baru untuk subjectivity dan polarity
df['Subjectivity'] = df['Column1'].apply(getSubjectivity)
df['Polarity'] = df['Column1'].apply(getPolarity)

#memanggil
df[['Column1', 'Subjectivity', 'Polarity']]
```

Gambar 2.15 Query subjectivity dan polarity

Pada gambar di atas adalah query untuk penentuan berapa nilai polaritas serta subjektivitas untuk data tweet.

Setelah menentukan nilai-nilai polaritas serta subjektivitas dari kalimat data tweet, maka langkah selanjutnya adalah labeling data. Fungsi dari labeling data adalah untuk menentukan data yang berkategori negatif, atau netral, dan atau positif. Gambar 2.16 di bawah ini adalah bentuk querynya.

```
def getAnalysis(score):
    if score < 0:
        return 'Negatif'
    elif score == 0:
        return 'Netral'
    else:
        return 'Positif'
df['Label'] = df['Polarity'].apply(getAnalysis)

#show dataframe lagi
df[['Column1', 'Subjectivity', 'Polarity', 'Label']]
```

Gambar 2.16 Query untuk penentuan label

Gambar di atas merupakan query untuk pelabelan data dalam menentukan kategori netral, negatif, dan positif. Pada gambar 2.17 di bawah akan ditunjukkan nilai dari polaritas dan subjektivitas serta pelabelannya.

		0	Subjectivity	Polarity	Label
0	The Director General of Taxes is the craziest ...	0.500000	0.050000		Positif
1	Hello, what's the story?	0.000000	0.000000		Neutral
2	500 M is all money not sand huh !! If you are ...	0.625000	0.125000		Positif
3	From the police, they moved to the Director Ge...	0.500000	0.050000		Positif
4	PPATK Blocks Tens of Accounts related to Rafae...	0.450000	0.025000		Positif
...
6596	_pajak _1202 If you want to ask for a forgotte...	0.644444	-0.327778		Negatif
6597	Hi bro. DGT services that make calls to taxpay...	0.400000	0.000000		Neutral
6598	So, in terms of regulations, even though the A...	0.550000	0.183333		Positif
6599	_Veei Afternoon, Sis. DM is only intended for ...	0.750000	0.025000		Positif
6600	Get saving tips from the founder of the Smart ...	0.760714	0.269664		Positif

...6575 rows x 6 columns

Gambar 2.17 Hasil labeling

Gambar di atas menunjukkan hasil data yang negatif, netral, dan positif dari data tweet.

Langkah selanjutnya menghitung jumlah sentimen yang ditunjukkan pada gambar 2.18 di bawah ini.

```
df['Label'].value_counts()

Neutral      2666
Positif      2589
Negatif      1321
Name: Label, dtype: int64
```

Gambar 2.18 Hasil perhitungan jumlah sentimen

Pada gambar di atas manmpilkan hasil dengan jumlah neutral sebanyak 2666, hasil jumlah positif sebanyak 2589, dan hasil jumlah negatif sebanyak 1321 data.

3.2.11 Persentase Sentimen

Setelah melakukan proses labeling penentuan negatif, positif, dan netral, maka langkah selanjutnya adalah proses perhitungan berapa persentase dari sentimen tersebut. Query untuk menghitung berapa persen sentimen akan ditampilkan pada gambar 2.19 di bawah ini.


```

ntweets = df[df.Label == 'Negatif']
ntweets = ntweets['Column1']
ntweets
SumNeg=round((ntweets.shape[0] / df.shape[0]) * 100, 1)

ptweets = df[df.Label == 'Positif']
ptweets = ptweets['Column1']
ptweets
SumPos=round((ptweets.shape[0] / df.shape[0]) * 100, 1)

nttweets = df[df.Label == 'Netral']
nttweets = nttweets['Column1']
nttweets
SumNet=round((nttweets.shape[0] / df.shape[0]) * 100, 1)

print (SumNeg, SumPos, SumNet)
df
20.1 39.4 40.5

```

Gambar 2.19 Perhitungan persentase

Pada gambar di atas terlihat query untuk menghitung berapa persen dari sentimen dan hasilnya menunjukkan bahwa sebanyak 20.1% hasil untuk negatif, sebanyak 39.4% hasil untuk positif, dan sebanyak 40.5% untuk netral.

3.2.12 Plotting

Proses yang dilakukan selanjutnya adalah plotting menentukan titik-titik sebaran untuk subjectivity serta polarity. Query untuk plotting akan ditunjukkan pada gambar 2.20 di bawah ini.

```

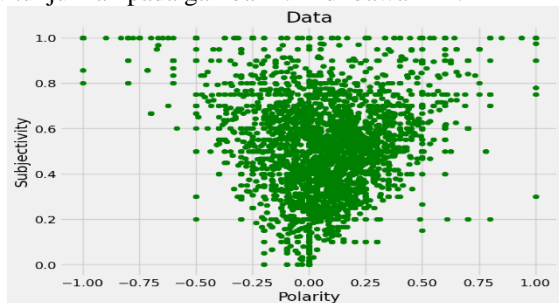
#Plotting
plt.figure(figsize=(8,6))
for i in range(0, df.shape[0]):
    plt.scatter(df["Polarity"][i], df["Subjectivity"][i], color='Green')

plt.scatter(x,y,color)
plt.title('Data')
plt.xlabel('Polarity')
plt.ylabel('Subjectivity')
plt.show()

```

Gambar 2.20 Query untuk plotting

Gambar di atas merupakan query untuk menentukan titik-titik sebaran yang berwarna hijau dan ditunjukkan pada gambar 2.21 di bawah ini.



Gambar 2.21 Hasil untuk plotting

Gambar di atas menunjukkan sebaran titik untuk subjectivity terlihat banyak pada titik 0.0 hingga titik 1.0, sedangkan sebaran titik untuk polarity terlihat banyak pada titik -0.50 hingga 0.75.

3.2.13 Grafik Batang

Tahap proses ini akan menampilkan data sentimen dalam bentuk grafik batang yang dapat dilihat querynya pada gambar 2.22 di bawah ini.

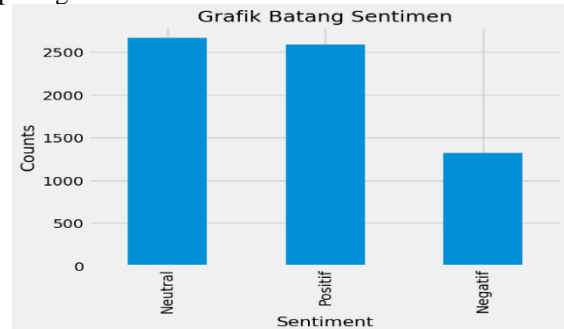
```

#plotting and visualizing grafiknya
plt.title('Grafik Batang Sentimen', fontsize=18, loc = 'center')
plt.xlabel('Sentiment', fontsize=16)
plt.ylabel('Counts', fontsize=16)
df['Label'].value_counts().plot(kind = 'bar')
plt.show()

```

Gambar 2.22 Query untuk grafik batang

Query pada gambar di atas akan menunjukkan data sentimen dalam bentuk grafik batang yang terlihat pada gambar 2.23 di bawah ini.



Gambar 2.23 Hasil untuk grafik batang

Gambar di atas menunjukkan bahwa data yang bersifat neutral memiliki grafik lebih tinggi dengan jumlah 2666 data yang diikuti oleh sentimen positif sebanyak 2589 jumlah data. Dan data yang bersifat negatif terlihat di gambar menunjukkan grafik yang terendah dengan jumlah 1321 data dari dua grafik lainnya.

3.2.14 Hasil Analisis

Setelah melalui beberapa tahapan rangkaian proses dan telah didapatkan hasil perhitungan dari sentimen, maka diperoleh hasil dari Analisis Sentimen Masyarakat terhadap Direktorat Jenderal Pajak yang dapat dilihat pada tabel 1 di bawah ini.

Tabel 1 Hasil Analisis

Dataset	Sentimen (%)		
	Positif	Negatif	Netral
Dirjen Pajak	39.4 %	20.1 %	40.5 %

Pada tabel di atas diperoleh hasil yang lebih banyak persentase pada sentimen netral sebanyak 40.5%, positif dengan persentase 39.4%, dan negatif dengan persentase 20.1%. Dimana hasil persentase yang lebih rendah adalah sentimen negatif, sedangkan lebih tinggi sentimen netral. Hasil tersebut dikarenakan respon dari para pengguna twitter terhadap dirjen pajak yang berupa tweet masih bersifat netral sesuai dengan pengkategorian sentimen berdasarkan proses analisis yang telah dilakukan.

4. Kesimpulan

Proses untuk analisis sentimen terhadap dirjen pajak melalui beberapa tahap sudah dilakukan dengan hasil analisis yang menunjukkan untuk sentimen paling tinggi adalah netral dengan persentase 40.5%, diikuti dengan sentimen yang positif dengan persentase 39.5%, dan negatif dengan persentase 20.1%. Berdasarkan persentase tersebut, sentimen masyarakat terhadap dirjen pajak masih tergolong netral, terlebih tanggapan mereka pada dirjen pajak termasuk bersifat positif lebih tinggi daripada negatif.

Ucapan Terimakasih

Dengan selesainya tahap penelitian sentimen masyarakat terhadap dirjen pajak yang telah dilakukan oleh penulis, maka penulis mengucapkan terimakasih banyak kepada pihak yang telah membantu dalam proses penelitian dan dukungan lainnya. Terimakasih kepada ibu Febe Maedjaja, bapak Sunneng Sandino Berutu, dan ibu Haeni Budiati yang telah membantu dan mendukung penulis dalam penelitian ini.

Daftar Pustaka

- [1] M. F. Fahrezi, and A. A. Permana (2022). Sentimen Analisis Opini Masyarakat Pada Sosial Media Twitter Terhadap Organisasi Aksi Cepat Tanggap Menggunakan Naïve Bayes Classifier. *JT : Jurnal Teknik*, 11(02), 113-121.
- [2] U. Kurniasih, and A. T. Suseno. (2022). Analisis Sentimen Terhadap Bantuan Subsidi Upah (BSU) pada Kenaikan Harga Bahan Bakar Minyak (BBM). *Jurnal Media Informatika Budidarma*, 6(4), 2335-2340.
- [3] M. Nurmalasari, N. A. Temesvari, and S. N. Maula. (2020). Analisis Sentimen terhadap Opini Masyarakat dalam Penggunaan Mobile-JKN untuk Pelayanan BPJS Kesehatan Tahun 2019. *Indonesian of Health Information Management Journal*, 8(1), 35-44.
- [4] T. Rivanie, R. Pebrianto, T. Hidayat, A. Bayhaqy, W. Gata, and H. B. Novitasari. (2021). Analisis Sentimen Terhadap Kinerja Menteri Kesehatan Indonesia Selama Pandemi Covid-19. *Jurnal Informatika*, 21(1), 1-10.
- [5] F. R. Rakhman, R. W. Ramadhani, and Y. A. Kuncoroyati. (2021). Analisis Sentimen Dan Opini Digital Kampanye 3M Di Masa Covid-19 Melalui Media Sosial Twitter. *Komunikologi: Jurnal Ilmiah Ilmu Komunikasi*, 18(01), 8.
- [6] Hot M. H. 2022. Analisis Sentimen Pengguna Twitter Terhadap Layanan Pajak Kendaraan Bermotor Menggunakan Algoritme Naive Bayes Classifier. *Skripsi*. Fakultas Matematika dan Ilmu Pengetahuan Alam. Universitas Lampung: Bandar Lampung.