

Analisis Komunitas Pengguna Jejaring Sosial Terhadap COVID-19 Di Indonesia

Muhammad Rezky Dwiafian^{#1}, Rizki Amanullah Hakim^{*2}, Ecky Prasetyo Nugroho^{#3}

Abstrak

COVID-19 merupakan satu wabak penyakit yang telah melanda seluruh dunia. Tambahan lagi, wabak penyakit ini telah menjadi pandemik dunia pada 12 Maret 2020. Dengan adanya Twitter yaitu sebuah jejaring sosial yang dimana pengguna bisa melakukan mikroblog seperti mengirimkan dan membaca pesan, masyarakat Indonesia pun memanfaatkan kesempatan tersebut untuk melakukan diskusi dengan kata kunci COVID-19. Dengan menggunakan metode seperti Betweenness dan Closeness Centrality, kita bisa melihat siapa yang paling berpengaruh dalam hal ini kata-kata apa sahaja yang sering dibincangkan dalam perbualan topik mengenai COVID-19.

Kata kunci : COVID-19, Twitter, Betweenness, Closeness, WordCloud, Sentiment, Analysis

I. PENGENALAN

Twitter adalah jejaring sosial dan mikroblog yang memungkinkan penggunanya untuk mengirim dan membaca pesan sebanyak 280 karakter yang dikenal dengan dengan kicauan (tweet). Twitter ini didirikan pada Maret 2006 oleh Jack Dorsey dan diberi julukan pesan singkat dari internet. Twitter di Indonesia sendiri memiliki pengguna aktif sebanyak 19,5 juta pengguna, yang membuat Indonesia sebagai pengguna Twitter terbesar kelima di bawah Inggris dan negara besar lainnya.

COVID-19 adalah sebutan untuk jenis varian virus SARS-CoV-2 yang merupakan pandemi yang terjadi sejak 2019 hingga sekarang. Virus ini menyerang manusia dengan menyerang sistem pernapasan manusia atau sering juga disebut dengan severe acute respiratory disease syndrome (SARS). Virus ini saat ini memiliki hampir 2 juta kasus diseluruh dunia.

Python adalah bahasa pemrograman tingkat tinggi yang dibuat oleh Guido van Rossum pada

tahun 1991. Python saat ini memiliki ribuan library yang digunakan untuk berbagai macam bidang, seperti IoT, Data Science, Research, dan lain sebagainya. Dalam pengembangan riset melalui media sosial terutama pada Twitter. Salah satu library yang sangat umum digunakan adalah library Tweepy. Tweepy adalah sebuah paket atau library dalam bahasa pemrograman Python yang memungkinkan kita melakukan interaksi dengan API yang telah disediakan Twitter dengan mudah.

Application Programming Interface yang digunakan pada project ini adalah Twitter API sendiri. API milik Twitter ini digunakan untuk memudahkan para developer untuk mengambil data dari Twitter dan mengolahnnya. Ada pula menegaskan tujuan proyek kami adalah untuk melakukan analisa opini pengguna media sosial Twitter pada topik "COVID". Kami berencana untuk mendapatkan sentimen, mendapatkan banyak tweet per hari, membandingkan sentimen yang muncul per hari, tingkat influencing dari akun Telkomsel, serta graph dari komunitas yang terhubung dengan topik "covid" tersebut.

II. METODE

NetworkX adalah suatu package atau library pada bahasa pemrograman Python yang berguna untuk mengeksplorasi serta menganalisa jaringan yang terbentuk dan algoritma dari sebuah jaringan. Fitur dari library ini seperti menggambarkan sebuah jaringan, entah *simple graph*, *undirected* dan *directed graph*.

Algoritma jaringan yang dapat diimplementasikan pada library ini seperti shortest paths, betweenness centrality, closeness centrality, dan serupa lainnya.

Pembersihan data dilakukan dengan melakukan lowercase pada teks, menghilangkan

beberapa tanda baca, link shortcut, angka-angka dan spasi newline (\n).

Closeness centrality adalah algoritma yang berguna untuk menghitung rerata jarak suatu node semua node lainnya dalam suatu jaringan [10]. Pengukuran ini menggambarkan kedekatan antara node satu dengan node lainnya. Semakin besar nilai closeness centrality maka node tersebut memiliki tingkat eksistensi paling tinggi dibandingkan dengan node lainnya.

Dalam menilai suatu nilai closeness centrality pada sebuah node dapat menggunakan rumus (1) sebagai berikut.

$$C(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(u,v)} \dots (1)$$

Rumus 1: Closeness Centrality

Penjelasan pada rumus tersebut, huruf u melambangkan sebuah node, huruf n melambangkan jumlah node yang terhubung pada graf mencakup node u . Sedangkan pada $d(u,v)$ adalah jarak terpendek antara node v dan node (u) lainnya.

Betweenness centrality adalah pengukuran dimana bagaimana sebuah informasi tersebut dari suatu node yang akan dikirim kepada jaringan yang lain yang mempunyai jalur yang pendek. Bukan itu juga, metode ini menghitung node yang mana mempunyai jalur yang terpendek. Secara singkatnya, di *Betweenness Centrality* akan menghitung node yang mempunyai nilai *Betweenness Centrality* mempunyai jalur yang terpenting dalam menghubungkan dari satu node ke node yang lain. Tambahan lagi, daripada hasil pengukuran *betweenness centrality* ini, kita dapat mencari tahu node mana yang menjadi penghubung atau mempunyai peranan yang penting dalam menghubungkan antara dua jaringan yang berbeda.

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Rumus 2: Betweenness Centrality

Seperti pada rumus di atas, jalur dari s ke t ialah total jalur yang melalui node v .

Fine-grained sentiment analysis adalah sebuah metode dalam menentukan sentimen/konteks berdasarkan kalimat yang terdapat pada suatu teks. Metode ini memiliki 2 lapisan, yang pertama adalah klasifikasi sentimen: positif dan negatif. Dan yang kedua adalah asosiasi sentimen: untuk menghubungkan sentimen dengan suatu topik atau suatu brand [6].

Selain itu, alur untuk dalam menyelesaikan proyek adalah seperti berikut:



Gambar 1: Alur Penyelesaian Proyek.

Seperti yang ditunjukkan pada *Gambar 1*, metode mempunyai 6 tahapan yaitu:

- Penemuan Topik Project
 - Penemuan topik project kami diawali dengan menentukan topik yang merupakan happening yang terjadi saat ini.
- Persiapan Library
 - Mencari library yang sesuai untuk digunakan dalam mendukung melakukan crawling data, visualisasi data dan analisis data. Library digunakan agar mempersingkat waktu dalam penyelesaian proyek.
- Crawling Data

- Bermula untuk mengambil data daripada API Twitter yang diberikan secara resmi dari Twitter. Dataset yang ada masih belum diolah.
- Pembersihan Data
 - Pembersihan data dilakukan agar data-data yang tidak diperlukan akan dibuang supaya dalam proses menganalisa data dapat dipermudah dan cepat.
- Analisa Data
 - Analisa data yang kami lakukan pada dataset yang telah kami crawl sebelumnya, dilakukan dengan mencari nilai-nilai betweenness centrality dan closeness centrality. Dari pencarian nilai tersebut, maka kami akan melakukan
 - Sebagai tambahan, kami juga meneliti bagaimana sentimen yang terbentuk dari kumpulan dataset kami, membandingkan antara sentimen positif dan negatif, beserta kata yang paling sering digunakan untuk tiap-tiap sentimen.
- Visualisasi
 - Kami melakukan visualisasi data hasil dari analisa dan pengamatan kami menggunakan Gephi yang digenerate menggunakan tools NetworkX.
 - Untuk visualisasi sentimen yang didapatkan, kami menggunakan Word Cloud untuk menampilkannya.

III. DATA CRAWLING

Data crawling adalah cara untuk mendapatkan dataset daripada API Twitter. Dengan adanya library Tweepy yang disediakan oleh Python, kita bisa dengan mudah untuk mendapatkan data dari Twitter dengan parameter yang akan ditentukan.

Untuk parameter yang ditentukan, kami akan memilih dengan ada kata kunci seperti “covid”.

	userid	screenname	retweeted	createdat	cleantext	sentiment
0	1595045810	AnotherAniart	0	2021-03-28 23:59:43	rt klaputrajaya wanita hamil kumpulan berisiko...	0
1	409925087	RDVSPTR	0	2021-03-28 23:59:40	rt putheta cek satpam teman yg tersebar wag w...	1
2	1274974402040741377	Cr07Fans1	0	2021-03-28 23:59:40	rt abawaras patut apresiasi larangan mudik yg ...	0
3	1357708239168217090	RahmanR33500117	0	2021-03-28 23:59:38	rt djlkemenhub kawanlaut kemenhub penerapan a...	0
4	906066821723201536	ftahshw	0	2021-03-28 23:59:32	rt klaputrajaya wanita hamil kumpulan berisiko...	0
...
495	4437918318	akmalhiseftu	0	2021-03-28 23:15:44	rt theghrack salam sy dgn rendah hati doakan k...	0
496	596216213	jamal_bawean	0	2021-03-28 23:15:39	rt drkoko belasan siswa positif covid ikuti pe...	1
497	1210882111973933058	santuypan90	0	2021-03-28 23:15:38	rt segipascal terhubung utbk suka badut suka n...	1
498	834572846072418305	kotapontianak_	0	2021-03-28 23:15:36	informasi data covid kota pontianak data diper...	1
499	125718610046660473	Polsek_Klatara	0	2021-03-28 23:15:25	yuk pahami tujuan vaksinasi covid tconvirjx	0

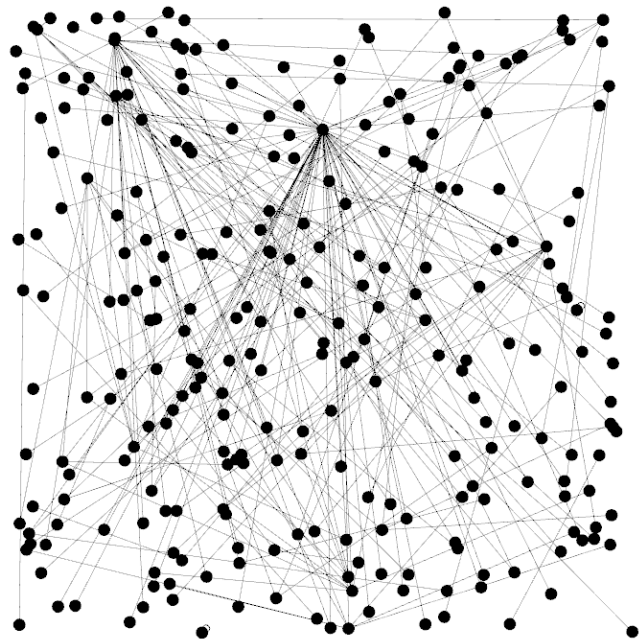
500 rows x 6 columns

Gambar 2: Dataset dengan kata kunci “covid”

Berdasarkan gambar di atas, dataset kami memiliki 500 tweets yang diambil secara acak dari tanggal 1 Maret hingga 29 Maret 2021. Hasil crawling data tersebut, memiliki beberapa atribut seperti id_ tweet, screen name pengguna, isi tweet, dan tanggal created tweet tersebut. Dari data tersebut dilakukan preprocessing, dimana kami tambahkan atribut cleaned_text dan sentiment, dimana masing-masing kolom tersebut berisi isi tweet yang sudah dibersihkan kalimatnya dan penentuan sentimen dari isi tweet yang sudah dibersihkan.

IV. IMPLEMENTASI METODE

Dengan menerapkan parameter yang telah ditentukan, maka visualisasi graph untuk dataset yang diambil dari Twitter adalah seperti berikut:



Gambar 3: Graph dengan kata kunci 'covid' pada Gephi



Gambar 4: Graph dengan kata kunci 'covid' pada NetworkX

Setelah melihat graph tersebut, graph tersebut mempunyai karakteristik *undirected graph*. Hal ini bermaksud, setiap node yang terhubung dengan edges mempunyai tanda arah ke setiap nodes.

Berdasarkan graph tersebut, diketahui bahawa setiap node hanya berpusat pada satu node sahaja. Hal ini bermaksud, setiap pengguna diwakili dengan nodes mempunyai edges yang mempunyai pusat kepada node dengan edges paling banyak. Hal ini kerana, setiap tweets yang ada pada setiap pengguna mempunyai kata kunci yang sama yaitu 'covid'.

Untuk melihat hasil akun dengan betweenness centrality terbesar adalah berikut:

Betweenness Centrality	
KKMPutrajaya	4161.250000
DGHisham	1676.750000
JPenerangan	864.142857
Jchris_68	758.547619
Ahmad_KanekiKen	471.333333
tbgsem	471.333333
KelingMenua8	438.178571
ParamesPolaris	389.178571
lyra_fara	389.178571
JapenSarawak	251.464286

Gambar 5: Hasil Betweenness Centrality

Berdasarkan Gambar 5, kita ketahui nilai betweenness centrality tertinggi adalah 4161.25 pada akun *KKM Putrajaya*. Hal ini bermaksud jika dilihat dari Gambar 2, tentang node yang memiliki edges terbanyak, mempunyai betweenness yang tinggi dimana semua nodes sangat bergantung dengan node tersebut.

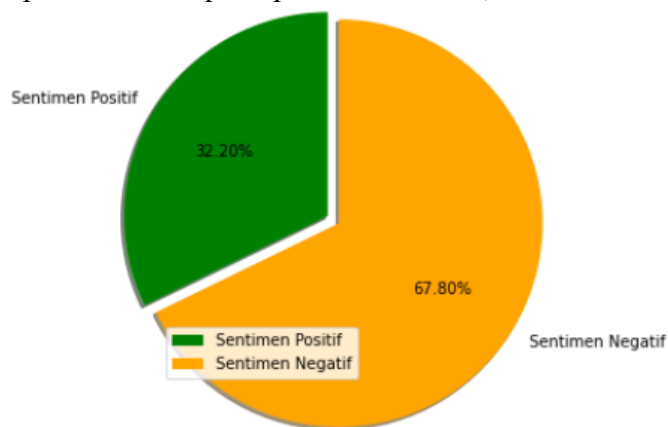
Sedangkan untuk *closeness centrality* tiap akun dalam dataset dapat dilihat pada Gambar 3. Untuk akun yang memiliki skor closeness centrality paling tinggi adalah yang merupakan pemilik skor tertinggi betweenness centrality, menandai bahwa akun tersebut adalah akun yang paling menginfluence pada kata kunci "covid"

Closeness Centrality	
KKMPutrajaya	0.207143
Jchris_68	0.161423
tbgsem	0.155808
Ahmad_KanekiKen	0.155808
lyra_fara	0.149316
ParamesPolaris	0.149316
KelingMenua8	0.149316
AnotherAznTart	0.132725
haziqProperty	0.132725
vouzmeyouz7	0.132725

Gambar 6: Hasil closeness centrality

Untuk hasil analisa penentuan sentiment yang terbentuk di dalam komunitas tersebut, kami menggunakan metode *Fined-grained sentiment analysis* yang bekerja pada tingkat kalimat yang terbentuk. Sentimen yang kami dapatkan kami

representasikan pada pie chart berikut;



Gambar 7: Pie chart sentimen

Untuk penentuan kata-kata positif dan negatif kami menggunakan referensi kalimat untuk digunakan untuk hasil penentuan kami untuk kata-kata positif dan negatif [5]. Visualisasi kami generate menggunakan Word Cloud ini, kami gunakan untuk menampilkan kalimat yang sering muncul pada tiap sentimen positif dan negatif dapat dilihat pada Gambar 8 dan 9.



Gambar 8: Visualisasi kalimat yang sering muncul pada sentimen positif



Gambar 9: Visualisasi kalimat sering muncul pada sentimen negatif

KESIMPULAN

Graf yang terbentuk dari komunitas yang terbentuk dengan menggunakan topik kata kunci ‘covid’ di Indonesia, kita bisa menemukan beberapa informasi mengenai akun yang paling influence, menggambarkan komunitas yang terbentuk serta mendapatkan sentiment yang terbentuk dari topik ‘covid’ pada media sosial Twitter.

Tidak hanya itu, dari kumpulan komunitas tersebut dapat diambil beberapa informasi penting seperti akun yang memiliki tingkat influence paling tinggi pada kata kunci ‘covid’, kemudian bagaimana akun-akun dapat menjadi jembatan bagi akun lain, direpresentasikan menggunakan graf yang terbentuk. Serta bagaimana sentiment yang terbentuk dari kumpulan jejaring sosial tersebut, hingga menganalisa kecenderungan berbagai pendapat pada topik ‘covid’ ini, berputar di masyarakat khususnya netizen twitter di Indonesia.

Saran untuk penelitian kedepannya, dapat dikembangkan lagi suatu model yang dapat membangun sebuah analisis komunitas yang terbentuk dengan beracuan pada sentiment masyarakat, menggunakan metode centrality yang lebih bervariasi untuk menentukan nodes paling besar kemampuan influencenya

DAFTAR PUSTAKA

- [1] B. Susanto, H. Lina, and A. R. Chrismanto, "Penerapan Social Network Analysis dalam Penentuan Centrality Studi Kasus Social Network Twitter," *Jurnal Informatika*, vol. 8, no. 1, 2012.
- [2] B. A. Huberman, D. M. Romero, and F. Wu, "Social Networks that Matter: Twitter Under the Microscope," *SSRN Electronic Journal*, 2008.
- [3] P. Kallas, "Top 15 Most Popular Social Networking Sites and Apps [2021] @DreamGrow," *DreamGrow*, 04-Jan-2021. [Online]. Available: <https://www.dreamgrow.com/top-15-most-popular-social-networking-sites/>. [Accessed: 01-Mar-2021].
- [4] Eka A, Babyta, "STRATEGI PENENTUAN PERINGKAT BRAND BERDASARKAN JEJARING SOSIAL PERCAKAPAN PADA MEDIA SOSIAL TWITTER (STUDI KASUS PADA TELKOMSEL, XL DAN INDOSAT)". *Telkom University Open Library Repositories*, 2016.
- [5] C. Rio, "Analisis Sentimen ID [2019]" [Online]. Available:<https://github.com/riochr17/Analisis-Sentimen-ID>. [Accessed: 28-Mar-2021].
- [6] Li, Wei, "Coarse Grained vs Fine-Grained Sentiment Analysis," [Online]. Available:<https://www.linkedin.com/pulse/coarse-grained-vs-fine-grained-sentiment-analysis-wei-li/>. [Accessed: 1-Apr-2021].
- [7] Labs, HK "Bermain Twitter dengan R Part 3 - Pengantar Sentiment Analysis" [Online]. Available:<https://hkalabs.com/bermain-twitter-dengan-r-part-3-pengantar-sentiment-analysis/>. [Accessed: 29-Mar-2021].
- [8] Labs, HK "Bermain Twitter dengan R Part 4 - Pengantar Sentiment Analysis" [Online]. Available:<https://hkalabs.com/bermain-twitter-dengan-r-part-4-pengantar-algoritma-klasifikasi-sentimen/>. [Accessed: 28-Mar-2021].
- [9] Labs, HK "Bermain Twitter dengan R Part 5 - Pengantar Sentiment Analysis" [Online]. Available:<https://hkalabs.com/bermain-twitter-dengan-r-part-5-sentiment-analysis-gojek-vs-grab-vs-uber/>. [Accessed: 2-Apr-2021].
- [10] Azkia, Wildan Al, "Sentiment Analysis in R" [Online]. Available:<https://medium.com/@wildanalazkia/sentiment-analysis-in-r-8924856e4b62>. [Accessed: 2-Apr-2021].
- [11] Septiar, Irvan, "Introduction Sentiment Analysis Mudah" [Online]. Available:<https://medium.com/@irvanseptiar/introduction-sentiment-analysis-mudah-5785f88e435d>. [Accessed: 1-Apr-2021].
- [12] Twitter, "Intro to Tweet JSON" *Twitter Developer*. [Online]. Available:<https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/overview/intro-to-tweet-json#tweetobject>. [Accessed: 3-Apr-2021].