

Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap “Figure Pemimpin” Menggunakan Python

Putu Pasek Okta Mahawardana^{a1}, Gusti Arya Sasmita^{a2}, I Putu Agus Eka Pratama^{a3}

^aProgram Studi Teknologi Informasi, Fakultas Teknik, Universitas Udayana
Bukit Jimbaran, Bali, Indonesia, telp(0361)701806

e-mail: ¹pasekoka35@gmail.com, ²aryasasmita@unud.ac.id, ³eka.pratama@unud.ac.id

Abstrak

Media sosial adalah sebuah media yang digunakan untuk bersosialisasi dan bertukar informasi oleh para penggunanya dengan menggunakan internet. Ada banyak sekali manfaat dari media sosial, beberapa kegunaan media sosial seperti berkenalan dengan teman baru, mengetahui informasi olahraga, ekonomi, pariwisata dan juga untuk urusan politik. Salah satunya merupakan figure pemimpin kandidat calon presiden 2024, sehingga penulis ingin mengetahui informasi apa yang dapat diambil dari opini masyarakat pada media sosial twitter tentang figure pemimpin kandidat calon presiden 2024. Adapun masalah ini dapat diatasi dengan melakukan penelitian di bidang Analisis Sentiment, yang mana merupakan bidang penelitian yang berfokus kepada studi komputasi atas opini, tingkah laku, dan emosi terhadap suatu entitas yang dituangkan dalam bentuk teks. Penelitian ini dilakukan untuk mengetahui bagaimana hasil analisis sentimen terkait respon masyarakat dari kabar calon presiden 2024 dan mengklasifikasikannya menjadi tiga kelas yaitu positif, negatif, dan netral menggunakan Python. Hasil penelitian ini sentimen dengan kelas yang positif sebesar 21,6% dengan jumlah 108, netral sebesar 78% dengan jumlah 390 dan negatif sebesar 0.2% dengan jumlah 2. Dapat disimpulkan bahwa penelitian ini cenderung netral.

Kata Kunci: Calon Presiden, Social Media, Analisis Sentiment, Text Mining, Python.

Abstract

Social media is a medium used to socialize and exchange information by its users using the internet. There are many benefits of social media, several uses of social media such as meeting new friends, knowing sports, economics, tourism information and also for political matters. One of them is the figure of the leader of the presidential candidate for 2024, so the writer wants to know what information can be taken from public opinion on Twitter social media about the figure of the leader of the presidential candidate for 2024. This problem can be overcome by conducting research in the field of Sentiment Analysis, which is a field of research that focuses on the computational study of opinions, behavior, and emotions towards an entity as outlined in the form of text. This study was conducted to find out how the results of sentiment analysis related to public response to the news of the 2024 presidential candidate and classify them into three classes, namely positive, negative, and neutral using Python. The results of this study are positive class sentiments of 21.6% with a total of 108, neutral by 78% with a total of 390 and negative by 0.2% with a number of 2. It can be concluded that this study tends to be neutral.

Keywords: Presidential Candidate, Social Media, Sentiment Analysis, Text Mining, Python.

1. Introduction

Indonesia merupakan negara demokrasi di mana seluruh warga negara Indonesia memiliki hak yang sama untuk pengambilan keputusan yang dapat mengubah hidup mereka[1]. Kata “Demokrasi” berasal dari dua kata, yaitu dari kata “Demos” dan “Kratos”. “Demos” yang berarti rakyat, “Kratos” yang berarti pemerintahan, jadi Demokrasi dapat diartikan sebagai

pemerintah rakyat, atau lebih dikenal dengan pemerintahan dari rakyat, oleh rakyat, dan untuk rakyat.

Demokrasi sendiri sudah berjalan melalui beberapa media massa. Saat ini, media berbasis web merupakan salah satu media yang paling membumi dan paling cepat berkembang di Indonesia. Dulu, hanya orang-orang tertentu yang dapat menyuarakan pendapatnya melalui media cetak, TV, dan radio dengan tujuan agar dapat dilihat oleh banyak orang, kini setiap orang dapat mengkomunikasikan pandangannya dan pandangannya harus dapat dilihat oleh banyak orang karena kemajuan teknologi media online di Indonesia.

Pada saat ini ekonomi global mengalami perubahan besar dikarenakan tingkat kemajuan teknologi yang semakin pesat. Telah banyak revolusi industri yang telah terjadi, namun perubahan yang sangat besar bagi perekonomian global saat ini adalah revolusi industri 4.0 dimana teknologi masa kini memberikan banyak kemudahan dan lebih efisien digunakan manusia seperti penggunaan berbagai media sosial sebagai contoh Instagram, Twitter, Facebook dan sebagainya.

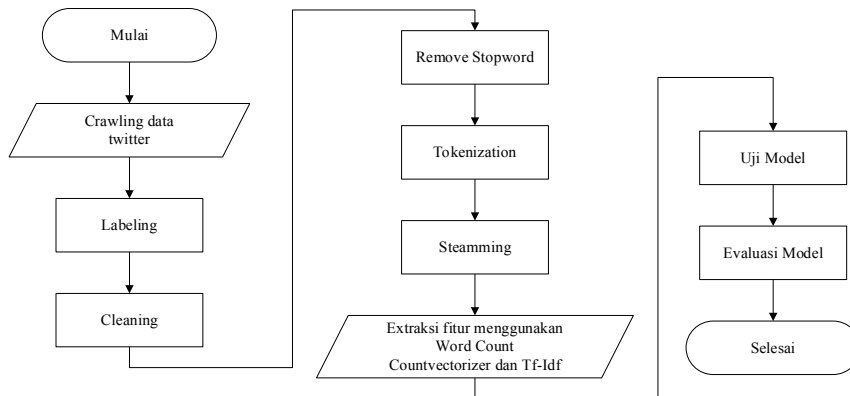
Sosial media merupakan media yang digunakan untuk berbaur dan bertukar data oleh kliennya dengan memanfaatkan web. Ada banyak manfaat dari sosial media, beberapa pemanfaatan media berbasis web seperti mengumpulkan teman baru, mengetahui permainan, aspek keuangan, data industri perjalanan dan juga untuk masalah politik. Saat ini, sosial media yang paling terkenal adalah Twitter, Facebook, dan Instagram.

Twitter sendiri merupakan sebuah aplikasi berbagi foto dan video, serta memiliki fitur-fitur seperti retweet, pengambilan foto dan video, serta membagikannya ke beberapa jaringan sosial lainnya. Salah satu isu yang sedang banyak diperbincakan akhir-akhir ini yaitu tentang munculnya “Kandidat Calon Presiden 2024” pada media sosial. Pada tahun 2024 bagi Negara Indonesia merupakan tahun politik, hal ini dikarenakan bahwa pada tahun 2024 tersebut dilaksanakan pemilihan Presiden dan Wakil Presiden RI periode 2024 – 2029. Berdasarkan Lembaga Puspoll Indonesia merilis survei terkait elektabilitas calon presiden dalam Pemilihan Presiden 2024. Ada Anies Baswedan yang nantinya dalam penelitian ini menganalisis sentimen ketiga tokoh politik tersebut.

Pada penelitian ini dilakukan analisis sentimen menggunakan metode Naïve bayes dalam penentuan polaritas sentimen. Analisis sentimen sendiri merupakan alat untuk memproses koleksi hasil pencarian yang bertujuan untuk mencari atribut dari suatu produk dan bagaimana proses dalam mendapatkan hasil pendapatnya (Lawrence et al, 2014). Dengan menerapkan metode Naïve Bayes Classifier dalam melihat sentimen masyarakat di media sosial twitter terhadap tokoh politik yang akan maju dalam pemilihan presiden Indonesia pada tahun 2024 diharapkan dapat memberikan hasil opini dari masyarakat yang akan membantu masyarakat dalam menentukan pilihan mana presiden yang tepat untuk masyarakat Indonesia.

2. Metodologi Penelitian

Metodologi penelitian menjelaskan tahapan yang dilakukan dalam perancangan analisis sentimen berdasarkan opini dari media sosial twitter terhadap “Figure Pemimpin” menggunakan python. Tahapan atau metodologi penelitian dijelaskan sebagai berikut.



Gambar 1. Alur Penelitian

Pada gambar 1 dijelaskan crawling data merupakan proses dimana mendapatkan data dari postingan atau tweet pengguna Twitter kemudian setelah semua data dikumpulkan selanjutnya tahap proses labelling data untuk menentukan sentimen terhadap postingan pengguna Twitter yang didapatkan. Langkah kedua, dilakukan proses preprocessing berguna sebagai menyeleksi data serta mengubahnya menjadi data yang lebih terstruktur. Pada sistem preprocessing terdapat 4 fase yang diselesaikan, yaitu Cleaning, Eliminasi Stopword, Tokenization dan Stemming. Pada tahap Cleaning, berguna sebagai pembersih kata-kata yang diharapkan tidak mengurangi keributan seperti html, koneksi, dan konten. Selain kata-kata yang tidak boleh diabaikan pada tahap ini, juga menghilangkan tanda centang seperti titik (.), koma (,) dan juga tanda aksentuasi lainnya. Selain menghilangkan kata dan tanda aksentuasi, menahan Case Collapsing juga mengubah semua kata menjadi huruf kecil (huruf kecil). Tahap selanjutnya adalah tahap Eliminasi Stopword, pada tahap ini kata-kata yang kurang penting atau tidak penting akan dihapus, seperti kata-kata: I, dan, atau. Kemudian, pada saat itu, memasuki tahap ketiga, tepatnya tahap Tokenization, yang digunakan untuk membedakan kata-kata yang terkandung dalam teks menjadi beberapa susunan yang dipersingkat dengan spasi atau ditambah dengan karakter yang unik.

Tahap terakhir dalam sistem preprocessing adalah tahap Stemming, pada tahap ini mengubah kata yang diikat kembali ke struktur uniknya. Langkah ketiga adalah proses ekstraksi komponen pada siklus ini, termasuk pembuatan yang dilakukan. Langkah keempat adalah menguji model sebagai estimasi esteem eksekusi order yang telah selesai. Langkah terakhir, setelah uji model selesai, menilai model dengan melihat pada tingkat sentimen analisis.

3. Kajian Pustaka

Bagian kajian pustaka membahas teori dasar ilmu yang digunakan dalam penelitian analisis sentimen berdasarkan opini dari media sosial twitter terhadap "Figure Pemimpin" menggunakan python. Teori didapat berdasarkan hasil studi literatur yang berkaitan dengan pengembangan analisis sentimen.

3.1. Sentimen Analisis

Sentiment analysis atau analisis sentimen dalam bahasa Indonesia adalah sebuah teknik atau cara yang digunakan untuk mengidentifikasi bagaimana sebuah sentimen diekspresikan menggunakan teks dan bagaimana sentimen tersebut bisa dikategorikan sebagai sentimen positif maupun sentimen negatif. Hasil sistem prototipe mencapai tinggi presisi (75-95% tergantung pada data) dalam mencari sentimen pada halaman web dan artikel berita, [3]

3.2. Data Mining

Menurut [4], Data mining merupakan gabungan dari ragam bidang ilmu antara lain basis data, information retrieval, statistika, algoritma dan machine learning. Bidang keilmuan ini sudah berkembang telah lama namun mempunyai peran penting pada masa sekarang ini dimana keluar kepentingan untuk memperoleh informasi yang lebih dari sebuah data transaksi maupun fakta yang telah dikumpulkan selama bertahun-tahun. Data mining salah satu cara mendapatkan informasi tersembunyi pada sebuah basis data dan tergolong bagian pada proses Knowledge Discovery in Databases (KDD) guna mendapatkan informasi atau pola yang berfungsi pada data.

3.3. Natural Language Processing (NLP)

Natural Language Processing (NLP) adalah salah satu bidang ilmu komputer yang merupakan cabang kecerdasan buatan dan bahasa (linguistik) yang berkaitan dengan interaksi antar komputer dan bahasa alami manusia, seperti bahasa indonesia atau bahasa inggris. Natural Language Processing adalah bagian dari pembelajaran mesin (machine learning) yang berkaitan dengan pembelajaran teks. Pembelajaran Natural Language Processing (NLP) bertujuan untuk membuat mesin yang dapat mengerti dan memahami makna bahasa manusia kemudian memberikan respon yang sesuai [5].

3.4. Media Sosial

Jejaring sosial merupakan situs dimana setiap orang bisa membuat web page pribadi, kemudian terhubung dengan teman-teman untuk berbagi informasi dan berkomunikasi. Jejaring sosial terbesar antara lain facebook, myspace, plurk, twitter, dan instagram. Jika media

tradisional menggunakan media cetak dan media broadcast, maka media sosial menggunakan internet. Media sosial mengajak siapa saja yang tertarik untuk berpartisipasi dengan memberi kontribusi dan feedback secara terbuka, memberi komentar, serta membagi informasi dalam waktu yang cepat dan tak terbatas.

Media sosial adalah sebuah media online, dengan para penggunanya bisa dengan mudah berpartisipasi, berbagi dan menciptakan isi meliputi blog, jejaring sosial, wiki, forum dan dunia virtual. Blog, jejaring sosial dan Wiki merupakan bentuk media sosial yang paling umum digunakan oleh masyarakat di seluruh dunia.

3.5. Twitter

Twitter merupakan platform media sosial umum digunakan user untuk berkomunikasi dan menyebarkan informasi berupa tweets. Tweets dapat di jadikan sumber data penting untuk melakukan penelitian Neuro-Linguistic Programming (NLP) seperti analisa sentimen, deteksi polaritas dan prediksi emoji [6].

3.6. Python

Python merupakan salah satu dari bahasa pemrograman yang sering digunakan oleh programmer atau pembuat program dalam membuat program mereka. Python memiliki karakteristik sintaks yang tidak terlalu rumit. Sehingga Python menjadi salah satu bahasa pemrograman tingkat tinggi yang mudah untuk digunakan. Dalam menulis sebuah kode program menggunakan bahasa pemrograman Python, terdapat beberapa aturan yang harus dipenuhi. Hal ini untuk mengantisipasi terjadinya error atau masalah pada program yang dibuat. Aturan sintaks Python yang pertama adalah dalam penulisan Statement atau perintah.

4. Hasil dan Pembahasan

Hasil dan pembahasan penelitian menjelaskan implementasi analisis sentimen berdasarkan opini dari media sosial twitter terhadap “Figure Pemimpin” menggunakan python. Berikut adalah tampilan dari implementasi analisis sentimen.

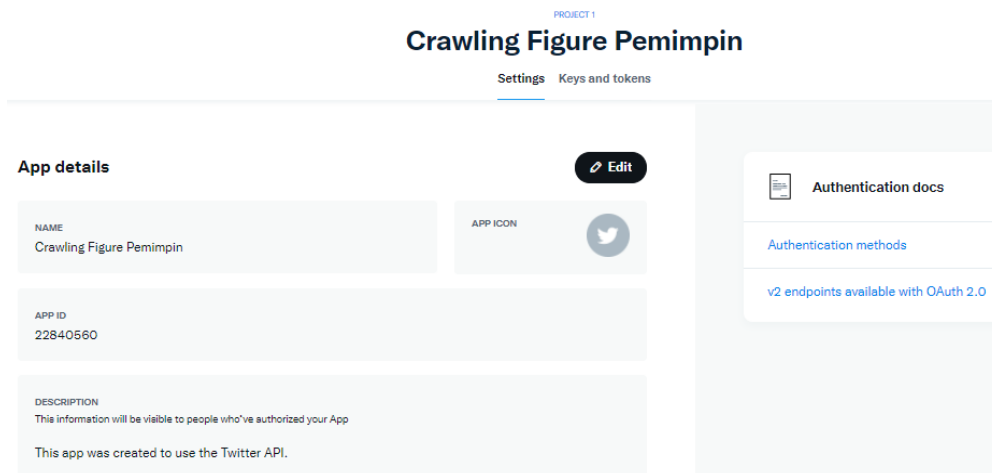
4.1. Pengumpulan Data

Tahap awal yang dilakukan adalah mengumpulkan data tweet berbahasa Indonesia dengan pencarian kata kunci “Anies Baswedan” menggunakan Twitter API. Terlebih dahulu dilakukan Import Libraries yang disediakan oleh Python. Library yang digunakan dalam pengumpulan data yang digunakan adalah library tweepy yang dapat mengakses API twitter secara langsung di console ataupun script. Library sys adalah library yang digunakan untuk menyediakan akses ke beberapa variabel yang digunakan atau dikelola oleh penerjemah, library matplotlib.pyplot digunakan untuk membuat fungsi visualisasi kedalam bentuk grafik.

```
# Import Libraries
from textblob import TextBlob
import sys
import tweepy
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import os
import nltk
import pycountry
import re
import string
```

Gambar 2. Library Python

Untuk dapat mengambil data dari *twitter*, terlebih dahulu mendaftar di *twitter Developer*, dilakukan pengisian semua *form* yang ditampilkan. Setelah selesai proses mendaftar akan diberikan *consumer key*, *consumer secret*, *access token* dan *access token secret* yang digunakan untuk mengakses data pada *twitter*.



Gambar 3. Mendaftar pada *Developer Twitter*

Setelah didapatkan *Token* dari *Developer twitter* selanjutnya dilakukan autentikasi *token* dengan *code python* seperti pada gambar 4.

```
# Authentication
api_key = "HjAVN0rcVtR2UtAh4eYfE8uRW"
api_secret_key = "QbW6XXut9w96y3BTMaxZG3w2MpA3jawc94EH92ohhQh0DI4iEW"
access_token = "857077190382440449-ETIutqKzPtvrv1gCdCsm8Xobc_jpeCZZ"
access_token_secret = "1UBi87IjKapYAHFbw9Hkyajuc1GCbLODiZs2VEVDxVJK6"
auth = tweepy.OAuthHandler(api_key, api_secret_key)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth, wait_on_rate_limit=True)
```

Gambar 4. *Code Python Aunthentikasi Token*

Setelah terkoneksi dengan *twitter*, program akan meminta data tentang kata kunci maupun hashtag yang dicari, kemudian dimasukkan kedalam variabel *keyword* dan akan meminta jumlah data *tweet* yang akan dianalisis. Kemudian dimasukkan kedalam variabel *noOfTweet* yang berupa integer. Variabel *tweet* akan melakukan operasi pengambilan data yang telah disimpan pada variabel *searchTerm* dan *noOfSearchTerm*. Kata kunci yang digunakan dalam pencarian data adalah "Anies Baswedan". Penggunaan *Query* tersebut dikarenakan Anies Baswedan merupakan hasil terkuat dari Lembaga Puspoll Indonesia merilis survei terkait elektabilitas calon presiden dalam Pemilihan Presiden 2024. Pada penelitian kali ini versi API Twitter yang digunakan tidak mendukung dalam pengambilan data dalam waktu dikarenakan setiap hari data twitter selalu update dan diperbaharui dengan otomatis.

```
keyword = input("Masukkan Hashtag atau Query yang dicari: ")
noOfTweet = int(input("Masukkan jumlah Tweet : "))
tweets = tweepy.Cursor(api.search_tweets, q=keyword).items(noOfTweet)
```

Gambar 5. *Code Python Crawling Data*

Data yang terkumpul sebanyak 500 *tweet* dikumpulkan dan disimpan kedalam file excel dengan format *.csv*. Tidak ada batasan dalam jumlah data yang diambil namun dengan 500 data diharapkan dapat mewakili hasil opini masyarakat secara umum. Data yang tersimpan terdiri dari tanggal pembuatan *tweet*, *user twitter* yang memposting *tweet* mengenai kuliah daring dan kuliah online serta *tweet* yang diposting.

	Tweet
0	RT @kebijakananies: Dua Tahun Berturut-turutJakarta Raih STA 2021 Kalahkan Kota Besar DuniaANIES BASWEDAN:“Penghargaan berturut-turut i...
1	PSI Usulkan Pembentukan Pansus Proyek Sumur Resapan Milik Anies Baswedan, Dianggap Meresahkan Warga... https://t.co/geXmWVxaOw
2	RT @GeiszChalifah: Anies Baswedan: Saya tak membalas kata-kata dengan kata- kata. Saya menjawab kata2 dengan karya-karya, lalu hasilnya: DU...
3	Relawan dari Solo Hadiahkan Lagu untuk Anies Baswedan Republika Online Mobile https://t.co/fDbx04nrzC
4	RT @kebijakananies: Dua Tahun Berturut-turutJakarta Raih STA 2021 Kalahkan Kota Besar DuniaANIES BASWEDAN:“Penghargaan berturut-turut i...
5	RT @GeiszChalifah: Anies Baswedan: Saya tak membalas kata-kata dengan kata- kata. Saya menjawab kata2 dengan karya-karya, lalu hasilnya: DU...
6	RT @kebijakananies: Dua Tahun Berturut-turutJakarta Raih STA 2021 Kalahkan Kota Besar DuniaANIES BASWEDAN:“Penghargaan berturut-turut i...
7	RT @FerdinandHaeen3: Jk gagal Juni 2022, mk Anies Baswedan hrs dipenjarakan karena telah merugikan keuangan negara sangat besar...!!! Rakyat...
8	RT @FerdinandHaeen3: Jk gagal Juni 2022, mk Anies Baswedan hrs dipenjarakan karena telah merugikan keuangan negara sangat besar...!!! Rakyat...
9	@GunRomli Akhir-akhir ini sangat bahagia melihat penderitaan haternya Bapak Anies Baswedan menderita.
10	RT @MCAOps: Raihan Sustainable Transport Award (STA) dua tahun beruntun adalah bukti pengakuan dunia terhadap inovasi Provinsi DKI Jakarta...
11	RT @GunRomli: Anies Baswedan Disentil Guntur Romli Soal Formula E: Udah Habis Ratusan Miliar, Ternyata Kajiannya Belum https://t.co/BPaOBf...
12	RT @GunRomli: Memuluskan jalan: butuh anggaran, menggali jalan: butuh anggaran, memuluskan lagi jalan: lagi2 butuh anggaran, melubangi jala...
13	RT @FerdinandHaeen3: Beri dia penghargaan karena telah berhasil menjadikan tukang mi ayam jadi kontraktor. Ini prestasi besar Anies Baswedan...
14	RT @GunRomli: Anies Baswedan Disentil Guntur Romli Soal Formula E: Udah Habis Ratusan Miliar, Ternyata Kajiannya Belum https://t.co/BPaOBf...
15	RT @Lady_Zeebo: Turki bangkrut hancur lebur, gara-gara dipimpin oleh IM (Ikhwanul Muslimin). DKI Jakarta hancur, gara-gara dipimpin ole...
16	RT @Lady_Zeebo: Turki bangkrut hancur lebur, gara-gara dipimpin oleh IM (Ikhwanul Muslimin). DKI Jakarta hancur, gara-gara dipimpin ole...
17	RT @GeiszChalifah: Anies Baswedan: Saya tak membalas kata-kata dengan kata- kata. Saya menjawab kata2 dengan karya-karya, lalu hasilnya: DU...
18	RT @FerdinandHaeen3: Jk gagal Juni 2022, mk Anies Baswedan hrs dipenjarakan karena telah merugikan keuangan negara sangat besar...!!! Rakyat...
19	RT @Lady_Zeebo: Turki bangkrut hancur lebur, gara-gara dipimpin oleh IM (Ikhwanul Muslimin). DKI Jakarta hancur, gara-gara dipimpin ole...

Gambar 6. Hasil Pengumpulan data dengan *Twitter API*

4.2. *Praprocessing Data*

Tahapan *praprocessing* data perlu dilakukan karena beberapa kalimat *tweet* yang didapatkan tidak sepenuhnya menggunakan kata baku dan menggunakan bahasa indonesia yang baik. *Preprocessing* dilakukan menggunakan bantuan *library* pada bahasa pemrograman *Python*.

4.2.1. *Case Folding*

Case Folding adalah proses merubah data *tweet* menjadi *lowercase*. Berikut merupakan contoh data penelitian yang dilakukan proses *case folding*.

```
# ----- Case Folding -----
# gunakan fungsi Series.str.lower() pada Pandas
tw_list['text'] = tw_list['text'].str.lower()
print('Case Folding Result : \n')
print(tw_list['text'].head(5))
print('\n\n')
```

Gambar 7. Kode Program Tahap *Case Folding*

No	Tweet	Case Folding
1	RT @keuangannews_id: Mayjen (Purn) Deddy Budiman: Terdapat Upaya Menjegal Anies Baswedan Menjadi Capres 2024 https://t.co/jkAPNRSn7l	rt id mayjen purn deddy budiman terdapat upaya menjegal anies baswedan menjadi capres 2024 https://t.co/jkAPNRSn7l
2	RT @Rossali24295306: Anies Baswedan Didukung JK dan Habib Rizieq di Pilpres 2024...? #BelaAniesPresiden2024	rt anies baswedan didukung jk dan habib rizieq di pilpres 2024...? #belaaniespresiden2024

Tabel 1. Hasil *Praprocessing Case Folding*

4.2.2. Tokenizing

Tokenizing dalam penelitian ini merupakan tahapan dalam memecah *string* atau *input* terhadap suatu teks yang telah melewati tahap *case folding* berdasarkan tiap kata yang menyusunnya dan menghilangkan URL, @mention dan hashtag. Tahap *tokenization* dilakukan dengan menggunakan fungsi *nltk_tokenize()*, *library* pada bahasa pemrograman *Python3* yang bernama NLTK.

```
import string
import re #regex Library
```

Gambar 8. Import Library yang dibutuhkan

Library string digunakan untuk memuat satu karakter atau lebih yang ada pada data *tweet*. Terlebih dahulu diimport *library re* untuk melakukan tahapan *Regular Expression (regex)* atau deretan karakter yang digunakan untuk pencarian teks dengan menggunakan pola (*pattern*). Dengan menggunakan *library regex* dapat memudahkan dalam mencari string tertentu dari teks yang banyak. Selain itu pada tahap ini juga dilakukan proses *removing number*, *whitespace* dan *punctuation* (tanda baca).

```
def remove_tweet_special(text):
    # remove tab, new line, ans back slice
    text = text.replace('\t'," ").replace('\n'," ").replace('\u'," ").replace('\'," ")
    # remove non ASCII (emoticon, chinese word, .etc)
    text = text.encode('ascii', 'replace').decode('ascii')
    # remove mention, link, hashtag
    text = ' '.join(re.sub("[@#][A-Za-z0-9+](\w+:\w+\S+)", " ", text).split())
    # remove incomplete URL
    return text.replace("http://", " ").replace("https://", " ")
tw_list['text'] = tw_list['text'].apply(remove_tweet_special)

#remove number
def remove_number(text):
    return re.sub(r"\d+", "", text)
tw_list['text'] = tw_list['text'].apply(remove_number)

#remove punctuation
def remove_punctuation(text):
    return text.translate(str.maketrans("", "", string.punctuation))
tw_list['text'] = tw_list['text'].apply(remove_punctuation)

#remove whitespace leading & trailing
def remove_whitespace_LT(text):
    return text.strip()
```

Gambar 9. Code Python Tahap Tokenizing

No	Tweet	Tokenizing
1	RT @MissKoral: Anies Baswedan for the next Presiden 2024...insya Allah #IndonesiaHarusBersamaOrangBaik https://t.co/wdKwelD4f7	anies, baswedan, for, the, next, presiden, insya, allah
2	@InSrikandi_____ Semoga dilancarkan jalan pak Anies Baswedan untuk jadi presiden RI #BelaAniesPresiden202	semoga, dilancarkan, jalan, pak, anies, baswedan, untuk, jadi, presiden, ri.

	https://t.co/JT1UW9yXdf
--	---

Tabel 2. Hasil *Praprocessing Tokenizing*

Pada tahap ini juga dilakukan proses normalization, yaitu mengubah kata yang tidak lengkap, kesalahan dalam pengetikan (typo) kedalam kata yang normal dan dapat dipahami dengan baik.

```

normalizad_word_dict = {}

for index, row in normalizad_word.iterrows():
    if row[0] not in normalizad_word_dict:
        normalizad_word_dict[row[0]] = row[1]

def normalized_term(document):
    return [normalizad_word_dict[term] if term in normalizad_word_dict else term for term in document]

tw_list['tweet_normalized'] = tw_list['tweet_tokens_WSW'].apply(normalized_term)
    
```

Gambar 10. Code Python Tahap Normalization

No	Tweet	Normalization
1	semoga, dilancarkan, jalan, pak, anies, baswedan, untuk, jadi, presiden, ri.	semoga, dilancarkan, jalan , anies, baswedan, presiden, ri.

Tabel 3. Hasil Proses Normalization

4.2.3. Filtering

Proses membuang kata yang tidak memiliki arti. Proses *filtering* disebut dengan *Stopword Removal*. Pada tahap ini menggunakan nltk. NLTK (*Natural Language ToolKit*) adalah *Library* yang disediakan oleh oleh *Python* untuk membangun program analisis teks. Terlebih dahulu dilakukan proses *install library* nltk pada *anaconda prompt*.

Selain list stopwords Indonesia yang disediakan oleh library nltk, ditambahkan list kata yang tidak dibutuhkan dalam analisis sentimen dengan cara menambahkan secara langsung kata pada `list_stopword.extends` agar dapat dihapus oleh sistem.

```

from nltk.corpus import stopwords

# ----- get stopwords from NLTK stopwords -----
# get stopwords indonesia
list_stopwords = stopwords.words('indonesian')
# ----- manually add stopwords -----
# append additional stopwords
list_stopwords.extend(["yg","make", "dg", "gua","gw", "rt", "dgn", "ny", "d", "klo",
'kalo', 'amp', 'bian', 'bikin', 'bilang',
'gak', 'ga', 'krn', 'nya', 'nih', 'sih',
'si', 'tau', 'tdk', 'tuh', 'utk', 'ya',
'jd', 'jgn', 'sdh', 'aja', 'n', 't',
'nyg', 'hehe', 'bgst', 'pen', 'u', 'nan', 'loh', 'rt',
'wkuk', 'wkuk', 'guys', 'kan',
'&amp;', 'yah'])

# convert list to dictionary
list_stopwords = set(list_stopwords)

#remove stopwords pada list token
def stopwords_removal(words):
    return [word for word in words if word not in list_stopwords]
tw_list['tweet_tokens_WSW'] = tw_list['tweet_tokens'].apply(stopwords_removal)
print(tw_list['tweet_tokens_WSW'].head())

normalizad_word_dict = {}
def normalized_term(document):
    return [normalizad_word_dict[term] if term in normalizad_word_dict else term
for term in document]
tw_list['tweet_normalized'] = tw_list['tweet_tokens_WSW'].apply(normalized_term)
    
```

Gambar 11. Code Python Tahap Filtering

No	Tweet	Filtering
----	-------	-----------

1	semoga, dilancarkan, jalan, pak, anies, baswedan, untuk, jadi, presiden, ri.	semoga, dilancarkan, jalan , anies, baswedan, presiden, ri.
---	--	---

Tabel 4. Hasil Praprocessing Filtering

4.2.4. Stemming

Stemming adalah tahap mencari *root* (dasar) kata dari tiap kata hasil *filtering* dengan menghapus kata imbuhan di depan maupun imbuhan di belakang kata. Tahap *stemming* dilakukan dengan menggunakan bantuan *library* pada bahasa pemrograman *Python3* yang bernama *Sastrawi*.

```

stemmer = factory.create_stemmer()

# stemmed
def stemmed_wrapper(term):
    return stemmer.stem(term)
term_dict = {}
for document in tw_list['tweet_normalized']:
    for term in document:
        if term not in term_dict:
            term_dict[term] = ''

print(len(term_dict))
print("-----")
for term in term_dict:
    term_dict[term] = stemmed_wrapper(term)
    print(term, ":", term_dict[term])

print(term_dict)
print("-----")

# apply stemmed term to dataframe
def get_stemmed_term(document):
    return [term_dict[term] for term in document]
tw_list['tweet_tokens_stemmed'] = tw_list['tweet_normalized'].swifter.apply(get_stemmed_term)
print(tw_list['tweet_tokens_stemmed'])
tw_list.to_excel("craw/Anies Baswedan Proses.xlsx")
    
```

Gambar 12. Code Python Tahapan Stemming

Kata Imbuhan	Kata Dasar
Kalahkan	Kalah
Penghargaan	Harga
Usulkan	Usul
Pembentukan	Bentuk
dianggap	Anggap
Meresahkan	Resah
dilancarkan	Lancar
Pendukung	Dukung

Tabel 5. Hasil Tahapan Stemming

4.3. Pelabelan Data

Pelabelan data hasil crawling dan telah melalui tahapan preprocessing dilakukan dengan menggunakan library python textblob dengan melihat polarity, subjectivity yang dimiliki oleh teks tweet yang telah dikumpulkan. Textblob adalah salah satu library yang disediakan

oleh Python untuk pemrosesan dibidang Natural Language Processing yang dapat memberikan tag kata, ekstraksi kata, penerjemahan kata dan sentiment analysis. Hasil objek textblob dapat digunakan untuk melakukan proses pembelajaran bahasa alami namun saat ini textblob hanya tersedia dalam bahasa inggris oleh sebab itu pada penelitian ini yang menggunakan data bahasa Indonesia dilakukan translate ke dalam bahasa inggris terlebih dahulu. Penentuan kelas positif, netral dan negatif didasari oleh nilai polaritas. Nilai polaritas pada analisis sentimen berada pada rentang 1 sampai -1 yang menunjukkan kelas sentimen data.

tweet_tokens_stemmed	polarity	subjectivity	sentiment	neg	neu	pos	compound
[berturuturut, 'jakarta', 'raih', 'sta', 'kalah', 'kota', 'dunia', 'anies', 'baswedan', 'harga', 'berturuturut']	0	0	neutral	0	1	0	0
[psi, 'usuf', 'bentuk', 'panas', 'proyek', 'sumur', 'resap', 'miisk', 'anies', 'baswedan', 'anggap', 'resah', 'warga']	0	0	neutral	0	1	0	0
[anies, 'baswedan', 'balas', 'katakata', 'karyakarya', 'hasil', 'du']	0	0	neutral	0	1	0	0
[rawan, 'solo', 'hadiah', 'lagu', 'anies', 'baswedan', 'republik', 'online', 'mobile']	0	0	neutral	0	1	0	0
[jk, 'gagal', 'juni', 'mk', 'anies', 'baswedan', 'his', 'perjara', 'ngi', 'uang', 'negara', 'rakyat']	0	0	positive				
[akhirakhir, 'bahaga', 'berita', 'haternya', 'anies', 'baswedan', 'denta']	0	0	neutral	0	1	0	0
[raih, 'sustainable', 'transport', 'sewad', 'sta', 'lmtan', 'bukti', 'aku', 'duma', 'inovasi', 'provinsi', 'dki', 'jakarta']	0	0	positive				
[anies, 'baswedan', 'sempil', 'guntur', 'romli', 'formulu', 'uduh', 'habis', 'ratas', 'militer', 'kaj']	0	0	neutral	0	1	0	0
[mulus, 'jalan', 'butuh', 'anggap', 'gali', 'jalan', 'butuh', 'anggap', 'mulus', 'jalan', 'butuh', 'anggap', 'lubang', 'jala']	0	0	neutral	0	1	0	0
[harga, 'hasil', 'jadi', 'tukang', 'mi', 'ayam', 'kontraktor', 'prestasi', 'anies', 'baswedan']	0	0	neutral	0	1	0	0
[zeebo, 'turki', 'bangkrut', 'hancur', 'lebur', 'garagara', 'dipimpin', 'im', 'khwanul', 'muslimin', 'dki', 'jakarta', 'hancur', 'garagara', 'dipimpin']	0	0	neutral	0	1	0	0
[anies, 'baswedan', 'bagi', 'ribu', 'sertifikat', 'tanah', 'warga', 'dki', 'goodheiver']	0	0	neutral	0	1	0	0
[jakarta, 'home', 'for', 'all', 'by', 'anies', 'rasyd', 'baswedan', 'the', 'first', 'of', 'three', 'articles']	0.25	0.33333	neutral	0	1	0	0
[anies, 'baswedan', 'bebas', 'gali', 'jakarta', 'pendit', 'nyases', 'genombol', 'arab', 'kussa', 'mpu']	0	0	neutral	0	1	0	0
[tdf, 'rame', 'umy', 'logis', 'sacara', 'seminar', 'good', 'green', 'government', 'umy', 'undang', 'anies', 'baswedan', 'ganjar', 'pranowo']	0.25	0.45	positive				
[viral, 'emakemak', 'dita', 'nyanyi', 'lagu', 'anies', 'baswedan', 'isi']	0	0	neutral	0	1	0	0
[viral, 'emakemak', 'dita', 'nyanyi', 'lagu', 'anies', 'baswedan', 'isi']	0	0	neutral	0	1	0	0
[asmara, 'retweet', 'keras', 'gus', 'pimpin', 'anies', 'baswedan', 'moga', 'dijabab', 'allah', 'swt']	0	0	neutral	0	1	0	0
[jakarta, 'unifying', 'home', 'by', 'anies', 'rasyd', 'baswedan', 'the', 'second', 'of', 'three', 'articles']	0	0	neutral	0	1	0	0
[piveneh, 'berita', 'mest', 'anies', 'baswedan', 'teases', 'jokowi', 'jubir']	0	0	neutral	0	1	0	0
[kritik, 'hina', 'anies', 'baswedan', 'satu', 'anies', 'baswedan', 'lapor', 'jokowi']	0	0	neutral	0	1	0	0
[anies, 'baswedan', 'kenzo', 'harga', 'pajak', 'mend', 'gk', 'dki', 'jakarta', 'nilai', 'hasil', 'jalan', 'inklusi', 'uang', 'ba']	0	0	positive				
[gubernur, 'anies', 'baswedan', 'putus', 'jakarta', 'ppkm', 'level']	0	0	neutral	0	1	0	0
[telak, 'sentil', 'amat', 'anies', 'baswedan', 'kerja', 'bohong']	0	0	neutral	0	1	0	0

Gambar 13. Hasil Pelabelan Tweet dengan *TextBlob*

4.4. Hasil dan Visualisasi Klasifikasi Sentimen

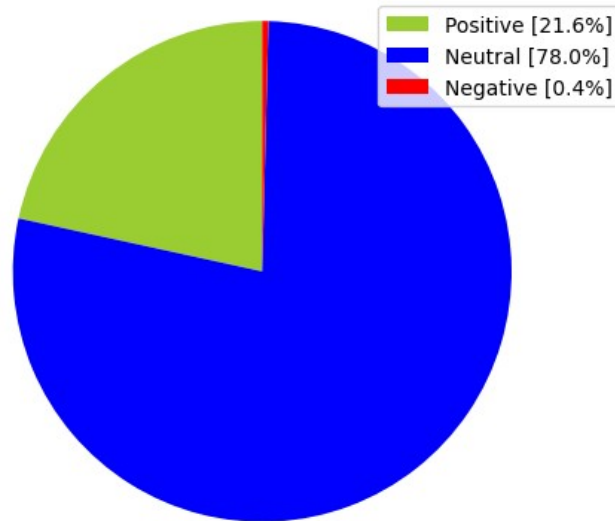
Selanjutnya data akan dibagi dengan 80% data latih dan 20% data uji. Sentimen pada data training sejumlah 1600 dibagi secara manual sesuai kelasnya. 400 data lainnya akan digunakan sebagai data uji.

Positif	Netral	Negatif	Jumlah
108	390	2	500
21,6%	78%	0,4%	100%

Tabel 6. Hasi Pelabelan Data dengan *Textblob*

Didapatkan hasil akhir dari pelabelan dengan menggunakan *library textblob* sebanyak 500 data *tweet* adalah 108 *tweet* yang masuk dalam kelas positif, 390 *tweet* kelas netral dan 2 *tweet* kelas negatif.

Sentiment Analysis Result for keyword= Anies Baswedan



Gambar 14. Grafik Analisis Sentimen

5. Kesimpulan

Penggunaan analisis sentimen terhadap opini masyarakat Indonesia di media sosial Twitter merupakan sebuah teknik atau cara yang digunakan untuk mengidentifikasi bagaimana sebuah sentimen diekspresikan menggunakan teks dan bagaimana sentimen tersebut bisa dikategorikan sebagai sentimen positif maupun sentimen negatif.

Daftar Pustaka

- [1] F. Nurhuda, S. Widya Sihwi, and A. Doewes, "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," *J. Teknol. Inf. ITSmart*, vol. 2, no. 2, p. 35, 2016, doi: 10.20961/its.v2i2.630.
 - [2] B. Liu, "Opinion spam detection," *Opin. Anal. Online Rev.*, no. May, pp. 79–94, 2012, doi: 10.1142/9789813100459_0007.
 - [3] J. Nasukawa, T. and Yi, "Sentiment Analysis: Capturing Favorability Using Natural Language Processing. Proceedings of the 2nd International Conference on Knowledge Capture," 2003.
 - [4] Hasmawati, J. Nangi, and M. Muchtar, "Aplikasi prediksi penjualan barang menggunakan metode k- nearest neighbor (knn) (studi kasus tumaka mart)," *semantik*, vol. 3, no. Jul-Des 2017, pp. 151–160, 2019.
 - [5] M. ZAMIL, "Klasifikasi Kalimat Ofensif pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," 2019.
 - [6] N. Chadha, R. C. Gangwar, and R. Bedi, "Current Challenges and Application of Speech Recognition Process using Natural Language Processing: A Survey," *Int. J. Comput. Appl.*, vol. 131, no. 11, pp. 28–31, 2015, doi: 10.5120/ijca2015907471.
-