

ANALISIS SENTIMEN MASYARAKAT TERHADAP VIRUS CORONA BERDASARKAN OPINI DARI TWITTER BERBASIS WEB SCRAPER

ROBI KURNIAWAN¹, AULIA APRILIANI²

^{1,2}Mahasiswa Program Pascasarjana Sistem Komputer STMIK Handayani

^{1,2}Tenaga Pengajar Jurusan Teknik Informatika Universitas Kristen Indonesia (UKI) Paulus

Jl. Adhyaksa Baru No.1 Panakukkang, Kota Makassar, 90231

Jl. Perintis Kemerdekaan Km. 13 Daya, Kota Makassar, 90245

E-mail: ¹robikurniawan@student.handayani.ac.id

²auliaapriliani@student.handayani.ac.id

Abstrak

Indonesia menjadi salah satu Negara yang pengguna aktif harian twitternya cukup tinggi, berdasarkan hal tersebut twitter dapat dijadikan sebagai media untuk melakukan analisis sentimen terhadap topik corona. Analisis sentimen merupakan salah satu cabang dari *text mining* yang melakukan proses klasifikasi pada dokumen atau teks. Penelitian ini bertujuan untuk mengetahui bagaimana dampak virus corona di Indonesia sesuai opini masyarakat melalui twitter. Pengumpulan data dilakukan dengan teknik *web scraper* yang menghasilkan 1000 *record* sejak tanggal 20 Januari sampai 1 Februari 2020, data yang telah di *scraping* kemudian dianalisis mengikuti tahapan *text mining* yaitu *case folding*, *tokenizing* dan *filtering*. Hasil dari penelitian ini menunjukkan persentase opini masyarakat terhadap virus corona yaitu 79% negatif, 11% Netral dan 10% Positif.

Kata kunci : corona, analisis sentimen, twitter;

I. PENDAHULUAN

Pengguna aktif harian twitter di Indonesia saat ini cukup banyak, Indonesia diklaim menjadi salah satu Negara yang pertumbuhan pengguna aktif harian twitternya paling besar berdasarkan laporan finansial twitter kuartal ke-3 tahun 2019 (Annual Report Twitter, 2019). Melihat bagaimana perkembangan twitter tersebut maka twitter menjadi salah satu media yang dapat digunakan untuk melakukan analisis sentimen terhadap berbagai topik. Penelitian ini melakukan suatu analisis sentimen terhadap topik yang saat ini sering menjadi *trending topic* di twitter yaitu “Corona”.

Virus corona atau dikenal sebagai pandemi Covid-19 pertama kali di deteksi di kota wuhan Desember 2019 (Wikipedia, 2019), sejak penyebarannya virus ini

juga langsung dibicarakan masyarakat twitter, saat ini virus corona tengah menjadi perhatian dunia internasional. Banyaknya jumlah korban dan cepatnya persebaran virus membuat masyarakat khawatir dan muncul berbagai opini tentang virus corona termasuk di Indonesia. Opini ini lah yang kemudian di analisis untuk diketahui polaritasnya dengan analisis sentimen.

Analisis sentimen adalah bagian dari teks mining yang dapat mengelompokkan polaritas dari teks, pengelompokkan tersebut dilakukan untuk melihat bagaimana polaritas dari suatu tweet apakah opini yang diberikan bersifat positif, negatif atau netral (Fanissa, Fauzi and Adinugroho, 2018). Juga untuk mengetahui siapa pemberi opini yang banyak mendapatkan respon dari pengguna twitter. Analisis Sentimen dapat digunakan untuk mengungkap opini publik terhadap suatu isu, kepuasan pelayanan, kebijakan, *cyber bulliyng*, memprediksi harga saham, dan analisis pesaing berdasarkan data tekstual.

Penelitian mengenai analisis sentimen telah banyak dilakukan sebelumnya untuk berbagai keperluan. Diantaranya adalah melakukan analisis sentimen masyarakat terhadap acara televisi dengan melakukan *retweet analysis* untuk dijadikan sebagai acuan dalam suatu penilaian terhadap suatu acara televisi (Berlian, Herdiani and Astuti, 2019). Penelitian selanjutnya yaitu penggunaan web crawler untuk menghimpun tweets dengan metode pre-processing text mining, penelitian tersebut menghasilkan suatu aplikasi yang dapat mengubah tweets yang telah dihimpun menjadi data yang siap diolah sesuai kebutuhan (Aditya, 2015). Dari kedua penelitian terkait yang dibahas diatas, terdapat beberapa perbedaan yaitu, penelitian ini melakukan teknik web scraping untuk menarik seluruh data atau tweets dengan kata kunci ‘corona’ baik itu *reply*, *like* dan *reetweet*. Data yang telah discraping kemudian dianalisis untuk mengetahui bagaimana opini masyarakat terhadap “virus corona”.

Artikel ini mendiskusikan tentang bagaimana analisis sentimen masyarakat terhadap tweet yang akan menghasilkan persentase tweet positif, negatif dan netral. Sehingga kita bisa mengetahui bagaimana dampak virus corona di Indonesia sesuai opini masyarakat melalui twitter.

II. METODE PENELITIAN

Objek Penelitian

Objek yang diteliti adalah opini masyarakat mengenai virus corona melalui jejaring sosial twitter. Data yang digunakan yaitu berupa tweet (*mention, reply, like, reetweet*) berbahasa Indonesia dengan kata kunci corona sejak tanggal 20 Januari sampai 1 Februari 2020 sebanyak 1000 record.

Teknik Pengolahan dan Analisis Data

Teknik Pengolahan dan analisis data menggunakan teknik *text mining* atau juga dikenal sebagai *data mining teks*. *Text mining* merupakan suatu kegiatan menggali informasi, kegiatan tersebut dilakukan oleh seorang *user* yang berinteraksi dengan sekumpulan dokumen menggunakan *tools* analisis (Imam and Fajtriab, 2015). *Text mining* digunakan untuk memperoleh informasi yang berguna dari sekumpulan teks yang ada pada *twitter*. Adapun tahapan yang dilakukan dalam penelitian ini yaitu:

- a. Studi Literatur dan Pemahaman: Pada tahap ini peneliti mencari dan mempelajari berbagai referensi berupa textbook, artikel ilmiah dan jurnal yang berkaitan dengan penelitian.
- b. Pengumpulan Data: Pengumpulan data dilakukan dengan teknik scraper yaitu mengekstraksi data sejak tanggal 20 Januari sampai 1 Februari 2020 sebanyak 1000 record dengan kata kunci corona yang berbahasa Indonesia pada jejaring sosial twitter. Hasil pencarian kemudian akan menampilkan semua tweet termasuk *mention, reply, like dan reetweet*. Semua tweet yang ditampilkan kemudian di simpan pada database untuk di analisis. Dengan teknik web scraper akan mempermudah untuk proses pengambilan data dari internet.
- c. *Text Preprocessing*: Tahap pre-processing mencakup proses untuk mempersiapkan data yang akan digunakan pada *text mining*. Sehingga dataset telah bersih dan siap untuk digunakan. Tahap *text preprocessing* pada penelitian ini menggunakan beberapa tahapan, diantaranya: *case folding, tokenizing* dan *filtering* (Informatikalogi.com, 2017).

Case Folding: Proses mengkonversi keseluruhan teks dalam tweet menjadi bentuk standar (huruf kecil atau lowercase).

Tokenizing: Proses memecah sekumpulan karakter dalam suatu tweet menjadi satuan kata.

Filtering: Proses mengambil kata-kata penting dan membuang kata-kata yang tidak penting pada tweet (*stopwords*) misalnya kata penghubung seperti “dan”, “atau”, “kemudian” dan seterusnya atau kata-kata yang tidak berpengaruh pada proses klasifikasi.

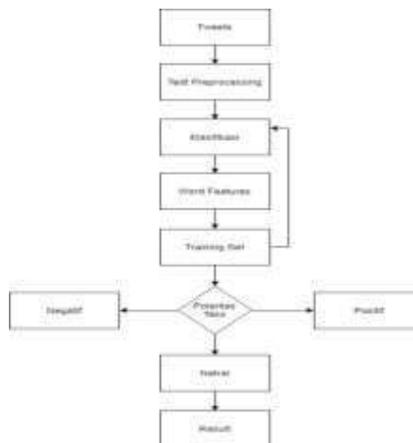
- d. Klasifikasi: Pada tahap ini tweet yang telah melalui tahapan *text preprocessing* kemudian di klasifikasi sesuai kelasnya (*sentiment class*) untuk menentukan polaritas dari text tersebut apakah text tersebut termasuk dalam opini positif, negatif atau netral (Indrayuni, 2019).

Secara garis besar berikut alur kerja penelitian:



Gambar 1. Alur Penelitian

Berikut gambaran proses analisa tweet hingga menghasilkan suatu polaritas teks apakah positif, negatif atau netral.



Gambar 2. Proses Analisis Teks

III. HASIL DAN PEMBAHASAN

Mengacu pada tahapan pengolahan dan analisis data, setelah proses pengumpulan data dengan teknik scraper selama kurang lebih 13 hari, data yang tersimpan pada database kemudian diolah dengan tahapan *text preprocessing*. Berikut tahapan *text preprocessing* yang dilakukan dengan mengambil satu *sample* tweet yang menjadi trending.

Tabel 1. Tahapan *text preprocessing*

TAHAPAN	INPUT	OUTPUT
Case Folding	@305_btrack MASYA ALLAH...!!!Dibalik virus corona sekarang di china sanggup berdesak desakan untuk sholat Jum'at sampai tumpah ruah kejalan..ALLAHUKBAR..!!.. #NegaraDijajahKoruptur#NegaraDijajahKoruptur pic.twitter.com/rdAECGtb2t	@305_btrack masya allah...!!!dibalik virus corona sekarang di china sanggup berdesak desakan untuk sholat jum'at sampai tumpah ruah kejalan..allahukbar..!!.. #negaradijajahkoruptur#negaradijajahkoruptur pic.twitter.com/rdaecgtb2t
Tokenizing	masya allah dibalik virus corona sekarang di china sanggup berdesak desakan untuk sholat jumat sampai tumpah ruah kejalan allahuakbar	Masya, allah, dibalik, virus, corona, sekarang, di, china, sanggup, berdesak, desakan, untuk, sholat, jumat, sampai, tumpah, ruah, kejalan, allahuakbar
Filtering	Masya, allah, dibalik, virus, corona, sekarang, di, china, sanggup, berdesak, desakan, untuk, sholat, jumat, sampai, tumpah, ruah, kejalan, allahuakbar	dibalik, virus, corona, sekarang, china, sanggup, berdesak, desakan, sholat, jumat, sampai, tumpah, ruah, jalan,

Setelah tahap *text preprocessing* dilakukan terjadi perubahan bentuk data yang terstruktur sembarangan menjadi bentuk data yang terstruktur sesuai dengan kebutuhan dalam proses text mining. Data yang sudah terstruktur dikategorisasikan dan dikelompokkan per kata kemudian dicocokkan dengan kamus data untuk mengetahui polaritas text apakah positif, negatif atau netral. Kamus data yang digunakan, sebagai berikut:

Tabel 2. Contoh Kamus Kata positif dan Negatif

POSITIF		NEGATIF	
anggukan	bala bantuan	anarkis	campur aduk
anggun	balas budi	ancam	candu
animo	balas jasa	ancaman	canggung
anjuran	bangga	aneh	cape
anteng	bangkit	bahaya	cape hati
antusiasme	loyalitas	bahayakan	capek
anugrah	luar biasa	bakteri	curiga
apresiasi	lucu	baku hantam	curigaan
bahagia	mabrur	bangsat	dajal
bahagiakan	mahir	banjir	daki
bahagian	maju	bantah	dakwa
bahu-membahu	makan	bantai	dakwaan
baik	makmur	cacat	dalih
baik-baik	mampu	caci	dampak
bajik	mandiri	caci maki	korban

Setelah pencocokan dengan kamus kata, selanjutnya di lakukan *training set*. Dari hasil *training set* tersebut menghasilkan polaritas teks. Sebuah teks/tweet dianggap positif atau negatif jika kata tersebut mempunyai keterkaitan dengan kamus kata, dan dianggap netral jika teks/tweet tidak terdapat kata yang terkait dalam kamus kata.

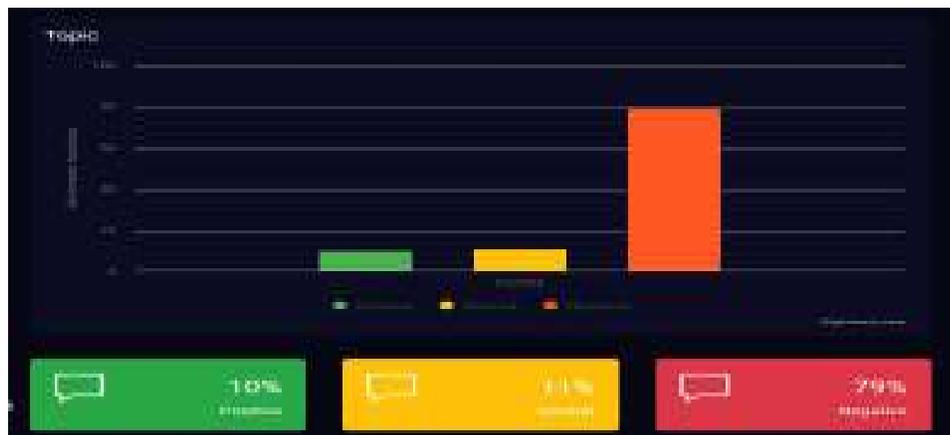
Tabel 3. Contoh Klasifikasi Tweet sesuai *Sentiment Class*

No	Tweet	Reetweet	Label
1	@godisko Untung udah bisa konsumsi kunyit madu tiap hari viral curcumin bisa tangkal virus corona	6	Positif

2	@officialfirman bayangin china keluar jutaan dolar buat riset obat corona dan kita cuma modal 2000 buat beli tolak angin	91	Netral
3	@cnnindonesia korban meninggal akibat virus corona tembus 1770 orang	54	Negatif

PEMBAHASAN

Penelitian ini menarik sebanyak 1000 record data dari twitter dengan kata kunci corona. Dari total keseluruhan record dihasilkan polaritas tweet sebagai berikut:



Gambar 3. Tampilan Dashboard Persentase Polaritas tweet

Gambar 3 menunjukkan analisis persentase polaritas dari tweets. Hasil persentase tersebut diperoleh dari hasil analisis *reply*, *like* dan *retweet* pada suatu tweet sejak tanggal 20 Januari sampai 1 Februari 2020. Dari hasil analisis tersebut diketahui bahwa opini masyarakat terhadap topik corona lebih banyak mengarah pada opini Negatif dengan total persentase 79%, opini Netral sebanyak 11% dan opini positif 10%.

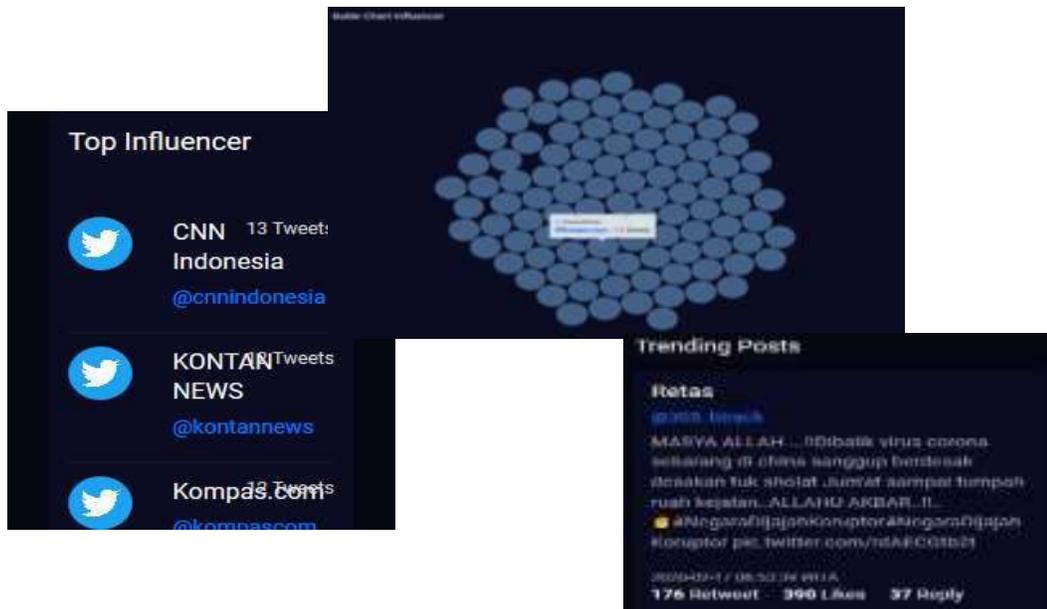
Hasil penelitian ini tidak hanya mengetahui presentase polaritas tweet tapi juga mengetahui informasi-informasi terkait akun (*Top Influencer*) dan tweet (*Top Post*) mana saja yang banyak mendapat respon masyarakat pada media sosial twitter.

Tabel 4. Top 10 Influencer

Top Influencer	Tweets	Top Influencer	Tweets
@cnnindonesia	13	@tribunnews	7
@kontannews	12	@cnbcindonesia	6
@kompascom	12	@kumparan	6
@mag_berita	8	@vibiznews	6
@katatodaycom	7	@jpncom	6

Tabel 5. Top 5 Post

Top Post	Reetweet	Like	Reply
@305_btrack	176	390	37
@officialfirman	91	135	8
@cnnindonesia	54	104	2
@abah_balagado	50	159	3
@yusuf_dumdum	48	1054	81



Gambar 4. Visualisasi Top Influencer dan Top Post

IV. KESIMPULAN

Hasil dari pengamatan sentimen masyarakat melalui media sosial twitter menghasilkan sebuah polaritas sentimen yang menunjukkan opini masyarakat cenderung negatif terhadap topik corona yaitu mencapai 79%, untuk polaritas tweet netral 11% dan positif hanya 10%. Dari hasil pengamatan juga diketahui rata-rata yang menjadi top influencer dan top post yang sering mendapat respon dari masyarakat adalah akun-akun berita. Sehingga dapat disimpulkan bahwa corona memberikan dampak negatif berdasarkan percakapan dan hasil sentiment negatif yang lebih tinggi pada media sosial twitter.

DAFTAR PUSTAKA

- Aditya, B.R., 2015. Penggunaan Web Crawler Untuk Menghimpun Tweets dengan Metode Pre-Processing Text Mining. *JURNAL INFOTEL - Informatika Telekomunikasi Elektronik*, 7(2), p.93.
- Anon 2019. *Annual Report Twitter*. [online] Available at: <<https://investor.twitterinc.com/financial-information/annual-reports/default.aspx>> [Accessed 10 Jan. 2020].
- Berlian, T.F., Herdiani, A. and Astuti, W., 2019. Analisis Sentimen Opini Masyarakat Terhadap Acara Televisi pada Twitter dengan Retweet Analysis dan Naïve Bayes Classifier. 6(2), pp.8660–8669.
- Fanissa, S., Fauzi, M.A. and Adinugroho, S., 2018. Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking | Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, [online] 2(8), pp.2766–2770. Available at: <<http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/1962>>.
- Imam, A. and Fajtriab, H., 2015. Implementasi Text Mining pada Mesin Pencarian Twitter untuk Menganalisis Topik - Topik Terkait ‘KPK dan Jokowi’. *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS 2015*, pp.570–581.
- Indrayuni, E., 2019. Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes. *Jurnal Khatulistiwa Informatika*, 7(1), pp.29–36.
- Informatikalogi.com, 2017. *Text Preprocessing*. [online] Available at: <<https://informatikalogi.com/text-preprocessing/>> [Accessed 25 Jan. 2019].
- Wikipedia, 2019. *Pandemi Korona Virus*. [online] Available at: <https://id.wikipedia.org/wiki/Pandemi_koronavirus_2019–2020> [Accessed 27 Jan. 2020].