



INTRODUCING

Ganesha Sentiment Analyzer

OLEH:
TIM ASENTITIES

ILHAM FIRDAUSI PUTRA
HAFIZH BUDIMAN
RESTU WAHYU KARTIKO

SEKOLAH TEKNIK ELEKTRO
DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG
2018

Cepat, efektif, dan tepat membantu
pemerintah mendapatkan umpan balik

Karya (GaneshaSentimentAnalyzer.exe, 35MB)
dapat dites pada:
[https://drive.google.com/
open?id=](https://drive.google.com/open?id=)



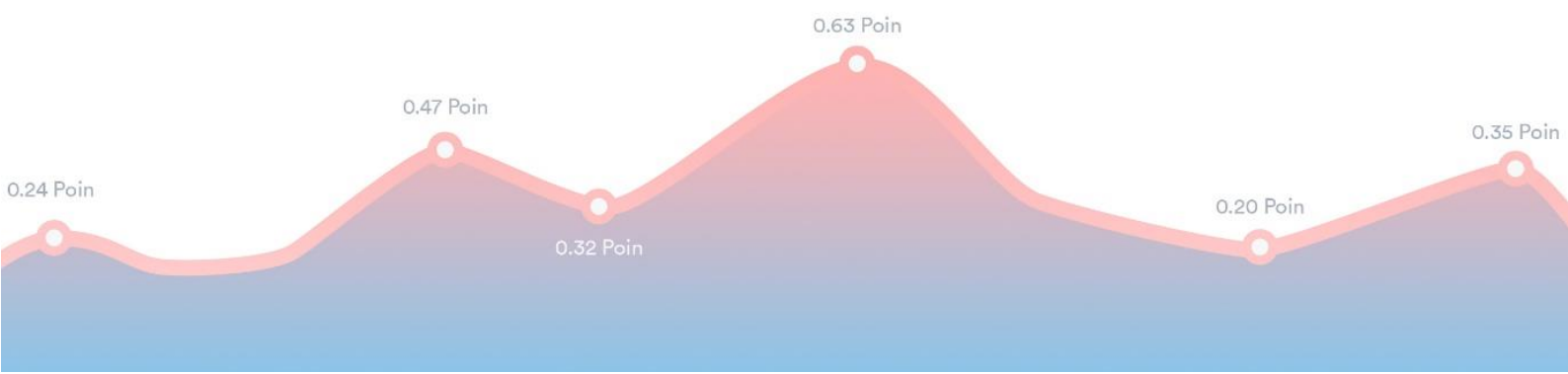
LATAR BELAKANG IDE

Kemajuan bidang teknologi informasi sedang berkembang dengan sangat pesat. Kelebihan dari penggunaan teknologi informasi adalah dalam hal kecepatan, kemudahan akses, biaya yang lebih terjangkau, dan sentralisasi data yang memudahkan dalam hal pengamatan dan evaluasi. Perkembangan ini turut mempengaruhi hampir seluruh faktor kehidupan masyarakat mulai dari berbelanja, memesan tiket, bahkan dalam segi pemerintahan. Penerapan teknologi informasi yang tepat dapat membantu meningkatkan produktifitas suatu sektor tidak terkecuali sektor pemerintah.

E-government atau electronic government adalah penggunaan teknologi informasi oleh pemerintah untuk meningkatkan mutu pelayanan dan kualitas program pemerintahan sehingga menjadi lebih berkualitas, praktis, dan efisien. Contoh E-goverment di Indonesia seperti e-procurement di Surabaya, One Stop Service(OSS) di Sragen, dan lain-lain.

Sektor pariwisata menyumbang hampir 10% untuk pendapatan nasional[6]. Pertumbuhan wisatawan mancanegara yang datang keindonesia mencapai 25% tiap tahunnya. Untuk itu, perbaikan sarana dan prasarana menjadi aspek penting untuk mengimbangi pertumbuhan pariwisata indonesia. Sedikitnya informasi tentang keluhan pengunjung menjadi salah satu kendala dalam peningkatan sarana dan prasarana suatu objek wisata. Padahal, pengunjunglah yang lebih tau tentang apa yang mereka inginkan mengenai suatu objek wisata. Metode survei langsung atau menyebar kuesioner kepada pengunjung membutuhkan waktu yang relatif lama dan biaya yang cukup besar. Belum lagi ketika objek survei merupakan wisatawan mancanegara.

Berdasarkan paparan masalah diatas, kami memberikan sebuah solusi yaitu dengan GaneshaSentimentAnalyzer. Aplikasi ini akan mencari semua data yang berhubungan dengan suatu objek wisata dalam tweet di Twitter oleh wisatawan mancanegara dalam rentang waktu yang akan ditentukan. Aplikasi ini selanjutnya dapat menganalisa data, menyimpan hasil analisa, dan menggambarkannya. Dengan ini, aplikasi diharapkan dapat membantu proses evaluasi objek wisata dan program pemerintah menjadi lebih mudah dan cepat.





TUJUAN DAN MANFAAT

Pembuatan aplikasi ini bertujuan untuk membantu pemerintah dalam mencari dan menyaring ulasan oleh wisatawan mancanegara mengenai sebuah objek wisata atau program pemerintah di Indonesia dalam rentang waktu yang akan ditentukan. Ulasan ini akan digunakan untuk meningkatkan mutu dan kualitas program pemerintah beserta sarana dan prasarana suatu objek wisata menjadi lebih baik.

tweets

Aug 1



Proses cepat dan tepat.

Dikarenakan sentimen didapatkan dari data-data di Twitter, ulasan dan umpan balik dapat diperoleh dengan cepat. Selain itu, proses analisa yang kami gunakan berakurasi tinggi. Proses ini memiliki akurasi dengan skor F1 0.96[1]. Jauh lebih baik dari penilaian manusia individual.



Mengetahui penilaian wisatawan tentang suatu objek wisata atau program pemerintah.

Survei ini mencari pendapat suatu wisatawan tentang objek wisata melalui tweet yang mereka buat. Dari tweet tersebut, kita bisa menyimpulkan apakah suatu objek wisata dalam rentang waktu tertentu dinilai baik atau kurang baik oleh para pengunjung.

Contoh: Pemerintah memiliki program Flores Exotic Tours yang bertujuan untuk meningkatkan pengunjung mancanegara ke Flores mulai pada tanggal 3 Maret 2018 hingga 3 Agustus 2018. Dengan memasukkan kata kunci Flores/Flores Exotic Tours, banyak sampel 1000, dan tanggal pencarian 3 Maret 2018 hingga 3 Agustus 2018, kita dapat melihat penilaian dan ulasan dari berbagai turis yang datang berkunjung dan membuat tweet mengenai hal ini.



Mencari bahan evaluasi tentang suatu objek wisata atau program pemerintah.

Aplikasi kami akan dapat mengenali suatu tweet dari wisatawan yang bermakna negatif. Tweet yang bersifat negatif dapat dijadikan sebagai bentuk lain dari keluhan untuk bahan evaluasi. Yang mana kemudian pemerintah bisa menindaklanjuti hal tersebut.

Selain itu, data dan analisa yang didapat akan disimpan dalam sebuah file. Hal ini memungkinkan pemerintah dapat membandingkan hasil suatu program dalam suatu objek wisata. Menjadikan program-program pemerintah selanjutnya lebih bermutu dan berkualitas.

Contoh: Dengan melakukan analisis sentimen menggunakan GaneshaSentimentAnalyzer pada 2 program pemerintah atau objek wisata pada rentang waktu tertentu, misal Flores Exotic Tours dan Visit Bali 2018. Pemerintah akan dapat membandingkan hasil analisis pada 2 program tersebut dan melakukan evaluasi.



Hemat biaya dan mudah.

Aplikasi kami bersifat seperti survei elektronik, namun, pengisi survey melakukannya secara tidak sadar karena data yang kami ambil adalah tweet dari wisatawan. Dengan aplikasi ini, biaya yang dikeluarkan jauh lebih terjangkau dari yang semula survei harus dilakukan secara manual dan menghabiskan banyak biaya akomodasi dan logistik.



METODOLOGI PENGEMBANGAN DAN ANALISIS KEBUTUHAN DESAIN KARYA

GaneshaSentimentAnalyzer akan mengambil data-data yang diperlukan untuk menganalisa sentimen dari sebuah keyword dalam rentang waktu yang ditentukan. Data-data selanjutnya akan dianalisa dan diklasifikasikan menjadi netral, positif, atau negatif. Data-data yang sudah diklasifikasikan selanjutnya akan divisualisasikan dan disimpan dalam sebuah file.

Dalam pengambilan data, GaneshaSentimentAnalyzer harus dapat berkomunikasi dengan platform Twitter dan menggunakan API-nya. Untuk hal ini kami menggunakan Tweepy[2]. Tweepy adalah sebuah python library yang dapat menghubungkan python dengan Twitter dan API-nya. Tweet yang didapat selanjutnya akan dibersihkan dari tweet-tweet sampah yang berisi spam-spam link dan hal-hal lain yang dapat mengganggu hasil analisis. Hal ini dilakukan dengan menggunakan library python re.

Dalam mengklasifikasi tweet, GaneshaSentimentAnalyzer harus dapat menganalisa dan mengklasifikasikannya menjadi netral, positif, atau negatif dengan akurasi yang tinggi. Untuk hal ini kami menggunakan VADER (Valence Aware Dictionary and sEntiment Reasoner)[1]. VADER adalah alat analisis sentimen berbasis leksikon dan aturan yang secara khusus disesuaikan dengan sentimen yang diungkapkan dalam media sosial.

Selanjutnya, GaneshaSentimentAnalyzer harus dapat menggambarkan dan menyimpan hasil analisa data-data yang sudah divisualisasikan. Untuk hal ini kami menggunakan NumPy[3] dan Matplotlib[4]. NumPy adalah python library yang menambahkan dukungan untuk array dan matriks yang multi dimensi dan besar, bersama dengan koleksi fungsi matematika untuk beroperasi pada array-nya. Matplotlib adalah python library yang berguna untuk menggambarkan data dalam 2 dimensi. Lalu kami menggunakan python library csv untuk menyimpan data-data yang didapatkan dalam bentuk file berekstensi csv.

Terakhir, GaneshaSentimentAnalyzer harus dapat dijalankan di platform umum. Untuk hal ini, kami menggunakan PyInstaller. PyInstaller adalah sebuah program yang dapat mengubah file .py menjadi sebuah executable. Executable ini nantinya akan dapat dijalankan dalam semua komputer bersistem operasi Windows tanpa harus memiliki python dan modul-modulnya terinstal.





Batasan Karya yang Dikembangkan

1. GaneshaSentimentAnalyzer dapat menganalisa data dengan akurasi skor F1 0.96^[1]. Jauh lebih baik dari penilaian manusia individual.
2. GaneshaSentimentAnalyzer dapat mengambil data-data yang diperlukan untuk menganalisa sentimen dari sebuah *keyword* dalam rentang waktu yang ditentukan.
3. GaneshaSentimentAnalyzer dapat menggambarkan hasil analisa sentimen dari sebuah *keyword* dan menyimpannya dalam sebuah *file*.
4. GaneshaSentimentAnalyzer dapat dijalankan pada semua perangkat komputer dengan sistem operasi *Windows* tanpa harus memiliki *python* terinstal dan segala modul-modulnya.
5. GaneshaSentimentAnalyzer harus terhubung dengan internet.
6. Berdasarkan tujuan dari dikembangkannya GaneshaSentimentAnalyzer, tweet yang disaring hanyalah tweet-tweet berbahasa Inggris.

Implementasi

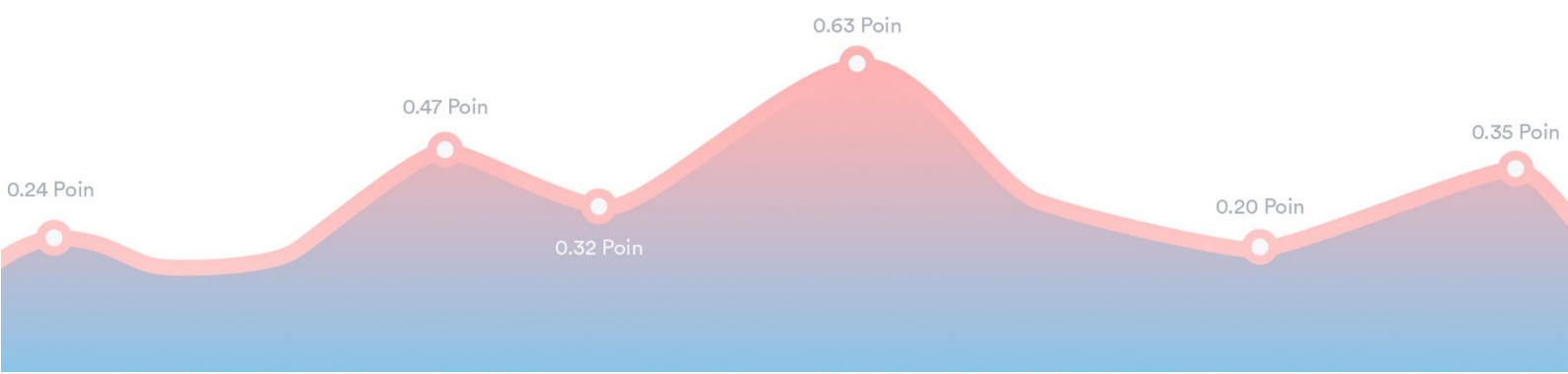
Langkah pertama implementasi adalah melakukan import semua *library* yang dibutuhkan serta inialisasi awal *key* dan *token* yang dibutuhkan untuk autentifikasi Tweepy nantinya.

```
# Ganesha Sentiment Analyzer
# Ilham Firdausi Putra, 13516140
# Hafizh Budiman, 13516137
# Restu Wahyu Kartiko, 13516155
# Informatics Engineering, Bandung Insitute of Technology
# for E-Government FindIT 2018

import tweepy
import re
import nltk
import pyforms
import csv
import numpy as np
import matplotlib.pyplot as plt
from pyforms import BaseWidget
from pyforms.Controls import ControlText
from pyforms.Controls import ControlTextArea
from pyforms.Controls import ControlButton
from nltk.sentiment.vader import SentimentIntensityAnalyzer

consumer_key = 'w5hzNHmPM7vRDrtQR71sCA4HA'
consumer_secret = 'QshW17aF7jGw6Tf4e6IDA2i1pVepZoEACNEYLsptRU4U6CPiaB'
access_token = '611656833-0a546jfvS1Ay0iq4msT9fPXs2u8L19aF9DQVQKdh'
access_token_secret = 'wX6Fmkst5o6IpMo8ecSekft5czQLCnI1jZ9146yDejCcD'
```

Gambar 1. Import *library* dan inialisasi *key*.





Ganesha Sentiment Analyzer

Implementasikan konstruktor dalam kelas GUI (PyForms) yang berisi inisialisasi elemen-elemen dari *window GUI* yang akan dibuat input *keyword* dari lokasi pariwisata yang ingin di analisis sentimennya, banyak sampel, tanggal awal & akhir pencarian data (YYYY-MM-DD). Masukkan input tersebut sebagai argumen ke dalam fungsi *ambilTweet* yang akan mengambil data dan mengeluarkan output hasil-hasil sentimen.

```
class GaneshaSentimentAnalyzer(BaseWidget):  
  
    def __init__(self):  
        super(GaneshaSentimentAnalyzer, self).__init__('Ganesha Sentiment Analyzer')  
  
        self.formset = [('_location', '_num_of_samples',  
                        '_start_date', '_end_date'), '_button', '_results']  
  
        # Definition of the forms fields  
        self._location = ControlText('Lokasi Pariwisata', 'Default value')  
        self._num_of_samples = ControlText('Banyak Sample')  
        self._start_date = ControlText('Tanggal Awal Pencarian')  
        self._end_date = ControlText('Tanggal Akhir Pencarian')  
        self._results = ControlTextArea('Hasil Sentiment Analysis')  
        self._button = ControlButton('Tekan tombol ini untuk mulai menganalisa!')  
  
        #Define the button action  
        self._button.value = self.__buttonAction
```

Gambar 1. Konstruktor kelas

Implementasikan *method* *buttonAction* yang akan menjalankan algoritma utama apabila tombol ditekan setelah mengisi input pada kolom teks yang tersedia.

```
def __buttonAction(self):  
    """Button action event"""  
    # def ambilTweet(lokasiPariwisata, banyakSampel, tanggalAwal, tanggalAkhir):  
    lokasiPariwisata, banyakSampel, tanggalAwal, tanggalAkhir = self._location.value, self._num_of_samples.value, self._start_date.value, self._end_date.value  
    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)
```

Gambar 2. Implementasi *Method* *buttonClick* yang akan memanggil *main*





Ganesha Sentiment Analyzer

Pada fungsi `ambilTweet`, mula-mula autentifikasi Tweepy, penulis ke file, array hasil, daftar tweet yang masuk, dan pengecek menggunakan VADER akan dipersiapkan.

```
def ambilTweet(lokasiPariwisata, banyakSampel, tanggalAwal, tanggalAkhir):
    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)

    csvFile = open('hasilcrawl.csv', 'w')
    csvWriter = csv.writer(csvFile)

    hasil = []
    tweetList = []
    pengecek = SentimentIntensityAnalyzer()
```

Gambar 3. Awal fungsi `ambilTweet`.

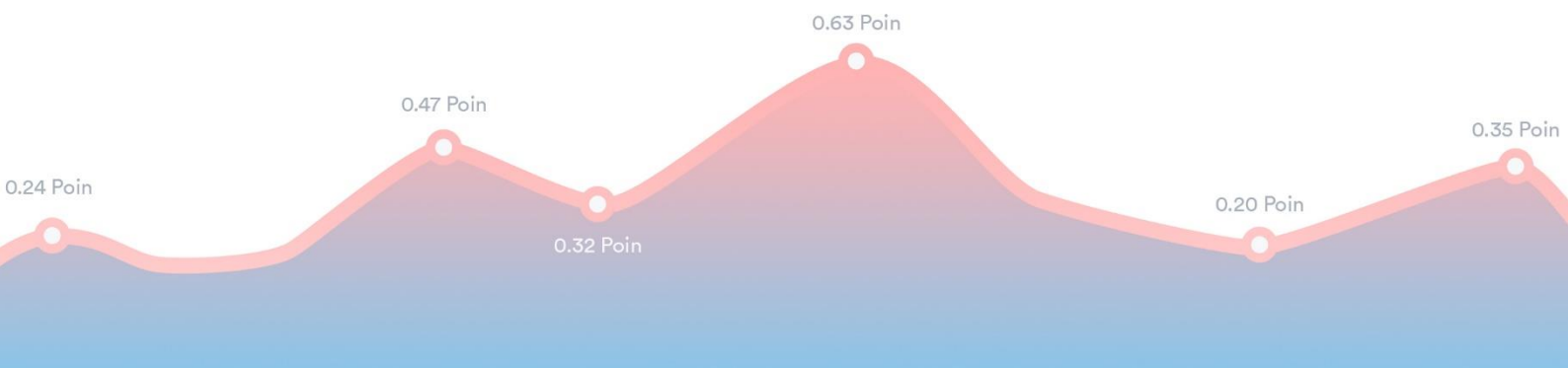
Selanjutnya, dimulai proses *looping* pengambilan data tweet. Data tweet yang masuk akan dibersihkan, lalu di cek adanya *keyword* setelah dibersihkan, tanggal, dan kemungkinan duplikatnya.

```
def bersihkanTweet(tweet):
    tweet = tweet.lower()
    tweet = re.sub('((www\.[^\s]+)|(https?://[^\s]+))', 'URL', tweet)
    tweet = re.sub('@[^\s]+', 'AT_USER', tweet)
    tweet = re.sub('[\s]+', ' ', tweet)
    tweet = re.sub(r'#([^\s]+)', r'\1', tweet)
    tweet = tweet.strip('\n')
    return tweet
```

Gambar 4. Pembersih tweet.

```
for tweet in tweepy.Cursor(api.search, q=lokasiPariwisata, lang="en").items():
    tweetBersih = bersihkanTweet(tweet.text)
    tanggal = str(tweet.created_at)
    tanggal = tanggal[0:10]
    if (tanggal < tanggalAwal):
        break
    if (tanggal > tanggalAkhir) or not(lokasiPariwisata.lower() in tweetBersih) or
    (tweetBersih in tweetList):
        continue
```

Gambar 5. Awal looping.





Ganesha Sentiment Analyzer

Jika semua syarat terpenuhi, tweet akan dimasukan ke daftar tweet dan dicek nilai polaritas tweet tersebut. Dari pengecekan nilai polaritas, tweet akan di print ke console dengan hasil analisisnya. Nilai polaritas tersebut akan dimasukan ke array hasil, dan tweet bersama nilai polaritasnya akan disimpan ke file .csv. Diakhir, akan dilakukan pengecekan banyak sampel yang sudah diambil dan tanggal yang sudah diproses. Setelah proses looping selesai, penulis ke file ditutup dan array hasil dikembalikan.

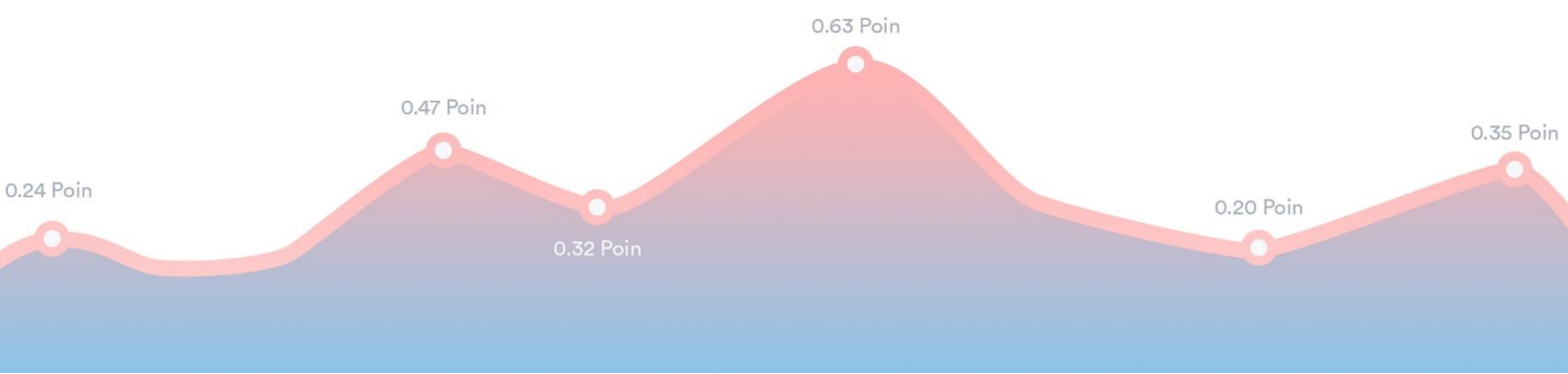
```
tweetList.append(tweetBersih)
print('~~~~~')
print(len(tweetList))
print (tweet.created_at, tweet.text)
nilaiSentimen = {}
nilaiSentimen = pengecek.polarity_scores(tweetBersih)
keynya = "compound"

if(nilaiSentimen[keynya] < 0):
    print('-->Tweet bernilai negatif!')
elif(nilaiSentimen[keynya] > 0):
    print('-->Tweet bernilai positif!')
else:
    print('-->Tweet bernilai netral!')
hasil.append(nilaiSentimen[keynya])
csvWriter.writerow([tweetBersih.encode('utf-8'),nilaiSentimen[keynya]])
print('~~~~~')
if (len(tweetList) >= int(banyakSampel)):
    break

csvFile.close()
return hasil
```

Gambar 6. Proses analisis

Output dari fungsi ambilTweet() tersebut adalah sebuah array yang berisi sentimen-sentimen wisatawan mancanegara. Output tersebut selanjutnya akan dirangkum dan divisualisasikan menggunakan NumPy dan Matplotlib.





```
banyakNeutral = 0
banyakPositif = 0
banyakNegatif = 0
for x in result:
    if(x < 0):
        banyakNegatif+=1
    elif(x > 0):
        banyakPositif+=1
    else:
        banyakNeutral+=1

print('Hasil Sentiment Analysis di ' + lokasiPariwisata)
print('Banyak Positif = ', banyakPositif , ' ( ' , (float(banyakPositif/int(
    banyakSampel)) * 100) , ' % )' )
print('Banyak Netral = ', banyakNeutral , ' ( ' , (float(banyakNeutral/int(
    banyakSampel)) * 100) , ' % )' )
print('Banyak Negatif = ', banyakNegatif , ' ( ' , (float(banyakNegatif/int(
    banyakSampel)) * 100) , ' % )' )

plt.style.use('classic')
my_xticks = ['Negative','Neutral','Positive']
y_pos = np.arange(3)
x = np.array([0,1,2])
y = np.array([banyakNegatif,banyakNeutral,banyakPositif])

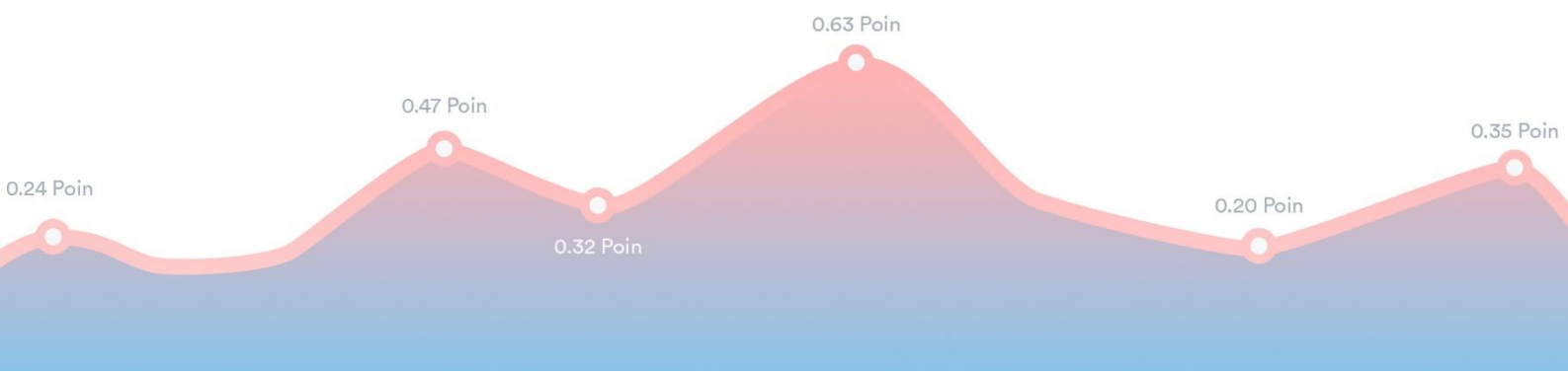
plt.bar(y_pos,y,align='center')
plt.ylabel('Banyak Data')
lokasiPariwisata = 'Sentiment Analysis Chart in ' + lokasiPariwisata
plt.title(lokasiPariwisata)
plt.xticks(x, my_xticks)
plt.show()
```

Gambar 7. Merangkum dan menggambarkan hasil

Terakhir, ubah file .py ini menjadi sebuah *executable*. Sehingga aplikasi ini akan dapat dijalankan dalam semua komputer bersistem operasi Windows tanpa harus memiliki python dan modul-modulnya terinstal.

```
#Execute the application
if __name__ == "__main__":
    pyforms.start_app( GaneshaSentimentAnalyzer )
```

Gambar 8. Implementasi *main* untuk mengeksekusi *program*





Ganesha Sentiment Analyzer

Contoh-contoh hasil penggunaan dengan *keyword* bali dengan banyak sampel 200 dan rentang waktu 2018-03-01 sampai 2018-03-02:

Tweet Positif Keluaran di Kolom Teks.

```
~~~~~
19
2018-04-05 13:24:39
thank you AT_USER at crown bali villa-seminyak for awesome honeymoon. . . honeymoon... URL -->Tweet bernilai positif!
~~~~~
```

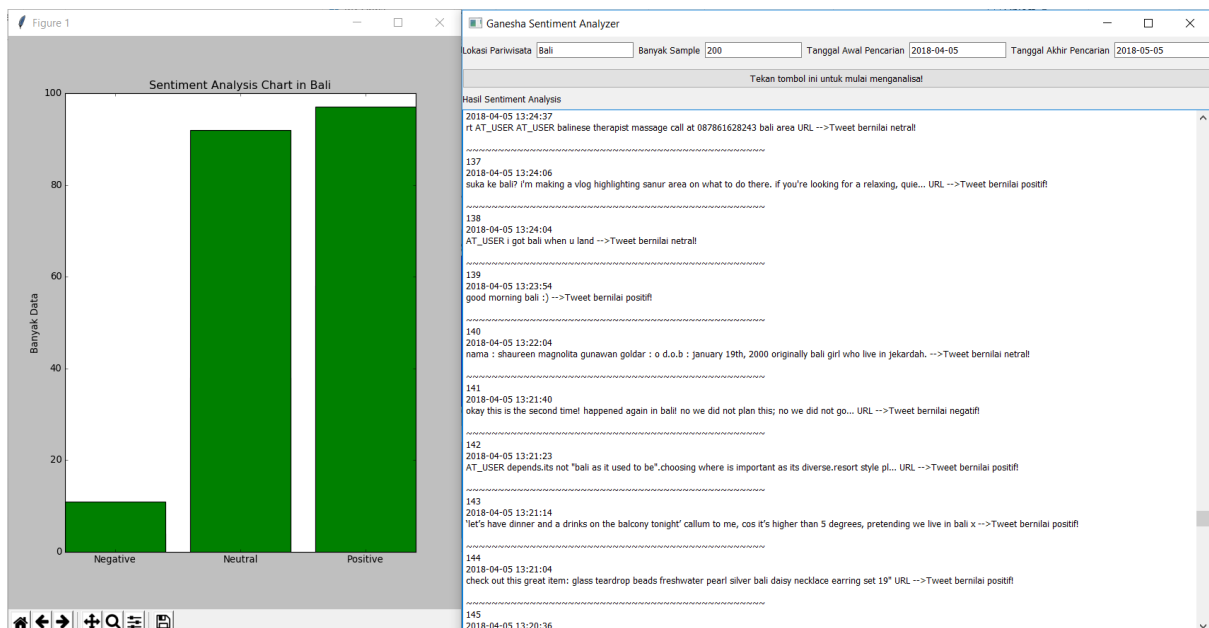
Tweet Negatif Keluaran di Kolom Teks.

```
~~~~~
141
2018-04-05 13:21:40
okay this is the second time! happened again in bali! no we did not plan this; no we did not go... URL -->Tweet bernilai negatif!
~~~~~
```

Tweet Netral Keluaran di Kolom Teks.

```
~~~~~
160
2018-04-05 13:09:04
latest update - bali surf report, east coast to west coast 4th – 5th april 2018 URL bali surf -->Tweet bernilai netral!
~~~~~
```

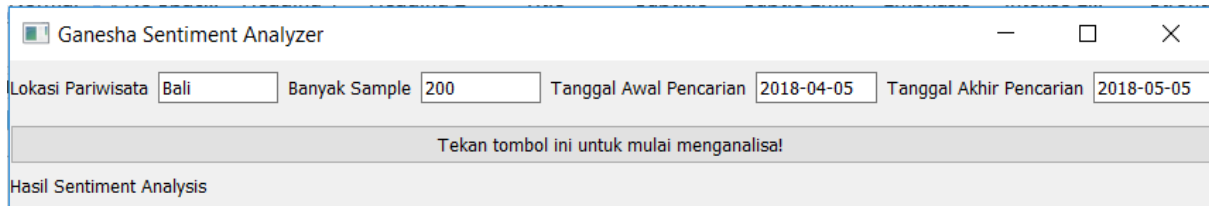
Hasil Akhir Keluaran di GUI.





Ganesha Sentiment Analyzer

Detail Input pada Window GUI Aplikasi



Ganesha Sentiment Analyzer

Lokasi Pariwisata Banyak Sample Tanggal Awal Pencarian Tanggal Akhir Pencarian

Tekan tombol ini untuk mulai menganalisa!

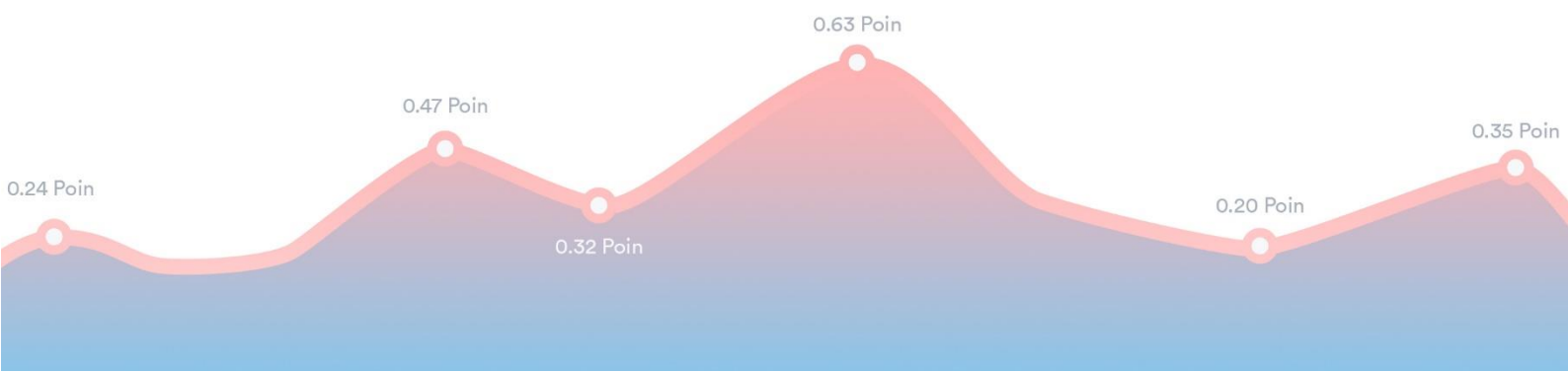
Hasil Sentiment Analysis

Detail Hasil Akhir

Hasil Sentiment Analysis di Bali
Banyak Positif = 97 (48.5 %)
Banyak Netral = 92 (46.0 %)
Banyak Negatif = 11 (5.5 %)

Hasil lengkap output ke file .csv & di console dapat dilihat pada:

<https://drive.google.com/open?id=1Kjp325YkOoUP4i5R6-pOkQPxYScPDoat>





Screenshot Mock-up Interface.

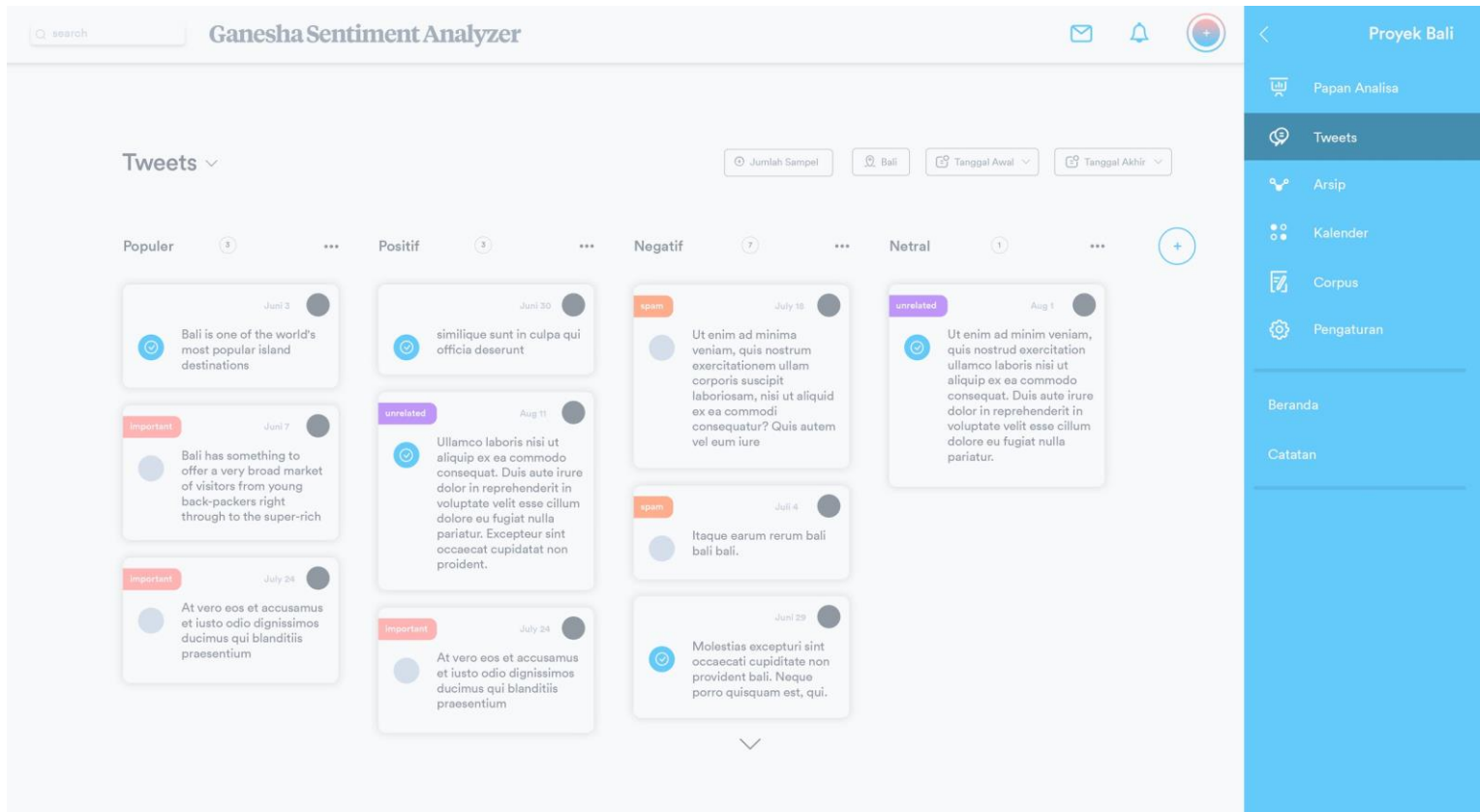
Dalam versi yang siap diterbitkan nantinya, agar penggunaan aplikasi mudah dan nyaman, aplikasi dapat direalisasikan dengan sebuah *graphical user interface* (GUI) yang tampak seperti berikut



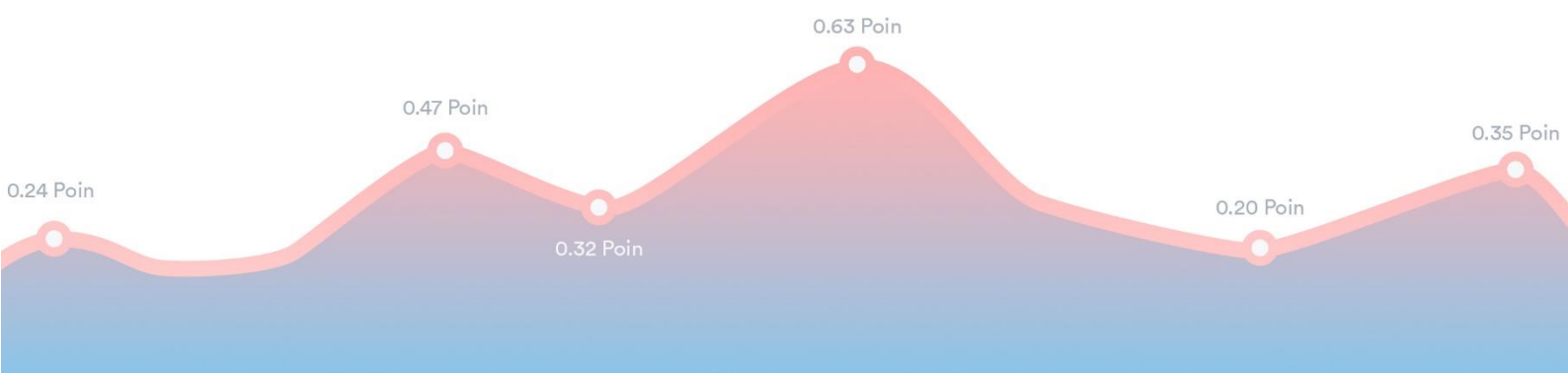
Gambar 8. Mock-up interface papan hasil analisis.



Ganesha Sentiment Analyzer



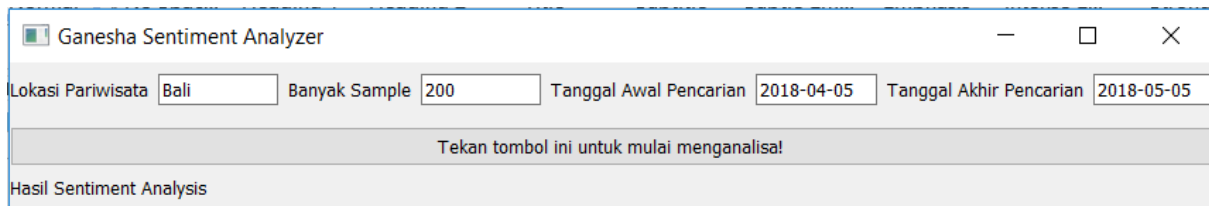
Gambar 9. Mockup interface visualisasi kategori tweet.





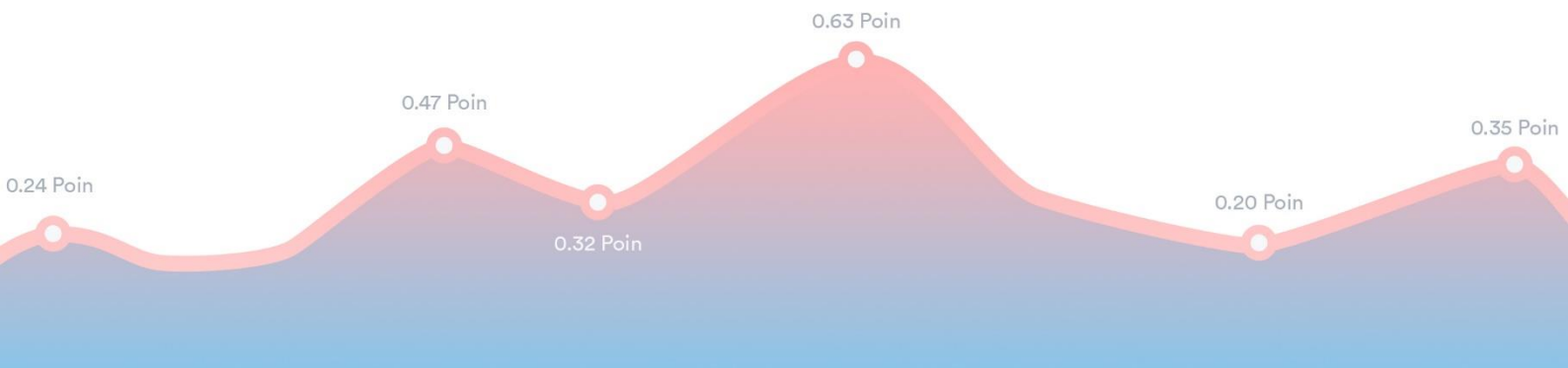
Dokumentasi Cara Penggunaan.

1. Pastikan komputer anda tersambung dengan jaringan internet
2. Buka file *executable* GaneshaSentimentAnalyzer.exe dengan menekan 2 kali pada *file* tersebut.
3. Ketikkan lokasi pariwisata yang ingin dianalisis, kemudian tekan tombol *enter*
4. Ketikkan jumlah sampel yang dikehendaki, kemudian tekan tombol *enter*
5. Ketikkan tanggal awal sampel yang dikehendaki, kemudian tekan tombol *enter*
6. Ketikkan tanggal akhir sampel yang dikehendaki, kemudian tekan tombol *enter*



Gambar 10. Contoh masukan program.

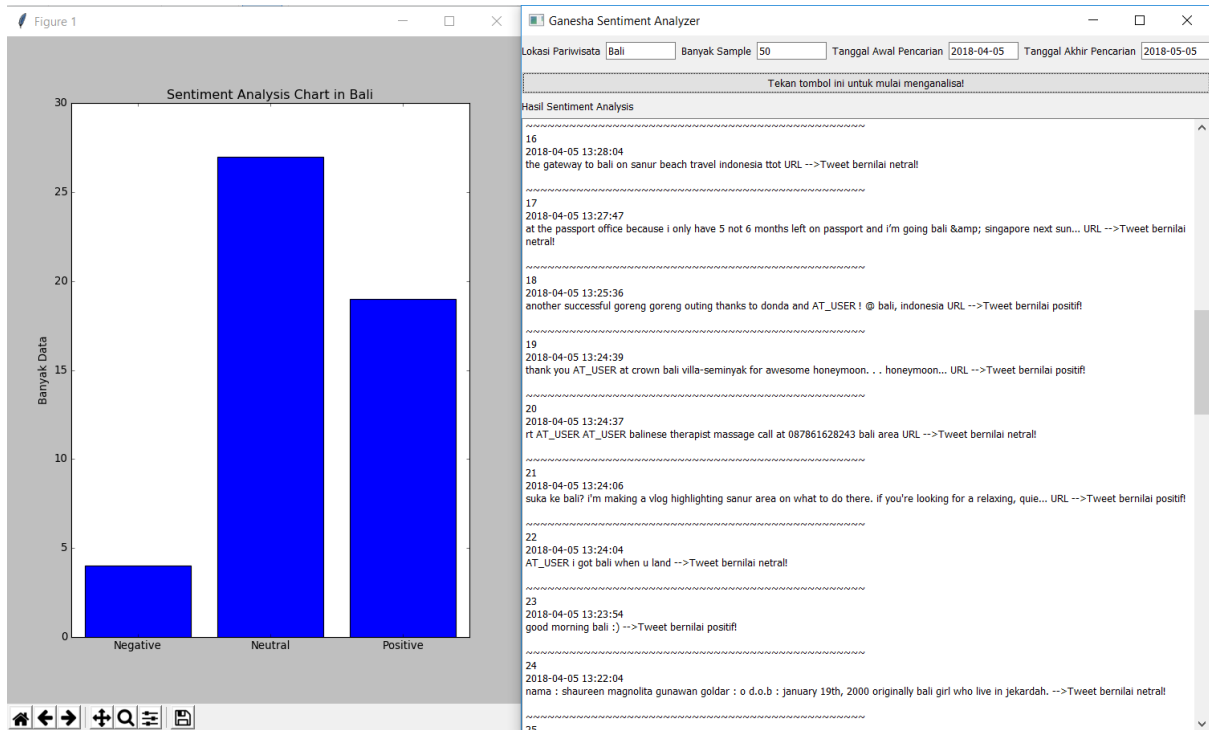
7. Program akan melakukan pencarian berdasarkan lokasi pariwisata tersebut dan menganalisis serta mengklasifikasikan *tweet* berdasarkan masukan pengguna dan melakukan penilaian ke dalam 3 kategori yakni: positif, netral, dan negatif. Proses ini akan memakan waktu tergantung banyaknya jumlah sampel dan spesifikasi komputer serta koneksi internet pengguna. Setelah proses selesai, program akan menuliskannya ke layar dan ke dalam *file* hasilcrawl.csv.
8. Program juga akan menuliskan persentase jumlah 3 kategori tersebut, dan menggambarkannya dalam bentuk *bar chart*.





Ganesha Sentiment Analyzer

Gambar 11. Contoh keluaran program berupa tweet dan persentase kategori, dan *bar chart*





Referensi

1. Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.
2. Novalić, Ahmet. (2013, January 23rd). Introduction to tweepy, Twitter for Python. Diambil dari <https://www.pythoncentral.io/introduction-to-tweepy-twitter-for-python/>.
3. Halaman resmi NumPy. Diambil dari <http://www.numpy.org/>.
4. Halaman resmi Matplotlib. Diambil dari <https://matplotlib.org/>.
5. Halaman resmi PyInstaller. Diambil dari <http://www.pyinstaller.org/>.
6. Halaman resmi PyForms, Diambil dari <https://pyforms.readthedocs.io/en/v3.0/>.
7. Maulana, Arief. (2017, 24 Januari). Menpar Arief Yahya, "Pariwisata Penyumbang PDB, Devisa, dan Lapangan Kerja yang Mudah dan Murah di Indonesia". Diambil dari <http://www.unpad.ac.id/2017/08/menpar-arief-yahya-pariwisata-penyumbang-pdb-devisa-dan-lapangan-kerja-yang-mudah-dan-murah-di-indonesia/>.

