



EXPLORATORY DATA ANALYSIS



ANGGOTA KELOMPOK

1. Medifa Puspaningrum (2106634944)
2. Musarrofah Kurnia (2106652543)
3. Nadia Amelia Kirana (2106653224)
4. Zahrah Aulia Putri (2106724896)

Pendahuluan

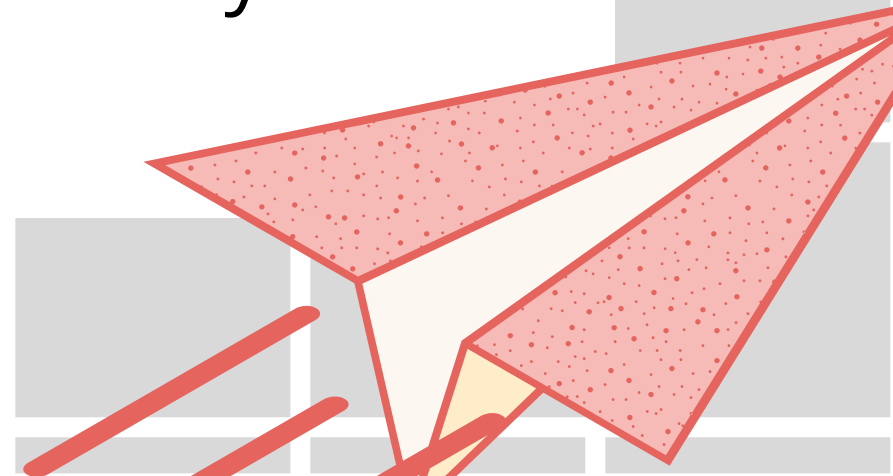


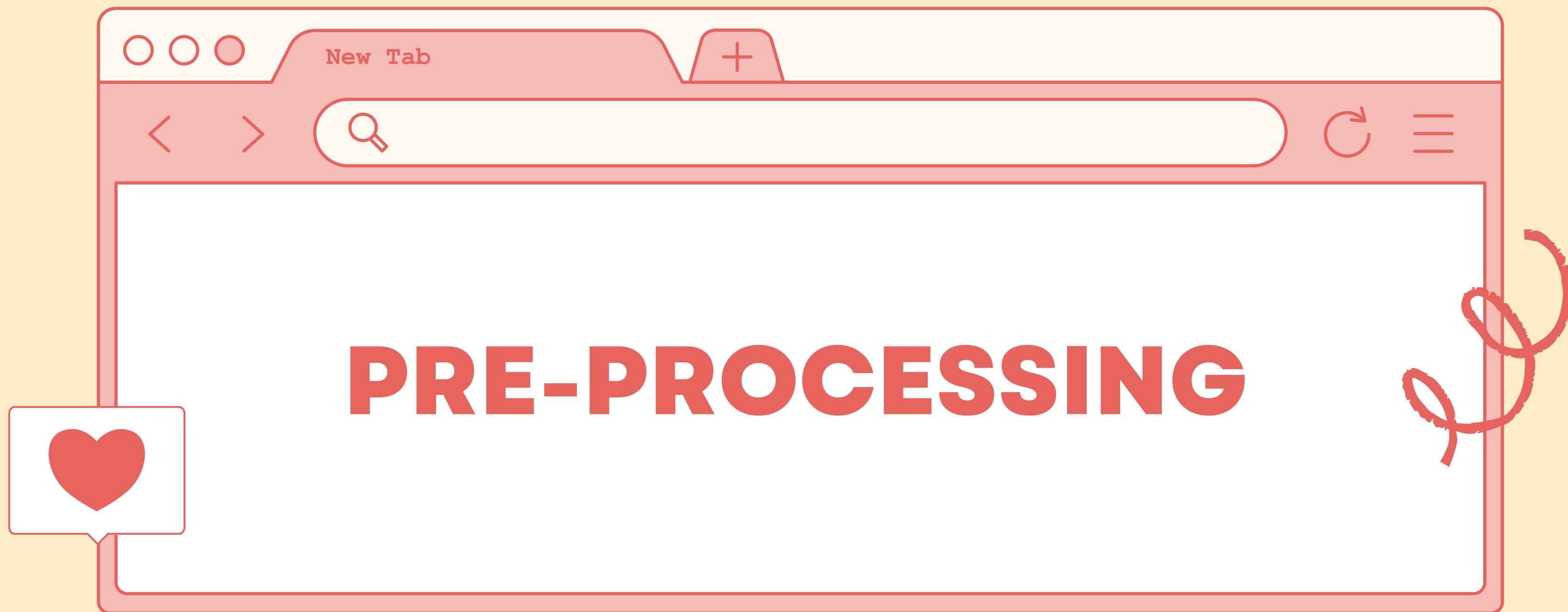
Permasalahan yang dibahas oleh kelompok kami adalah mengenai "tiket borobudur OR borobudur kontra OR borobudur setuju". Permasalahan ini merupakan topik hangat yang kerap diperbincangkan di kalangan masyarakat. Oleh karena itu, kelompok kami memutuskan untuk membahas serta menganalisis topik ini.

Sumber data kelompok kami berasal dari Twitter yang diambil secara real. kelompok kami ingin melihat bagaimana pendapat masyarakat mengenai topik yang kami pilih. Setiap pembahasan di masyarakat akan memiliki pro, kontra, dan netral di dalamnya.

Kelompok kami memilih platform media Twitter karena Twitter merupakan media yang menyediakan wadah untuk berbagi pendapat secara publik.

Ukuran data yang kami ambil sebanyak 2000 data. Menurut kelompok kami, jumlah tersebut sudah cukup untuk melihat variasi komentar atau pendapat masyarakat mengenai topik ini.





Data Gathering

```
▶ qry = 'tiket borobudur OR borobudur kontra OR borobudur setuju'
N = 2000 # jumlah N tweet yang ingin diambil
T = []
for post in tqdm(tweepy.Cursor(api.search_tweets, q=qry, lang='id', tweet_mode='extended').items(N)):
    T.append(post)

"Done"
```

Pengumpulan data dengan kata kunci “tiket borobudur OR borobudur kontra OR borobudur setuju” dan mengambil sebanyak 2000 data.

Tokenisasi



```
# tokenisasi
from nltk.tokenize import word_tokenize
tokens = [word_tokenize(i) for i in tweet_bersih]
tokens[:3]
```

Tokenisasi untuk memecah kalimat ke dalam kepingan - kepingan kata.

Slangfix dan Stopwords

```
# slangfix
df=open('data/slang.txt','r',encoding="utf-8", errors='replace')
slangS = df.readlines(); df.close()
slangS = [t.strip('\n').strip() for t in slangS]

slangS = [t.split(":") for t in slangS]
slangS = [[k.strip(), v.strip()] for k,v in slangS]
slangS = {k:v for k,v in slangS}

slangS['ringan2'] = 'ringan'
slangS['pikir2'] = 'pikir'
slangS['dar'] = 'dari'
slangS['twit'] = 'tweet'
slangS['denga'] = 'dengan'
slangS['sodara'] = 'saudara'
slangS['tuju'] = 'setuju'
slangS['bangkitbersamaet'] = 'bangkit bersama'

[ ] for j in range(len(tokens)):
    for i,t in enumerate(tokens[j]):
        if t in slangS.keys():
            tokens[j][i] = slangS[t]

tokens[:5]
```

```
# stopwords
df=open('data/stopwords_id.txt','r',encoding="utf-8", errors='replace')
id_stop = df.readlines()
df.close()
id_stop = [t.strip().lower() for t in id_stop]

id_stop = set(id_stop)

[ ] id_stop.add('rt')
    id_stop.add('nya')
    id_stop.add('kah')
    id_stop.add('iya')
    id_stop.add('sih')
    id_stop.add('deh')
    id_stop.add('ya')

[ ] # stopwords
    for j in range(len(tokens)):
        tokens[j] = [t for t in tokens[j] if t not in id_stop]
    tokens[30:40]
```

Slangfix digunakan untuk mengubah kata - kata yang mengalami kesalahan penulisan, seperti sodara, twit, dan lainnya. Stopwords removal digunakan untuk menghilangkan kata yang tidak digunakan dalam pengolahan data.

Lemmatization

```
▶ from Sastrawi.Stemmer.StemmerFactory import StemmerFactory  
stemmer = StemmerFactory().create_stemmer()
```

```
[ ] # memberikan kata dasar  
for j in range(len(tokens)):  
    tokens[j] = [stemmer.stem(t) for t in tokens[j]]
```

lemmatization untuk melakukan normalisasi pada kata yang berdasarkan pada bentuk dasar

Permasalahan :

Karena kami menggunakan bahasa Indonesia, pada slangfix dan stopwords banyak kata yang ditambahkan secara manual dan pada Lemmatization terdapat beberapa kata yang tidak benar, seperti kata pengunjung menjadi unjung

Pengolahan dan Analisis Data

```
[ ] # menggabungkan string-string menjadi satu kalimat agar bisa di cek polarisasinya
tokens2 = []
for j in range(len(tokens)):
    word = " ".join(tokens[j])
    tokens2.append(word)
```

```
▶ from textblob import TextBlob
tweet_positif = []
tweet_netral = []
tweet_negatif = []
for tweet in tokens2:
    analisis = TextBlob(tweet)

    if analisis.sentiment.polarity > 0.0:
        tweet_positif.append(tweet)
    elif analisis.sentiment.polarity == 0.0:
        tweet_netral.append(tweet)
    else:
        tweet_negatif.append(tweet)
```

```
[ ] # hasil sentimen
print('Hasil Analisis Sentimen')
print('Positif: ', len(tweet_positif), 'dengan persentase {} %'.format(100*len(tweet_positif)/len(tokens2)))
print('Netral: ', len(tweet_netral), 'dengan persentase {} %'.format(100*len(tweet_netral)/len(tokens2)))
print('Negatif: ', len(tweet_negatif), 'dengan persentase {} %'.format(100*len(tweet_negatif)/len(tokens2)))
```

```
Hasil Analisis Sentimen
Positif:  3 dengan persentase 0.15 %
Netral:  1984 dengan persentase 99.2 %
Negatif:  13 dengan persentase 0.65 %
```

Proses analisis data yang digunakan adalah sentiment analysis. Analisis ini digunakan untuk melihat reaksi publik terhadap topik, yaitu naiknya harga tiket untuk naik ke Candi Borobudur. Dari hasil penggunaan fungsi .sentiment.polarity didapatkan hasil sentimen berupa:

- Positif: 3 dengan persentase 0.15%
- Netral: 1984 dengan persentase 99.2%
- Negatif: 13 dengan persentase 0.65%

Proses visualisasi data menggunakan Voyant Tools

Hasil Visualisasi Data

catatan:

tuju = setuju

unjung = pengunjung

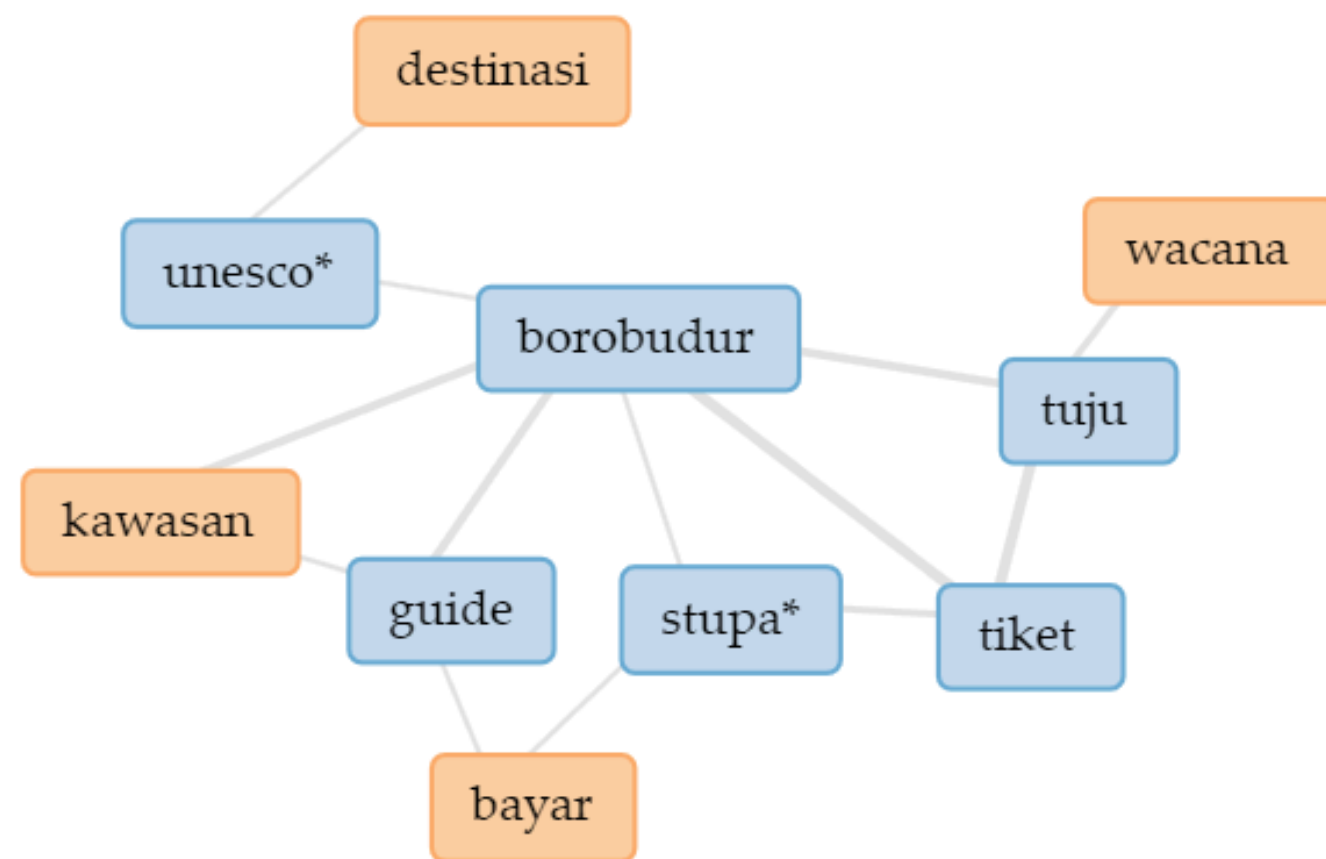
```

▶ filename = 'tweet_positif_borobudur_2022.txt'
  with open(filename, 'w') as f:
    for T in tweet_positif:
      f.write(T+'\n')

  "Tweets Saved!"

```

Visualisasi Tweet Positif



Interpretasi:

Dapat dilihat, terdapat hubungan/link antara kata 'bayar' dengan kata 'guide' dan 'stupa'. Setelah membaca secara sekilas raw tweets, kami memahami bahwa terdapat kesetujuan terhadap naiknya harga tiket untuk naik ke Candi Borobudur. Mereka setuju karena semakin banyak stupa-stupa di Candi Borobudur yang rusak. Selain itu, dengan harga tiket yang direncanakan, akan disediakan tour guide dan fasilitas lainnya di kawasan Borobudur. Di sisi lain, Candi Borobudur termasuk warisan budaya yang diakui oleh UNESCO. Menurut netizen, warisan budaya yang berada di luar negeri memang memiliki harga yang mahal untuk dikunjungi sehingga mereka setuju jika harga tiket Candi Borobudur dinaikkan.

Interpretasi:

Dapat dilihat bahwa kata 'ibadah', 'wisata', dan 'tutup' merupakan kata yang dengan ukuran paling besar yang mencerminkan bahwa kata tersebut banyak digunakan. Setelah membaca sekilas raw tweets, dapat dipahami bahwa tweet-tweet yang termasuk netral ini memberikan opini bahwa seharusnya Candi Borobudur ditutup saja dan tidak dijadikan tempat wisata, melainkan sebagai tempat ibadah bagi saudara-saudara beragama Buddha.

```
[ ] filename = 'tweet_netral_borobudur_2022.txt'
    with open(filename, 'w') as f:
        for T in tweet_netral:
            f.write(T+'\n')
```

"Tweets Saved!"

Visualisasi Tweet Netral



```
[ ] filename = 'tweet_negatif_borobudur_2022.txt'
with open(filename, 'w') as f:
    for T in tweet_negatif:
        f.write(T+'\n')
```

"Tweets Saved!"

Visualisasi Tweet Negatif



Interpretasi:

Berdasarkan visualisasi yang ada, kita dapat mengetahui bahwa kata 'tuju' merupakan kata dengan ukuran yang paling besar. Berikutnya, yang paling besar ukurannya adalah kata 'dimahalin' dan 'random'. Hal ini mencerminkan kata-kata tersebut sering digunakan. Berdasarkan raw tweets, kami memahami bahwa para user Twitter yang memiliki pandangan kontra/tidak setuju terhadap kebijakan harga tiket naik Candi Borobudur yang dimahalkan menganggap kebijakan tersebut sebagai sesuatu yang random atau tidak jelas.

Kesimpulan

Kesimpulan yang kami dapatkan dari data ini adalah reaksi/sentimen masyarakat terhadap tiket Candi Borobudur yang melonjak naik. Terdapat pro, kontra, dan netral di kalangan masyarakat terkait permasalahan tersebut. Banyak yang mengatakan bahwa ini merupakan cara untuk menjaga kelestarian Candi Borobudur. Akan tetapi, menaikkan harga tiket akan mengurangi pengunjung yang datang dan berdampak pada ekonomi masyarakat.

Terima Kasih

