Analisis Sentimen Netizen Twitter terhadap Aktor Johnny Depp

Ilham Bintang Nurmansyah - 13520102
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jalan Ganesha 10 Bandung
E-mail (gmail): ilhambintangn@gmail.com

***Abstract*— Analisis Sentimen adalah proses penggunaan *text analytics* untuk mendapatkan berbegai sumber data dari internet maupun media sosial. Tujuannya untuk memperoleh opini para pengguna media sosial yang menggunakan platform tersebut. Analisis Sentimen merupakan salah satu bidang dari NLP (*Natural Language Processing*) yang memiliki sistem untuk mengenali dan mengekstrak opini dalam bentuk teks. Dalam paper ini dibahas mengenai analisis sentimen dari tweet para pengguna twitter terhadap Johnny Depp yang dewasa ini sedang viral karena kasus dengan istrnya Amber Heard.**

***Kata kunci*: analisis sentimen, Johny Depp, twitter, tweet**

I. PENDAHULUAN

Twitter adalah media sosial berbentuk mikroblog daring yang memungkinkan seorang penggunanya untuk mengirim dan membaca pesan berisi teks sepanjang 280 karakter, gambar, dan video yang dikenal dengan sebutan tweet. Twitter sudah ada sejak tahun 2006, namun sempat redup hingga akhirnya ramai digunakan kembali ditahun 2020. Twitter terkenal karena kebebasannya dalam beropini tanpa batas. Apalagi baru-baru ini, Twitter dibeli oleh Elon Musk dengan tujuan membuat twitter menjadi media sosial yang sangat bebas untuk berpendapat apapun.

John Christopher Depp II atau yang biasanya dipanggil Johnny Depp adalah seorang aktor, produser, dan musisi yang berasal dari Amerika Serikat. Ia lahir pada tahun 1963 dan sekarang memiliki 2 orang anak. Aktor berusia 50 tahun ini populer karena perannya sebagai Jack Sparrow di film "Pirates of Caribean". Johnny mendapatkan trofi pertamanya pada ajang SAG Awards untuk kategori penampilan terbaik pemeran pria tahun 2003. Prestasi terbearnya adalah ketika memenangkan Aktor Terbaik kategori Motion Picture Musical or Comey di ajang Golden Globe Awards 2008.

Pada tahun 2016, Amber Heard, istri Johnny saat itu, menggugat cerai dirinya. Hal ini dikarenakan istrinya menuduh Johnny melakukan kekerasan dalam rumah tangga. Kasus ini membuat Johhny ditahan untuk sementara waktu. Johnny juga dituntut sebesar 7 juta US dollar oleh pengadilan. Mereka berdua akhirnya bercerai pada tahun 2017.

Fig. 1. Johnny Depp & Amber Heard

Pada tahun 2019, Depp menuntut Heard sebesar 50 juta US dolar atas pencemaran nama baik pada artikel WaPo. Lalu pada 2020 lalu, Sebuah rekaman telepon mengenai Heard memukuli Depp terungkap. Sejak saat itu, orang orang mulai kembali memercayai Depp. Kasus ini masih berlanjut dan menjadi bahan pembicaraan hangat hingga sekarang.

II. LANDASAN TEORI

A. Analisis Sentimen

Analisis sentimen, yang juga dirujuk sebagai *opinion mining* adalah sebuah pendekatan dari NLP (*Natural Language Processing)* yang mengidentifikasi nada emosional dibalik sebuah tulisan. Hal ini cukup populer digunakan ketika sebuah perusahaan atau organisasi ingin mengetahui pendapat dari konsumennya mengenai sebuah produk atau pelayanan yang diberikan.

Sentimen sendiri dalam KBBI berarti pendapat atau pandangan yang diadasarkan pada perasan yang berlebih-lebihan terhadap sesuatu (bertentangan dengan pertimbangan pikiran)

Analisis sentimen dapat membantu organisasi mengumpulkan *insight* (pandangan) dari suatu text yang tidak terstruktur yang biasanya didapat dari sumber-sumber online seperti website, email, social media, tiket keluhan, komentar, dan lain lain. Algoritma menggantikan pemrosesan data secara manual dengan menerapkan metode berbasis *rule, automatic,*

atau *hybrid*. Sistem berbasis *rule* akan melakukan analisis sentimen berdasarkan aturan berbasis lekskon yang telah ditentukan sebelumnya, sementara sistem *automatic* akan mempelajari data dengan teknik *machine learning*, sedangkan sistem *hybrid* akang menggabungkan kedua pendekatan.

Selain bisa mengidentifikasi sentimen, *opinion mining* juga dapat mengekstaksi polaritas (jumlah positif atau negatif), subjek, dan pemegang opini dalam teks. Selanjutnya analisis sentimen dapat diterakan pada berbagai cakupan seperti tingkat dokumen, paragraf, kalimat, dan sub-kalimat.

Analisis sentimen memiliki 4 tipe yaitu *fine-grained*, *emotion detection, intent-based,* dan *aspect-based. Fine-grained sentiment* memberikan tingkat polaritas yang lebih tepat dengan memecahnya ke dalam beberapa kategori lebih lanjut yang biasanya sangat positif hingga sangat negatif. Hal ini dapat dianggap sebagai opini yang setara dengan *ratings* pada skala bintang 5. *Emotion detection* mengidentifikasi emosi tertentu yang tergolong pada positif dan negatif. Misalnya seperti kebahagiaan, frustasi, kemarahan, dan kesedihan. *Intent-based* mengenali tindakan di balik teks selain opini. Contohnya seperti komentar di media sosial yang mengungkapkan rasa frustasi. Analisis *aspect-based* mengumpulkan komponen spesifik yang disebutkan secara positif maupun negatif. Seperti misalnya seseorang yang memberikan ulasan terhadap suatu produk yang mengatakan bahwa masa pakai baterai terlalu singkat. Lalu sistem akan mengembalikan bahwa entimen negatif bukan tentang produk secara *overall*, tapi hanya tentang bagian baterainya saja.

B. Twitter

Twitter adalah sosial media berbentuk mikroblog daring yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbentuk teks dengan maksimal 280 karakter. Pengguna juga bisa membagikan foto dan video dengan mengungahnya kedalam tweet. Twitter sudah ada sejak tahun 2006 dan masih menjadi sosial media yang ramai digunakan hingga sekarang. Twitter terkenal karena kebesannya dalam menuliskan apapun. Baru-baru ini, Twitter dibeli oleh Elon Musk demi menjunjung tinggi *free speech* di dalam sosial media ini.

C. Regular Expression

Regular Expression (regex) adalah pola yang digunakan untuk melakukan pencocokkan string. Biasanya *Regular Expression* digunakan untuk algoritma *string matching* untuk mencari dan mengganti operasi pada string atau validasi input. Regex banyak digunakan diberbagai hal yang berhubungan dengan pengolahan teks atau string.

Syntax dari regex disusun dari huruf, angka, dan simbol yang masing memiki fungsi dan arti yang berbeda penggunaannya. Sayangnya tisak semua bahasa pemrogramman menggunakan regex yang sama, tapi semua regex yang ada saling mirip satu sama lain.

Syntax *character classes* adalah syntax regex yang cocok dengan salah satu dari serangkaian karakter. Contohnya ada di tabel sebagai berikut

TABLE I. CHARACTER CLASS REGEX

Character	Description	Example	Matches
[some_string]	cocok dengan semua karakter single pada string	[ae]	“a” in “gray”
[^some_string]	cocok dengan semua karakter single yang tidak ada pada string	[^aei]	“r” in “aeri”
[first – last]	cocok dengan semua karakter single pada range first hingga last	[A-Z]	“A” in “A123”
.	cocok dengan karakter single apapun selain \n	a.e	“ave” in “averous”
\w	cocok dengan semua karakter kata	\w	“i”,”d”,”a” in “id.a”
\W	cocok dengan semua karakter non kata	\W	“ “,”.” In “ABC A.A”
\s	cocok dengan semua karakter *white-space*	\w\s	“D “ in “D ABC”

Selanjutnya adalah bagian *Anchors*. *Anchors* menyebabkan cocok atau tidak cocok tergantung pada posisi didalam string.

TABLE II. ANCHORS REGEX

Assertion	Description	Example	Matches
^	harus cocok dengan awal string dalam baris	^\d{3}	“901” in “901-333”
$	harus cocok dengan akhir string	-\d{3}$	“-333”in “-901-333”

Lalu ada *Quantifiers. Quntifiers* menentuan berapa banyak contoh elemen yang dapat berupa karakter, grup ,dan kelas karakter harus ada dalam string input agar kecocokan terjadi.

TABLE III. QUANTIFIER REGEX

Character	Description	Example	Matches
*	mencocokkan elemen sebelumnya nol kali atau lebih	\d.*\.\d	“.0”,”19.9”,”219.9”

+	mencocokkan elemen sebelumnya satu kali atau lebih	“be+”	“bee”in “been”
?	mencocokkan elemen sebelumnya nol atau satu kali	“rai?n”	“ran”,”rain”
{n}	mencocokkan elemen sebelumnya tepat n kali	“,\d{3}	“,043” in “1,043.6”
{n,}	mencocokkan elemen sebelumnya setidaknya n kali	“\d{2,}”	“166”
{n,m}	mencocokkan elemen sebelumnya dalam rentang n hingga m	\d{3,5}	“15123”

D. *String Matching*

String matching adalah algoritma yang bertugas untuk mencari kecocokkan diantara dua buah string. Biasanya string A yang akan dicocokkan pada string B akan memiliki ukuran yang lebih kecil.

Diasumsikan sebuah string S berukuran m, yang dalam hal ini S = $x_0x_1...x_{m-1}$, maka prefiks dari string S merupakan substring S[0...k] dan sufiks dari string S adalah substring S[k..m-1], dengan k adalah index diantara 0 dan m-1.

Misalkan sebuah string S = “andrew”. Maka semua kemungkinan prefiks S adalah “a” , “an”, “and”, “andr”,”andre”, “andrew” dab semua kemungkinan sufiksnya adalah “w”, ”ew”, ”rew”, “drew”, “ndrew”, “andrew".

Beberapa contoh algoritma yang biasa digunakan untuk melakukan *string matching* adalah algoritma *brute force*, *Knuth-Morris-Pratt(KMP) algorithm* , dan *Boyer-Moore algorithm.*Selain algoritma diatas, masih ada algoritma lainnya namun tidak diajarkan di perkuliahan ini.

III. Pembahasan

Untuk melakukan analisis sentimen, pertama kita harus mengambil data *tweet* dari netizen Twitter. Untuk mengambil *tweet* secara otomatis bisa menggunakan API dari Twitter, cuman bila menggunakan cara itu akan banyak limitasi yang ditemui, sehingga solusi alternatifnya adalah menggunakan library python yang bernama Snscrape. Snscrape dapat mengambil tweet dengan melakukan querry yang kita masukkan dalam code, kemudian mengcopy semua hasil querry yang ditemukan kedalam file .csv. Tweet yang bisa diambil menggunakan Snscrape adalah 3200 tweet terakhir dari setiap pengguna.

Untuk mulai mengambil data, kita memerlukan python3 versi terbaru yang bisa di unduh di website resminya. Setelah itu kita juga harus menginstall pip versi terbaru yang bisa di download di website resminya. Setelah itu kita bisa menginstall library Snscrape menggunakan command “pip install snscrape”. Ini adalah versi Snscrape yang paling stabil untuk saat ini.

Setelah tools diatas terinstall, maka kita akan mulai melakukan pengambilan data mengenai Johnny Depp. Tweet yang diambil akan dibatasi dalam berbahasa inggris saja, agar tidak perlu dibersihkan lagi ketika akan melakukan analisis sentimen, karena library yang digunakan hanya mendukung dalam bahasa inggris.

Tweet yang akan diambil adalah tweet berbahasa Inggris, dimulai pada tanggal 14-05-2022 dan diakhiri pada tanggal 21-05-2022 (saat makalah ini dibuat). Tujuan dari penggunaan batas waktu adalah agar tweet yang diambil tidak terlalu banyak, dan juga agar tweet yang didapat masih relevan dengan masalah yang sedang terjadi.

```
import snscrape.modules.twitter as sntwitter

import pandas as pd

import csv

# melakukan scraping data tweet dari twiter

# melakukan querry dengan keyword johnny
depp dari tanggal 21 may 2022

query = "johnny depp lang:en since:2022-05-
17 -filter:links -filter:replies"

f = open('tweets.csv', 'a',newline =
'',encoding='utf8')

writer = csv.writer(f)


  for tweet in
sntwitter.TwitterSearchScraper(query).get_it
ems():

      rawtweet = tweet.content

      parsed = rawtweet.replace('\n', ' ')

      writer.writerow([parsed])


f.close()
```

Setelah dilakukan *scraping* data, maka akan dilakukan pembersihan terhadap tweet yang sudah di *scraping*. Ada beberapa langkah dalam melakukan pembersihan data pada tweet ini menggunakan *regular expression*

1. Menghilangkan mention (@) dan hashtag(#)

   Mention dan hastag harus dihilangkan pada string karena mention berisi nama pengguna lain sedangkan

hashtag tidak diperlukan dalam melakukan analisis sentimen

2. Menghilangkan link atau URL pada tweet

   Sama seperti hashtag, URL tidak diperlukan untuk analisis sentimen, sehingga sebaiknya URL dihilangkan dari tweet.

3. Menghilangkan angka, tanda baca, dan emoji

   Angka dan tanda baca tidak diperlukan untuk melakukan analisis sentimen, sedangkan emoji dianggap sebagai noise dalam string sehingga harus dihilangkan.

4. Menghilangkan kata "Johnny Depp"

   Kata "Johnny Depp" tidak diperlukan dalam melakukan analisis sentimen, sehingga sebaiknya dihilangkan dari string.

5. Menghilangkan *stopwords*

   *Stopword*s adalah sekumpulan kata yang umum digunakan dalam berbagai bahasa, tidak hanya bahasa inggris. *Stopwords* baiknya dihilangkan agar kita bisa fokus pada kata kata yang penting saja.

6. Melakukkan *Stemming*

   Stemming adalah teknik NLP yang menurunkan infleksi dalam kata-kata ke bentuk dasarnya, sehingga membantu dalam *preprocessing* text.

```
# fungsi clean untuk membersihkan tweet menggunakan regex
def clean(tweet):
      # fungsi sub menggantikan string yang match dengan " "
      # menghilangkan @ dan #
      tweet = re.sub("([@#][A-Za-z0-9_]+)"," ",tweet)
      #  menghilangkan URL dan link
      tweet = re.sub(r'http\S+', '', tweet)
      # menghilangkan angka, tanda baca, dan emoji
      tweet = re.sub(r"\d+", "", tweet)
      tweet = re.sub(r"[^\w\s]", "", tweet)
      # menghilangkan kata "Johnny Depp" dan "Amber Heard"
      tweet = re.sub('[Jj]ohnny', "", tweet)
      tweet = re.sub('[Dd]epp', "", tweet)
      tweet = re.sub('[Aa]mber', "", tweet)
      tweet = re.sub('[Hh]eard', "", tweet)

      tweet = re.sub(r"\s+", " ", tweet)
    return tweet
```

```
  #fungsi clean_stopwords untuk membersihkan stopwords dari string
  def clean_stopwords(string):
      stop_words = set(stopwords.words('english'))
      word_tokens = word_tokenize(string)
      filtered_sentence = [w for w in word_tokens if not w.lower() in stop_words]
      sentence = ' '.join(filtered_sentence)
      return sentence

  # process untuk melakukan clean regex dan stopwords dan juga melakukan stemming
  def process(rawtweet):
      tweet = clean(rawtweet)
      tweet = tweet.lower()
      tweet = clean_stopwords(tweet)

      stemmer = PorterStemmer()
      stemmer.stem(tweet)
                return tweet
```

Tweet yang berhasil di dapatkan adalah sekitar 2974 tweet, dan ini adalah gambar yang berisi contoh *raw tweet* atau tweet yang belum di olah dan dibersihkan sama sekali.

	tweet
0	I don't give a F**** about Johnny Depp & C...
1	None of the news outlets that published the st...
2	How did we not notice how stupid Johnny Depp is?
3	taking back anything i said about johnny depp ...
4	my quotes look like: me: johnny depp is ugly ...
5	Tweeting this everyday until trial on Monday. ...
6	I can't believe that there are people who De L...
7	snl made so many jokes abt the addictions of s...
8	My words don't carry much weight BUUUUT Bungie...
9	Damn. Amber Heard and Johnny Depp need some da...
10	Amber Heard tried driving Johnny Depp to suici...
11	disney and johnny depp story uper video bnao @...
12	I have followed this Johnny Depp vs Amber Hear...
13	Please stop talking about Johnny depp and ambe...
14	ngl i'm tired of seeing Johnny depp and Amber ...
15	Amber and Johnny Depp, it didn't work out go y...
16	Regardless how well things seem to go for the ...
17	Johnny depp fans can go to hell for all I care
18	Things I never thought would happen. My dad re...
19	Johnny Depp really at court like how R Kelly w...
20	Sorry but half of the girls dick riding for Jo...
21	Let's talk about #AmberHeardForPrisonInmate if...
22	Woahh scott morrison just slapped Chris Rock a...
23	I don't feel like Johnny Depp is going to win ...
...	
2971	Robert Downey Jr career went from so low to su...
2972	We all respect you for your honesty. Best of l...
2973	things i don't give a fucking shit about: - a...
2974	And as far as real friends, tell all my film c...

Fig. 2. Data tweet sebelum di bersihkan

Setelah dilakukan *cleaning* pada tweet, bisa dilihat bahwa tweet sudah bersih dari kata "Johny Depp" ,"Amber Heard", sudah menghilangkan semua angka yang ada pada tweet, menghapus semua URL yang memiliki awalan "https". Tweet juga sudah dibersihkan dari *stopwords* dan sudah dilakukan *stemming*.

	tweet
0	dont give f amp co care racist murders black a...
1	none news outlets published statement support ...
2	notice stupid
3	taking back anything said may involved probabl...
4	quotes look like ugly support serial kivllers ...
5	tweeting everyday trial monday stupid ah brain...
6	cant believe people de lima leni supports yikes
7	snl made many jokes abt addictions many women ...
8	words dont carry much weight buuuut bungie nex...
9	damn need damn therapy
10	tried driving suicide almost succeeded took ev...
11	disney story uper video bnao
12	followed vs case much listened every witness t...
13	please stop talking thank
14	ngl im tired seeing internet
15	didnt work go separate ways already
16	regardless well things seem go lovely darling ...
17	fans go hell care
18	things never thought would happen dad really s...
19	really court like r kelly women camping get se...
20	sorry half girls dick riding werent even born ...
21	lets talk believe giving person attempt drunk ...
22	woahh scott morrison slapped chris rock took s...
23	dont feel like going win case however still
...	
2971	robert downey jr career went low suvh heights ...
2972	respect honesty best luck case
2973	things dont give fucking shit vs elon musk inf...
2974	far real friends tell film class classmates lo...

Fig. 3. Data tweet setelah dibersihkan

Langkah terakhir yang akan dilakukan adalah melakukan analisis sentimen itu sendiri dari tweet tweet yang sudah berhasil di proses. Tweet akan dilakukan analisis sentimen menggunakan library textblob dan akan dibagi menjadi tiga kategori yaitu tweet netral, positif, dan negatif.

```
# membaca file tweet yang sudah di proses

from textblob import TextBlob

file = pd.read_csv('processed_tweet.csv', encoding='utf8')

list_of_tweets = file['0'].toList()

tweet_positif = []

tweet_negatif = []

tweet_netral = []

tweet_error = []

# melakukan analisis sentimen dan menghitung polaritas

for i in range(len(list_of_tweets)):

    try:

        text = Textblob(list_of_tweets[i])

        polaritas = text.sentiment.polarity

        if polaritas > 0:

tweet_positif.append(list_of_tweets[i])

        elif polaritas < 0:

tweet_negatif.append(list_of_tweets[i])

        else:

tweet_netral.append(list_of_tweets[i]

    except:

        tweet_error.append(list_of_tweets[i]
```

Hasilnya, 39,3% netizen memiliki sentimen yang positif terhadap Johnny Depp, lalu 28,8% netizen memiliki sentimen yang negatif. Lalu sebesar 31,6% netizen memiliki sentimen yang netral, dan terakhir data yang error ada sebanyak 0,26%

```
Positif:  39.294117647058826 % ( 1169 tweet)
Negatif:  28.840336134453782 % ( 858 tweet)
Netral:  31.596638655462183 % ( 940 tweet)
Error:  0.2689075630252101 % ( 8 tweet)
Total:  2975
```

Fig. 4. Hasil polaritas dari data tweet

Dengan hasil seperti ini, dapat disimpulkan bahwa sebagian besar netizen bersentimen positif pada Johnny Depp selama kurang lebih seminggu kebelakang. Data ini bisa saja tidak akurat karena terdapat kemungkinan adanya bug pada program atau data yang kurang banyak. Hasil dari analisis sentimen ini

bisa digunakan untuk melihat berapa banyak netizen atau orang orang diluar sana yang membela Johnny Depp dalam kasus ini.

Sebaiknya, bila akan dilakukan analisis sentimen berikutnya diharapkan menggunakan data yang lebih besar lagi (lebih dari 5000 tweet) dan juga melakukan *data scraping* tidak hanya dari platform Twitter, tetapi juga dari platform lain seperti Facebook, Instagram, Telegram, Tiktok dll. Lalu untuk penggunaan *regular expression* juga bisa di perketat lagi seperti menghilangkan nama nama lain yang berkemungkinan muncul. Data yang diambil juga bisa lebih berkualitas lagi, tidak hanya sekedar random tweet karena bisa saja tweet yang diambil adalah tweet dari *buzzer* dan bisa juga tweet *spam*.

IV. Ucapan Terimakasih

Pertama-tama penulis mengucapkan puji dan syukur kepada Tuhan Yang Maha Esa karena atas rahmat dan berkah-Nya penulis dapat menyelesaikan makalah ini dengan baik.

Selanjutnya penulis mengucapkan terima kasih kepada orang tua penulis yang sudah memberikan dukungan terbaik kepada penulis dalam menuntut ilmu. Penulis juga mengucapkan terima kasih kepada dosen Dr. Ir. Rinaldi, M.T yang telah membimbing selama perkuliahan pada matakuliah ini.

Kemudian penulis mengucapkan terima kasih kepada Bapak dan Ibu dosen pengajar IF2210 Strategi Algoritma yang telah membimbing dan memberikan ilmu kpada penulis, khususnya ilmu mengenai strategi algoritma yang digunakan dalam pembuatan makalah ini.

Daftar Pustaka

[1] http://informatika.stei.itb.ac.id/~rinaldi.munir

[2] S. N. Aprisadianti, "Analisis Sentimen Twitter terhadap *Content Creator* Sisca Kohl Menggunakan *Regular Expression"* ,2021

[3] https://www.techtarget.com/searchbusinessanalytics/definition/opinion-mining-sentiment-mining

[4] https://www.insider.com/johnny-depp-amber-heard-relationship-timeline-2020-7

[5] https://www.zyte.com/learn/what-is-web-scraping/

[6] https://docs.microsoft.com/en-us/dotnet/standard/base-types/regular-expression-language-quick-reference

[7] https://lp2m.uma.ac.id/2022/02/21/analisis-sentimen-sentiment-analysis-definisi-tipe-dan-cara-kerjanya/

[8] https://www.geeksforgeeks.org/removing-stop-words-nltk-python/

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 2022

Ilham Bintang Nurmansyah 13520102