

ANALISIS SENTIMEN TERHADAP PARA KANDIDAT PRESIDEN 2024 BERDASARKAN NETIZEN PENGGUNA TWITTER DENGAN METODE *DATA MINING* DAN *TEXT MINING*

¹Raihanda Luthfiansyah dan ²Budi Wasito

^{1,2}Program Studi Sistem Informasi

¹raihandaluthfiansyah@gmail.com

²budi.wasito@kwikkiangie.ac.id

Abstrak

Dalam era digital modern saat ini, internet telah menjadi bagian tak terpisahkan dari kehidupan manusia, dan salah satu fenomena yang muncul adalah jejaring sosial media. Twitter merupakan salah satu platform jejaring sosial yang paling populer di Indonesia dan digunakan oleh berbagai lapisan masyarakat, termasuk dalam mengutarakan opini terhadap calon kandidat presiden Indonesia pada pemilihan umum 2024. Oleh karena itu, dilakukan penelitian dengan tema Analisis Sentimen pengguna Twitter terhadap kandidat pencalonan presiden 2024 untuk mengetahui bagaimana opini publik tentang kandidat-kandidat tersebut dalam bentuk sentimen positif, negatif, atau netral, sehingga dapat diketahui calon kandidat yang memiliki image positif, negatif, dan netral di mata masyarakat pengguna Twitter. Data Mining merupakan sebuah proses untuk mencari informasi yang berguna dalam penyimpanan yang besar berdasarkan metodologi CRISP-DM (Cross Industry Standard Process For Data Mining). Text Mining sama halnya seperti Data Mining dalam hal tujuan dan menggunakan proses yang sama, namun pada text mining input prosesnya adalah kumpulan file data yang tidak terstruktur (dokumen Word, file PDF, cuplikan teks, file XML, dan sebagainya). Sumber data dalam penelitian ini adalah data tweet pada platform Twitter pada bulan Februari 2023. Data ini diambil langsung dengan cara Crawling Dataset melalui Application Programming Interface (API) pada website developer.twitter.com. Penelitian ini membandingkan hasil Analisis Sentimen menggunakan Machine learning berbasis aplikasi dekstop Orange Data Mining dan bahasa pemrograman Python. Hasil sentimen positif, negatif dan netral nantinya dihitung masing-masing total keseluruhan pada masing-masing kandidat Presiden. Hasil seluruh proses data mining - text mining pada Orange Data Mining dan Python disajikan kedalam bentuk Graphic User Interface (GUI) sebagai tampilan dashboard. Hasil analisis sentimen terhadap Ganjar Pranowo, Prabowo Subianto dan Anies Baswedan berdasarkan Orange Data Mining menunjukkan skor positif, yakni masing-masing sebesar 61,80%, 33,60% dan 25,80%. Sedangkan, berdasarkan Python masing-masing sebesar 64,40%, 32,20% dan 28,34%. Baik skor menggunakan Orange Data Mining maupun Python secara konsisten menunjukkan nama kandidat Ganjar Pranowo memiliki skor tertinggi

Kata Kunci: *Data Mining, Text Mining, Application Programming Interface (API), Twitter, Orange Data Mining, Python, Graphic User Interface (GUI), Crawling Dataset.*

1. PENDAHULUAN

Pada era digital modern saat ini, banyak segala hal yang tidak dapat dilepaskan dari dunia internet. Awal penciptaan internet hanya digunakan secara rahasia oleh tentara Amerika Serikat (bernama *Advanced Research Project Agency Network* atau ARPANET), sebagai salah satu wadah untuk berkomunikasi dan bertukar data antar basis militer yang lambat laun berkembang untuk menjadi Internet terbuka dan bebas digunakan yang kita kenal saat ini. Seiring perkembangan Internet, seorang peneliti bernama Tim Berners-

Lee menciptakan *World Wide Web* (WWW), yaitu sebuah konsep *hyperlink* dan *Uniform Resource Locator* (URL) untuk mempermudah para pengguna dalam mengakses *website* melalui *Browser*.

Dengan muncul dan berkembangnya WWW secara masif, tersedianya berbagai macam *website* yang menyediakan muatan konten yang beragam dan banyak dimanfaatkan untuk kebutuhan para penggunanya, seperti sarana informasi, komunikasi, hiburan dan bisnis. Salah satu *website*

yang paling berkembang pesat ialah situs jejaring sosial media, pada situs ini pengguna dapat berkomunikasi antara satu pengguna dengan pengguna lain, berbagi foto, berbagi status komentar, serta bertukar atau berbagi informasi dari satu pengguna ke pengguna yang lain. Ada berbagai macam jejaring sosial media yang berkembang pesat setelah adanya kehadiran WWW ini, seperti Friendster, mIRC, MySpace, YahooMessenger, Facebook, Twitter.

Twitter merupakan salah satu contoh jejaring sosial media yang sangat populer digunakan generasi muda ataupun juga generasi tua. Pada *platform* jejaring sosial ini berbagai macam kalangan masyarakat berada dalam satu media yang sama, mulai dari tokoh terkenal didunia sampai pengguna biasa contohnya seperti artis, atlit olahraga, pemimpin organisasi/perusahaan, pengguna *reguler* (orang biasa) dan tokoh politik tertentu (Presiden, dewan pemerintahan, menteri, pemimpin partai politik dan tokoh-tokoh politik penting lainnya).

Dengan kepopuleran dan beragam lapisan masyarakat pengguna Twitter, hal tersebut menjadikan *platform* ini banyak dimanfaatkan oleh para netizen khususnya netizen Indonesia untuk mengutarakan opini atau pemikiran mereka terhadap suatu pandangan tertentu yang sedang terjadi atau sedang viral. Salah satu contohnya adalah pendapat publik terhadap beberapa tokoh politik kandidat Presiden Indonesia yang nantinya mencalonkan pada pemilihan umum 2024.

Menurut lembaga survei Indo Barometer mengeluarkan *statement*, terdapat 3 urutan nama kandidat besar yang bakal mencalonkan dirinya menjadi Presiden Indonesia tahun 2024, Nama-nama tersebut ialah Ganjar Pranowo, Prabowo Subianto dan Anies Baswedan. Dengan seiring nuansa diberlakukannya pemilihan presiden ini, banyak masyarakat Indonesia yang sedang gencar-gencarnya mengomentari calon-calon Presiden tersebut menggunakan Twitter. Berbagai macam komentar dilontarkan masyarakat Indonesia didalam sebuah tweet / postingan, baik Pro ataupun Kontra terhadap salah satu calon yang ada.

Sejumlah komentar tersebut, banyak sekali isi informasi didalamnya untuk dapat dijadikan sebagai penelitian dengan tema Analisis Sentimen pengguna Twitter dalam kandidat pencalonan presiden 2024. Analisis Sentimen ini

dimaksudkan untuk mengetahui bagaimana opini publik tentang kandidat-kandidat yang akan berkompetisi dalam pemilihan Presiden pada tahun 2024 kedalam bentuk sentimen Positif, Negatif ataupun Netral. Sentimen ini nantinya digunakan untuk melihat mana calon kandidat Presiden yang memiliki *Image* positif, negatif dan netral dimata masyarakat pengguna Twitter.

1.1 IDENTIFIKASI MASALAH

Pada latar belakang yang telah dijelaskan oleh peneliti, berikut adalah identifikasi masalah yang dapat ditemukan didalam penelitian, sebagai berikut :

1. Belum digunakannya data tweet secara maksimal untuk menemukan sebuah pola yang dapat dijadikan sebagai *insight* yang bersumber dari opini pada sosial media Twitter dalam menilai opini dari sebuah tweet.
2. Belum adanya komparasi perhitungan hasil analisis sentimen calon Presiden 2024 pada *platform* Twitter antara aplikasi Orange Data Mining dan bahasa pemrograman Python.

1.2 BATASAN MASALAH

Berdasarkan identifikasi masalah yang telah dijabarkan, berikut adalah batasan masalah yang menjadi batasan didalam penelitian :

1. Belum digunakannya data tweet secara maksimal untuk menemukan sebuah pola yang dapat dijadikan sebagai *insight* yang bersumber dari opini pada sosial media Twitter dalam menilai opini dari sebuah tweet.
2. Belum adanya komparasi perhitungan hasil analisis sentimen calon Presiden 2024 pada *platform* Twitter antara aplikasi Orange Data Mining dan bahasa pemrograman Python.

1.3 TUJUAN PENELITIAN

Tujuan dari penelitian ini adalah menganalisis data historis dari sebuah tweet / retweet untuk mendapatkan gambaran nilai Sentimen / emosi masyarakat Indonesia yang memberikan komentar pada *platform* Twitter terhadap calon presiden Indonesia 2024.

2. TINJAUAN PUSTAKA

2.1 Data Mining

Menurut Pan-Ning Tan, et.al (2019:24) definisi *data mining* adalah sebuah proses

otomatis mencari informasi yang berguna dalam penyimpanan data yang besar.

Menurut Pan Ning Tan et al (2019:25) ada dua tipe proses dalam *knowledge discovery in database* (KDD) yaitu :

- a. **Data preprocessing** bertujuan untuk mengubah data mentah menjadi format yang digunakan untuk analisis, tahapan dari *data preprocessing* termasuk menggabungkan data dari beberapa sumber, membersihkan data untuk menghapus observasi duplikat dan *noise*, dan menyeleksi catatan dan fitur yang relevan dalam *data mining*.
- b. **Data postprocessing** bertujuan untuk memastikan bahwa hasil valid dan berguna yang akan diintegrasikan pada pengambilan keputusan seperti filter pola, visualisasi dan interpretasi pola.

Menurut Pan Ning Tan, et.al (2019:29), tugas *data mining* dibagi menjadi dua kategori yaitu:

- a. **Predictive Tasks**
Tujuan dari tugas tersebut untuk memprediksi nilai dari atribut tertentu berdasarkan nilai atribut lainnya. Atribut yang akan diprediksi biasa disebut dengan variabel tidak bebas atau target sedangkan atribut yang dilakukan untuk prediksi disebut dengan variabel bebas atau bersifat penjelasan.
- b. **Descriptive Tasks**
Tujuan dari tugas tersebut untuk mendapatkan pola (korelasi, tren, kluster, lintasan dan anomali) yang merangkum hubungan yang mendasari data. Untuk tugas deskriptif bersifat eksploratif dan sering membutuhkan teknik *postprocessing* untuk memvalidasi dan menjelaskan hasilnya.

Menurut Pan Ning Tan, et.al (2019:29), *data mining* memiliki beberapa metode pengolahan yaitu:

- a. **Prediksi (Predictive)**.
Teknik prediksi digunakan apabila suatu nilai memiliki atribut yang berbeda, contoh algoritmanya seperti Algoritma *Linear Regression*, *Neural Network* dan lain-lain.

- b. **Asosiasi (Association)**
Teknik Asosiasi digunakan untuk hubungan antar data, contoh algoritmanya seperti Algoritma Apriori.
- c. **Klastering (Clustering)**
Teknik *clustering* digunakan untuk pengelompokan data dalam suatu kelompok tertentu, contoh algoritmanya seperti *K-Means*, *K-Medoids*, *Self Organization Map(SOM)*, *Fuzzy C-Means*.
- d. **Klasifikasi (Classification)**
Teknik klasifikasi mengelompokkan data yang mempunyai variabel tertentu berbeda dengan *clustering* yang tidak memiliki variabel yang dependen, contoh algoritma seperti *ID3* dan *K Nearest Neighbour*.

2.2. Text Mining

Pengertian *Text Mining* menurut Ramesh Sharda, Dursun Dellen et al (2018:278) adalah (juga dikenal sebagai penambangan data teks atau penemuan pengetahuan dalam database teks) adalah proses semi otomatis ekstraksi pola (informasi dan pengetahuan yang berguna) dari sumber data yang tidak terstruktur dalam jumlah besar.

Menurut Ramesh Sharda, Dursun et al (2018:282) juga menjelaskan pengertian Pemrosesan bahasa alami (*Natural Language Processing* atau NLP) adalah komponen penting dalam *text mining* dan merupakan sub-bidang dari kecerdasan buatan dan linguistik komputasi. NLP mempelajari masalah "pemahaman" bahasa manusia alami, dengan tujuan mengkonversi teks-teks dalam bahasa manusia (seperti dokumen teks) menjadi representasi yang lebih formal (dalam bentuk data numerik dan simbolik) yang lebih mudah dimanipulasi oleh program komputer.

2.3 Analisis Sentimen

Pengertian Analisis Sentimen menurut Bing Liu (2020:1) adalah bidang studi yang menganalisis opini, sentimen, penilaian, sikap, dan emosi orang terhadap entitas dan atributnya yang diungkapkan dalam teks tertulis.

Dalam melakukan Analisis Sentimen, digunakan Lexicon. Menurut Ramesh Sharda (2018:310) pengertian Lexicon adalah

Sebuah leksikon pada dasarnya katalog kata-kata, sinonim mereka, dan arti mereka untuk suatu bahasa tertentu.

2.4 Machine Learning

Pengertian *Machine Learning* menurut Andriy Burkov (2019:3) adalah cabang ilmu komputer yang berhubungan dengan membangun algoritma yang, untuk dapat berguna, mengandalkan kumpulan contoh fenomena tertentu. Contoh-contoh ini dapat berasal dari alam, dibuat oleh manusia, atau dihasilkan oleh algoritma lain.

Pengertian *Machine Learning* menurut Aurelien Geron (2023:4) adalah ilmu (dan seni) dalam memprogram komputer agar dapat belajar dari data.

a. *Supervised Learning*

Dalam *Supervised learning*, set pelatihan yang disediakan untuk algoritma mencakup solusi yang diinginkan, yang disebut label.

b. *Unsupervised Learning*

Data pelatihan tidak diberi label. Sistem mencoba belajar tanpa guru.

e. *Semi-Supervised Learning*

Karena memberi label pada data biasanya memakan waktu dan biaya yang cukup besar, maka seringkali Anda akan memiliki banyak contoh data yang tidak memiliki label, dan hanya sedikit contoh data yang sudah diberi label. Beberapa algoritma dapat menangani data yang sebagian telah diberi label.

f. *Self-Supervised Learning*

Pendekatan lain dalam *machine learning* melibatkan pembuatan kumpulan data yang sepenuhnya diberi label dari kumpulan data yang sepenuhnya tidak diberi label.

3. METODE

3.1 Teknik Pengumpulan Data

Pada pengumpulan data, peneliti menggunakan metode penelitian Kuantitatif. Metode penelitian Kuantitatif dihasilkan dari sebuah data yang berbentuk perhitungan, baik angka ataupun kuantitas. Data yang digunakan peneliti ialah data sekunder yang bersumber pada *platform* Twitter, berupa data

real aktivitas *posting* pengguna Twitter Indonesia yang diambil berdasarkan *Crawling dataset* menggunakan API Twitter, dengan menentukan *Query Word List* (ganjarpranowo, aniesbaswedan, prabowo) dengan total dataset sebanyak 1.500 *Records* pada bulan Februari tahun 2023.

3.2 Teknik Analisis Data

Dalam penelitian ini, peneliti menggunakan teknik analisis yang dilakukan berdasarkan metode CRISP-DM (*Cross Industry Standard Process For Data Mining*) yaitu sebuah metode *data mining* yang masih digunakan saat ini.

a. *Business Understanding*

Fase pertama yang berfokus pada pemahaman objektif proyek beserta kebutuhan dari perspektif bisnis, kemudian pengetahuan tersebut dikonversi menjadi data mining.

b. *Data Understanding*

Fase kedua yang berawal dengan pengumpulan data serta pemahaman data untuk mengidentifikasi masalah kualitas data untuk mengetahui *insight* dan membentuk hipotesis untuk informasi tersembunyi.

c. *Data Preparation*

Fase ketiga yang membentuk dataset final yang akan digunakan dalam alat *modelling* dari data mentah, tugas persiapan data akan dilakukan secara berulang-ulang seperti transformasi dan pembersihan data untuk alat *modelling*.

d. *Modelling*

Fase keempat yang menentukan teknik *modelling* yang akan diaplikasikan beserta *parameter* yang akan dikalibrasikan pada nilai optimal.

e. *Evaluation*

Fase kelima yang terdiri dari model yang sudah dibuat untuk dilakukan perspektif analisis data, sebelum diimplementasikan memerlukan evaluasi model dan meninjau tahapan yang dieksekusi untuk membentuk model tersebut sesuai dengan tujuan bisnis.

f. **Deployment**

Fase keenam yang merupakan tahap implementasi dari sebuah model yang

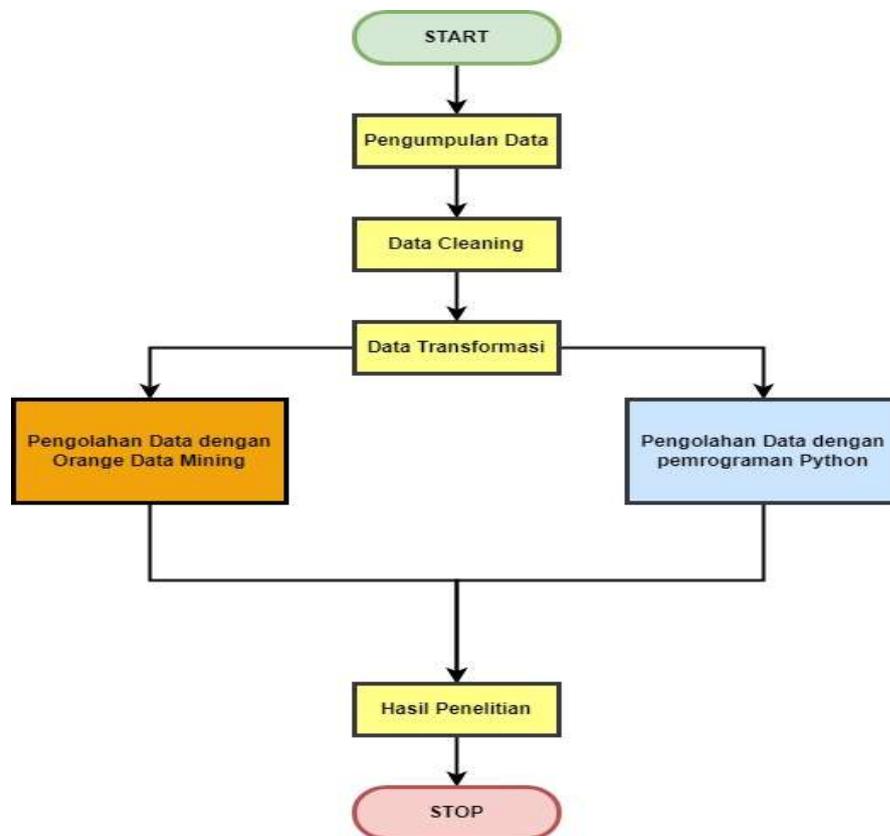
sudah dibuat yang berguna untuk meningkatkan pengetahuan tentang data.

3.3 Penerapan Algoritma

Pada penelitian ini, peneliti melakukan pembahasan tentang bagaimana penerapan sebuah algoritma didalam melakukan *Text Mining* dan menilai isi sentimen yang terkandung didalam sebuah teks. Selain itu, penerapan algoritma digunakan untuk

membandingkan hasil akhir data sentimen yang didapatkan dari aplikasi Orange Data Mining dan kode manual Python. Berikut penerapan algoritma tersebut yang dapat digambarkan kedalam bentuk diagram :

Gambar 1
Alur Penerapan Algoritma



3.4 Teknik Pengukuran Data

3.4.1 Sentiment Rating

Menurut Bing Liu (2020:24) *Sentiment Rating* adalah Penilaian sentimen. Dalam aplikasi praktis, kita biasanya menggunakan jenis penilaian diskrit untuk mengungkapkan

intensitas sentimen. Sebagai contoh, lima level (misalnya, 1-5 bintang) sering digunakan.

Tabel 1
Nilai Sentiment Rating

Nilai Sentimen	Keterangan
Jika nilai (> 1)	Positif
Jika nilai nol (0)	Netral
Jika nilai (> -1)	Negatif

4. HASIL

4.1 Exploratory Data Analysis (EDA)

Berikut ini adalah atribut data tweet pengguna Twitter yang *dicrawling* datanya memakai *Application Programming Interface* (API) Twitter pada bulan Februari 2023:

Tabel 2
Nilai Sentiment Rating

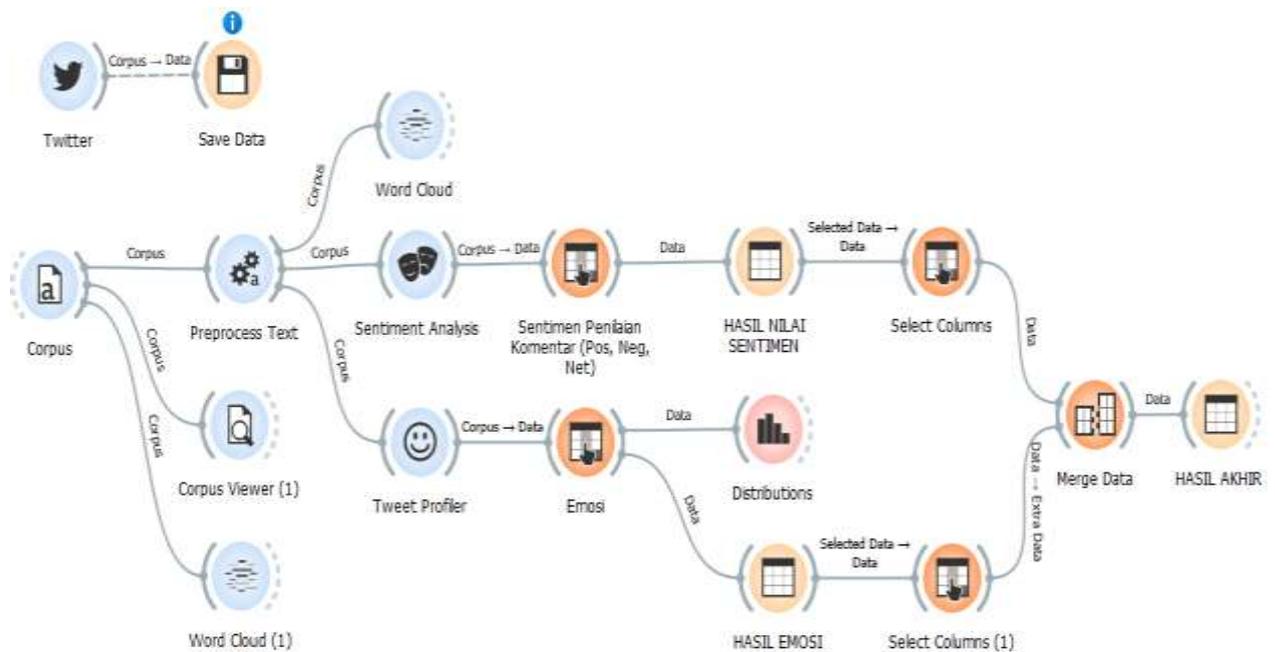
No	Atribut	Keterangan
1	Content	Isi Tweet
2	Author	Username akun Twitter
3	Date	Tanggal dibuatnya Tweet
4	Language	Bahasa dalam Tweet
5	Location	Lokasi Akun
6	Number of Likes	Jumlah Like Tweet
7	Number of Retweets	Jumlah berbagi/ <i>share</i> / Retweet
8	In Reply To	Membalas sebuah Tweet / RT @
9	Author Name	Nama akun penulis
10	Author Description	Keterangan biografi akun penulis
11	Author Tweets Count	Jumlah Tweet akun penulis
12	Author Following Count	Jumlah mengikuti / following penulis
13	Author Followers Count	Jumlah pengikut / followers penulis
14	Author Listed Count	Jumlah list penulis
15	Author Verified	Identifikasi verified akun penulis
16	Longitude	Lokasi geografis akun
17	Latitude	Lokasi geografis akun

Dataset Twitter tersebut dilakukan *Cleaning* atau *Preprocess* menggunakan aplikasi pengolah angka (Microsoft Excel). Dengan menggunakan Orange Data Mining dan Python untuk memperoleh pola kandungan informasi analisis sentimen positif, netral dan negatif.

Dapat dilihat pada Gambar 2 adalah keseluruhan proses analisis dari awal *Crawling Dataset* hingga menghasilkan hasil akhir sentimen pada aplikasi Orange Data Mining. Proses-proses tersebut dilakukan dengan memanfaatkan *widjet* yang tersedia didalam aplikasi.

4.2 Workflow Orange Data Mining

Gambar 2
Workflow Analisis

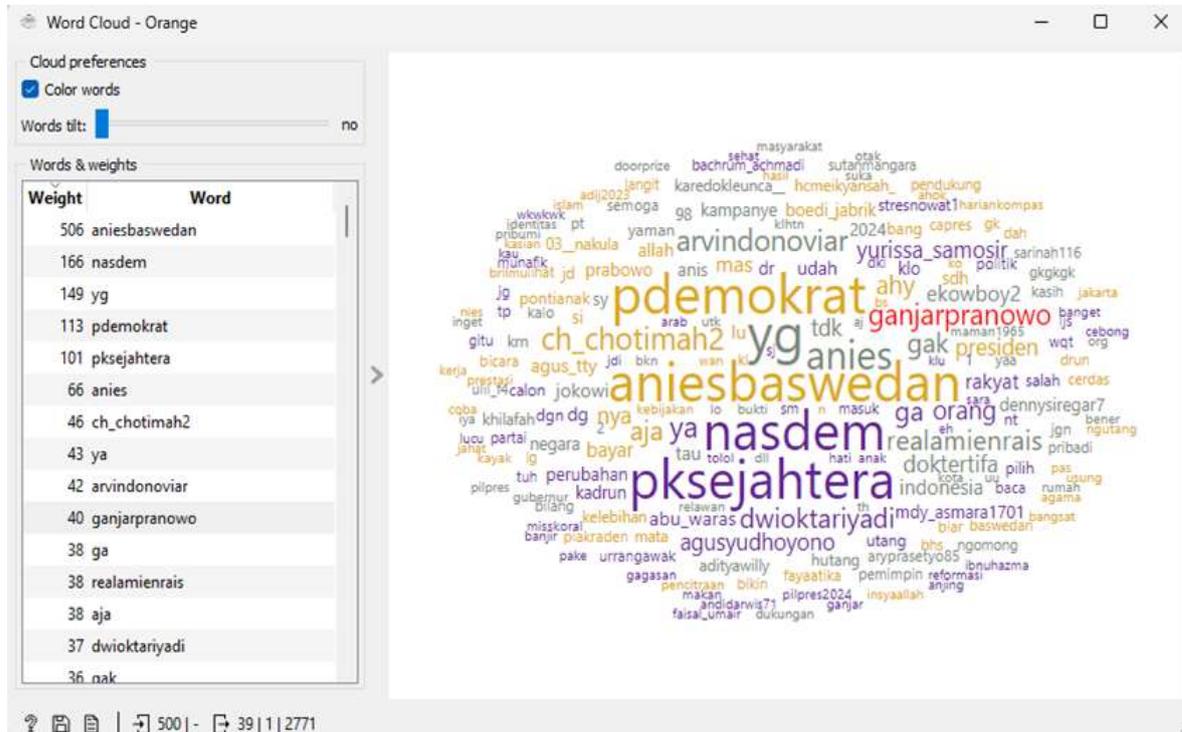


4.2.1 Hasil *Word Cloud* Orange Data Mining

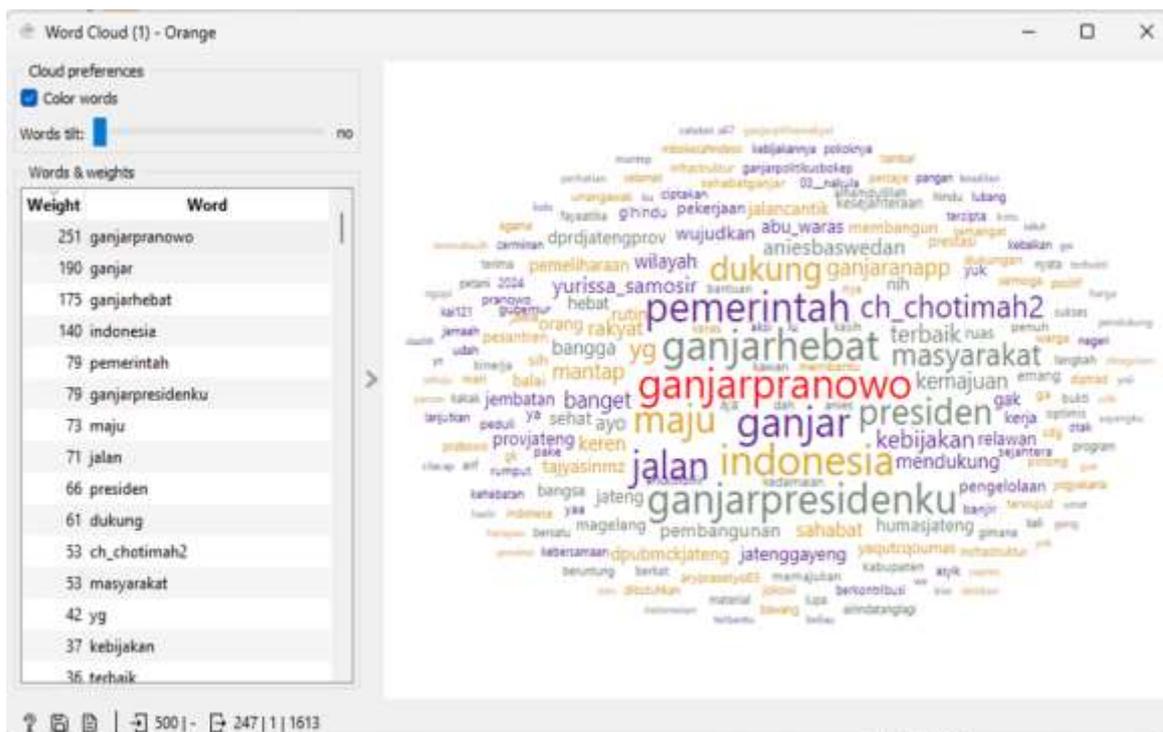
Hasil Proses *Word Cloud* digunakan sebagai representasi visual dari kumpulan kata yang disusun secara acak dan diatur dalam sebuah gambar. Semakin sering

sebuah kata muncul dalam teks, semakin besar ukuran kata tersebut, seperti terlihat pada Gambar 3, Gambar 4 dan Gambar 5.

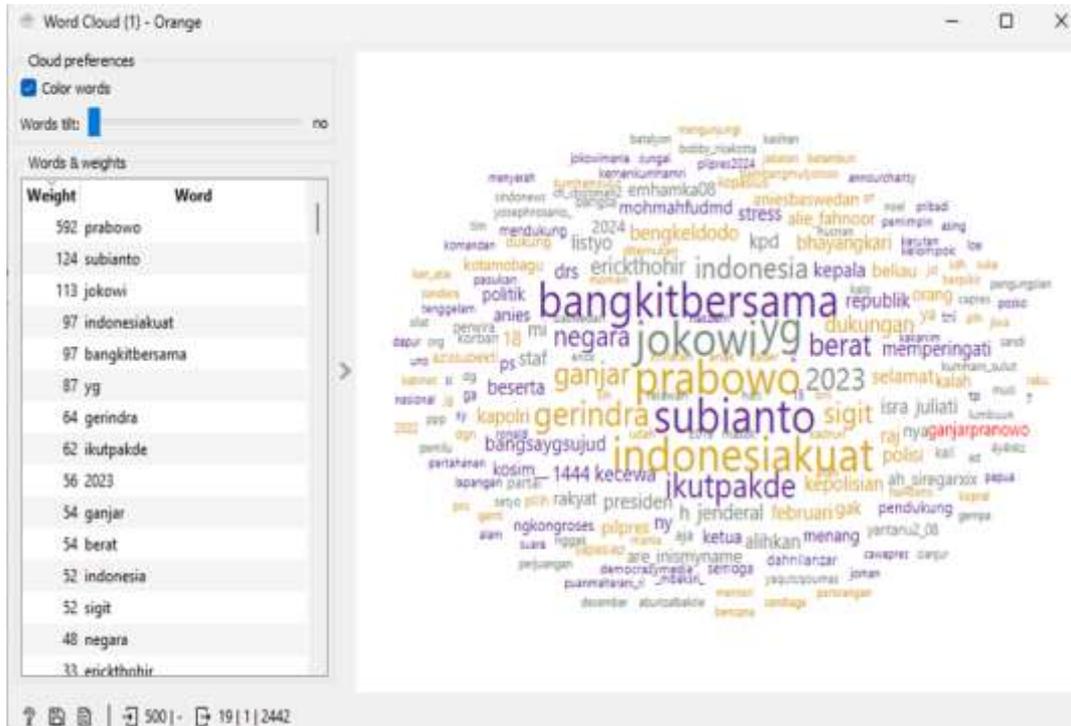
Gambar 3
Word Cloud Anies Baswedan



Gambar 4
Word Cloud Text Ganjar Pranowo



Gambar 5
Word Cloud Prabowo Subianto



4.2.2 Hasil Emotion

Proses *Tweet Profiler* berguna untuk menentukan *Mood Emotions* dalam sebuah teks tweet dengan cara mengambil informasi tentang sentimen dari *server* untuk setiap tweet (atau dokumen) yang diberikan. *Widget* mengirimkan data ke *server*, dimana model menghitung probabilitas dan/atau skor *emotion*.

Setelah semua nama kandidat dicari masing-masing hasilnya, peneliti melanjutkan proses membandingkan hasil pada setiap nama kandidat presiden dengan bantuan Microsoft Excel. Berikut hasil akhirnya :

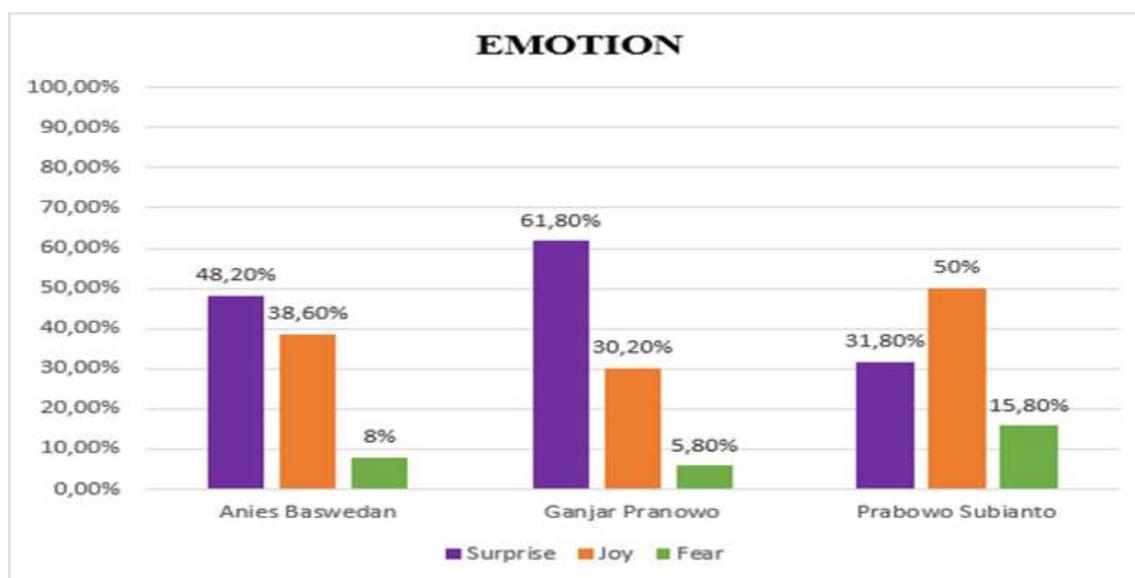
Gambar 6
Hasil Total Sentimen Pada Setiap Nama Kandidat Presiden

EMOTION	Anies Baswedan	Ganjar Pranowo	Prabowo Subianto
Surprise	48,20%	61,80%	31,80%
Joy	38,60%	30,20%	50%
Fear	8%	5,80%	15,80%

Dapat dilihat pada Gambar 6 bahwasannya nama kandidat Presiden Ganjar Pranowo meraih nilai *emotion Surprise* paling banyak diantara dua nama kandidat lain, sedangkan nilai *emotion Joy* dan Fear

paling tinggi dimiliki oleh kandidat Presiden Prabowo Subianto diangka 50% dan 15.80%. Nilai-nilai tersebut dapat dilihat kedalam bentuk *bar plot diagram*, sebagai berikut :

Gambar 7
Hasil Total Emotion Dalam Bentuk Bar Plot Diagram



4.2.3 Hasil Analisis Sentimen Orange Data Mining

Proses *Sentiment Analysis* digunakan untuk menentukan sebuah sentimen yang terdapat dalam kalimat atau teks tweet dengan melabelinya sebagai positif sentimen, negatif sentimen dan juga netral sentimen.

Peneliti membandingkan nilai sentimen pada setiap nama kandidat Presiden dengan bantuan Microsoft Excel. berikut hasil akhirnya :

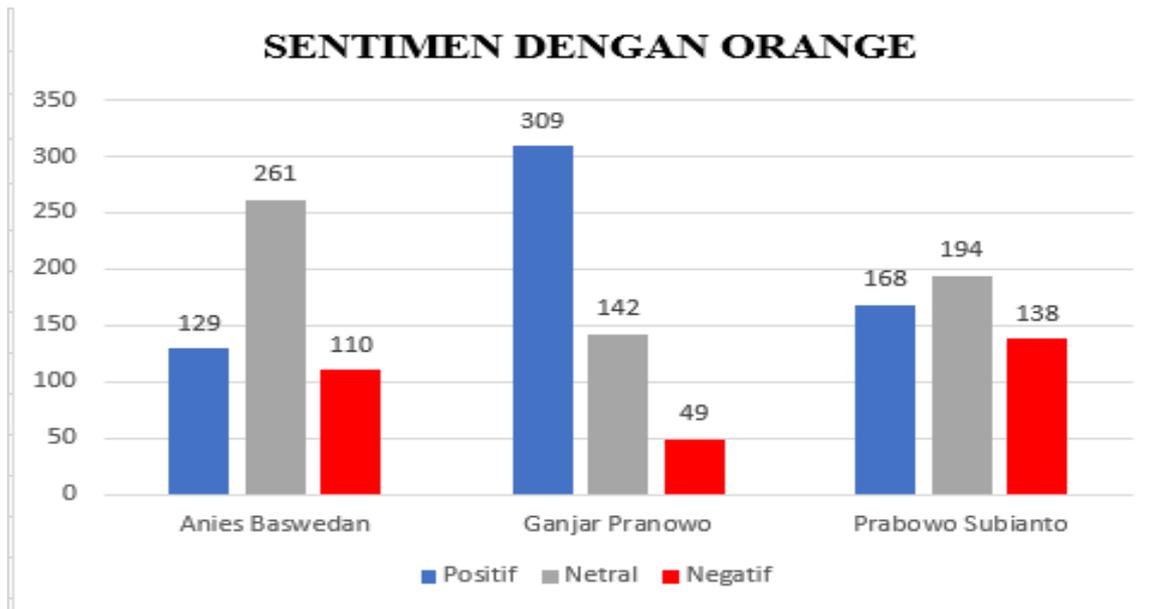
Gambar 8
Hasil Total Sentimen Pada Setiap Nama Kandidat Presiden

HITUNGAN ORANGE			
SENTIMEN	Anies Baswedan	Ganjar Pranowo	Prabowo Subianto
Positif	129	309	168
Netral	261	142	194
Negatif	110	49	138

Dapat dilihat pada Gambar 8 bahwasannya nama kandidat Presiden Ganjar Pranowo meraih nilai sentimen positif terbanyak diangka 309 dibandingkan oleh kedua nama kandidat Presiden lain. Selain itu, nilai sentimen netral meraih nilai paling banyak diangka 261 oleh nama kandidat

Presiden Anies Baswedan dan sebaliknya nilai sentimen negatif meraih nilai paling banyak pada nama kandidat Presiden Prabowo Subianto diangka 138. Nilai-nilai tersebut dapat dilihat kedalam bentuk *bar plot diagram*, sebagai berikut :

Gambar 9
Hasil Total Sentimen Dalam Bentuk *Bar Plot Diagram*



4.3 Python

Peneliti menggunakan bahasa pemrograman Python yang dijalankan menggunakan *Intergrated Development Environment* (IDE) Google Colab untuk mencari skor sentimen pada setiap nama

kandidat Presiden menggunakan beberapa *library*, salah satunya Lexicon dan Sastrawi. Data yang digunakan adalah hasil *Crawling dataset* tweet Twitter pada bulan Februari 2023. Berikut penjelasan dari setiap tahapannya dan *source code* program yang dibuat.

4.3.1 Source Code

```
#Import Library
!pip install nltk Sastrawi
import pandas as pd
import numpy as np
import re
import string
import nltk
nltk.download('stopwords')
nltk.download('punkt')
from nltk.corpus import stopwords
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import unicodedata
import html
```

```

import Sastrawi
from wordcloud import WordCloud, STOPWORDS
import matplotlib.pyplot as plt
from PIL import Image



---


#Import Dataset
# Membaca file dan menampilkan
df=pd.read_csv('dataset_anies_baswedan_3.csv', encoding = "utf-8")
df.head()

#Tahapan Preprocess (Pembersihan)
def preprocess_text(text):
# Remove twitter username
text = re.sub(r'@[^\s]+', "", text)

# Lowercase transformation
text = text.lower()

# Remove accents
text = unicodedata.normalize('NFKD', text).encode('ascii', 'ignore').decode('utf-8')

# Parse HTML
text = html.unescape(text)

# Remove URLs
text = re.sub(r'http\S+', "", text)

# Filtering stopwords (bahasa Indonesia)
stop_words = set(stopwords.words('indonesian'))
tokens = nltk.word_tokenize(text)
tokens = [word for word in tokens if not word in stop_words]

# Filtering numbers
tokens = [word for word in tokens if not word.isnumeric()]

# Filtering regular expression
tokens = [word for word in tokens if not re.match(r'^\w\s', word)]

# Stemming
factory = StemmerFactory()
stemmer = factory.create_stemmer()
tokens = [stemmer.stem(word) for word in tokens]

# Join tokens back into a string
text = ''.join(tokens)

return text



---


#Menjalankan Preprocess
# Menggabungkan kelas dan menjalankan fungsi Preprocess

```

```
df['content'] = df['content'].apply(preprocess_text)
df.head(10)
```

```
#Import kamus kata positif dan negatif
# membuat kamus kata positif
with open('positive_words_id.txt', 'r') as file:
    kata_positif = file.readlines()
kata_positif = [kata.strip() for kata in kata_positif]

# membuat kamus kata negatif
with open('negative_words_id.txt', 'r') as file:
    kata_negatif = file.readlines()
kata_negatif = [kata.strip() for kata in kata_negatif]
```

```
#Mencari Sentimen
def hitung_sentimen(text):
    # Tokenisasi teks
    tokens = nltk.tokenize.word_tokenize(text)
    # Inisialisasi nilai sentimen
    nilai_sentimen = 0
    # Loop untuk setiap kata dalam teks
    for kata in tokens:
        # Jika kata ada dalam kamus kata positif, tambahkan 1 ke nilai sentimen
        if kata in kata_positif:
            nilai_sentimen += 1
        # Jika kata ada dalam kamus kata negatif, kurangi 1 dari nilai sentimen
        elif kata in kata_negatif:
            nilai_sentimen -= 1
    return nilai_sentimen
```

```
#Memposisikan kalimat sentimen
df['sentimen'] = df['content'].apply(hitung_sentimen)
# Menghitung jumlah komentar dengan sentimen positif
jumlah_positif = len(df[df['sentimen'] > 0])
```

```
# Menghitung jumlah komentar dengan sentimen negatif
jumlah_negatif = len(df[df['sentimen'] < 0])
```

```
# Menghitung jumlah komentar dengan sentimen netral
jumlah_netral = len(df[df['sentimen'] == 0])
```

```
df.head(80)
```

```
#Membuat WordCloud keseluruhan
# Menggabungkan seluruh konten komentar menjadi satu string
all_text = " ".join(df['content'].values)
```

```
# Membuat objek WordCloud
```

```

wordcloud = WordCloud(width=800, height=500, random_state=42,
max_font_size=100).generate(all_text)

# Menampilkan wordcloud
plt.figure(figsize=(12, 8))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.show()

```

```

#Membuat list seluruh kata positif dari dataset
list_kata_positif = []
for kata in df['content']:
    for k in kata.split():
        if k in kata_positif:
            list_kata_positif.append(k)

# Menggabungkan seluruh kata positif menjadi satu string
string_kata_positif = ''.join(list_kata_positif)

# Membuat wordcloud
wordcloud = WordCloud(width=800, height=500, min_font_size=10, background_color='white',
colormap='Blues').generate(string_kata_positif)

# Mengatur warna tulisan menjadi biru tua
blue_color = '#1f77b4' # kode warna RGB untuk biru tua
wordcloud.recolor(color_func=lambda *args, **kwargs: blue_color)

# Menampilkan wordcloud
plt.imshow(wordcloud)
plt.axis('off')
plt.tight_layout(pad=0)
plt.show()

```

```

#Membuat list kata-kata dalam komentar dengan sentimen negatif
list_kata_negatif = []
for index, row in df.iterrows():
    if row['sentimen'] < 0:
        kata = row['content'].split()
        for k in kata:
            list_kata_negatif.append(k)

# Menggabungkan list kata menjadi satu string
string_kata_negatif = ''.join(list_kata_negatif)

# Membuat fungsi untuk menentukan warna tiap kata pada word cloud
def color_func(word, font_size, position, orientation, random_state=None, **kwargs):
    return "hsl(0, 100%, 50%)"

# Membuat wordcloud

```


4.3.3 Hasil Analisis Sentimen Python

Proses *Sentiment Analysis* digunakan untuk menentukan sebuah sentimen yang terdapat dalam kalimat atau teks tweet dengan melabelinya sebagai positif sentimen, negatif sentimen dan juga netral sentimen.

Peneliti membandingkan nilai sentimen pada setiap nama kandidat Presiden dengan bantuan Microsoft Excel. berikut hasil akhirnya :

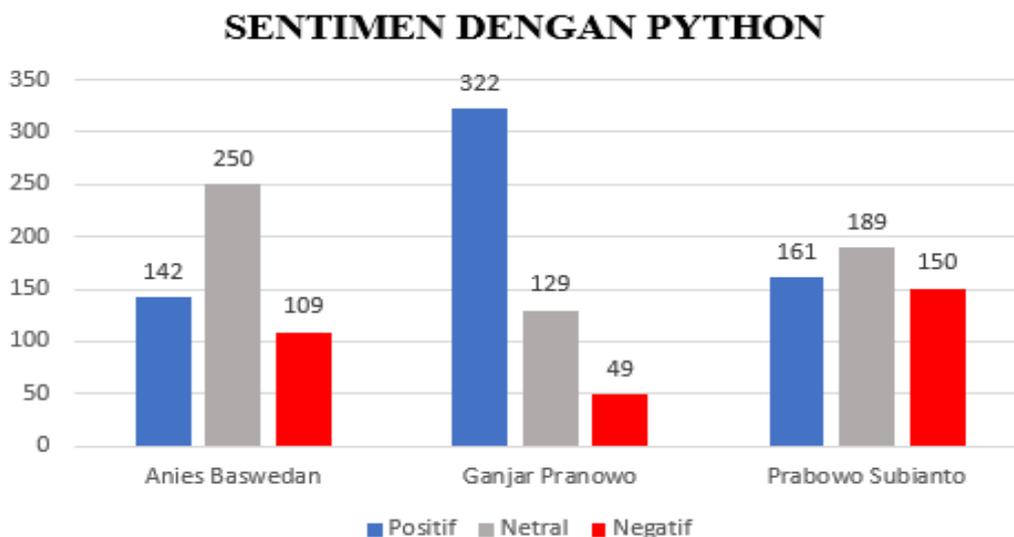
Gambar 12
Hasil Total Sentimen Python Pada Setiap Nama Kandidat

HITUNGAN PYTHON			
SENTIMEN	Anies Baswedan	Ganjar Pranowo	Prabowo Subianto
Positif	142	322	161
Netral	250	129	189
Negatif	109	49	150

Dapat dilihat pada Gambar 12 bahwasannya nama kandidat Presiden Ganjar Pranowo meraih nilai sentimen positif terbanyak diangka 322 dibandingkan oleh kedua nama kandidat Presiden lain. Selain itu, nilai sentimen netral meraih nilai paling banyak diangka 250 oleh nama kandidat

Presiden Anies Baswedan dan sebaliknya nilai sentimen negatif meraih nilai paling banyak pada nama kandidat Presiden Prabowo Subianto diangka 150. Nilai-nilai tersebut dapat dilihat kedalam bentuk *bar plot diagram*, sebagai berikut :

Gambar 13
Hasil Total Sentimen Python Dalam Bentuk Bar Plot Diagram



5. PEMBAHASAN

Hasil dari penelitian ini adalah penerapan metode *Data Mining* dan *Text Mining* untuk mencari sentimen Positif, sentimen Netral dan sentimen Negatif dari sebuah tweet yang dikemukakan atau dilontarkan para netizen Twitter Di Indonesia, mengenai calon kandidat Presiden Indonesia pada tahun 2024. Dalam mencari sebuah sentimen, digunakan Analisis Sentimen, Lexicon dan Natural Language Processing (NLP). Dengan Penggunaan aplikasi Orange Data Mining dan Pemrograman Python untuk dapat melihat perbedaan jumlah sentimen Positif, sentimen Netral dan sentimen Negatif yang dihasilkan masing-masing *tools/alat* yang nantinya dapat dijadikan sebuah informasi atau *Insight* pengetahuan yang bermanfaat bagaimana arah pandangan opini masyarakat (Pro dan Kontra) di sosial media. Berdasarkan jurnal-jurnal penelitian terdahulu, peneliti menjadikan sumber data sekunder, seperti (Artha Dini Akmal, Lip Permana, dkk, 2022; Ratu Aghnia Raffaidy Wiguna dan Andri Irfan Rifai, 2021; Septian Fendyputra Pratama, Ricsa Andrean, dkk, 2019; Cahyo Prianto, Nisa Hanum Harani, dkk, 2019; Neng Resti Wardani, Sudin Saepudin, dkk, 2022). Nama kandidat, tahun pengambilan *dataset*, atribut, perbandingan *tools/alat* dan lain-lain yang disajikan dalam penelitian ini berbeda, tetapi kesamaan dengan penelitian-penelitian terdahulu adalah aplikasi pengolahan data mining dan pencarian sentimen yang sama oleh peneliti gunakan sekarang.

6. Kesimpulan

Dari temuan yang telah diteliti, peneliti dapat menyimpulkan hal-hal sebagai berikut:

1. Berdasarkan hasil komentar pada media sosial Twitter sebanyak 1.500 tweet dari para netizen didapatkan *insight* tentang perbandingan sentimen positif, negatif terhadap Ganjar Pranowo, Prabowo Subianto dan Anies Baswedan.
2. Hasil Analisis Sentimen terhadap Ganjar Pranowo, Prabowo Subianto dan Anies Baswedan berdasarkan Orange Data

Mining menunjukkan skor positif, yakni masing-masing sebesar 61,80%, 33,60% dan 25,80%. Sedangkan, berdasarkan Python masing-masing sebesar 64,40%, 32,20% dan 28,34%. Baik skor menggunakan Orange Data Mining maupun Python secara konsisten menunjukkan nama kandidat Ganjar Pranowo memiliki skor tertinggi.

Dalam penelitian karya akhir ini, peneliti menemukan beberapa saran berdasarkan tahapan yang telah dilakukan, sehingga dapat dijadikan referensi bagi peneliti selanjutnya untuk melakukan perbaikan dan menemukan ide-ide lain dimasa depan, berikut beberapa saran yang dapat diberikan :

1. Untuk peneliti berikutnya, proses mengambil dataset tweet dapat menggunakan bahasa pemrograman Python dan R, serta mencari *Query Word List* unik dengan menyesuaikan perkembangan tren yang sedang terjadi untuk mendapatkan keragaman informasi.
2. Penggunaan aplikasi selain Orange Data Mining untuk peneliti selanjutnya dalam mencari analisis sentimen, seperti Rstudio, KNIME, Weka, SAS, RapidMiner atau IBM Watson Studio.
3. Rancangan tampilan *dashboard* atau *Graphic User Interface (GUI)* yang untuk menampilkan hasil analisis oleh peneliti selanjutnya, dapat memanfaatkan streamlit atau shiny dashboard.
4. Menambahkan kamus *lexicon* untuk kata *slang/gaul* agar dapat melengkapi pencarian sentimen positif, negatif dan netral.
5. Menambahkan hasil analisis emosi pada bahasa pemrograman Python.

DAFTAR PUSTAKA

- 1) Artha Dini Akmal., Iip Permana., Hidayatul Fajri., Yuliarti.(2022). OPINI MASYARAKAT TWITTER

- TERHADAP KANDIDAT BAKAL CALON PRESIDEN REPUBLIK INDONESIA TAHUN 2024, Universitas Negeri Padang.
- 2) Block, G., Cibraro, P., Felix, P., Dierking, H., & Miller, D. (2014). *Designing Evolvable Web APIs with ASP.NET: Harnessing the power of the web*. O'Reilly Media
 - 3) Burkov, A. (2019). *The Hundred-Page Machine Learning Book*. Andriy Burkov.
 - 4) Cahyo Prianto., Nisa Hanum Harani., Indra Firmansyah.(2019).*Analisis Sentimen Terhadap Kandidat Presiden Republik Indonesia Pada Pemilu 2019 di Media Sosial Twitter*, Politeknik Pos Indonesia.
 - 5) Geewax, J. (2021). *API Design Patterns*. Shelter Island, NY: Manning Publications.
 - 6) Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems* Edition: 3. O'Reilly Media.
 - 7) Kumar,Umesh., D.P, Kothari.(2022) *Research Methodology Techniques And Trends*,Boca Raton: CRC Press.
 - 8) Laudon, K. C., & Laudon, J. P. (2019). *Management Information Systems: Managing the Digital Firm* Edition: 16. New York, NY: Pearson.
 - 9) Liu, B. (2020). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions* Edition: 2. Cambridge University Press.
 - 10) Larose, Daniel T., Chantal D, Larose.(2015).*Data Mining and Predictive Analysis*. New Jersey: John Wiley & Sons,Inc.
 - 11) Matthes, E. (2019). *Python Crash Course: A Hands On Project Based Introduction To Programming* Edition: 2nd Edition. San Francisco, CA: No Starch Press.
 - 12) Ning-Tan,Pang.,Michael Steinbach.,Anuj Karpatne.,Vipin Kumar.(2019). *INTRODUCTION TO DATA MINING SECOND EDITION*. United Kingdom: Pearson Education Limited.
 - 13) Orange Data Mining, About Orange. sumber <https://orangedatamining.com/faq/> (diakses pada 15 Februari 2023).
 - 14) Orange Data Mining, Interactive Data Visualization. sumber https://orangedatamining.com/home/interactive_data_visualization/ (Diakses pada 15 Februari 2023).
 - 15) Orange Data Mining, Visual Programming. sumber <https://orangedatamining.com/home/visual-programming/> (Diakses pada 15 Februari 2023).
 - 16) Ratu Aghnia Raffaidy Wiguna., Andri Irfan Rifa.(2021).*Analisis Text Clustering Masyarakat Di Twitter Mengenai Omnibus Law Menggunakan Orange Data Mining*, Universitas Padjadjaran, Universitas Internasional Batam.
 - 17) Syafrizal Helmi Situmorang., Muslich Lufti.(2014).*Analisis Data Untuk Riset Manajemen dan Bisnis, Edisi Ke-3*, Medan: USU Press.
 - 18) Sharda, R., Delen, D., & Turban, E. (2018). *Business Intelligence, Analytics, and Data Science A Managerial Perspective* FOURTH EDITION. Harlow, England: Pearson.

- 19) Sweigart, A. (2019). Automate the Boring Stuff with Python, 2nd Edition: Practical Programming for Total Beginners. San Francisco: No Starch Press.
- 20) Septian Fendyputra Pratama., Ricsa Andrian., Aryo Nugroho.(2019) Analisis Sentimen Twitter Debat Calon Presiden Indonesia Menggunakan Metode Fined-Grained Sentiment Analysis, Universitas Narotama Surabaya.
- 21) Twitter, New user FAQ. sumber <https://help.twitter.com/en/resources/new-user-faq> (Diakses pada 15 Februari 2023).
- 22) Twitter, Retweet FAQs. sumber <https://help.twitter.com/en/resources/new-user-faq> (Diakses pada 15 Februari 2023).
- 23) Undang-undang, Undang-Undang No.42 Tahun 2008 tentang Pemilihan Umum Presiden dan Wakil Presiden. sumber <https://peraturan.bpk.go.id/Home/Details/37655/uu-no-42-tahun-2008> (Diakses pada 15 Februari 2023).
- 24) Undang-undang, Undang-Undang No.7 Tahun 2017 tentang Pemilihan Umum. sumber <https://peraturan.bpk.go.id/Home/Details/37644/uu-no-7-tahun-2017> (Diakses pada 15 Februari 2023)
- 25) Undang-undang, Undang-Undang No.1 Tahun 2022 tentang Perubahan atas Undang-Undang Nomor 7 Tahun 2017 tentang Pemilihan Umum. Sumber <https://peraturan.bpk.go.id/Home/Details/232762/perpu-no-1-tahun-2022> (Diakses pada 15 Februari 2023)
- 26) Wizner, William.(2020), Python For Data Science, Independent Published.
- 27) Wallace, P. M. (2019). Introduction to information systems Edition: Fourth. Pearson.