Penerapan *Social Network Analysis* dalam Menentukan User Paling Berpengaruh pada Topik #VaksinCOVID19 di Twitter

Mila Putri Kartika Dewi1   
Teknik Informatika  
Telkom UniversityBandung, Indonesia  
miladewi@student.telkomuniversity.ac.idAlvi Rahmy Royyan2  
Teknik Informatika  
Telkom UniversityBandung, Indonesia  
alvirahmyr@student.telkomuniversity.ac.id

*Abstract*— Twitter merupakan salah satu layanan jejaring sosial yang populer, dimana penggunanya dapat memposting *tweet* yang mereka tulis tentang kegiatan maupun fenomena yang sedang terjadi, sehingga *tweet* tersebut dapat menjadi topik pembicaraan antar pengguna Twitter. Dari berbagai interaksi atau hubungan yang terjadi antar pengguna, maka akan membentuk sebuah jaringan, dimana jaringan tersebut dapat berupa jaringan sosial dan penyebaran informasi terkait postingan *tweet* pada topik pembicaraan tertentu. Penelitian ini akan menerapkan metode Social Network Analysis (SNA) dengan menentukan nilai *centrality* menggunakan perhitungan *Degree Centrality.* Studi kasus pada penelitian ini akan mengangkat topik yang berkaitan dengan vaksinasi untuk Indonesia, tujuanya adalah untuk menentukan *user* paling berpengaruh dalam persebaran informasi, sehingga informasi yang akan dihasilkan dapat digunakan untuk kegiatan sosialisasi pemerintah Indonesia dalam program Vaksinasi Covid-19.

Keywords—twitter, social network analysis, *degree centrality,* vaksinasi.

# Pendauluan

Media sosial adalah sebuah media yang memungkinkan pengguna dapat membuat dan berbagi isi dalam jaringan atau komunitas sosial di dunia virtual. Era modern saat ini, sebuah situs jejaring sosial tak bisa lepas dari kehidupan sehari-hari, karena seseorang dapat terhubung dengan orang-orang yang ada di belahan bumi lainnya secara mudah dan cepat, dengan kata lain mendekatkan mereka yang jauh dari pandangan. Tren menggunakan media sosial sangat meningkat seperti di salah satu web *microblog* yang terkenal, yaitu Twitter. Web *microblog* adalah web yang memungkinkan penggunanya untuk bertukar elemen kecil dari konten seperti kalimat pendek, gambar individu, atau tautan video [1]. Menurut [2], jumlah pengguna aktif Twitter telah mencapai 192 juta dengan 500 juta cuitan yang dikirim tiap hari, hal ini jelas menunjukkan banyaknya topik percakapan menarik yang terjadi bersamaan. Untuk menemukan sebuah topik, biasanya para pengguna Twitter menggunakan *hashtag* atau tagar (#) untuk mengidentifikasikan dan memudahkan pengguna jika ingin mencari topik tertentu yang ingin dibaca. Tidak heran, saat ini banyak orang menjadikan Twitter sebagai salah satu media untuk berekspresi dan mengeluarkan pendapat atas suatu masalah yang terjadi di lingkungan sekitarnya, serta ruang berinteraksi dengan orang yang menurut mereka anggap penting atau menarik.

Sama halnya dengan media sosial lain, Twitter juga menyediakan fitur untuk penggunanya agar dapat berinteraksi dengan cara mengikuti (*follow*) siapapun yang dikehendakinya dan juga berhenti mengikuti mereka (*unfollow*) apabila dirasa mengganggu. Interaksi yang terjadi antara satu akun dengan akun pengguna lain, belum tentu menunjukkan bahwa mereka tahu dengan baik tentang seseorang tersebut. Oleh karena itu, dari berbagai macam interaksi atau hubungan yang terjadi antar pengguna Twitter, akan terbentuk sebuah jaringan, dimana jaringan tersebut dapat berupa jaringan pertemanan, jaringan penyebaran informasi terkait postingan *tweet* dengan topik pembicaraan tertentu, maupun jaringan lainnya. Studi terkait hal ini telah banyak dikembangkan dengan menggunakan metode *Social Network Analysis*  (SNA).

*Social Network Analysis* (SNA) secara teori dapat dideskripsikan sebagai ilmu yang mempelajari hubungan antar individu dengan bantuan teori graf dan berfokus pada pembelajaran struktur sosial yang terbentuk. Dimana dalam struktur jaringan yang terbentuk terdiri dari *node* dan *edge*. *Node* pada jaringan dapat berupa seorang individu/aktor, serta *edge* merupakan hubungan/interaksi yang terjadi antar *user* lain atau sekelompok *user*. Proyek kali ini akan mengembangkan penelitian terkait *Social Network Analysis* (SNA) dengan studi kasus yang diambil dari Twitter dengan topik mengenai vaksinasi untuk Indonesia. Penelitian ini dilakukan dengan tujuan untuk mendeskripsikan *user* yang paling berpengaruh pada jejaring sosial berdasarkan topik yang berkaitan dengan vaksinasi untuk Indonesia di Twitter, dimana dalam penerapannya akan menentukan nilai *centrality* dengan metode *Degree Centrality*. Tahapan yang akan dilakukan pada penelitian ini adalah melakukan *crawling* data pada Twitter (*data collection*), menganalisis data (*data analysis*) dan melakukan visualisasi pada data. Dari proses yang telah dilakukan diharapkan informasi yang dihasilkan mungkin dapat digunakan untuk kegiatan sosialisasi program pemerintah Indonesia, yaitu vaksinasi Covid-19. Dengan membangun sistem ini pula, sangat dimungkinkan untuk kemudian dianalisis terkait dengan berbagai segi atau bidang ilmu lain terhadap jejaring sosial yang terbentuk.

Jurnal ini akan disusun dengan struktur pada bagian II menjelaskan tentang studi terkait, bagian III terkait metodologi yang digunakan dalam penelitian, bagian IV memaparkan hasil dari penelitian dan pada bagian V berisi kesimpulan serta saran dari penulis.

# Studi terkait

*Online social networks* dan analisis media sosial adalah bidang penelitian populer masa kini dalam ilmu jaringan. Fokus utama dalam penelitian jejaring sosial dapat berupa prediksi tautan [3], prediksi koneksi sosial [4], prediksi kepribadian untuk para pengguna *microblog* [5], bahkan seperti analisis sentimen dan *opinion mining* [6]. Pada penelitian Syed K. Tanbeer et al [7], memperkenalkan gagasan baru tentang teman yang berpengaruh untuk database jejaring sosial dan mempresentasikan DIFSoN model untuk menemukan teman (atau entitas) yang berpengaruh di jaringan sosial.

Kimura et al [8], telah mempresentasikan model kaskade untuk menemukan node yang paling berpengaruh dalam penyebaran informasi di jaringan sosial. Menggunakan struktur algoritma *SR-community* mereka mencari fitur-fitur yang memainkan peran penting dalam mengidentifikasi node yang berpengaruh. Eksperimen mereka menggunakan media sosial dengan jaringan data yang besar membuktikan struktur *SR-community* dapat berkorelasi lebih kuat dengan *greedy solution* daripada algoritma Newman dan Leicht *community*.

# Metodologi

Penentuan *user* paling berpengaruh pada media sosial Twitter dengan topik yang berkaitan dengan vaksinasi untuk Indonesia akan dilakukan beberapa tahapan pengerjaan, dari proses pengumpulan data, *preprocessing* data, menganalisis data, dan visualisasi data. Berikut merupakan penjelasan dari setiap proses yang akan dilakukan:

## Crawling

Untuk mengambil data dari Twitter, digunakan sebuah teknik *crawling* data otomatis. *Crawling* adalah proses pengambilan data yang berukuran besar maupun kecil yang berada di dalam halaman web yang dapat disimpan di penyimpanan lokal dan data diambil berdasarkan sejumlah kata kunci yang dicari [9]. Proses *crawling* data pada Twitter dilakukan dengan menggunakan Twitter *Application Programming Interface* (API). Twitter API didapatkan dari akun *developer* Twitter yang telah mendapatkan autentifikasi, serta proses *crawling* akan dilakukan dengan menggunakan bahasa pemrograman Python. Data yang akan diambil sebanyak 3179 data dengan *keyword* #VaksinNasional, #VaksinUntukIndonesia, #VaksinDorongPemulihanRI, #VaksinUntukKita, dan #VaksinPulihkanNegeri. Deskripsi data yang terkumpul berupa *username* dari setiap akun, *tweet* dengan *hashtag* atau tagar yang berkaitan dengan vaksinasi untuk Indonesia, serta *username* dari *mention* yang ada pada setiap *tweet*.

## Pre-processing

*Pre-processing* merupakan tahapan yang paling penting dilakukan sebelum data diproses lebih lanjut. Umumnya data yang telah berhasil di *crawling* masih bersifat *noisy,* seperti terdapatnya data yang duplikat serta data kosong, sehingga perlu dilakukan penghapusan pada data tersebut agar tidak mengganggu proses pengolahan data pada tahap berikutnya.

## Degree Centrality

*Degree Centrality* merupakan salah satu metode perhitungan dalam menentukan sentralitas pada Social Network Analysis (SNA). *Degree Centrality* adalah indeks paparan yang mewakili seberapa baik *node* individu terhubung ke jaringan menggunakan jumlah kontak langsung dari node [10]. Berdasarkan teorinya, *Degree Centrality* merupakan jumlah koneksi yang dimiliki oleh setiap *node*, sehingga jumlah koneksi yang lebih banyak akan berpengaruh pada jaringan yang dibangun. Dari *graph* yang terbentuk dibedakan menjadi 2 jenis yaitu *directed graph* (graf berarah / memiliki arah) dan *undirected graph* (graf tidak terarah / tidak memiliki arah), sehingga dalam penelitian ini akan digunakan *undirected graph* sebagai representasi dari hasil pengolahan data. Dalam menentukan nilai *Degree Centrality* atau bobot suatu *node* dapat dilakukan perhitungan dengan bentuk normalisasi sebagai berikut:

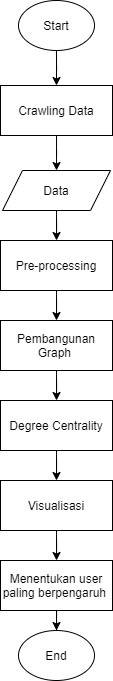


Dimana:

* = *Degree Centrality* *node* ke- i.
* = Jumlah *edge* yang terbentuk pada *node* i.
* n = Jumlah *node*.

# Rancangan sistem

Rancangan sistem yang dikembangkan dalam membangun sistem *Social Network Analysis* (SNA) untuk menentukan *user* paling berpengaruh terhadap postingan *tweet* dengan *hashtag* yang berkaitan dengan vaksinasi untuk Indonesia adalah sebagai berikut:



1. Rancangan Sistem

Dari penjelasan *flowchart* diatas pada proses *crawling* data akan digunakan data sebanyak 3179 dengan spesifikasi atribut seperti *username*, *tweet*, *mention*. Kemudian data akan dilakukan *pre-processing* dengan menghilangkan data yang teridentifikasi duplikat dan data kosong yang bernilai NaN. Data yang telah siap diproses kemudian akan ditentukan *node* dan *edge* nya, pada penelitian ini akan menggunakan *usernnmae* sebagai *node* dan *mention* sebagai *edge*, kemudian data akan dibangun dengan menggunakan graph tidak berarah, setalah graf terbentuk kemudian akan menentukan nilai *degree* berdasarkan perhitungan pada rumus (1.1), nilai *degree* ditentukan untuk menentukan sentralitas pada setiap akun dengan jumlah interaksi yang paling besar. Proses akhir yang dilakukan oleh sistem adalah memvisualisasikan hasil dari data yang telah dibangun menggunakan metode *graph*.

## Data

Berikut merupakan contoh tampilan dari data yang telah berhasil dilakukan *crawling*:

1. Dataset #VaksinCOVID19

|  |  |  |
| --- | --- | --- |
| Usernmae | Tweet | Mention |
| wakilbupatipml | Vaksin tahap pertama sebelumnya sudah dilakukan Bupati Agung, Rabu, (10/3/2021) yang lalu di Puskesmas Kebondalem @PemkabPemalang #VaksinasiNasional #VaksinCovid19 https://t.co/0O0H8MjqK8' | PemkabPemalang |
| wakilbupatipml | Bupati Pemalang,\xc2\xa0 @muktiagungwbw, hari ini, Kamis (25/3/2021) melakukan vaksin covid-19 tahap kedua di Pasar Petarukan #VaksinCovid19 #pemalang @ganjarpranowo https://t.co/raXfbDa7I7' | muktiagungwbw |
| wakilbupatipml | Bupati Pemalang,\xc2\xa0 @muktiagungwbw, hari ini, Kamis (25/3/2021) melakukan vaksin covid-19 tahap kedua di Pasar Petarukan #VaksinCovid19 #pemalang @ganjarpranowo https://t.co/raXfbDa7I7' | ganjarpranowo |
| papipungky | Numpang tanya, kalau cekdi https://t.co/MwhVn5Wq7L dan dapatnya begini, apakah bisa langsung ke faskes vaksinasi covid atau gimana ya? Terima kasih sebelumnya. @pedulilindungi #VaksinCovid19 #VaksinasiNasional https://t.co/KI4vAjares' | pedulilindungi |
| RILISonline | Dia mengaku, tangannya baru bisa digerakkan pada Kamis (24/3/2021). Tetapi, datang lagi gejala lainnya.\n@KemenkesRI #VaksinCovid19 \n https://t.co/uEijMQoc6p' | KemenkesRI |

# Hasil Penelitian

Didapatkan hasil penelitian sebagai berikut:

## Degree Centrality

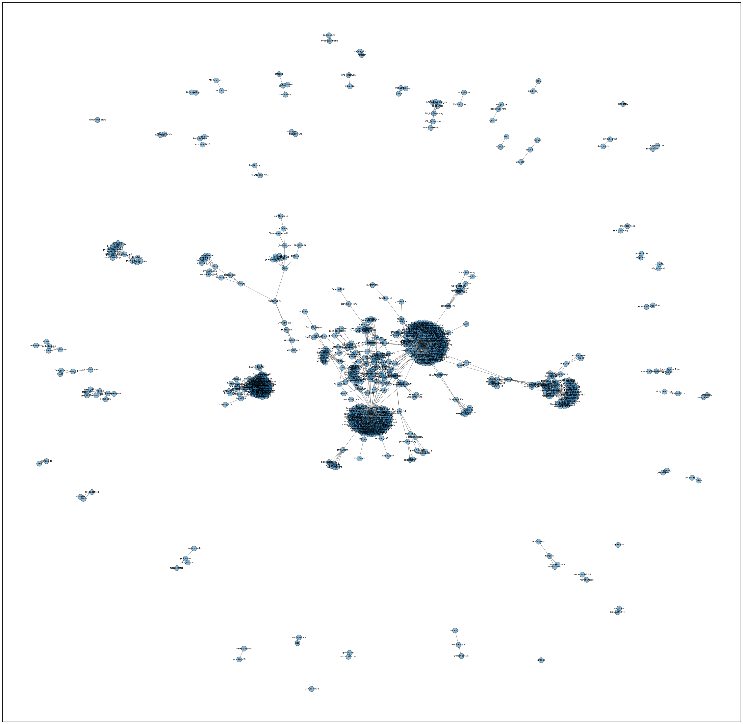
Berikut top 5 *influential user* dari data Twitter berkaitan dengan vaksinasi untuk Indonesia yang telah diproses menggunakan metode *Degree Centrality*:

1. Influential User

|  |  |  |
| --- | --- | --- |
| ***No*** | ***Username*** | ***Degree Centrality*** |
| 1 | KemenkesRI | 158 |
| 2 | sikecilmarmut | 125 |
| 3 | Vaksin\_Kita | 66 |
| 4 | ontynad | 65 |
| 5 | Vaksin\_Update | 65 |

## Visualisasi

Visualisasi yang dilakukan pada data yang telah diproses menggunakan bahasa pemrograman Python dengan memanfaaatkan *tools* NetworkX untuk menggambarkan jaringan yang terbentuk secara kompleks. Berikut merupakan hasil visualisasi data yang digunakan:



1. Visualisasi Jaringan

##### Acknowledgment

Dalam penelitian ini, dapat menunjukkan bahwa metode *Degree Centrality* dapat menemukan *influential user* yang berhubungan dengan *hashtag* atau tagar yang berkaitan dengan vaksinasi untuk Indonesia di Twitter. Berdasarkan data yang digunakan, menghasilkan akun @KemenkesRI sebagai *influential user* dalam tagar-tagar vaksinasi khususnya untuk Indonesia. Apabila ditelusuri lebih lanjut, akun @KemenkesRI memiliki banyak relasi dengan akun lain dan akun tersebut selalu giat menyebarkan informasi mengenai program vaksinasi untuk Indonesia mengingat akun @KemenkesRI ini merupakan akun resmi milik Kementerian Kesehatan Republik Indonesia.

Pada penelitian selanjutnya, dapat diteliti menggunakan berbagai macam metode lainnya dengan penambahan fitur lain untuk menentukan pengguna paling berpengaruh di dalam sebuah jaringan media sosial.

##### References

[1] A. M. Kaplan and M. Haenlein, “The early bird catches the news: Nine things you should know about micro-blogging,” *Bus. Horiz.*, vol. 54, no. 2, pp. 105–113, 2011, doi: 10.1016/j.bushor.2010.09.004.

[2] Ying Lin, “10 Twitter Statistics Every Marketer Should Know in 2021 [Infographic],” Jan. 25, 2021. https://id.oberlo.com/blog/twitter-statistics (accessed Mar. 01, 2021).

[3] D. Liben-Nowell and J. Kleinberg, “The link-prediction problem for social networks,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 58, no. 7, 2007, doi: 10.1002/asi.20591.

[4] S. Utz and J. Jankowski, “Making ‘Friends’ in a Virtual World: The Role of Preferential Attachment, Homophily, and Status,” *Soc. Sci. Comput. Rev.*, vol. 34, no. 5, 2016, doi: 10.1177/0894439315605476.

[5] Q. Zu, B. Hu, N. Gu, and S. Seng, “Human Centered Computing: First International Conference, HCC 2014 Phnom Penh, Cambodia, November 27-29, 2014 Revised Selected Papers,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 8944, doi: 10.1007/978-3-319-15554-8.

[6] G. Petz, M. Karpowicz, H. Fürschuß, A. Auinger, V. Stříteský, and A. Holzinger, “Reprint of: Computational approaches for mining user’s opinions on the Web 2.0,” *Inf. Process. Manag.*, vol. 51, no. 4, 2015, doi: 10.1016/j.ipm.2014.07.011.

[7] S. K. Tanbeer, C. K. S. Leung, and J. J. Cameron, “DIFSoN: Discovering influential friends from social networks,” 2012, doi: 10.1109/CASoN.2012.6412389.

[8] M. Kimura, K. Yamakawa, K. Saito, and H. Motoda, “Community analysis of influential nodes for information diffusion on a social network,” 2008, doi: 10.1109/IJCNN.2008.4633974.

[9] S. Chakrabarti, “Crawling the Web,” *Min. Web*, pp. 17–43, 2003, doi: 10.1016/b978-155860754-5/50003-3.

[10] M. De Laat, V. Lally, L. Lipponen, and R. J. Simons, “Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis,” *Int. J. Comput. Collab. Learn.*, vol. 2, no. 1, 2007, doi: 10.1007/s11412-007-9006-4.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**