

Analisis Sentimen pada Postingan Media Sosial X Tentang Palestina setelah Peristiwa Tragedi Rafah Melalui Pendekatan Regular Expression

Dhidit Abdi Aziz - 13522040

Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jalan Ganesha 10 Bandung
E-mail: dhiditabdi@gmail.com

Abstract—Penyerangan Israel terhadap Palestina masih terus berjalan dan semakin memburuk. Pada akhir bulan Mei 2024, Israel memulai penyerangan terhadap Rafah, Palestina yang merupakan zona aman dan tempat pengungsian warga sipil. Kampanye dan demonstrasi terus digalakkan baik melalui turun ke jalan ataupun digital agar Palestina mendapatkan haknya. Untuk mengetahui efektifitas metode kampanye pada Twitter, dilakukan analisis sentimen pada cuitan teratas yang membahas tentang Palestina sejak sebulan terakhir. Digunakan metode *regular expression* untuk melakukan pembersihan pada data agar hasil yang diperoleh menjadi lebih akurat. Didapatkan hasil bahwa sentimen positif pengguna Twitter terhadap Palestina masih dibawah 50% sehingga perlu dilakukan peningkatan kembali terkait kuantitas dan kualitas kampanye.

Keywords—*Palestina; Sentimen; Regular Expression*

I. PENDAHULUAN

Pada Senin, 6 Mei 2024, Israel memerintahkan warga Palestina untuk mengevakuasi sebagian Rafah, Palestina. Pernyataan evakuasi tersebut tidak disertai alasan spesifik, juga tidak menyebutkan apakah akan ada tindakan ofensif yang akan terjadi. Dalam waktu sekitar 20 hari, setidaknya 945.000 warga sipil Rafah harus mengungsi. Mayoritas dari mereka harus berjalan kaki melewati medan perang yang sangat berbahaya. Terlebih lagi, Israel menutup gerbang perbatasan Rafah yang bertetangga dengan Mesir. Padahal, gerbang Rafah merupakan satu-satunya akses bantuan kemanusiaan untuk bisa mencapai wilayah selatan Gaza. Situasi pertempuran yang begitu berbahaya juga membuat UNRWA tidak dapat mendistribusikan bantuan kemanusiaan ke wilayah Rafah [1].

Pada Minggu, 26 Mei 2024, terjadi serangan udara terhadap kamp pengungsian Tel Al-Sultan di Rafah. Peristiwa ini menewaskan sedikitnya 45 orang dan 249 lainnya terluka. Selain karena terkena ledakan, mayoritas korban terkena kobaran api yang menjalar dengan cepat di pengungsian. Akibatnya, jumlah korban sipil Palestina akibat penyerangan Israel terus bertambah banyak. Hingga 27 Mei 2024, angka korban jiwa mencapai 36.050 jiwa, sementara 81.026 lainnya terluka.

Dunia segera mengecam keras serangan Israel yang dianggap sangat brutal tersebut. Hal ini utamanya disebabkan bom Israel menghantam tenda-tenda pengungsi di area yang ditetapkan sebagai zona aman. Kampanye dan demonstrasi semakin meluas baik itu di jalanan maupun melalui media sosial.

X atau Twitter turut menjadi salah satu media kampanye untuk meningkatkan kepedulian masyarakat dunia terhadap peristiwa tragedi Rafah. X memiliki jumlah pengguna aktif yang mencapai 500 juta per bulan serta menjangkau 12,6% dari pengguna internet dunia. Terlebih lagi, mayoritas pengguna mengakses X dengan tujuan untuk mengikuti berita terbaru [2]. Dengan demikian, X merupakan salah satu media kampanye yang efisien dan efektif.

Keberhasilan suatu kampanye perlu diukur untuk mengetahui efektifitas dari metode yang telah digunakan. Untuk mengetahui sukses tidaknya kampanye Rafah pada aplikasi Twitter, diperlukan analisis terhadap sentimen publik dengan rentang waktu sebulan sejak terjadinya tragedi. Hasil dari analisis tersebut dapat digunakan sebagai dasar dalam pengambilan keputusan untuk menentukan metode kampanye yang lebih mutakhir.

II. LANDASAN TEORI

A. Analisis Sentimen

Analisis sentimen atau *sentiment analysis* adalah salah satu proses komputasi dengan menganalisis teks digital yang digunakan untuk menentukan apakah kata-kata atau kalimat yang disampaikan memiliki makna atau emosional pesan. Sentimen analisis akan memahami dan mengelompokkan emosi dalam kategori positif, negatif, dan netral. Dalam analisis sentimen, mencakup langkah-langkah mulai dari pemrosesan teks hingga pengklasifikasian sentimen. Sistem analisis sentimen menggabungkan *natural language processing* (NLP) dan *machine learning* [3]. Terdapat beberapa tipe analisis sentimen:

1) *Fine-graded sentiment analysis*

Biasa digunakan untuk mengukur opini, misalnya penilaian produk pada *e-commerce* berupa skala bintang 1-5

- 2) *Emotion detection analysis*
Untuk mendeteksi perasaan dan emosi dengan cara mencocokkan teks terhadap daftar kata yang berkorespondensi terhadap suatu emosi
- 3) *Fine-grained analysis*
Melakukan analisis dengan memecah kalimat menjadi beberapa bagian penyusunnya.
- 4) *Aspect-based sentiment analysis*
Mirip seperti *fined-graded analysis* sebab menggunakan masukan untuk mencari sentimen negatif atau positif.
- 5) *Intent analysis*
Melakukan analisis apakah suatu pernyataan berupa pertanyaan, keluhan, saran, pendapat, berita, dll.
- 6) *Multilingual sentiment analysis*
Digunakan untuk menganalisis kata-kata dengan berbagai bahasa

Analisis sentimen sangat bermanfaat untuk memberikan wawasan yang objektif dalam menganalisis pendapat pelanggan dengan cepat dalam skala masif. Dengan demikian, peningkatan kualitas produk ataupun layanan dapat tercapai berdasarkan umpan balik pelanggan yang tulus dan spesifik [4].

B. Pencocokan String

Pencocokan *string* merupakan algoritma untuk melakukan pencarian semua kemunculan *string* pendek *pattern* $[0..n-1]$ di *string* yang lebih panjang *text* $[0..m-1]$ [5].

Persoalan pencocokan *string* dirumuskan sebagai berikut [6]:

- (1) Teks, yaitu *string* dengan panjang m karakter
- (2) *Pattern*, yaitu *string* dengan panjang n karakter ($n < m$) yang akan dicari dalam teks
- (3) Akan dicari lokasi pertama di dalam teks yang bersesuaian dengan *pattern*

Terdapat beberapa algoritma pencocokan *string* berdasarkan arah pencariannya:

- (1) Kiri ke kanan
Bersesuaian dengan cara baca manusia pada umumnya. Contoh algoritma pada kategori ini adalah *regular expression*, *brute force*, dan *knuth-morris-prat*
- (2) Kanan ke kiri
Menghasilkan hasil praktikal yang lebih baik. Contoh algoritmanya adalah *booyer-moore*

- (3) Urutan tertentu

Menghasilkan hasil teoritis yang terbaik. Contoh algoritmanya adalah *colossi* dan *crochemore-perrin*

Permasalahan pencocokan *string* memiliki aplikasi yang sangat luas. Salah satu yang paling umum adalah pencarian suatu kata di dalam dokumen (misalnya menu Find di dalam Microsoft Word)

C. Regular Expression

Regular expression adalah pola yang coba dicocokkan oleh mesin *regular expression* dalam sebuah teks masukan. Sebuah pola dapat terdiri dari satu atau lebih karakter literal, operator, ataupun konstruksi. Disamping karakter huruf dan angka, berikut ini beberapa format masukan pola dasar pada *regular expression* [7]:

TABEL 2.1 Format Masukan Dasar Regex

Simbol	Kegunaan
\char	Untuk mencocokkan karakter yang memiliki makna khusus di regex
	Operator OR
[...]	Menerima salah satu dari karakter apapun yang ada di dalam kurung
[.-.]	Menerima karakter yang berada di rentang nilai dua karakter pada kurung
^	Menegasikan seluruh ekspresi yang berada di belakangnya
+	Menerima satu atau lebih karakter yang berada di depannya
*	Menerima nol atau lebih karakter yang berada di depannya
?	Menerima nol atau satu karakter yang berada di depannya
\d	Menerima satu digit angka apapun
\w	Menerima satu karakter huruf apapun
\s	Menerima satu karakter non spasi apapun

Pencocokan *string* dengan *regular expression* ini termasuk kategori kiri ke kanan. Sebab, pada *string text* akan dicari letak pertama dimana *substring* cocok dengan aturan/query *regular expression* masukan pengguna.

III. METODOLOGI

A. Batasan Masalah

Penulis menerapkan beberapa batasan dengan tujuan menyederhanakan permasalahan analisis sentimen publik terhadap Palestina pasca tragedi Rafah, yaitu

- (1) Data yang diperoleh maksimal sebanyak 1000 cuitan teratas dalam kurun waktu sebulan ke belakang

- (2) Pembersihan data hanya dilakukan untuk menghapus cuitan yang tidak mengandung kata kunci Palestina. Selain itu, kata pengotor seperti emoji, tag, dsb, dihapus. Cuitan yang menjadi kehilangan makna setelah dilakukan pembersihan sehingga mengganggu proses perhitungan sentimen akan diabaikan
- (3) Data cuitan yang diambil hanyalah yang berbahasa Inggris sebab keterbatasan kemampuan *library*
- (4) Data hanya dikategorikan menjadi tiga tipe sentimen, yaitu positif, negatif, dan netral

B. Pemodelan Masalah

Penulis melakukan ekstraksi data dari API X dalam format CSV. Kemudian, pembersihan dan perhitungan sentimen dilakukan dalam format TXT. Hasil akhir yang akan dianalisis berupa persentase sentimen positif, negatif, dan netral dari data cuitan secara keseluruhan

IV. PENGOLAHAN DATA

Pengolahan data penelitian ini memiliki beberapa tahapan, mulai dari pengambilan data, pembersihan data, dan analisis sentimen. Berikut ini penjelasan lebih lanjutnya:

A. Pengambilan Data

Penulis mengambil data 1000 cuitan teratas di dunia yang mengandung kata kunci “Palestina” selama sebulan sejak peristiwa tragedi Rafah – Senin, 6 Mei 2024 hingga Kamis, 6 Juni 2024. Sebab *library* yang akan digunakan untuk analisis sentimen saat ini baru berjalan optimal di teks berbahasa Inggris, maka diberikan batasan bahasa pada pengambilan data. Terakhir, dipilih tipe cuitan selain *retweet* untuk mencegah didapatkannya data yang kurang akurat.

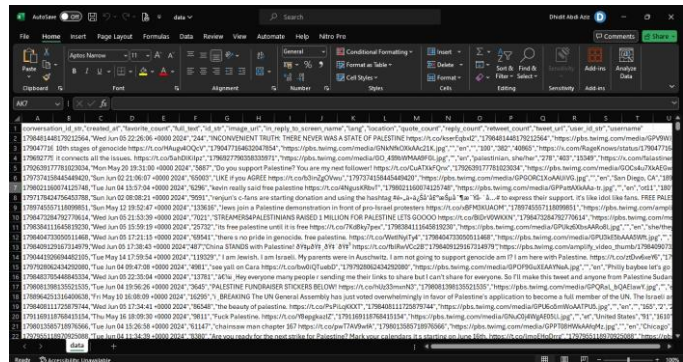
Penulis menggunakan salah satu package dari npm yaitu *tweet-harvest*. Package ini mampu mengumpulkan dan mengekstraksi data cuitan dari Twitter melalui *scrolling* secara otomatis di belakang layar. Berikut ini command yang penulis masukan berdasarkan keperluan pencarian dan batasan di atas

```
filename = 'rafah.csv'
search_keyword = 'Palestine OR
palestine -is:retweet since:2024-05-
06 until:2024-06-06 lang:en'
limit = 1000

!npx -y tweet-harvest@2.6.1 -o
"{filename}" -s "{search_keyword}" -
-tab "TOP" -l {limit} --token
{twitter_auth_token}
```

Setelah beberapa kali percobaan, pengumpulan data mendapatkan kendala berupa batasan dari API Twitter. Maka dari itu, data terbanyak yang berhasil diperoleh hanya sebanyak 741 cuitan. Data tersebut masih berupa csv yang berisi

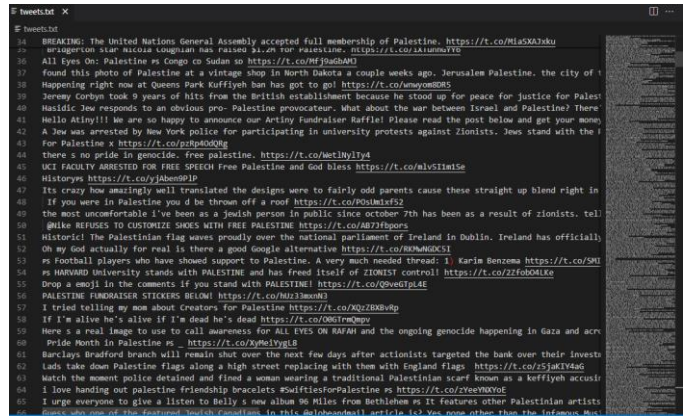
beragam atribut mulai dari tanggal, isi, pengirim, dan tautan terkait cuitan.



Gambar 3.1 Data CSV Hasil Tweet Harvest

Agar proses pemberihan data menjadi lebih cepat, data CSV diubah menjadi format TXT dan hanya diambil isi cuitannya saja melalui eksekusi program sebagai berikut

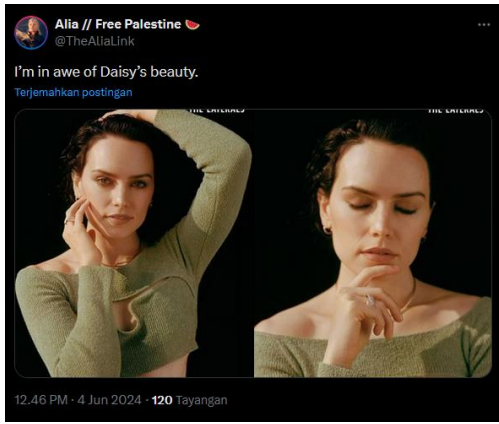
```
file_path = f"rafah.csv"
df = pd.read_csv(file_path,
delimiter=",")
for row in df["full_text"]:
    with open(f'tweets.txt', 'a',
encoding="utf-8") as out_file:
        out_file.write(row)
        out_file.write("\n")
```



Gambar 3.2 Data TXT Hasil Konversi

B. Pembersihan Data

Pada Gambar 3.2 dapat dilihat bahwa teks cuitan masih berisi “pengotor” seperti emoji, tautan, dan tag yang tidak terlalu penting untuk diikuti ke dalam proses analisis data. Selain itu, juga ditemukan kasus dimana cuitan berhasil masuk ke dalam data padahal tidak mengandung kata kunci “Palestina”. Hal ini terjadi sebab *tweet-harvest* juga mengikutkan cuitan dari pengguna dengan username yang mengandung kata kunci meskipun isi cuitannya tidak relevan seperti pada gambar di bawah ini sehingga perlu dihapus.



Gambar 3.3 Data Cuitan yang Tidak Relevan

Proses pembersihan data dilakukan dengan bantuan library *regular expression* python untuk menghapus cuitan yang tidak relevan, serta menghapus kata-kata pengotor pada cuitan yang sudah relevan. Berikut ini program yang dijalankan

```
def clean_text(text):
    # Menghapus tag dan hashtag
    text = re.sub('([@#][A-Za-z0-9_]+)', '', text)
    # Menghapus link
    text = re.sub(r'https\S+', '', text)
    # Menghapus emoji
    text = re.sub(r'[^\w\s]', '', text)
    return text

# Membuka file input
with open('tweets.txt', 'r', encoding='utf-8') as txt_file,
    open('clean.txt', 'w', encoding='utf-8') as clean_file:
    # Membaca setiap baris dari file
    for line in txt_file:
        # Hanya memasukkan cuitan yang relevan
        if (re.search('[Pp][Aa][Ll][Ee][Ss][Tt][Ii][Nn]', line)):
            cleaned_text = clean_text(line)
            clean_file.write(cleaned_text)
```

Berikut ini perbandingan cuitan sebelum dan sesudah dibersihkan

TABEL 3.1 Perbandingan Data Bersih dan Kotor

Sebelum	Sesudah
<p>🇵🇸 Zionist occupying forces abducted Palestinians mostly children during invasions in the various cities in #WestBank. Among the abductees; 1. The child Aboud Hassanein Qalqilya 2. The child Hadi Jadou' Qalqilya (1/ https://t.co/iu9ES6DfTr</p>	<p>Zionist occupying forces abducted Palestinians mostly children during invasions in the various cities in Among the abductees 1 The child Aboud Hassanein Qalqilya 2 The child Hadi Jadou Qalqilya 1</p>
<p>All out in Manchester this Saturday 8th June! Show your solidarity with Palestinians on the Naksa commemoration as they call for a Palestine free from Zionist brutality! pSPS Meet 12pm St Peter s Square #StopGazaGenocide #FreePalestine #StopArmingIsrael https://t.co/jPvtRjZbMC</p>	<p>All out in Manchester this Saturday 8th June Show your solidarity with Palestinians on the Naksa commemoration as they call for a Palestine free from Zionist brutality Meet 12pm St Peter s Square</p>
<p>Pack the courts for our #CUNY22 Tell @CUNY to meet the 5 Demands! Drop the charges for the CUNY 22! Free Palestine! June 12th 17th 18th 🕒 8:30 AM https://t.co/mu47B4VG9D</p>	<p>Pack the courts for our Tell to meet the 5 Demands Drop the charges for the CUNY 22 Free Palestine June 12th 17th 18th 830 AM</p>
<p>@Rayan_Assalam @slerudulier The Lab de Yotam Feldman un laboratoire de guerre à ciel ouvert The Palestine Laboratory. https://t.co/tPIDfRnpuu How Israel Exports the Technology of Occupation around the Worldde Anthony Loewenstein https://t.co/z9jjQKFDEQ And with Weapons from United States Complicity https://t.co/MqGDWWOo03</p>	<p>The Lab de Yotam Feldman un laboratoire de guerre à ciel ouvert The Palestine Laboratory How Israel Exports the Technology of Occupation around the Worldde Anthony Loewenstein And with Weapons from United States Complicity</p>
<p>@stairwayto3dom I get it! It's a reference to the love for sub-saharian slave trade Palestine had. Did I get it right? https://t.co/GfnO24LnPA</p>	<p>I get it Its a reference to the love for subsaharian slave trade Palestine had Did I get it right</p>

Dapat diamati bahwa tag, emoji, dan tautan berhasil dihilangkan pada data cuitan. Di samping itu, program juga berhasil menghapus beberapa cuitan yang tidak relevan

sehingga data berkurang dari yang semula 741 menjadi 499.

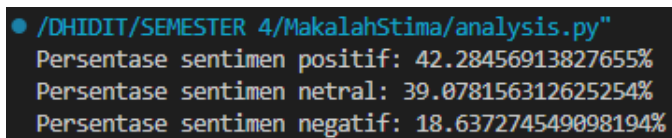
C. Analisis Sentimen

Penulis melakukan perhitungan sentimen pada data dengan menggunakan *library* Textblob, Python. *Library* ini memiliki fitur untuk menentukan tingkat polaritas sentimen dari suatu teks. Polaritas bernilai lebih dari nol bermakna teks tersebut bersentimen positif. Polaritas bernilai sama dengan nol berarti teks tersebut bersentimen netral. Sedangkan, polaritas bernilai kurang dari nol berarti teks tersebut bersentimen negatif.

Setiap cuitan pada data akan dikategorikan berdasarkan tiga tipe sentimen tersebut. Kemudian, akan dihitung persentase masing-masing kategori terhadap total keseluruhan. Berikut ini adalah program yang dijalankan:

```
from textblob import TextBlob
clean = open("clean.txt", "r",
encoding='utf-8')
positif = 0
negatif = 0
netral = 0
for line in clean:
    blob = TextBlob(line)
    polarity =
blob.sentiment.polarity
    if polarity < 0:
        negatif+=1
    elif polarity == 0:
        netral+=1
    else:
        positif+=1
print("Persentase sentimen positif:
" + str(positif/499*100) + "%")
print("Persentase sentimen netral: "
+ str(netral/499*100) + "%")
print("Persentase sentimen negatif:
" + str(negatif/499*100) + "%")
```

Berikut ini hasil dari eksekusi program



```
● /DHIDIT/SEMESTER 4/MakalahStima/analysis.py"
Persentase sentimen positif: 42.28456913827655%
Persentase sentimen netral: 39.078156312625254%
Persentase sentimen negatif: 18.637274549098194%
```

Gambar 4.1. Hasil Perhitungan Sentimen

Dari 499 data yang dianalisis, sebanyak 42% memiliki sentimen positif terhadap Palestina setelah terjadi peristiwa tragedi Rafah. Sebanyak 39% tidak menunjukkan kecondongan terhadap sisi manapun. Di samping itu, terdapat sekitar 19% yang memiliki sentimen negatif terhadap Palestina.

V. ANALISIS DAN PEMBAHASAN

Sentimen positif terhadap Palestina pasca peristiwa tragedi Rafah memang mendominasi dan menjadi mayoritas dibandingkan dengan sentimen netral dan negatif. Meskipun begitu, jumlah sentimen netral sebanyak 39% dan sentimen negatif sekitar 19% tidak dapat diabaikan begitu saja. Opini netral publik memiliki potensi bahaya yang tinggi apabila berhasil dipengaruhi oleh suatu pihak menjadi negatif. Kampanye untuk menyuarkan pembebasan Palestina harus terus digencarkan dengan metode yang lebih inovatif dan massa yang lebih banyak.

Data yang digunakan pada makalah ini kurang akurat sebab jumlah data apabila dibandingkan dengan populasi dunia sangatlah timpang. Di samping itu, pemanfaatan *library* dalam menghitung sentimen belum tentu berhasil melakukan perhitungan sentimen dengan sangat akurat sehingga presentase di atas memiliki galat dengan jumlah tertentu.

VI. KESIMPULAN DAN SARAN

Sebanyak 42% cuitan di Twitter memiliki sentimen positif terhadap Palestina setelah peristiwa tragedi Rafah. Sedangkan, 39% cuitan tidak menunjukkan kecondongan terhadap sisi manapun. Di samping itu, terdapat sekitar 19% yang memiliki sentimen negatif terhadap Palestina. Berdasarkan angka ini, kampanye untuk menyuarkan kebebasan Palestina masih harus terus diperjuangkan agar jumlah sentimen netral dapat diubah menjadi positif, serta menekan jumlah sentimen negatif.

VII. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada:

- (1) Allah SWT atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan makalah ini dengan baik
- (2) Kedua orang tua penulis yang senantiasa mendukung dan mendoakan penulis
- (3) Dr. Rinaldi Munir, Dr. Nur Ulfa Maulidevi, Dr. Rila Mandala, dan Bapak Monterico Adrian selaku dosen mata kuliah IF2211 Strategi Algoritma yang telah membimbing penulis dan memberikan banyak pengetahuan yang bermanfaat selama perkuliahan

VIII. REFERENCES

- [1] Y. B. Priambada, "Serangan Rafah, Simbol Lumpuhnya Dunia Menghentikan Israel," Kompas, 29 Mei 2024. [Online]. Available: <https://www.kompas.id/baca/riset/2024/05/29/serangan-rafah-simbol-lumpuhnya-dunia-menghentikan-israel>. [Diakses 11 Juni 2024].

- [2] B. Dean, "X (Twitter) Statistics: How Many People Use X?," Backlinko, 10 April 2024. [Online]. Available: <https://backlinko.com/twitter-users>. [Diakses 11 Juni 2024].
- [3] Y. Kurniawati, "Analisis Sentimen dan Jenisnya," BINUS University, 24 November 2023. [Online]. Available: <https://sis.binus.ac.id/2023/11/24/analisis-sentimen-dan-jenisnya/>. [Accessed 11 Juni 2024].
- [4] AWS, "Apa yang dimaksud dengan Analisis Sentimen?," AWS Amazon, 11 Juni 2024. [Online]. Available: <https://aws.amazon.com/id/what-is/sentiment-analysis/>. [Accessed 12 Juni 2024].
- [5] C. Charras and T. Lecroq, "Handboox of Exact String Matching Algorithms," [Online]. Available: <http://www-igm.univ-mlv.fr/%7Eelecroq/string/string.pdf>. [Accessed 12 Juni 2024].
- [6] R. Munir, "String Matching dengan Reegular Expression," 2024. [Online]. Available: <https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2022-2023/String-Matching-dengan-Regex-2019.pdf>. [Accessed 12 Juni 2024].
- [7] NTU, "Regular Expressions," NTU, Novembern 2018. [Online]. Available: <https://www3.ntu.edu.sg/home/ehchua/programming/howto/Regexe.html>. [Accessed 12 Juni 2024].

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 12 Juni 2024



Dhidit Abdi Aziz 13522040