

Penerapan *Regular Expression* dalam *Opinion Mining* pada Twitter untuk Survei Opini Politik

Shinta Ayu Chandra Kemala - 13516029

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika Institut Teknologi Bandung
Bandung, Indonesia

13516029@std.stei.itb.ac.id • shintaayuck@gmail.com

Abstrak—*Opinion Mining* merupakan suatu riset komputasional untuk mengenali opini dan emosi tekstual. *Regular Expression* merupakan sebuah sekuens huruf yang mendefinisikan suatu bahasa. Saat ini, Twitter adalah salah satu media sosial dengan populasi pengguna yang besar dengan iklim interaksi yang masih cenderung positif. Selain menjadi media komunikasi dua arah, Twitter saat ini menjelma menjadi media sosial tempat informasi tersebar dan ruang diskusi publik yang digemari tidak hanya di Indonesia, namun skala dunia. Salah satu topik diskusi yang selalu hangat di kalangan pengguna Twitter adalah isu politik, sehingga peluang bagi para pelaku politik menganalisis sentimen publik untuk menentukan langkah politik lewat Twitter merupakan suatu hal yang cukup pantas dipertimbangkan. Makalah ini akan membahas bagaimana mengklasifikasi sentimen opini pengguna Twitter terhadap suatu kata kunci dengan menerapkan *Regular Expression* untuk proses *Lexicon-Based Opinion Mining* dan bagaimana penerapannya dalam melakukan survei opini politik.

Keywords—*Regular Expression; Opinion Mining; Twitter;*

I. PENDAHULUAN

Manusia adalah makhluk sosial, di mana interaksi sosial merupakan salah satu kebutuhan manusia yang harus terpenuhi. Bentuk interaksi manusia tidak terbatas pada interaksi langsung seperti tatap muka dan berbicara, tetapi bisa juga dengan menggunakan media lain. Saat ini, media interaksi yang kerap kali digunakan adalah media sosial.

Media sosial mulai berkembang semenjak internet semakin lazim digunakan di kalangan masyarakat. Selain fiturnya yang lengkap serta mudah digunakan, aksesnya pun biasanya tidak berbayar. Salah satu media sosial populer di dunia adalah Twitter. Bukan hanya fiturnya yang simpel dan mudah digunakan, namun juga kultur berinteraksi Twitter yang menarik sehingga banyak kalangan memilih membuka akun Twitter.

Selain masyarakat biasa, publik figur seperti artis sampai tokoh politik pun memiliki akun Twitter pribadi. Cukup banyak akun ofisial kenegaraan milik para pemimpin negara, anggota parlemen, petinggi partai politik sampai akun resmi tim sukses pemilihan umum suatu pasangan calon dan simpatisan. Beberapa tahun belakangan ini Twitter bukan digunakan hanya untuk berkomunikasi dengan relasi, namun juga menjadi sarana penyebaran informasi dan ruang diskusi publik. Topik-topik

yang didiskusikan pun beragam, tidak terbatas hal ringan seperti keseharian maupun hiburan, namun juga hal-hal fundamental seperti ideologi, ekonomi sampai politik. Ruang diskusi publik seperti Twitter membuat masyarakat awam pun bisa menyampaikan opininya dengan mudah.

Pengumpulan data opini publik terhadap suatu hal merupakan proses yang sulit dan mahal. Dari mulai pencarian responden, pengisian dan pengumpulan data yang ditutup dengan pengolahan data merupakan proses yang panjang dan tetap belum tentu menggambarkan kondisi yang sebenarnya. Dengan pengumpulan data lewat Twitter dan pengolahan data menggunakan *opinion mining*, biaya dan waktu untuk melakukan riset opini bisa dipangkas. Karena itulah, Twitter merupakan salah satu media sosial yang menjadi acuan riset baik primer maupun sekunder bagi para pihak yang berperan dalam dinamika politik.

II. TEORI DASAR REGULAR EXPRESSION

A. Definisi *Regular Expression*

Regular expression, atau yang biasa disebut *regex*, adalah sebuah sekuens huruf yang mendefinisikan sebuah pola pencarian dalam teori bahasa formal di ilmu komputer. *Regular expression* merupakan salah satu *pattern matching* yang tidak mewajibkan *pattern* persis sama (*exact match*) namun tergantung pada aturan *pattern* yang dibuat sebelumnya. Pada penerapannya, *regular expression* seringkali digunakan untuk memvalidasi input sebuah *text* dengan format penulisan khusus, misalnya e-mail, nomor telepon, nomor anggota dan sebagainya.

B. *Regular Expression* pada Teori Automata

Pada teori automata, *regular expression* merupakan notasi yang dapat mendefinisikan tepat seluruh bahasa dalam suatu bahasa *regular*. Bila E adalah *regular expression*, maka $L(E)$ adalah bahasa yang tepat didefinisikan oleh E . *Regular expression* dijelaskan secara rekursif dengan basis :

1. Jika a adalah suatu simbol, maka a adalah *regular expression* dan $L(a) = \{a\}$.
2. Jika ϵ (*empty string*) adalah *regular expression*, maka $L(\epsilon) = \{\epsilon\}$.
3. Jika \emptyset (*empty set*) adalah *regular expression*, maka $L(\emptyset) = \emptyset$.

dan rekurens :

1. Jika E1 dan E2 adalah regular expression, maka E1 + E2 adalah regular expression dan $L(E1 + E2) = L(E1) \cup L(E2)$ (Union).
2. Jika E1 dan E2 adalah regular expression, maka E1E2 adalah regular expression dan $L(E1E2) = L(E1) L(E2)$ (Concatenation).
3. Jika E adalah regular expression, maka E* (kleene closure dari E) adalah regular expression dan $L(E^*) = (L(E))^*$.

Pada regular expression, berlaku *operator precedence*, di mana tanda * memiliki *precedence* paling tinggi, diikuti oleh *concatination* kemudian terakhir adalah *union*. Tanda kurung (*parentheses*) mempunyai arti seperti ekspresi matematika yang bisa digunakan saat ingin mengabaikan *precedence* yang ada.

Dalam penggunaan operator konkatinasi dan union, berlaku hukum aljabar yang mirip seperti hukum aljabar matematika. Konkatinasi memiliki sifat yang mirip dengan perkalian (\times), di mana perkalian memiliki sifat komutatif dan asosiatif sedangkan konkatinasi memiliki sifat asosiatif. Sedangkan union memiliki sifat sama seperti penjumlahan (+), yaitu memiliki sifat komutatif dan asosiatif. Identitas dari union adalah \emptyset , di mana $R + \emptyset = R$. Sedangkan identitas dari konkatinasi adalah ϵ , di mana $\epsilon R = R$.

C. Notasi Umum Regular Expression

Berikut adalah beberapa notasi umum regular expression :

.	Karakter apapun kecuali <i>newline</i> .
\.	Titik. Berlaku untuk karakter simbol lainnya. (\., *, \(, dll)
^	Prefix dari <i>string</i> .
\$	Suffix dari <i>string</i> .
\d	sebuah digit.
\w	sebuah karakter huruf [A-Za-z0-9_].
\s	sebuah spasi.
\D	Apapun selain sebuah digit.
\W	Apapun selain sebuah karakter huruf.
\S	Apapun selain sebuah spasi.
[]	Sebuah karakter apapun, selama ada di dalam kurung siku.
[^]	Karakter apapun kecuali yang berada dalam kurung siku.
	atau.

?	Nol atau satu elemen yang sama dengan elemen sebelumnya.
*	Nol atau lebih elemen yang sama dengan elemen sebelumnya.
+	Satu atau lebih elemen yang sama dengan elemen sebelumnya.
{n}	Tepat sejumlah n elemen sebelumnya.
{n,}	Minimal sejumlah n elemen sebelumnya.
{m,n}	Sejumlah antara m sampai n elemen sebelumnya.
(expr)	<i>Capturing group</i> .
(?:expr)	<i>Non-capturing group</i> .
(?=expr)	Diikuti oleh expr.
(?!expr)	Tidak diikuti oleh expr.

III. TEORI DASAR OPINION MINING

A. Definisi Opinion Mining

Pada paper yang ditulis oleh Dave et al, alat *opinion mining* yang ideal akan memproses sebuah himpunan hasil pencarian untuk sebuah item, *men-generate* sebuah list berisi atribut-atribut dari produk (kualitas, fitur, dan sebagainya) dan mengagregasi opini tentang setiap produk tersebut, apakah baik, campuran maupun buruk. Secara definisi, *opinion mining* merupakan sinonim dari *sentiment analysis*, *subjective analysis* dan sebagainya karena memiliki definisi yang cenderung sama. Sedangkan menurut Bing Liu, *opinion mining* adalah proses analisis teks untuk mendapatkan informasi tertentu dalam suatu kalimat yang berbentuk opini.

B. Opini

Kalimat opini pada dasarnya memiliki dua komponen, yaitu target (atau topik) dan sentimen. Target adalah objek yang dinilai sedangkan sentimen adalah bentuk penilaiannya. Dalam menganalisis opini, terdapat tiga level analisis, yaitu :

- level dokumen
di mana yang dianalisis adalah keseluruhan dokumen, dengan asumsi suatu dokumen yang dianalisis hanya membahas satu topik saja.
- level kalimat
akan diambil sentimen secara umum dalam kalimat yang dianalisis.
- level entitas dan aspek
akan dianalisis lebih dalam sentimen tiap item yang ada

Selain sentimen, terdapat dua faktor yang berkaitan erat dalam menentukan sentimen suatu opini, yaitu :

- subjektivitas
suatu opini dinilai dengan skala [0,1] yang menentukan apakah opini tersebut merupakan opini subjektif atau objektif. Opini objektif menyatakan kebenaran dan fakta, sedangkan opini subjektif menyatakan pendapat, pandangan maupun kepercayaan pribadi. Walaupun begitu, opini objektif mungkin memiliki sentimen dan opini subjektif mungkin tidak mengandung sentimen.
- emosi
suatu opini mungkin mengandung emosi yang lebih spesifik, seperti senang, sedih, kecewa, marah, bingung, takut dan sebagainya.

C. Penilaian Polaritas Sentimen

Dalam penilaian sentimen sebuah opini, skala polaritas sentimen secara kuantitatif adalah [-1,1] sedangkan dalam skala kualitatif, nilai polaritas sentimen adalah negatif, netral dan positif. Dalam penentuan polaritas sentimen sebuah opini, dapat digunakan berbagai metode. Secara umum, terdapat dua pendekatan dalam penilaian sentimen, yaitu dengan menggunakan pendekatan *Machine Learning* dan pendekatan *Knowledge-Based*.

Pendekatan *Machine Learning* memerlukan dataset yang berfungsi sebagai *data training*. Setelah itu, dataset tersebut akan diklasifikasikan dengan class tag serta melakukan proses training. Akurasi sangat baik namun performanya tergantung pada dataset yang digunakan. Beberapa metode yang termasuk pendekatan ini antara lain :

- Naive Bayes
- Maximum Entropy
- SVM
- Neural Network

Pendekatan *Knowledge-Based* atau berbasis pengetahuan pada level kata, di mana yang diproses adalah kata. Pendekatan ini bergantung pada kamus lexicon yang digunakan untuk penilaian fitur. Karena berbasis kata, terkadang makna kata yang ditangkap tidak sesuai dengan konteks kalimat, misalnya pada bentuk kalimat sarkasme yang tidak menggambarkan konteks kalimat asli. Namun sisi positif dari pendekatan ini adalah pengetahuan dapat ditambahkan kapanpun ke dalam domain.

Metode yang termasuk pendekatan ini antara lain :

- Lexicon-Based
- Pointwise Mutual Information

Pada metode Lexicon-based, kata pada kamus lexicon dipasangkan dengan nilai polaritasnya. Sebelum dilakukan analisis kata, perlu dilakukan pemilihan kata yang akan dianalisis. Pemilihan kata bisa dilakukan dengan melakukan *Part-Of-Speech tagging*.

Part of speech tagging adalah metode untuk mengklasifikasikan kata-kata sesuai kategori yang tepat,

misalnya adjektiva atau adverbial. *Part of speech tagging* merupakan hal yang biasa dilakukan pada *Natural Language Processing* (NLP). *Part of speech tagging* ini berguna untuk mengidentifikasi dan membedakan dua kalimat dengan konteks yang berbeda. Misal kalimat "*I love the scenario*" mengandung sentimen positif sedangkan "*The film is about their love story*" merupakan opini netral, meskipun keduanya sama-sama memiliki kata "love". Ini bisa terjadi karena kedua kata "love" memiliki tag yang berbeda. Pada kalimat pertama, "love" merupakan kata kerja (verba) sedangkan kata "love" pada kalimat kedua merupakan kata *attributive noun*.

Beberapa kamus yang digunakan dalam lexicon-based antara lain adalah SentiWordNet, AFINN-111 yang menggunakan bahasa Inggris sebagai acuan bahasanya.

IV. METODE OPINION MINING PADA TWITTER

Pada makalah ini, metode *opinion mining* yang digunakan adalah *opinion mining* dengan level dokumen, di mana satu *tweet* dianggap satu dokumen. Pendekatan yang digunakan adalah *Knowledge-Based* dengan metode yang digunakan adalah metode Lexicon-Based. Kamus yang digunakan berasal dari kamus SentiWordNet. Lalu subjektivitas dan emosi diabaikan detailnya, sehingga hanya berfokus pada polaritas opini. Konversi polaritas opini dari kuantitatif menjadi kualitatif sebagai berikut :

- Positif : $-1 \leq \text{Polaritas} < 0$
- Netral : $\text{Polaritas} = 0$
- Negatif : $0 < \text{Polaritas} \leq 1$

Secara umum, proses *opinion mining* pada Twitter dilakukan dalam empat tahapan, yaitu :

- pengumpulan data
- pembersihan data (*data cleaning*)
- klasifikasi data
- penarikan kesimpulan

A. Pengumpulan Data

Pada tahap pengumpulan data, proses dimulai dengan *data scraping* pada laman hasil pencarian Twitter www.search.twitter.com sesuai kata kunci yang diinginkan. *Data scraping* dilakukan dengan menggunakan Twitter API dan library Python yaitu Tweepy <http://tweepy.readthedocs.io/en/v3.5.0/api.html#tweepy-api-twitter-api-wrapper>.

Scraping tweet dilakukan secara otomatis untuk mengambil data tweet dengan jangka waktu data yang bisa ditentukan. Sebelum dapat melakukan *scraping*, sistem ini akan melakukan pengecekan *consumer key* dan *token access* agar sistem dapat mengambil data dari Twitter dengan memanfaatkan Twitter API. Syarat untuk mendapatkan *consumer key* dan *token access* adalah dengan membuat aplikasi Twitter di laman resmi Twitter.

B. Pembersihan Data

Setelah data berhasil dikumpulkan, data akan dibersihkan sehingga setiap tweet yang akan diolah dipastikan sudah siap diolah. Tweet dibersihkan dari karakter khusus seperti tanda baca (titik, koma, garis miring, underscore, dsb), emoticon (:), (:, :3, :v, dsb)), karakter non-huruf (hangul, hiragana, kanji, dsb), serta angka (0-9). Selain itu, link juga dihapus dari tweet. Pembersihan ini dilakukan untuk memastikan tweet yang akan olah sudah bersih dari karakter yang tidak akan berpengaruh pada proses analisis. Pada proses pembersihan, dimanfaatkan *Regular Expression* untuk mengenali mana saja elemen pada tweet yang harus dibuang. Pada pembersihan data ini tidak diperlukan *casefolding* (pengubahan seluruh tweet menjadi huruf kecil) karena pada *Regular Expression* disediakan fitur pengabaian huruf kecil dan kapital.

C. Klasifikasi Data

Proses selanjutnya adalah klasifikasi data. Data yang sudah bersih tersebut akan diklasifikasi berdasarkan klasifikasi kata yang ada. Pengecekan klasifikasi kata juga menggunakan *Regular Expression*. Klasifikasi kata dilakukan berdasarkan polaritas suatu tweet. Nilai polaritas ditentukan dari jumlah sentimen positif dikurangi jumlah sentimen negatif yang ada, lalu dirata-ratakan.

D. Penarikan Kesimpulan

Setelah diklasifikasikan, proses terakhir adalah penarikan kesimpulan. Pada proses ini, dihitung persentase tweet yang memiliki polaritas positif, negatif maupun netral serta lima sampel tweet teratas.

V. IMPLEMENTASI PROGRAM DAN ANALISIS

Pada bagian ini, akan dijelaskan strategi dan implementasi *Opinion Mining* menggunakan *Regular Expression* dalam bahasa Python. Batasan pengecekan program adalah menggunakan Bahasa Inggris, karena dataset referensi yang digunakan adalah Bahasa Inggris.

A. Persiapan

Sebelum memulai pembuatan program *opinion mining*, diperlukan sekumpulan data yang berisi kata beserta nilai polaritasnya. Proses pembuatan data nilai polaritas ini dibuat menggunakan SentiWordNet 3.0. Selain itu, diperlukan pembuatan akun developer Twitter serta aplikasi Twitter di laman dev.twitter.com. Setelah aplikasi dibuat bisa didapatkan *token access* dan *consumer key* yang akan digunakan dalam pengaksesan API Twitter.

B. Pseudo-Code

Berikut adalah *pseudo-code* dari program utama *opinion mining*.

```
Function Main() : void
ALGORITMA
//Setting API menggunakan data dari akun
```

```
Developer Twitter
api = TwitterClient()

//Memanggil fungsi get_tweets
tweets = api.get_tweets(query = Keyword
, count = 200)

//Mengambil data tweet
print(panjang tweet positif)
print(panjang tweet negatif)
print(panjang tweet netral)
print(isi maksimal 5 tweet positif)
print(isi maksimal 5 tweet negatif)
print(isi maksimal 5 tweet netral)
```

Berikut adalah *pseudo-code* dari fungsi *get_tweets* dari program *opinion mining*.

```
Function get_tweets(query, count) :
Dict[]

ALGORITMA
//Ambil data tweet menggunakan API
Twitter
fetched_tweets = self.api.search(q =
query, count = count)

//Parse tweet ke dalam dictionary
parsed_tweet['text'] = isi_tweet
parsed_tweet['sentiment'] =
get_tweet_sentiment(isi_tweet)

//Tambahkan tweet ke array
tweets.append(parsed_tweet)

return tweets
```

Berikut adalah *pseudo-code* dari fungsi *get_tweet_sentiment* dari program *opinion mining*.

```
Function get_tweet_sentiment(isi_tweet)
: String

ALGORITMA
//Lakukan cleaning tweet
analysis = clean_tweet(isi_tweet)

//Cek polaritas
for (semua kata di isi_tweet) :
    if (match(kata, kata_di_dataset) :
        sum += kata_di_dataset.polarity
//Klasifikasi polaritas
if sum > 0 :
    return 'Positif'
else if sum < 0 :
```

```

return 'Negatif'
else :
return 'Netral'

```

Berikut adalah *pseudo-code* dari fungsi `clean_tweet` dari program *opinion mining*.

```

Function clean_tweet(isi_tweet) : String

ALGORITMA
//Hapus seluruh karakter non huruf
return isi_tweet.remove((@[A-Za-z0-9]+)|([\^0-9A-Za-z \t]) |(\w+:\//\//S+))

```

C. Hasil Program

Berdasarkan program yang sudah dibuat, berikut adalah hasil yang didapatkan. Data tweet yang diambil sebanyak 100 untuk setiap percobaan.

Untuk Keyword = Kim Jong Un

```

Masukan keyword = Kim Jong Un

Persentasi Tweet Positif : 36.111111111111114 %
Persentasi Tweet Negatif: 11.111111111111111 %
Persentasi Tweet Netral: 52.777777777777778 %

```

Tweet Positif :
RT @gctje48: @erikmouthaanRTL @menbearpig1 LOL,

Na Kim Jong Un
maakt ook Erik Mouthaan
een 180 graden draai ;-)) <https://t.co/JmfT5xcLc8>
RT @FoxNews: Maj. Gen. Don Alston: "This meeting with [@POTUS] is gonna be a
bout the most important day in [Kim Jong Un's] life, publicly,--
This site was already destroyed in their final test. Kim jong un is as good
or better a con man than... <https://t.co/Xb1otHVnfp>
RT @RealJack: In 1 week, the Trump Administration did more for the American
people than Obama did in around 416 weeks.

Brought home hostag--
RT @realDonaldTrump: The highly anticipated meeting between Kim Jong Un and
myself will take place in Singapore on June 12th. We will both...
RT @realDonaldTrump: I am pleased to inform you that Secretary of State Mike
Pompeo is in the air and on his way back from North Korea with...
RT @TheBaxterBean: BREAKING: Trump gives five stars on Airbnb to 'very honor
able' Kim Jong Un, over Americans imprisoned in North Korea htt...
@phlubup Trump fans don't believe any news source, but they have no problem
at all believing North Korea.

New Yo... <https://t.co/EJTzWbx5FX>
@erikmouthaanRTL @menbearpig1 LOL,

Na Kim Jong Un
maakt ook Erik Mouthaan
een 180 graden draai ;-)) <https://t.co/JmfT5xcLc8>
RT @Zimrico: Nothing terrifies the Left more than the prospect of @POTUS suc
ceeding.

<https://t.co/X9GL8zSBIw>

Tweet Negatif:
@celtofcanada Well, some other stuff happened while you were off grid. Turns
out Kim Jong Un was an alien from a di... <https://t.co/H1mCixOPr7>
RT @DropTheTidePod: It's hysterical that the left are now criticizing Trump
for thanking Kim Jong Un for releasing prisoners...

Feels like...
RT @KFOX14: North Korea said Saturday that it will dismantle its nuclear tes
t site on May 23-25, in a dramatic event that would set up lead...
RT @Glorfindel2: Chuck Schumer reveals his stupidity when he rags on Trump f
or diplomatically flattering Kim Jong Un. In an illogical way,--

Untuk Keyword = Donald Trump

Masukan keyword = Donald Trump

```

Persentasi Tweet Positif : 44.444444444444444 %
Persentasi Tweet Negatif: 18.055555555555557 %
Persentasi Tweet Netral: 37.5 %

```

Tweet Positif :
RT @BigshotEddie57: @sarahkendzior: "Well, all of these people are all so in
timately connected with Donald Trump and the people you just na...
RT @RealJamesWoods: Clear that she may be getting challenged in her cognitiv
e abilities, which is no laughing matter. Henceforth I will not...
RT @joncoopertweets: Eddie Devine voted for Donald Trump because he thought
he'd be good for American business. Now, the Trump administrati...
RT @Amy_Siskind: More on Vekselberg:
"The FBI warned four years ago that a foundation controlled by the Russian o
ligarch who allegedly reim...
RT @manbeck57: @55true4u The Greatest Living PRESIDENT Ever!!!!POTUS DONALD
JOHN TRUMP !!!THE ONLY CHANGE IS TO DE-elite the ones who share...
This article is really long, but worth reading. <https://t.co/nPGSENLYe1> via
@intelligencer
RT @krassenstein: John McCain seems to be the EXACT inverse of Donald Trump.

For every year he was being tortured as a POW in a prison cam...
RT @bill_aclair: What a drag... it turns out that #RudyGiuliani really *is*
@realDonaldTrump 's bitch. <https://t.co/ipR4tBTh7o> #TrumpLiesM...
RT @GeorgeTakei: Another new low for Donald. <https://t.co/OZObLYFJh1>
RT @JacobAWohl: Many liberals I talk to say they are completely ashamed of t
hemselves for voting for Hillary Clinton, now that they see wha...

Tweet Negatif:
RT @CREWcrew: The FBI warned four years ago that a foundation controlled by
the Russian oligarch who allegedly reimbursed Donald Trump's pe...
RT @michaelbeatty3: ANOTHER
RESISTANCE 🇺🇸 BLUE WAVE
KICK IN THE NUTS 🤔🤔

*BREAKING:
🇺🇸North Korea reveals it will destroy its
nuclear test...
RT @GeorgeTakei: Childish and classless as always, Donald. <https://t.co/OZObLYFJh1>
RT @ChrisJZullo: Remember when Donald Trump said only guilty people take the
fifth amendment? Rudy Giuliani is now saying that Donald Trump...
RT @dbongino: Liberals: "Donald Trump is a dangerous tyrant who must be stop
ped at all costs."

Same Liberals: "Surrender your firearms to...
RT @1776Stonewall: This lazy ignorant liberal here, doubted my story that Ri
chard Nixon wrote Donald Trump a letter on running for office i...
RT @Brasilmagic: It's all fake, for the cameras. By the way, I kiss my fathe
r on the cheeks. <https://t.co/35pK5YbhyW>
RT @marstu67: take-away: Trump may be an idiot in the WH. but he's a smart c
onman. <https://t.co/wvMMEyP6s>
RT @bulldoghill: John McCain: 'Vladimir Putin Is an Evil Man'
In an excerpt from his new memoir, Sen. McCain discusses his longstanding op
p...
RT @CNNPolitics: President Trump misleads about military pay raises <https://t.co/kWKOSFD116> <https://t.co/QjPviI0ZAT>

Untuk Query = Joko Widodo

```
Masukan keyword = Joko Widodo

Persentasi Tweet Positif : 7.6923876923876925 %
Persentasi Tweet Negatif: 0.0 %
Persentasi Tweet Netral: 92.3876923876923 %

Tweet Positif :
Retweeted Bali Antara News (@antaranews_bali):

Presiden Jokowi main bola basket bersama atlet pelajar: Presiden Jo... https://t.c
o/HrIWZxuQbc
RT @antaranews_bali: Presiden Jokowi main bola basket bersama atlet pelajar: Pre
siden Joko Widodo bermain bola basket bersama atlet pelajar...
Main Basket Bersama Presiden Joko Widodo, Menpora Kenakan Kaos Asian Games deng
an Motif Avengers... https://t.co/Y6BsVbgoFi
Retweeted Joko Widodo (@jokowi):

Main basket lagi setelah 37 tahun nggak pegang bola basket. Kegiatan seru bareng
... https://t.co/IgSjkwTS6k

Tweet Negatif: _
```

D. Analisis Hasil Percobaan

Berdasarkan hasil percobaan dari kedua teori tersebut, dapat dilihat bahwa hasil yang didapat sudah cukup baik untuk data berbahasa Inggris dan sangat buruk untuk data berbahasa Indonesia.

Untuk data berbahasa Inggris, beberapa tweet dengan konteks yang berbeda dengan polaritas kata yang dikandungnya tidak terdeteksi, meskipun beberapa berhasil mengenali konteks kalimat. Contoh dari tweet yang kurang tepat dikenali konteksnya oleh program adalah :

“Chuck Schumer reveals his stupidity when he rag on Trump for diplomatically flattering Kim Jong Un. In an illogical way”

yang ditemukan pada keyword “Kim Jong Un”, di mana secara bahasa memang benar bernada negatif, namun nada negatif tersebut tidak ditujukan pada Kim Jong Un melainkan pada Chuck Schumer. Sedangkan di sisi lain, data yang memiliki konteks agak berbeda dari pemilihan katanya, seperti:

“Many liberals I talk to say they are completely ashamed of themselves for voting for Hillary Clinton”

diklasifikasikan sebagai data tweet positif untuk Donald Trump. Walaupun kalimat yang ada bernada negatif, namun ujaran negatif tersebut ditujukan pada Hillary Clinton yang merupakan lawan politik Donald Trump sehingga klasifikasi yang terjadi merupakan contoh data yang sukses dianalisis.

Untuk data ketiga, yaitu dengan kata “Joko Widodo”, nama presiden Indonesia saat ini, hasilnya tidak memuaskan. Kesesuaiannya sangat jauh dari realita karena memang data referensi yang digunakan adalah data dari bahasa Inggris, sehingga program tidak bisa memproses teks berbahasa Indonesia. Untuk saran pengembangan, bisa dilakukan translasi terhadap tweet yang diterima terlebih dahulu sebelum dilakukan klasifikasi. Namun cara ini tetap kurang efektif karena banyak kata yang dalam bahasa Indonesia memiliki konteks negatif namun saat ditranslasi menjadi hilang makna awalnya. Selain itu kata tidak baku (*slang*) juga biasanya sulit

ditranslasi sehingga akan tetap sama dengan bahasa Indonesia dan merusak keseluruhan kalimat yang ditranslasi.

VI. PENERAPAN *OPINION MINING* UNTUK SURVEI OPINI POLITIK

Dengan penggunaan *opinion mining* berbasis Twitter, data yang didapat bisa dimanfaatkan untuk berbagai hal, antara lain:

- Mengetahui sentimen publik terhadap suatu tokoh politik
Dengan mengetahui sentimen publik terhadap suatu tokoh politik, secara kasar dapat dilihat *branding* apa yang diterima baik atau malah ditolak masyarakat, bagaimana tindakan seorang tokoh bisa diterima masyarakatnya, sampai menentukan elektabilitas.
- Mengetahui sentimen publik terhadap suatu kubu politik
Dengan mengetahui sentimen publik terhadap suatu kubu politik, bisa dianalisis isu apa saja yang sebaiknya diangkat kubu tersebut atau kubu lawan, seperti apa kelompok masyarakat yang potensial mendukung suatu kubu dan sebagainya.
- Mengetahui sentimen publik terhadap suatu kebijakan
Bagi pemimpin negara, uji coba suatu kebijakan terkadang menghabiskan biaya yang tidak sedikit demi mendapat sentimen publik. Namun dengan pelepasan isu di media sosial lalu menganalisis umpan balik dari masyarakat, dapat digambarkan secara kasar bagaimana kira-kira tanggapan masyarakat bila kebijakan tersebut diterapkan.
- Mengetahui sentimen publik terhadap suatu isu publik
Melakukan *opinion mining* lewat media sosial terhadap isu kenegaraan seperti kurs naik, perjanjian antar negara dan semacamnya merupakan salah satu metode cepat dan mudah untuk pemerintah menerima umpan balik dari masyarakat, terlebih karena pemerintah tidak bisa selalu memantau reaksi masyarakat di kehidupan nyata.

Selain dalam bidang politik, beberapa bidang yang bisa memanfaatkan *opinion mining* adalah bidang ekonomi, pemasaran, pemerintahan, hukum dan sebagainya. *Opinion mining* bisa digunakan untuk mendapatkan segmentasi pasar, peluang penjualan suatu produk, strategi pemasaran yang tepat sasaran, uji kebijakan publik, uji materi, dan masih banyak lagi. Secara garis besar, *opinion mining* berguna untuk menganalisis suatu hal yang berhubungan dengan masyarakat luas.

Selain untuk menganalisis sentiment suatu kelompok, *opinion mining* juga bisa berfungsi untuk menilai kredibilitas suatu sistem yang seharusnya netral. Misal untuk penilaian kredibilitas media dapat dilihat dari seberapa netral pemberitaan yang dilakukan, karena media yang kredibel adalah yang mampu membuat pemberitaan yang berimbang.

VII. KESIMPULAN

Regular Expression adalah suatu sekuens huruf yang mendefinisikan sebuah pola pencarian dalam teori bahasa formal di ilmu komputer serta digunakan untuk pengecekan *string* tanpa *exact matching*. *Opinion mining* merupakan proses analisis teks untuk mendapatkan informasi tertentu dalam suatu kalimat yang berbentuk opini, dan menghasilkan derajat polaritas suatu kalimat, baik itu positif, netral maupun negatif.

Proses *opinion mining* dilakukan menggunakan metode Lexicon-based, yaitu bagian dari pendekatan berbasis pengetahuan berbentuk kamus. Pencocokan kata pada kamus menggunakan *regular expression* berdasarkan data tweet suatu topik dari Twitter yang diambil dengan memanfaatkan Twitter API dan Library Python bernama Tweepy.

Opinion mining hanya bisa dilakukan pada data berbahasa Inggris, dengan memanfaatkan kamus dari SentiWordNet. Data uji yang digunakan terdiri dari tiga kata kunci, yaitu “Kim Jong Un”, “Donald Trump” dan “Joko Widodo”. Pengujian menghasilkan hasil yang cukup valid pada dua kata kunci pertama yang memiliki data uji dengan mayoritas Bahasa Inggris, namun hasil untuk data ketiga, di mana data uji mayoritas menggunakan Bahasa Indonesia, memiliki validitas sangat rendah karena pencocokan pada kamus tidak berjalan dengan baik.

Metode *opinion mining* pada Twitter bisa diaplikasikan pada beberapa sektor yang membutuhkan analisis sentimen, seperti marketing maupun politik. Terkhusus pada bidang politik, *opinion mining* bisa berfungsi sebagai parameter awal bagi para pelaku politik, baik pemerintah, pihak oposisi maupun pihak lain untuk mengetahui bagaimana reaksi masyarakat dan seperti apa sebuah informasi tentang kebijakan maupun *branding* sebuah pihak diterima di masyarakat. Hasil analisis ini bisa dimanfaatkan untuk menentukan langkah politik yang akan diambil selanjutnya.

ACKNOWLEDGMENT

Penulis mengucapkan terima kasih kepada Allah SWT atas karunia dan hidayah-Nya sehingga Penulis bisa menyelesaikan proses pembuatan makalah ini. Penulis juga mengucapkan terima kasih kepada Ibu Nur Ulfa Maulidevi yang telah membimbing penulis dalam memahami materi perkuliahan mata kuliah IF2211 Strategi Algoritma, kepada orang tua penulis yang selalu mendukung penulis sehingga dapat menyelesaikan makalah ini.

REFERENCES

- [1] Aho, Alfred V. (1990). van Leeuwen, Jan, ed. Algorithms for finding patterns in strings. Handbook of Theoretical Computer Science, volume A: Algorithms and Complexity. The MIT Press. pp. 255–300.
- [2] <https://bmrnd.me/social-sentiment-analysis/> diakses tanggal 13 Mei 2018.
- [3] <https://mti.binus.ac.id/2017/10/04/1900/> diakses tanggal 13 Mei 2018.
- [4] <http://docs.tweepy.org/en/v3.5.0/> diakses tanggal 13 Mei 2018
- [5] <http://nmis.isti.cnr.it/sebastiani/Publications/LREC10.pdf> diakses tanggal 13 Mei 2018

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, 13 Mei 2018



Shinta Ayu Chandra Kemala - 13516029