

Analisis Media Sosial

Topik: 'Cipta Kerja'

Analisa media sosial dilakukan dengan mengambil dan menganalisa data yang diperoleh dari twitter. Untuk mengambil, mengolah, dan visualisasi hasil analisa menggunakan Bahasa python dengan library tweepy.

MENGUMPULKAN DATA

Data dari twitter dikumpulkan selama 3 hari (21-23 Oktober 2020). Struktur data yang didapatkan dari Twitter antara lain: id, username, isi tweet, tanggal dan lokasi tweet tersebut dibuat, dll.

A. Import semua library yang dibutuhkan

```
import tweepy
import pandas as pd
import sqlite3
import numpy as np
```

B. Persiapan Key dan Authentication Handler

```
#persiapkan key
consumer_key = "XXXXXXXXXXXXXXXXXXXXX"
consumer_secret = "XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX"
access_token = "XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX"
access_token_secret = "XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX"

#buat handler auth
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth, wait_on_rate_limit=True)
```

C. Mengambil data berdasarkan kata kunci

```
search_words = ["cipta kerja", "ciptaker", "omnibus law"]
date_since = "2020-10-21"
date_until = "2020-10-23"
nomor = []
items = []
waktu = []
akun = []
lokasi = []
```

```

for tweet in tweepy.Cursor(api.search,q=search_words, count=100,
lang="id", since= date_since, until = date_until).items(10):
    nomor.append(tweet.id)
    items.append(tweet.text)
    waktu.append(tweet.created_at)
    akun.append(tweet.user.screen_name)
    lokasi.append(tweet.user.location)

#memasukkan dalam dataframe
df = pd.DataFrame({'ID' : nomor, 'Tweets' : items, 'Waktu' : waktu, 'Akun' : akun, 'Lokasi' : lokasi})
print (df)

```

OUTPUT:

	ID	...	Lokasi
0	1319426925522161664	...	
1	1319424690234904577	...	Indonesia
2	1319422818178584578	...	
3	1319421172656730112	...	Tasikmalaya, Jawa Barat
4	1319419774221258753	...	Indonesia
..
760	1318733744698662913	...	
761	1318733744459579392	...	
762	1318730579324514305	...	
763	1318719833559629825	...	Indonesia
764	1318719672422858752	...	bogor

[765 rows x 5 columns]

Berdasarkan hasil pengambilan data diperoleh tweet sebanyak 765 dengan menggunakan keyword: "cipta kerja", "ciptaker", "omnibus law"

D. Simpan Data Pada SQL

```

conn = sqlite3.connect(' data_to_sql.db')
c = conn.cursor()
c.execute('CREATE TABLE data_tweet (ID, Tweets, Waktu, Akun, Lokasi)')
conn.commit()
df.to_sql('data_tweet', conn, if_exists='replace', index = False)

```

Hasil dari code di atas adalah memberikan output sebuah database .sql yang menampung data tweets yang sudah dikumpulkan. Berikut adalah tampilan struktur table data_tweet yang diakses dengan aplikasi DB Browser SQLite.



Tables (1)	
data_tweet	
ID	INTEGER
Tweets	TEXT
Waktu	TIMESTAMP
Akun	TEXT
Lokasi	TEXT

Gambar di bawah ini merupakan tampilan 10 data pertama pada table data_tweet.

	ID	Tweets	Waktu	Akun	Lokasi
	Filter	Filter	Filter	Filter	Filter
1	1319426925522161664	Semua bidang terdampak karena Omnibus Law. ...	2020-10-22 23:54:27	FeraDiana14	
2	1319424690234904577	RT @prastow: Yth @CNNIndonesia, mungkin per...	2020-10-22 23:45:34	nunasyeban	Indonesia
3	1319422818178584578	Semua bidang terdampak karena Omnibus Law. ...	2020-10-22 23:38:08	AkinaYoshi1	
4	1319421172656730112	RT @kumparan: Naskah UU Cipta Kerja yang ...	2020-10-22 23:31:35	nana040461	Tasikmalaya, Jawa Barat
5	1319419774221258753	Ketua MPR Bambang Soesatyo bicara soal prose...	2020-10-22 23:26:02	jpnncom	Indonesia
6	1319417977238810624	Semua bidang terdampak karena Omnibus Law. ...	2020-10-22 23:18:54	yesitadamy	
7	1319414422171910144	RT @prastow: Yth @CNNIndonesia, mungkin per...	2020-10-22 23:04:46	Ari3Pras	Bumi Nusantara
8	1319414261135806464	Semua bidang terdampak karena Omnibus Law. ...	2020-10-22 23:04:08	AgungVenti	
9	1319412233433018369	Semua bidang terdampak karena Omnibus Law. ...	2020-10-22 22:56:04	Uafathar	
10	1319411750584811520	RT @prastow: Yth @CNNIndonesia, mungkin per...	2020-10-22 22:54:09	_Bakemono	

PRA-PROSES DATA

Pada bagian pra-proses dilakukan pembersihan data dengan mengaplikasikan berbagai metode seperti menghapus tanda baca, repetisi karakter, username, hashtag, link, dan sebagainya.

A. Import semua library yang dibutuhkan

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import tweepy
import re,string
import pandas as pd
import sqlite3
import matplotlib.pyplot as plt
import numpy as np
```

B. Buka file sql dari data yang sudah dikumpulkan

```
con = sqlite3.connect(file)
df = pd.read_sql_query("SELECT * from data_tweet", con)
con.close()
```

C. Membersihkan data

```
clean_tweets = []
tweets = []

for tweet in data['Tweets']:
    def hapus_tanda(tweet):
        tanda_baca = set(string.punctuation)
        tweet = ''.join(ch for ch in tweet if ch not in tanda_baca)
        return tweet

    def hapus_katadouble(s):
        pattern = re.compile(r"(\1{1,})", re.DOTALL)
        return pattern.sub(r"\1\1", s)

    #mengecilkan huruf
    tweet=tweet.lower()
    #hapus link
    tweet = re.sub(r'\\u\\w\\w\\w\\w', '', tweet)
    tweet=re.sub(r'http\S+', '',tweet)
    #hapus @username
    tweet=re.sub('@[^\s]+', '',tweet)
    #hapus #tagger
    tweet = re.sub(r'#([^\s]+)', r'\1', tweet)
    #hapus tanda baca
    tweet=hapus_tanda(tweet)
    #hapus angka dan angka yang berada dalam string
    tweet=re.sub(r'\w*\d\w*', '',tweet).strip()
    #hapus repetisi karakter
    tweet=hapus_katadouble(tweet)
    #stemming
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    tweet = stemmer.stem(tweet)
    clean_tweets.append(tweet)
```

ANALISA SENTIMEN

Analisa sentiment dilakukan untuk melihat bagaimana interpretasi dari data tweet yang telah dikumpulkan, apakah cenderung memberikan respon atau opini positif ataupun negatif. Adapun untuk melakukan analisa sentiment ini, perlu disiapkan file:

- kata_positif.txt -> mengandung kumpulan kata-kata bernilai positif
- kata_negatif.txt -> mengandung kumpulan kata-kata bernilai negatif

A. Import semua library yang dibutuhkan

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import tweepy
import re,string
import pandas as pd
import sqlite3
import matplotlib.pyplot as plt
import numpy as np
```

B. Buka file sql dari data yang sudah dikumpulkan

```
con = sqlite3.connect(file)
df = pd.read_sql_query("SELECT * from data_tweet", con)
con.close()
```

C. Membersihkan data

```
clean_tweets = []
tweets = []

for tweet in data['Tweets']:
    def hapus_tanda(tweet):
        tanda_baca = set(string.punctuation)
        tweet = ''.join(ch for ch in tweet if ch not in tanda_baca)
        return tweet

    def hapus_katadouble(s):
        pattern = re.compile(r"(\.|\{|\}|,)", re.DOTALL)
        return pattern.sub(r"\1\1", s)

    #mengecilkan huruf
    tweet=tweet.lower()
    #hapus link
    tweet = re.sub(r'\\u\\w\\w\\w\\w', '', tweet)
    tweet=re.sub(r'http\S+', '',tweet)
    #hapus @username
    tweet=re.sub('@[^\s]+', '',tweet)
    #hapus #tagger
    tweet = re.sub(r'#([^\s]+)', r'\1', tweet)
    #hapus tanda baca
    tweet=hapus_tanda(tweet)
    #hapus angka dan angka yang berada dalam string
    tweet=re.sub(r'\w*\d\w*', '',tweet).strip()
    #hapus repetisi karakter
    tweet=hapus_katadouble(tweet)
```

```
#stemming
factory = StemmerFactory()
stemmer = factory.create_stemmer()
tweet = stemmer.stem(tweet)
clean_data.append(tweet)
```

D. Analisis Sentiment

```
pos_list= open("./kata_positif.txt","r")
pos_kata = pos_list.readlines()
neg_list= open("./kata_negatif.txt","r")
neg_kata = neg_list.readlines()

Sentiment = []
for item in clean_data:
    count_p = 0
    count_n = 0
    for kata_pos in pos_kata:
        if kata_pos.strip() in item:
            count_p +=1
    for kata_neg in neg_kata:
        if kata_neg.strip() in item:
            count_n +=1

    Sentiment.append(count_p - count_n)
```

E. Visualisasi

```
#Menghitung mean, median, dan standart deviasi
print ("Nilai rata-rata: "+str(np.mean(Sentiment)))
print ("Median: "+str(np.median(Sentiment)))
print ("Standart Deviasi: "+str(np.std(Sentiment)))

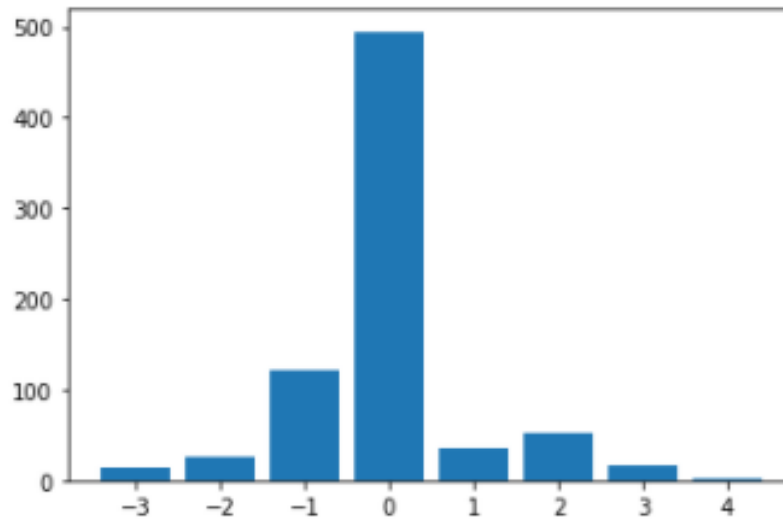
labels, counts = np.unique(Sentiment, return_counts=True)
plt.bar(labels, counts, align='center')
plt.gca().set_xticks(labels)
plt.show()
```

OUTPUT:

Nilai rata-rata: -0.030065359477124184

Median: 0.0

Standart Deviasi: 1.0099559925260357



Hasil analisa sentiment menunjukkan bahwa opini atau statement yang dihasilkan pada tweet yang diambil selama 3 hari (21-23 Oktober 2020) mengenai Cipta Kerja cenderung negatif, dengan nilai rata-rata sebanyak -0,03, median 0, dan standart deviasi 1,009.