

Analisis Sentimen pada Cuitan tentang Performa Erling Haaland di Musim Ini dengan Pendekatan Regular Expression

Shidqi Indy Izhari - 13521097
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jalan Ganesha 10 Bandung
E-mail: 13521097@std.stei.itb.ac.id

Abstract—Twitter merupakan salah satu sosial media yang cukup banyak digunakan khalayak umum saat ini untuk berbagai hal mulai dari menceritakan kegiatan sehari-hari, membagikan suatu insight baru, hingga melontarkan opini tentang suatu hal tertentu. Makalah ini akan ditujukan kepada sentimen opini warga internasional terhadap Erling Haaland, salah satu *rising star* dalam sepak bola dunia.

Keywords—*component; Erling Haaland, Twitter, analisis sentimen, regular expression*

I. PENDAHULUAN

Twitter adalah platform media sosial yang populer di mana pengguna dapat mengirim dan membaca pesan singkat yang disebut "tweet" atau dalam Bahasa Indonesia, cuitan. Dalam batasan 280 karakter, pengguna dapat berbagi pemikiran, informasi, berita, dan opini tentang berbagai topik. Twitter memungkinkan pengguna untuk mengikuti akun lain dan berinteraksi dengan mereka melalui mencuit, membalas, menyukai, dan menyebarkan cuitan. Platform ini digunakan oleh individu, selebriti, organisasi, dan media sebagai saluran komunikasi langsung dengan audiens mereka. Dengan sifatnya yang real-time, Twitter menjadi sumber berita aktual dan pemantauan tren global yang cepat serta tempat diskusi publik yang luas.

Erling Haaland merupakan seorang pemain sepak bola profesional asal Norwegia yang dikenal sebagai salah satu penyerang muda paling menjanjikan di dunia yang kini bermain untuk klub bola Manchester City. Haaland memiliki postur fisik yang kuat dan kokoh, kecepatan yang luar biasa, dan naluri mencetak gol yang tajam. Ia telah mencatatkan performa yang mengesankan di klub Borussia Dortmund dan timnas Norwegia. Haaland sering kali menjadi sorotan media dan penggemar sepak bola karena kontribusinya yang luar biasa dalam mencetak gol dan kemampuan menghadapi pertahanan lawan.

Kontribusi gol Haaland pada musim ini bersama Manchester City telah memecahkan tiga buah rekor sekaligus. Hal ini merupakan suatu hal yang tidak biasa karena dia berhasil mencapainya dalam musim debutnya bersama Manchester City. Salah satu rekor yang dicapai adalah ia merupakan pencetak gol terbanyak Premier League dalam satu

musim. Kesuksesannya yang mengagumkan inilah yang memicu banyaknya opini dari berbagai kalangan penikmat olahraga sepak bola.



Gambar 1.1 Potret Erling Haaland

Sumber: https://www.ksat.com/topic/Erling_Haland/

II. LANDASAN TEORI

A. Analisis Sentimen

Analisis sentimen adalah proses mengumpulkan, memahami, dan mengevaluasi opini, pandangan, atau emosi yang terkandung dalam sebuah teks untuk memperoleh pemahaman tentang sentimen umum terkait suatu topik yang sedang dibahas. Metode analisis sentimen digunakan untuk menganalisis besar jumlah data teks, seperti cuitan di media sosial, ulasan produk, atau komentar pengguna, dengan tujuan mengidentifikasi dan mengukur sentimen positif, negatif, atau netral yang terkait dengan topik yang sedang diteliti.

Dalam analisis sentimen, terdapat beberapa teknik dan algoritma yang digunakan. Salah satu metode yang umum adalah penggunaan algoritma berbasis mesin pembelajaran, di mana data latih yang telah dilabeli dengan sentimen positif, negatif, atau netral digunakan untuk melatih model dan mengklasifikasikan teks yang belum diketahui sentimennya. Selain itu, metode berbasis aturan seperti penggunaan regular

expression juga digunakan untuk mengidentifikasi sentimen dalam teks.

Hasil dari analisis sentimen dapat memberikan wawasan yang berharga dalam memahami persepsi dan respon pengguna terhadap suatu topik. Informasi ini dapat digunakan untuk mengukur kepuasan pengguna, mengidentifikasi tren atau isu yang sedang viral, memonitor citra merek atau produk, atau mengarahkan pengambilan keputusan bisnis. Dengan analisis sentimen, kita dapat menggali pandangan dan opini yang tersembunyi dalam teks yang luas, sehingga memberikan pemahaman yang lebih baik tentang bagaimana orang bereaksi terhadap suatu topik atau peristiwa.

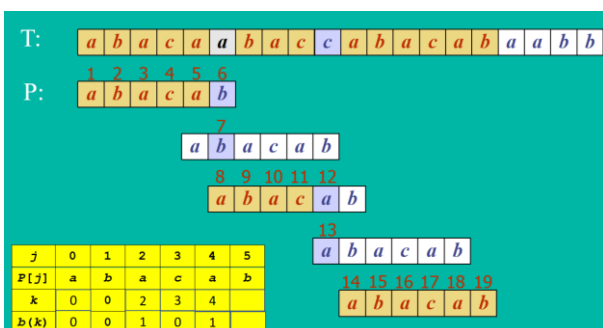
B. String Matching

String merupakan sekumpulan karakter yang disusun dalam sebuah array. Sesuai dengan sifat array, sebuah string juga memiliki index untuk setiap karakter yang dimilikinya. Indeks ini dimulai dari 0 hingga n-1, dengan n adalah Panjang dari string tersebut

String matching adalah proses mencocokkan sebuah pola atau substring tertentu dalam sebuah teks atau string yang lebih besar. Tujuan dari string matching adalah untuk menemukan atau mengidentifikasi keberadaan pola yang dicari dalam string yang sedang dianalisis. Metode string matching banyak digunakan dalam pengolahan teks, pengenalan pola, dan analisis data untuk menemukan kesamaan atau korelasi antara string atau pola yang dicari dengan teks yang ada.

Salah satu metode dari dua yang umum digunakan dalam string matching adalah metode pencocokan eksak atau exact matching, di mana pola atau substring yang dicari harus cocok secara persis dengan string dalam teks. Pencocokan eksak dapat dilakukan dengan menggunakan algoritma sederhana seperti metode brute-force atau menggunakan algoritma yang lebih efisien seperti algoritma Knuth-Morris-Pratt (KMP) atau algoritma Boyer-Moore.

Metode yang digunakan selanjutnya adalah Regular Expression Matching, atau yang kerap dikenal sebagai Regex Matching. Metode ini menggunakan notasi atau aturan tertentu yang akan dijelaskan pada subbab selanjutnya



Gambar 2.1 Contoh String Matching dengan Algoritma KMP

Sumber:

<https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2020-2021/Pencocokan-string-2021.pdf>

C. Regular Expression

Seperti yang telah disebutkan sebelumnya, Regular Expression memiliki notasi tertentu agar ia dapat berjalan. Berikut adalah penjelasan lebih lanjut tentang *pattern* yang dimaksud dalam bentuk tabel:

Construct	Definisi
[abcd]	a, b, c, atau d
[^abcd]	Selain a, b, c, dan d
[a-d]	a hingga d (a, b, c, d)
.	Wildcard, seluruh karakter kecuali \n
\w	karakter word (a, b, c, dsb)
\W	selain word (!, -, >, dsb)
\s	spasi ()
\S	selain spasi (f, !, <, dsb)
\d	digit decimal (1, 3, 10, dsb)
\D	selain digit decimail (A, !, >, dsb)

Tabel 2.1 Character Classes

Construct	Definisi
^	berada pada awal string atau baris
\$	berada pada akhir string atau sebelum \n
\A	berada pada awal string
\Z	berada pada akhir string
\b	ada di antara batasan \w dan \W

Tabel 2.2 Anchors

Construct	Definisi
\a	karakter bell
\b	backspace
\t	tab
\n	new line/enter
\e	escape
\v	vertical tab

Tabel 2.3 Character Escapes

Construct	Definisi
X*	karakter sebelumnya (X) harus berjumlah >= 0
X+	karakter sebelumnya (X) harus berjumlah >= 1
X?	karakter sebelumnya (X) harus berjumlah 0 atau 1
X{n}	karakter sebelumnya (X) harus berjumlah sebanyak n
X{n,}	karakter sebelumnya (X) harus berjumlah sebanyak >= n
X{n, m}	karakter sebelumnya (X) harus berjumlah sebanyak >= n tetapi < m
X{n}?	karakter sebelumnya (X) berjumlah tepat n
X{n,}?	karakter sebelumnya (X) berjumlah >= n, dengan n seminimal mungkin
X{n, m}?	karakter sebelumnya (X) harus berjumlah sebanyak >= n tetapi < m, dengan n seminimal mungkin

Tabel 2.4 Quantifiers

Construct	Definisi
(?=X)	memastikan apakah string selanjutnya adalah X
(?<=X)	memastikan apakah string sebelumnya adalah X
(?!X)	memastikan apakah string selanjutnya bukan X
(?<!X)	memastikan apakah string sebelumnya bukan X

Tabel 2.5 Lookarounds

III. IMPLEMENTASI DAN PEMBAHASAN

Penelitian ini dilakukan melalui beberapa tahapan yang terstruktur, yang meliputi pengambilan data, pengekraksian data, pembersihan data, dan yang terakhir pembacaan data berdasarkan sentimen. Setiap tahap memiliki peran penting dalam memastikan keberhasilan dan keakuratan penelitian. Berikut adalah penjelasan dari setiap bagian:

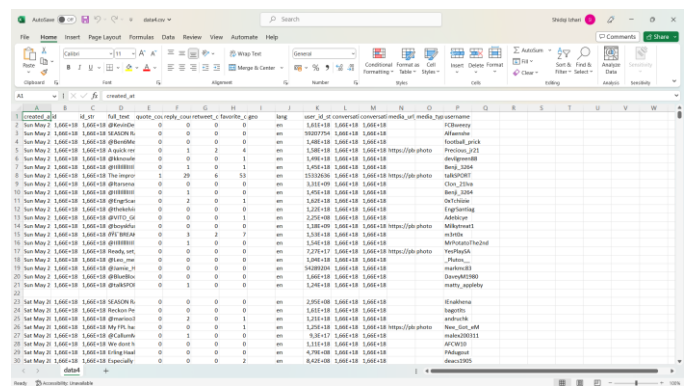
A. Pengambilan Data

Data di sini merupakan kumpulan cuitan dari berbagai akun di Twitter dalam Bahasa Inggris. Untuk mengambil cuitan tersebut, penulis menggunakan salah satu package dari npm,

yaitu tweet-harvest. Tweet-harvest adalah alat yang berguna untuk mengumpulkan dan mengekraks data dari Twitter menggunakan API Twitter. Untuk menggunakan package ini, perlu dipastikan bahwa Node sudah terinstall terlebih dahulu. Selanjutnya, penulis memasukkan *command* pada terminal seperti di bawah ini:

```
npx tweet-harvest@0.0.35 -s "haaland (erling OR player OR good OR bad OR overrated OR robot OR flop OR lucky OR key OR kdb OR de bruyne OR penalty OR tap in OR finishing OR pep) lang:en" -l 1500
```

-s merupakan query yang dimasukkan sesuai dengan notasi *searching* dalam Twitter, -l merupakan banyaknya cuitan yang akan dimasukkan ke dalam file .csv, yaitu pada kasus ini berjumlah kurang lebih 1500. Setelah kurang lebih 5 menit, program akan selesai berjalan dan mengeluarkan output berupa file csv yang berisi segala atribut dari data, mulai dari username, isi tweet, link terkait, hingga waktu cuitan.



Gambar 3.1 Hasil CSV

B. Ekstraksi Data

Untuk memudahkan dan mempercepat pembacaan data, file CSV akan diubah kedalam bentuk txt yang hanya berisi cuitan (atribut full_text). Berikut adalah kode pengekraksian data tersebut:

```
import csv

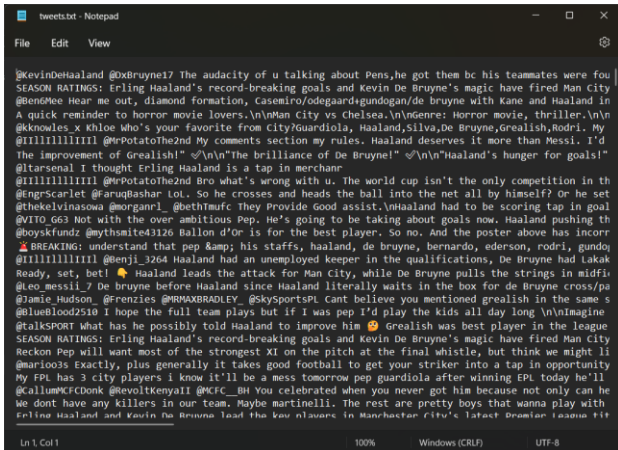
input_file = 'data4.csv'
output_file = 'tweets.txt'

# Membuka file input CSV
with open(input_file, 'r', encoding='utf-8') as csv_file, open(output_file, 'w', encoding='utf-8') as txt_file:
    reader = csv.DictReader(csv_file, delimiter=';')

    # Menulis full_text ke dalam file teks
    for row in reader:
        full_text = row['full text']
```

```
txt_file.write(full_text + '\n')
```

Kode dituliskan dalam Bahasa pemrograman Python dengan memanfaatkan *library* csv. Output file juga sudah terbentuk sebelumnya sehingga kode ini hanya membuka file dan menuliskan ke dalamnya.



Gambar 3.2 Isi dari file Tweets.txt

C. Pembersihan Data

Jika dilihat sekilas dari data pada file Tweets.txt yang telah dihasilkan, terdapat beberapa masalah seperti masih terdapatnya tag @ dan emoji yang akan mengganggu pada saat analisis data dilakukan. Maka dari itu, pembersihan data dilakukan agar hal-hal tidak penting tadi dapat dihilangkan. Pembersihan dilakukan dengan Bahasa pemrograman Python dengan bantuan Regular Expression yang sudah dalam bentuk *library*. Berikut adalah kodenya:

```
import re

input_file = 'tweets.txt'
output_file = 'cleaned_tweets.txt'

def clean_text(text):
    # Menghapus tag dan hashtag
    text = re.sub('([@#][A-Za-z0-9_]+)', '', text)

    # Menghapus link
    text = re.sub(r'http\S+|www\S+|https\S+', '', text)

    # Menghapus emoji
    text = re.sub(r'[^\w\s]', '', text)

    # Menghapus enter (\n)
    text = re.sub(r'\n', '', text)

    return text

# Membuka file input
```

```
with open(input_file, 'r',
encoding='utf-8') as txt_file,
open(output_file, 'w', encoding='utf-8')
as clean_file:
```

```
    # Membaca setiap baris dari file
    teks
    for line in txt_file:
        # Membersihkan teks dari emote,
        enter (\n), dan link
        cleaned_text = clean_text(line)

        # Menulis teks yang telah
        dibersihkan ke dalam file
        clean_tweets.txt
        clean_file.write(cleaned_text +
'\n')
```

Sebagai perbandingan, ini adalah perbandingan dari beberapa baris data sebelum dibersihkan dan setelah dibersihkan:

Sebelum dibersihkan
@KevinDeHaaland @DxBruyne17 The audacity of u talking about Pens,he got them bc his teammates were fouled,i can say take tap ins from Haaland he's not even top 5 in the world,he's not even the best player in City,KDB is,the same way u delete Messi WC i can ignore Haaland's potential UCL
@VITO_G63 Not with the over ambitious Pep. He's going to be taking about goals now. Haaland pushing the record up and De Bruyne with the assist
Ready, set, bet! 🏆 Haaland leads the attack for Man City, while De Bruyne pulls the strings in midfield. Guided by coach Pep Guardiola, City aims to triumph over Chelsea. Watch out for Sterling's forward play & Fernandez's midfield skills for Chelsea, guided by coach Lampard. https://t.co/NDjqfppjL
@Jamie_Hudson_ @Frenzies @MRMAXBRADLEY_ @SkySportsPL Cant believe you mentioned grealish in the same sentence as haaland, dias and de bruyne.
@CFC_Xed @ESPNFC Because Messi is Messi. Haaland is great but without good service he cant do much. No hate but for me De Bruyne is the best player in Man City

Setelah dibersihkan
The audacity of u talking about Penshe got them bc his teammates were fouledi can say take tap ins from Haaland hes not even top 5 in the worldhes not even the best player in CityKDB isthe same way u delete Messi WC i can ignore Haalands potential UCL
Not with the over ambitious Pep Hes going to be taking

about goals now Haaland pushing the record up and De Bruyne with the assist
Ready set bet Haaland leads the attack for Man City while De Bruyne pulls the strings in midfield Guided by coach Pep Guardiola City aims to triumph over Chelsea Watch out for Sterlings forward play amp Fernandezs midfield skills for Chelsea guided by coach Lampard
Cant believe you mentioned grealish in the same sentence as haaland dias and de bruyne
Because Messi is Messi Haaland is great but without good service he cant do much No hate but for me De Bruyne is the best player in Man City

```
print("Persentase sentimen positif: " +
str(countpos/(countpos+countneg+countnet)
*100) + "%")
print("Persentase sentimen negatif: " +
str(countneg/(countpos+countneg+countnet)
*100) + "%")
print("Persentase sentimen netral: " +
str(countnet/(countpos+countneg+countnet)
*100) + "%")
```

Setelah kode di atas dijalankan, barulah persentase sentimen dari seluruh data yang berjumlah kurang lebih 1500 akan terlihat. Berikut adalah hasilnya:

```
[Running] python -u "c:\Users\Shidq\OneDrive\Document
Persentase sentimen positif: 57.29303547963206%
Persentase sentimen negatif: 17.41130091984231%
Persentase sentimen netral: 25.295663600525625%

[Done] exited with code=0 in 1.577 seconds
```

Gambar 4.1 Hasil persentase sentimen

Dari 1522 data yang dianalisis, ditemukan bahwa sebanyak 57.3% cuitan menunjukkan sentimen positif terhadap performa Erling Haaland pada musim ini. Sementara itu, 17.41% cuitan menunjukkan ketidaksetujuan, dan 25.3% cuitan menunjukkan sikap netral atau mungkin tidak terkait dengan perdebatan yang ada. Dari 1522 data, didapatkan bahwasannya terdapat 57.3% cuitan yang memberikan sentimen positif terhadap performa Erling Haaland pada musim ini, 17.41% yang kurang setuju, dan 25.3% yang merasa tidak keduanya, atau mungkin cuitannya tidak bersifat argumentative.

Dominannya nilai sentimen positif menunjukkan bahwasannya banyak orang yang setuju akan betapa baiknya performa Haaland pada musim ini bersama Manchester City.

IV. KESIMPULAN DAN SARAN

Analisis sentimen dapat dilakukan menggunakan bantuan regular expression, salah satu materi yang diajarkan pada Mata Kuliah IF2211. Hal ini tentunya juga dibantu oleh beberapa Bahasa pemrograman beserta *library-library*nya.

Untuk saran kedepannya, penulis menyarankan untuk memperbanyak data yang akan diproses agar hasil persentase sentimen akan menjadi lebih akurat. Selain itu, pembersihan data atau pemfilteran yang lebih teliti dan presisi juga akan membantu memudahkan analisis sentimen.

Dalam praktiknya, ada banyak teknik dan pendekatan yang dapat digunakan dalam analisis sentimen, termasuk penggunaan algoritma machine learning dan pendekatan berbasis aturan. Setiap pendekatan memiliki kelebihan dan kelemahan tersendiri, dan pemilihan metode tergantung pada kebutuhan dan konteks spesifik yang dihadapi.

Terlihat bahwa terdapat perbedaan signifikan pada kedua text yang menghilangkan emoji, link, hingga tag untuk seseorang.

D. Analisis Sentimen Data

Bagian selanjutnya, atau bagian terakhir dari program ini adalah melakukan pembacaan sentimen pada data dengan memanfaatkan *library* Python, `textblob`. `TextBlob` memiliki fungsi yang dapat menentukan tingkat polaritas sentimen dari suatu string. Apabila polaritas bernilai > 0 , artinya string tersebut termasuk kepada sentimen positif, polaritas bernilai < 0 berarti sentimen *negative*, dan polaritas $= 0$ berarti sentimen bersifat netral. Berikut adalah pengimplementasian kodenya:

```
from textblob import TextBlob

# buka file teks yang berisi tweet
with open("cleaned_tweets.txt", "r",
encoding="utf-8") as f:
    tweets = f.readlines()

# lakukan analisis sentimen terhadap
setiap tweet
# deklarasi untuk perhitungan setiap
jenis sentimen
countpos = 0
countneg = 0
countnet = 0
for tweet in tweets:
    blob = TextBlob(tweet)
    sentiment = blob.sentiment.polarity

    if sentiment > 0:
        countpos+=1
    elif sentiment < 0:
        countneg+=1
    else:
        countnet+=1
```

UCAPAN TERIMA KASIH

Penulis ingin mengucapkan rasa syukur kepada Tuhan yang Maha Esa karena atas anugerah-Nya, penulis dapat menyelesaikan makalah Mata Kuliah IF2211 yang berjudul “Analisis Sentimen pada Cuitan tentang Performa Erling Haaland di Musim Ini dengan Pendekatan Regular Expression”. Tak lupa juga, penulis pun mengucapkan terima kasih kepada orang tua yang telah membantu penulis secara tenaga dan material. Dan yang terakhir, penulis juga ingin berterima kasih kepada Dr. Ir. Rinaldi Munir, M.T., selaku dosen pengajar Mata Kuliah IF2211, beserta dosen pengampu lainnya yang telah membimbing kami selama pembelajaran.

REFERENCES

- [1] <https://learn.microsoft.com/en-us/dotnet/standard/base-types/regular-expression-language-quick-reference>, diakses pada 21 Mei 2023
- [2] <https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2018-2019/String-Matching-dengan-Regex-2019.pdf>, diakses pada 21 Mei 2023
- [3] <https://infografis.sindonews.com/photo/22485/tiga-rekor-yang-bisa-dipecahkan-erling-haaland-musim-ini-1681329211>, diakses pada 21 Mei 2023
- [4] <https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2020-2021/Pencocokan-string-2021.pdf>, diakses pada 21 Mei 2023

- [5] <https://medium.com/@albertusheronius/twitter-sentiment-analysis-bahasa-indonesia-dengan-textblob-f34e1ffdcdaa>, diakses pada 21 Mei 2023
- [6] <https://docs.python.org/3/library/re.html>, diakses pada 21 Mei 2023

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, 22 Mei 2023



Shidqi Indy Izhari 13521097