

## JARINGAN KOMUNITAS BERBASIS SIMILARITAS TOPIK BAHASAN DAN EMOSI UNTUK MENGIDENTIFIKASI PERILAKU PENGGUNA TWITTER

Apriantoni<sup>1</sup>, Diana Purwitasari<sup>\*2</sup>, Agus Budi Raharjo<sup>3</sup>

<sup>1,2,3</sup>Institut Teknologi Sepuluh Nopember, Surabaya  
Email: <sup>1</sup>apriantoni.206025@mhs.its.ac.id, <sup>2</sup>diana@if.its.ac.id, <sup>3</sup>agus.budi@its.ac.id  
\* penulis korespondensi

(Naskah masuk: 25 Mei 2022, diterima untuk diterbitkan: 28 Februari 2023)

### Abstrak

Pandemi COVID-19 menyebabkan situasi krisis yang berdampak pada perubahan perilaku pengguna Twitter terkait pengalaman distres publik. Perubahan perilaku positif bisa berdampak positif. Namun, perubahan perilaku negatif bisa menjadi masalah jika terjadi secara masif, seperti meningkatnya kecemasan pengguna. Oleh karena itu, mengeksplorasi hubungan antara perilaku dan jaringan komunitas pengguna sangat penting untuk menemukan implikasi pandemi COVID-19 terhadap perubahan perilaku pengguna Twitter. Penelitian ini berkontribusi dalam mengidentifikasi perubahan perilaku pengguna berdasarkan model ekstraksi perilaku kolektif pada aktivitas *tweet* temporal. Mekanisme ini menggunakan topik bahasan dan emosi sebagai variabel ekstraksi untuk menghasilkan jaringan perilaku pengguna. Kemudian, jaringan perilaku tersebut dimodelkan dengan algoritma *DeepWalk Network Embeddings* untuk memetakan hubungan kedekatan perilaku antar pengguna dan *Density Peak Clustering Algorithm* untuk mengelompokkan komunitas pengguna berdasarkan kesamaan perilaku yang kuat. Dari analisis 121 pengguna aktif, periode sebelum COVID-19 memiliki 98 pengguna representatif yang didominasi oleh 33% perilaku komunitas terkait aktivitas pribadi dengan emosi senang. Di sisi lain, periode setelah COVID-19 memiliki 54 pengguna representatif yang didominasi oleh 65% perilaku komunitas terkait kesehatan dengan emosi marah. Perubahan perilaku kedua periode tersebut dipengaruhi oleh transisi pola jaringan terdistribusi ke pola jaringan *clique graph*, sehingga sentralisasi penyebaran informasi mempengaruhi potensi peningkatan perubahan perilaku pengguna pada jaringan komunitas. Hasil ini dapat digunakan untuk mengurangi potensi penyebaran perilaku negatif dengan memanfaatkan komunitas yang memiliki pengaruh perilaku positif dikalangan pengguna Twitter.

**Kata kunci:** COVID-19, Topik Bahasan, Emosi, Komunitas Pengguna, Perubahan Perilaku, Perilaku Kolektif

### COMMUNITY NETWORK BASED ON THE SIMILARITY OF DISCUSSION TOPICS AND EMOTIONS FOR IDENTIFYING TWITTER USER BEHAVIOR

#### Abstract

The COVID-19 pandemic caused a crisis that impacted behavior changes of Twitter users related to public distress experiences. Positive behavior changes could have a positive impact. However, negative behavior changes could have problems if it occur massively, such as increased user anxiety. Therefore, exploring the relationship between behavior and user community in the social networks is very important to find the implication of the COVID-19 pandemic on behavior changes of Twitter users. This study contributes to identify user behavior changes based on the collective behavior extraction model on temporal tweet activities. This mechanism used discussion topics and emotions as extraction variables to generate user behavior network. Then, the behavioral network was modeled by the DeepWalk Network Embeddings algorithm to map the behavioral closeness relationship between users and the Density Peak Clustering Algorithm to group user communities with strong behavioral similarities. Based on the analysis of 121 active users, before the COVID-19 period had 98 representative users, who were dominated by 33% of community behavior related to personal activities with happy emotions. On the other hand, after the COVID-19 period, 54 representative users were dominated by 65% of community behavior related to health with anger. Behavior changes in both periods are influenced by the transition from a distributed network pattern to a clique graph network pattern, so the centralization of information dissemination could affect the potential for increasing user behavioral changes in the community network. These findings could be used to reduce the potential for spreading negative behavior by leveraging communities with positive behavior influence among Twitter users.

**Keywords:** COVID-19, Discussion Topic, Emotion, User Community, Collective Behavior, Behavior Changes

## 1. PENDAHULUAN

Pandemi COVID-19 menyebabkan situasi krisis yang meningkatkan emosi marah pengguna Twitter terkait pengalaman distress publik (Basiri et al., 2021). Transisi emosional ini dapat mempengaruhi perilaku individu karena tingginya pertukaran informasi dalam jaringan sosial. Perubahan perilaku yang positif, seperti optimisme terhadap penyakit dan semangat untuk sembuh, dapat memberikan dampak positif. Akan tetapi, perubahan perilaku negatif dapat meningkatkan isu gangguan psikologis secara masif, seperti rasa kecemasan, ketakutan, atau kehilangan semangat. Kondisi ini juga dapat mempengaruhi perilaku pengguna lain dalam jaringan komunitas yang sama (Jaidka et al., 2021). Selain itu, implikasi dari interaksi pengguna yang meluas, baik dalam strata individu ataupun komunitas, menyebabkan informasi mudah menyebar pada jaringan komunitas lain. Hal ini dapat meningkatkan potensi perubahan perilaku pengguna pada komunitas lain (Cresci et al., 2020). Oleh karena itu, menggali hubungan antara pandemi COVID-19 dan perilaku pengguna menjadi hal yang penting untuk mengidentifikasi bagaimana implikasinya terhadap perubahan perilaku pengguna Twitter dalam jaringan komunitas.

Implikasi perilaku pengguna dalam jaringan komunitas sosial, misalnya terkait topik bahasan dan emosi, sangat penting untuk dipelajari karena dapat mempengaruhi penyebaran informasi secara dinamis (Gutierrez et al., 2021). Misalnya, pengguna aktif sering melakukan aktivitas *tweet*, berkomentar, atau membalas konten terkait topik kelangkaan fasilitas medis dengan perasaan cemas. Konten *tweet* tersebut kemudian menyebar di beranda temannya, sehingga mereka yang terkadang melihat konten tersebut merasa setuju dengan argumennya. Hal ini dapat dipengaruhi oleh adanya persepsi yang sama, misalnya dalam hal topik bahasan atau emosi. Kondisi ini dapat mempengaruhi stigma teman-temannya untuk membuat konten terkait topik kelangkaan fasilitas medis dengan perasaan cemas yang sama. Oleh karena itu, keterkaitan antara topik bahasan dan emosi dalam jaringan komunitas perlu dianalisis karena hal ini dapat mempengaruhi perubahan perilaku pengguna lain secara luas.

Beberapa penelitian saat ini telah membahas analisis perilaku pengguna pada interaksi jaringan sosial. Misalnya, penelitian yang mengekstraksi perilaku kolektif pengguna berdasarkan similaritas topik bahasan (Li et al., 2018). Penelitian tersebut menggunakan algoritma *DeepWalk* untuk memetakan kesamaan perilaku antar pengguna dan algoritma *Density Peak Clustering Algorithm* (DPCA) untuk melakukan klusterisasi perilaku pengguna. Eksperimen tersebut hanya berfokus pada topik bahasan sebagai variabel ekstraksi perilaku, sehingga hasil klusterisasi perilaku kolektif terbatas pada cakupan topik bahasan terkait. Penelitian lain mengidentifikasi perilaku komunitas pengguna berdasarkan kategorisasi emosi pengguna (Azer et al.,

2021). Eksperimen tersebut menemukan bahwa emosi dapat mempengaruhi pembentukan sembilan tipe perilaku berbeda. Akan tetapi, penelitian tersebut tidak menganalisis perilaku pengguna pada aktivitas *tweet* temporal. Oleh karena itu, ekstraksi perilaku kolektif berdasarkan aktivitas *tweet* temporal perlu dilakukan untuk mengidentifikasi bagaimana pengaruhnya terhadap perubahan perilaku pengguna dalam struktur jaringan komunitas.

Penelitian ini bertujuan untuk mengidentifikasi perubahan perilaku pengguna Twitter berdasarkan model ekstraksi perilaku kolektif pada aktivitas *tweet* temporal. Linier dengan hal tersebut, kontribusi dari penelitian ini didefinisikan sebagai berikut. Pertama, proses ekstraksi perilaku kolektif pada penelitian sebelumnya hanya berfokus pada variabel topik bahasan. Padahal, penelitian lain menemukan bahwa variabel emosi dapat menjadi faktor penting dalam mempengaruhi perilaku pengguna (Gutierrez et al., 2021). Oleh karena itu, perlu adanya analisis terkait kombinasi topik bahasan dan emosi sebagai variabel ekstraksi perilaku kolektif untuk meningkatkan kesamaan perilaku pengguna. Kedua, beberapa studi sebelumnya tidak mengeksplorasi dampak perubahan perilaku pengguna pada struktur jaringan komunitas berdasarkan aktivitas *tweet* temporal. Misalnya, bagaimana perilaku komunitas secara kontinu dapat mempengaruhi perubahan perilaku pengguna lain, atau sebaliknya. Memahami hubungan perilaku dalam jaringan komunitas *online* dapat membantu menemukan bagaimana implikasi interaksi antar pengguna pada aktivitas *tweet* secara temporal dapat memengaruhi perubahan perilaku pengguna lain.

Penelitian ini menganalisis perubahan perilaku pengguna Twitter menggunakan model ekstraksi perilaku kolektif pada jaringan sosial. Topik bahasan dan emosi dikombinasikan sebagai variabel ekstraksi untuk meningkatkan similaritas perilaku pengguna. Hasil kedua variabel tersebut selanjutnya dimodelkan dengan algoritma *DeepWalk* untuk memetakan kedekatan perilaku antar pengguna dan algoritma DPCA untuk mengelompokkan komunitas pengguna berdasarkan kesamaan perilaku yang dimiliki. Hasil ekstraksi perilaku kemudian dianalisis menggunakan jaringan berbasis kesamaan perilaku pada aktivitas *tweet* temporal untuk menemukan dominasi perilaku pada setiap komunitas pengguna. Hasil penelitian ini diharapkan dapat digunakan untuk mengidentifikasi perubahan perilaku berdasarkan klusterisasi perilaku pengguna Twitter selama masa pandemi COVID-19.

Garis besar makalah ini disusun sebagai berikut. Bagian 2 membahas penelitian sebelumnya terkait dengan mekanisme ekstraksi perilaku pengguna pada jaringan sosial. Bagian 3 menjelaskan metode yang diusulkan untuk mengekstraksi perilaku kolektif pengguna berdasarkan similaritas topik bahasan dan emosi. Bagian 4 memaparkan hasil identifikasi perubahan perilaku pengguna pada periode sebelum dan setelah COVID-19 berdasarkan hasil klusterisasi

perilaku pengguna. Terakhir, bagian 5 menguraikan kesimpulan dan peluang penelitian di masa depan.

## 2. TINJAUAN PUSTAKA

Ekstraksi perilaku pada jaringan sosial biasanya terkendala dengan tingginya dimensi dan variasi data. Beberapa penelitian telah melakukan identifikasi perilaku dengan teknik ekstraksi perilaku kolektif untuk mengatasi hal tersebut. Penelitian sebelumnya melakukan ekstraksi perilaku kolektif pengguna berdasarkan similaritas topik bahasan. Penelitian tersebut menunjukkan bahwa ekstraksi perilaku kolektif dapat digunakan untuk memetakan kesamaan perilaku dalam komunitas pengguna Twitter secara representatif (Li et al., 2018). Penelitian lain mengelompokkan perilaku pengguna di beberapa wilayah berdasarkan aktivitas *tweet* secara temporal. Topik bahasan dijadikan sebagai variabel untuk proses ekstraksi kesamaan perilaku setiap pengguna. Kemudian, metode Hidden Markov Models (HMMs) digunakan untuk mengekstraksi kesamaan perilaku pada setiap wilayah. Studi tersebut menghasilkan informasi bahwa similaritas topik bahasan dapat mempengaruhi perubahan perilaku kolektif pengguna di setiap wilayah yang berbeda (Bello-Orgaz et al., 2020). Penelitian selanjutnya membahas ekstraksi perilaku kolektif berdasarkan reaksi pengguna selama pandemi COVID-19. Studi tersebut menunjukkan bahwa komunitas pengguna dengan similaritas sentimen yang tinggi cenderung memiliki perubahan perilaku yang sama. Hal tersebut disebabkan adanya korelasi positif antara kesamaan sentimen dengan kesamaan perilaku pengguna (Cresci et al., 2020). Ketiga studi tersebut menunjukkan bahwa ekstraksi perilaku kolektif dapat menjadi solusi dalam proses identifikasi perilaku berdasarkan data *tweet* dengan dimensi dan variasi yang tinggi.

Dalam teknik ekstraksi perilaku, topik bahasan dapat dijadikan sebagai variabel identifikasi perilaku pengguna. Pemodelan topik adalah pendekatan yang digunakan untuk mengekstraksi topik bahasan dari sebuah dokumen. Di antara metode yang ada, salah satu yang paling umum digunakan adalah *Latent Dirichlet Allocation* (LDA) (Costa & Ortale, 2021). LDA menggunakan model Bayesian Hirarki untuk memetakan setiap topik bahasan sebagai distribusi probabilitas kata ke dalam kamus virtual. Evaluasi metode tersebut dilakukan dengan skor koherensi untuk mengukur relevansi semantik antara kata dengan topik terkait. Beberapa penelitian sebelumnya telah berfokus pada analisis perilaku berdasarkan proses pemodelan topik pada jaringan sosial. Misalnya, penelitian yang melakukan identifikasi isu-isu dominan selama wabah COVID-19 (Basiri et al., 2021), identifikasi perubahan emosi pengguna berdasarkan topik bahasan pada situasi krisis (Garcia & Berton, 2021), dan identifikasi dominan topik yang mempengaruhi partisipasi dan diskusi aktif pengguna (Mohammed & Ferraris, 2021). Ketiga studi tersebut menunjukkan bahwa mekanisme pemodelan topik

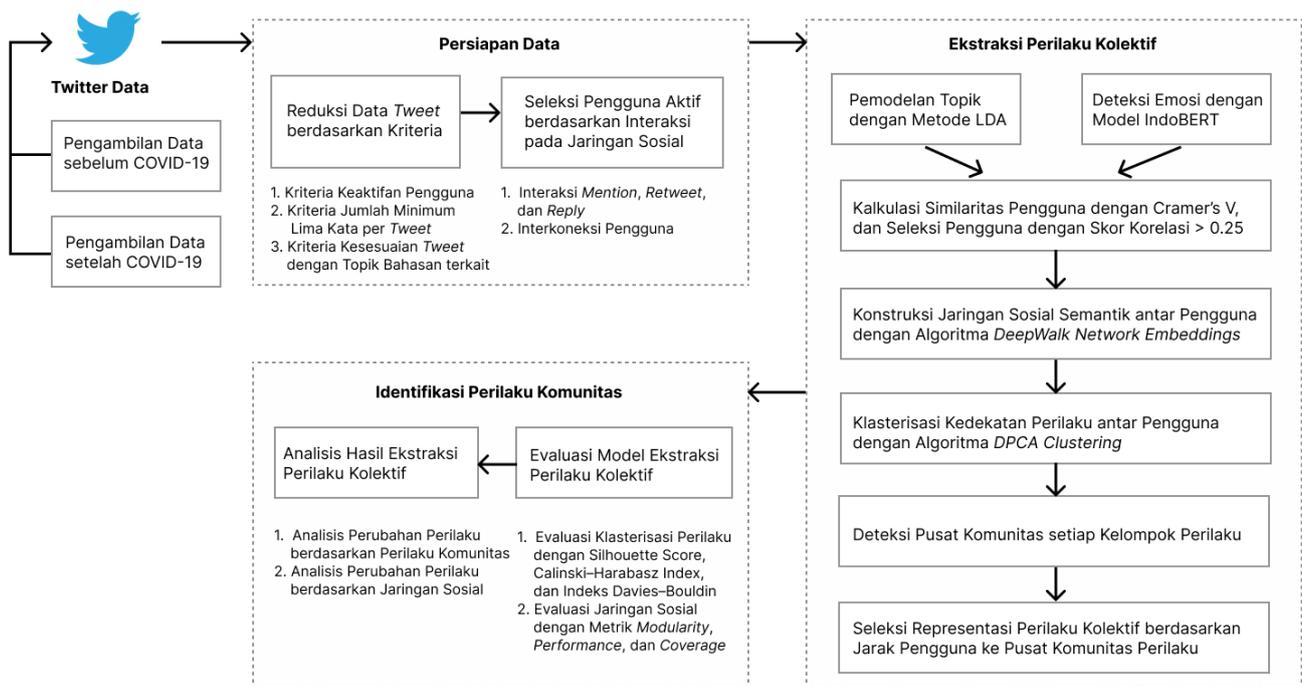
memberikan hasil yang baik untuk identifikasi perilaku berdasarkan similaritas topik bahasan.

Di sisi lain, emosi juga dapat dijadikan sebagai variabel identifikasi perilaku pengguna. Deteksi emosi adalah pendekatan untuk mengekstraksi emosi pengguna berdasarkan informasi implisit dari sumber data tekstual. Mekanisme deteksi emosi biasanya dilakukan dalam tiga pendekatan: berbasis leksikon, berbasis pembelajaran, dan kombinasi berbasis leksikon dan pembelajaran (Gutierrez et al., 2021). Misalnya, penelitian sebelumnya telah membahas hubungan antara emosi dan potensi perilaku bunuh diri menggunakan pendekatan berbasis leksikon, yaitu *NRC Affect Intensity Lexicon* dan *SentiStrength*. Studi tersebut menemukan bahwa emosi negatif dapat mempengaruhi perilaku negatif pengguna (Sarsam et al., 2021). Penelitian lain membahas tren dan kekhawatiran pengguna selama situasi COVID-19. Dalam studi tersebut, karakteristik emosi diekstraksi dengan arsitektur berbasis pembelajaran, yang disebut *Drone Emprit Academy* (DEA). Kemudian, hasil ekstraksi tersebut diklasifikasikan dengan metode Naïve Bayes. Studi tersebut menemukan bahwa emosi mempengaruhi perilaku pengguna secara dinamis dalam krisis longitudinal (Rahmanti et al., 2021). Penelitian selanjutnya melakukan deteksi emosi selama pandemi COVID-19 menggunakan pendekatan berbasis leksikon dan pembelajaran, yang dinamakan *CrystalFeel*. Studi tersebut menemukan bahwa dominasi emosi dalam diskusi *online* dapat mempengaruhi perilaku pengguna lain (Garcia & Berton, 2021). Ketiga studi tersebut menunjukkan bahwa emosi secara implisit dapat memengaruhi stigma pengguna lain. Selanjutnya, perubahan stigma tersebut dapat mempengaruhi perilaku pengguna lain.

Dibandingkan dengan studi-studi sebelumnya, penelitian ini melakukan ekstraksi perilaku kolektif menggunakan kombinasi variabel topik bahasan dan emosi untuk meningkatkan kesamaan perilaku antar pengguna. Kemudian, perilaku pengguna dianalisis dengan jaringan berbasis similaritas perilaku sosial untuk memudahkan eksplorasi perilaku dalam jaringan komunitas. Hasil mekanisme ini digunakan untuk mengidentifikasi bagaimana perubahan perilaku pengguna Twitter pada periode sebelum dan setelah COVID-19 berdasarkan aktivitas temporal.

## 3. METODE PENELITIAN

Penelitian ini mengusulkan mekanisme untuk mengidentifikasi perubahan perilaku pengguna Twitter berdasarkan jaringan berbasis similaritas topik bahasan dan emosi pada aktivitas *tweet* secara temporal. Mekanisme ini kemudian dianalisis melalui interaksi jaringan sosial untuk mengetahui bagaimana alur perubahan perilaku pengguna dalam struktur jaringan komunitas. Alur mekanisme yang diusulkan dalam penelitian ini dapat dilihat pada Gambar 1. Kemudian, notasi yang digunakan dalam mekanisme ini dapat dilihat pada Tabel 1.



Gambar 1. Alur Identifikasi Perubahan Perilaku berdasarkan Model Ekstraksi Perilaku Kolektif Pengguna

### 3.1. Persiapan Data

Penelitian ini menggunakan data *tweet* berbasis bahasa Indonesia dari aktivitas temporal pengguna Twitter pada periode sebelum dan setelah COVID-19 di area Jakarta (Purwitasari et al., 2021). Data periode sebelum COVID-19 berisi *tweet* pengguna dalam rentang 14 Desember 2019 hingga 13 Maret 2020. Di sisi lain, data periode setelah COVID-19 berisi *tweet* pengguna dalam rentang 14 Maret 2020 hingga 13 Juni 2020. Pembagian periode data ini bertujuan untuk memetakan perbedaan perilaku pengguna pada kedua periode tersebut (Valdez et al., 2020). Kemudian, pemilihan area pengambilan data ini didasarkan pada teknik *probability sampling* untuk meningkatkan proses analisis dalam penelitian ini. Hal ini karena area tersebut memiliki dampak tertinggi terhadap kasus COVID-19 di Indonesia selama periode pengambilan data. Hal ini juga didasarkan bahwa pengguna sosial media di lokasi perkotaan yang ramai dan strategis, misalnya wilayah perkantoran, memiliki potensi perubahan perilaku yang tinggi selama situasi krisis (Jaidka et al., 2021).

Data *tweet* temporal didasarkan pada aktivitas *tweet* secara kontinu dari pengguna aktif pada periode sebelum dan setelah COVID-19. Hal ini bertujuan untuk mendapatkan informasi terkait perilaku sehari-hari pengguna secara ekstensif pada kedua periode tersebut berdasarkan representasi tekstual dari data *tweet*. Jenis aktivitas *tweet* yang digunakan hanya berfokus pada interaksi *reply*, *retweet*, dan *mention*. Data temporal ini dapat memberikan informasi terkait perilaku *tweet* secara berkelanjutan, sehingga hasil penelitian ini menjadi relevan untuk mengidentifikasi bagaimana transisi perubahan perilaku pengguna pada kedua periode tersebut (Valdez et al., 2020).

Tabel 1. Notasi yang Digunakan dalam Ekstraksi Perilaku Kolektif

Notasi	Penjelasan Notasi
$U$	Himpunan pengguna Twitter
$u_i$	Pengguna Twitter dengan index ke 1, 2, 3, ..., $i$ .
$UT_i$	Teks Twitter yang dimiliki oleh setiap pengguna $U$
$UD_i$	Himpunan label topik bahasan dari setiap pengguna $U$
$topic_{ij}$	Label topik bahasan pada setiap <i>tweet</i> $UT_i$
$UE_i$	Himpunan label emosi dari setiap pengguna $U$
$emotion_{ij}$	Label emosi pada setiap <i>tweet</i> $UT_i$
$(u_i, u_k)$	Kombinasi pengguna dari himpunan $U$
$g_{ij}$	Matriks penyimpanan nilai korelasi pengguna $u_i$ dan $u_k$
$\phi$	Matriks representasi kedekatan perilaku antar <i>node</i>
$C$	Komunitas dari hasil klusterisasi matriks $\phi$
$UC$	Himpunan perilaku kolektif dari komunitas pengguna $C$
$UC_1$	Himpunan perilaku kolektif periode sebelum COVID-19
$UC_2$	Himpunan perilaku kolektif periode setelah COVID-19

### 3.2. Ekstraksi Perilaku Kolektif

#### a. Kontruksi Jaringan Perilaku Sosial

Pada bagian ini, konstruksi jaringan perilaku sosial dari pengguna  $U \in \{u_1, u_2, u_3, \dots, u_i\}$  terdiri atas dua langkah. Langkah pertama adalah ekstraksi perilaku berdasarkan teks Twitter  $UT_i$ . Langkah ini menggunakan kombinasi topik bahasan dan emosi sebagai variabel ekstraksi perilaku. Hal ini dimulai dengan proses pemodelan topik untuk mendapatkan representasi topik bahasan  $UD_i$  menggunakan metode LDA. Proses pemodelan ini menghasilkan label topik bahasan  $topic_{ij}$  pada setiap *tweet* pengguna. Hasil pemodelan topik kemudian dievaluasi dengan skor koherensi untuk mengukur seberapa baik relevansi kata dengan topik terkait (Costa & Ortale, 2021). Proses selanjutnya adalah deteksi emosi pengguna  $UE_i$  menggunakan model IndoBert. Pemodelan ini menghasilkan lima opsi label emosi  $emotion_{ij}$  pada setiap *tweet* pengguna, yaitu *anger*, *sadness*, *fear*,

*love*, dan *happy* (Hidayat et al., 2021). Hasil ekstraksi kedua variabel ini selanjutnya digunakan sebagai data masukan untuk pemetaan perilaku pengguna.

Langkah kedua adalah membangun jaringan perilaku sosial dari kombinasi pengguna  $(u_i, u_k) \in U$ . Dalam penelitian ini, hubungan kesamaan perilaku pengguna diukur dengan korelasi *Cramer's V*, dengan menggunakan label topik bahasan  $UD_i$  dan label emosi  $UE_i$ . Perhitungan korelasi ini bertujuan untuk mendapatkan similaritas perilaku pengguna yang kuat dan homogen. Jika koefisien *Cramer's V* antara pengguna  $u_i$  dan  $u_k$  lebih besar dari nilai ambang 0.25, maka hubungan kesamaan perilaku dianggap kuat. Representasi hubungan perilaku tersebut kemudian disimpan ke matriks  $g_{ij}$ . Sebaliknya, jika hasil koefisien *Cramer's V* antara pengguna  $u_i$  dan  $u_k$  kurang dari nilai ambang batas 0.25, maka hubungan kesamaan perilaku dianggap lemah. Representasi hubungan perilaku tersebut dapat diabaikan (Li et al., 2018). Relasi pengguna  $u_i$  dan  $u_k$  yang memiliki kesamaan perilaku kuat kemudian divisualisasikan ke dalam jaringan perilaku sosial untuk mempermudah proses analisis perilaku. Proses ini menghasilkan jaringan sosial yang dapat memetakan kesamaan perilaku antar pengguna berdasarkan similaritas topik bahasan dan emosi.

#### b. Deteksi Komunitas berbasis Perilaku Pengguna

Pada bagian ini, jaringan perilaku sosial dikelompokkan berdasarkan kedekatan perilaku antar pengguna. Proses ini terdiri dari dua langkah utama. Langkah pertama adalah ekstraksi informasi perilaku pengguna atau *node* dengan algoritma *DeepWalk*. Algoritma ini dipilih karena dapat merepresentasikan kedekatan perilaku antar *node* pada jaringan sosial berdasarkan informasi semantik pengguna (Li et al., 2018). Algoritma ini dapat mempelajari representasi laten setiap *node* dalam jaringan secara kontinu dan memetakan kedekatan antar *node* ke dalam ruang vektor. Pemrosesan dari algoritma ini menghasilkan matriks  $\phi$  yang merepresentasikan kedekatan perilaku antar *node* dalam jaringan sosial. Hasil pemodelan ini kemudian digunakan sebagai data masukan dalam proses klusterisasi kedekatan perilaku pengguna.

Langkah kedua adalah klusterisasi matriks  $\phi$  menggunakan algoritma DPCA. Algoritma ini dipilih karena hasil dari pemodelan *DeepWalk* berupa representasi vektor  $\phi$  dengan basis jaringan graf dan

densitas. Hal ini menyebabkan klusterisasi perilaku dengan algoritma berbasis jaringan graf dan densitas dapat memiliki hasil yang optimal (Li et al., 2018). Dalam klusterisasinya, algoritma ini menentukan pusat komunitas  $C$  berdasarkan densitas antar *node* pada jaringan perilaku sosial. Pusat komunitas  $C$  ini dapat merepresentasikan kedekatan perilaku antara anggota komunitasnya berdasarkan similaritas topik bahasan dan emosi yang kuat dan homogen.

#### c. Pemilihan Komunitas Perilaku Kolektif

Dalam penelitian ini, analisis pemilihan anggota komunitas  $UC$  didasarkan pada *node* dengan jarak terdekat ke pusat komunitas  $C$ . Strategi ini dapat memetakan kedekatan perilaku antar *node* yang mewakili keseluruhan perilaku pada anggota komunitas  $C$ . Hal ini dapat memberikan informasi kesamaan perilaku yang substansial dari komunitas tersebut. Untuk mendapatkan data yang sesuai terkait analisis perilaku kolektif, penelitian ini melakukan eliminasi data berdasarkan pengguna representatif terpilih dari komunitas pengguna  $UC$ . Data *tweet* yang berasal dari pengguna representatif  $UC$  dipertahankan, sedangkan data *tweet* lainnya diabaikan. Proses ini menghasilkan kumpulan data aktivitas *tweet* temporal dari himpunan komunitas pengguna  $UC_1$  dan  $UC_2$ .

## 4. HASIL DAN PEMBAHASAN

### 4.1 Persiapan Data

Intensitas *tweet* yang tinggi dapat mempersulit proses analisis perilaku pengguna (Li et al., 2018), sehingga perlu dilakukan proses reduksi data untuk memperkecil dimensinya. Untuk mengatasi masalah tersebut, penelitian ini melakukan tiga langkah reduksi data, seperti yang ditunjukkan pada Tabel 2. Langkah pertama adalah memilih pengguna aktif berdasarkan intensitas *tweet*. Proses ini diawali dengan mengakumulasi *tweet* dari semua pengguna. Selanjutnya, keaktifan pengguna dipilih berdasarkan nilai ambang batas dari nilai rata-rata *tweet* per hari, yaitu enam *tweet*. Data *tweet* dengan akumulasi di atas ambang batas tersebut dipertahankan, sedangkan data *tweet* dengan akumulasi di bawah nilai ambang batas tersebut dieliminasi. Proses ini menghasilkan kumpulan data pengguna aktif yang rutin melakukan

Tabel 2. Perbandingan Data Periode Sebelum dan Setelah COVID-19

Tahap Reduksi Data	Periode sebelum COVID-19			Periode setelah COVID-19				
	Jumlah <i>Tweet</i>	Jumlah Pengguna Aktif	Skor Koherensi	Jumlah Topik	Jumlah <i>Tweet</i>	Jumlah Pengguna Aktif	Skor Koherensi	Jumlah Topik
Data awal	1.420.738	73.529	-	-	1.617.101	60.238	-	-
Batas aktivitas <i>tweet</i> harian	772.635	293	0.32	#97 topik	305.165	383	0.41	#90 topik
Batas minimum lima kata	238.797	293	0.53	#74 topik	152.432	383	0.57	#53 topik
Kesesuaian topik bahasan	21.215	290	<b>0.69</b>	#32 topik	49.081	378	<b>0.72</b>	#28 topik

aktivitas *tweet* secara temporal, sehingga analisis dan pemetaan perilaku menjadi lebih relevan.

Langkah kedua adalah mereduksi data *tweet* dengan jumlah kata bawah lima. Hal ini dikarenakan *tweet* dengan jumlah kata yang terbatas, misalnya empat kata, menyebabkan representasi topik bahasan menjadi implisit (Wandabwa et al., 2021). Kondisi ini dapat mempersulit pencarian makna tekstual dari *tweet* saat proses pemodelan. Untuk mengatasi hal ini, *tweet* dengan akumulasi jumlah kata kurang dari lima kata dieliminasi. Hasil ini memberikan representasi *tweet* yang lebih konkret dan eksplisit, serta memiliki makna tekstual yang lebih mudah dipahami.

Langkah ketiga adalah mereduksi data *tweet* berdasarkan kesesuaian topik bahasan. Langkah ini terdiri dari tiga pemrosesan. Pertama, membatasi nilai probabilitas kata pada setiap kluster topik. Proses ini dimodelkan dengan LDA, menggunakan *minimum probability* = 0.01 sebagai ambang batas klusterisasi topik. Hal ini berarti jika suatu kata dalam kluster topik memiliki nilai probabilitas di bawah ambang batas tersebut, maka kata tersebut dieliminasi. Kedua, melakukan tinjauan *tweet* secara manual terhadap relevansi kata kunci setiap kluster topik. Kata-kata yang tidak relevan dengan topik dieliminasi, misalnya bahasa slang atau bahasa daerah. Ketiga, eliminasi data berdasarkan kesesuaian *tweet* dengan relevansi topik terkait. Setiap *tweet* diberikan label topik bahasan sesuai hasil pemodelan LDA, kemudian setiap kata dalam *tweet* dibandingkan similaritasnya dengan kata kunci dari label topik bahasan tersebut. Proses ini menggunakan nilai ambang batas sebesar 60% untuk membandingkan kata-kata dalam *tweet* dengan kata kunci pada kluster topik bahasan terkait. Selanjutnya, data *tweet* dengan persentase similaritas kata di bawah nilai ambang batas tersebut dieliminasi. Proses eliminasi ini menghasilkan data *tweet* dengan representasi yang relevan dengan topik terkait.

Untuk mengevaluasi hasil reduksi data, maka penelitian ini melakukan proses pemodelan topik beserta evaluasi skor koherensi, sesuai pada Tabel 2. Hasil ini menunjukkan bahwa proses reduksi data dapat mempengaruhi peningkatan skor koherensi. Selain itu, proses tersebut juga dapat mengurangi jumlah variasi topik. Hal ini berarti proses reduksi data dapat mengurangi intensitas dan meningkatkan kualitas data. Hasil ini penting untuk mendapatkan hasil ekstraksi perilaku yang optimal berdasarkan kondisi data *tweet* yang bersih dan relevan.

Interkoneksi pengguna dalam jaringan sosial dapat digunakan sebagai objek analisis untuk eksplorasi alur interaksi dan pertukaran informasi (Pascual-Ferrá et al., 2020). Berdasarkan hasil pada Tabel 2, penelitian ini hanya berfokus pada 123 pengguna yang *overlapping* pada kedua periode untuk mendapatkan hasil analisis perilaku yang relevan. Selanjutnya, proses pemilihan pengguna aktif diawali dengan mengakumulasi interaksi antar pengguna pada aktivitas *mention*, *retweet*, dan *reply*. Hasil akumulasi ini kemudian dinormalisasi ke dalam

matriks interaksi dan divisualisasikan ke dalam jaringan graf menggunakan pustaka *NetworkX*.

Tabel 3. Hasil Validasi Deteksi Emosi menggunakan IndoBERT

Metode	Periode sebelum COVID-19		Periode setelah COVID-19	
	Akurasi	F1-Score	Akurasi	F1-Score
SVM	0.78	0.77	0.78	0.77
Naive Bayes	0.79	0.80	0.79	0.80
Random Forest	<b>0.81</b>	<b>0.82</b>	<b>0.81</b>	<b>0.82</b>

Dalam proses konstruksi graf, 123 pengguna tersebut direpresentasikan sebagai aktor atau *node*, sedangkan aktivitas *mention*, *retweet*, dan *reply* antar pengguna direpresentasikan sebagai *edge*. Proses konstruksi ini menghasilkan graf yang dapat memetakan hubungan interaksi pengguna pada jaringan sosial. Untuk mendapatkan hasil pemetaan interkoneksi pengguna, penelitian ini menerapkan eliminasi data pada *node* terisolasi, yaitu *node* yang tidak memiliki relasi dengan *node* lain. Hal ini dikarenakan *node* terisolasi dianggap tidak memiliki kriteria yang sesuai untuk analisis perilaku berdasarkan struktur jaringan sosialnya. Proses eliminasi ini menghasilkan 121 *node* saling terhubung. Secara keseluruhan, proses reduksi data pada penelitian ini menghasilkan 121 pengguna aktif, dengan jumlah data sebanyak 23.647 *tweet* pada periode sebelum COVID-19 dan 7.333 *tweet* pada periode setelah COVID-19.

## 4.2. Ekstraksi Perilaku Kolektif

### a. Konstruksi Jaringan Sosial

Pada bagian ini, konstruksi jaringan sosial terdiri atas dua langkah. Langkah pertama adalah ekstraksi variabel topik bahasan dan emosi dari *tweet* pengguna Twitter. Dalam deteksi emosi, pemrosesan teks diawali dengan *case-folding*. Hasil ini digunakan pada proses pemodelan IndoBERT untuk mendeteksi label emosi pada setiap *tweet* (Hidayat et al., 2021). Data *tweet* pengguna tersebut kemudian diklasifikasi berdasarkan hasil ekstraksi *unigram* dan *bigram*, dengan perbandingan 80% data pelatihan dan 20% data pengujian pada masing-masing periode. Hasil klasifikasi tersebut kemudian divalidasi dengan metrik akurasi dan *f1-score*, karena *ground truth* data emosi pada penelitian ini memiliki lima jenis label atau bersifat *multi class*. Evaluasi kedua metrik ini dapat mengoptimalkan perbandingan prediksi benar atau salah dari masing-masing label secara akurat (Garcia & Berton, 2021). Hasil validasi ini dapat dilihat pada Tabel 3. Hasil ini menunjukkan bahwa metode *Random Forest* memiliki performa yang paling optimal. Hasil ini juga menunjukkan bahwa IndoBERT memiliki performa yang cukup baik dalam mendeteksi emosi, dengan nilai rata-rata setiap metode di atas 75%. Karena data *tweet* dalam penelitian ini bersifat *unsupervised*, maka hasil dari pemodelan IndoBERT digunakan untuk menentukan *ground truth* label emosi pada setiap data *tweet*.

Tabel 4. Hasil Pengelompokan Topik Bahasan menggunakan *Agglomerative Clustering*

Level Topik	Periode sebelum COVID-19	Periode setelah COVID-19
Level 0	Pekerjaan, Hobi, Aktivitas Pribadi, Aktivitas	Psikologis, Hobi dan Aktivitas Pribadi, Kesehatan, Hukum
Topik Utama	Ekonomi, Politik and Pemerintahan.	dan Kriminalitas, Politik dan Pemerintahan
Level 1	Pekerjaan, Kesehatan, Berita dan Informasi, Hukum,	Psikologis, Politik dan Pemerintahan, Hobi dan Aktivitas
Sub-topik 1	Hobi, Aktivitas Personal,	Pribadi, Pandemi COVID-19, Pekerjaan, Perilaku
	Aktivitas Jual-Beli, Politik dan Pemerintahan	Masyarakat, Media Sosial, Kriminalitas
Level 2	Pekerjaan, Kesehatan, Berita dan Informasi, Hobi,	Depresi, Aktivitas Ekonomi, Politik dan Pemerintahan, Hobi
Sub-topik 2	Hukum, Aktivitas Pribadi, Makanan, Seni, Investasi,	dan Aktivitas Pribadi, Aktivitas Keagamaan, COVID-19,
	Aktivitas Jual-Beli, Politik dan Pemerintahan.	Pekerjaan, Perilaku Masyarakat, Media Sosial, Kriminalitas

Dalam pemodelan topik, pemrosesan teks *tweet* juga diawali dengan proses *case-folding*. Proses selanjutnya adalah menghilangkan karakter khusus pada *tweet*, seperti URL, *email*, *username*, *hashtag*, atau karakter lain yang tidak berpengaruh positif terhadap hasil pemodelan (Garcia & Berton, 2021). Setelah itu, melakukan tokenisasi dan normalisasi *tweet* dengan mengubah kata tidak baku, misalnya bahasa slang atau bahasa daerah, menjadi kata baku yang sesuai dengan preferensi Kamus Besar Bahasa Indonesia (KBBI). Selanjutnya, menghapus *stopword* untuk menghilangkan kata-kata yang tidak memiliki makna penting, dengan menyesuaikan referensi dari KBBI. Kemudian, tahap *stemming* dilakukan untuk mengubah token kata ke bentuk kata dasar. Proses terakhir adalah *Part-of-Speech (POS) Tagging*. Klausa yang dipilih adalah kata benda, kata sifat, kata kerja, dan kata keterangan. Pemilihan klausa ini bertujuan untuk mendapatkan representasi kata yang penting dan bermakna dari sekumpulan teks *tweet* pengguna. Terakhir, hasil pemrosesan teks tersebut dimodelkan dengan LDA. Proses pemodelan ini menghasilkan skor koherensi yang dapat dilihat pada tahap reduksi data berdasarkan relevansi topik bahasan di Tabel 2. Hasil ini menunjukkan bahwa kedua periode memiliki hasil pengelompokan topik yang cukup baik, dengan nilai skor koherensi diatas 0.65. Hal ini berarti proses pemodelan topik memiliki kinerja yang baik dalam mengelompokkan *tweet* ke dalam representasi topik bahasan tertentu.

Berdasarkan hasil Tabel 2, pemodelan topik memiliki jumlah variasi topik yang cenderung tinggi. Hasil ini dapat mempersulit proses analisis karena cakupan topik yang meluas, sehingga penelitian ini melakukan proses pengelompokan topik untuk mengatasi hal tersebut (Apriantoni et al., 2021). Hal ini bertujuan untuk mengelompokkan topik-topik dengan kemiripan kata yang tinggi ke dalam satu klaster topik yang sama. Pengelompokan topik ini dimodelkan dengan metode *Agglomerative Clustering*, dimana pendekatan *cosine similarity* digunakan untuk menghitung similaritas kata antar topik (Costa & Ortale, 2021). Dari hasil tersebut, penelitian ini memilih batasan lima topik yang dijadikan sebagai representasi topik utama. Pemilihan ini didasarkan pada hasil evaluasi *Silhouette Score (SC)* yang memiliki skor optimal pada lima klaster pada kedua periode, dimana periode sebelum COVID-19 memiliki nilai SC sebesar 0.59 dan setelah COVID-19 memiliki nilai SC sebesar 0.62.

Hasil pengelompokan topik pada penelitian ini dapat dilihat di Tabel 4. Hasil ini menunjukkan adanya perbedaan topik utama pada kedua periode analisis. Pada periode sebelum COVID-19, dominan topik membahas pekerjaan dan ekonomi, di mana topik tersebut lazim dibahas di area perkantoran. Hasil ini sesuai dengan lokasi pengambilan data di daerah perkantoran, sehingga hasil topik bahasan juga mengarah ke pembahasan tersebut. Di sisi lain, pada periode setelah COVID-19, terjadi perubahan di topik utama. Topik terkait pekerjaan dan aktivitas ekonomi tidak lagi dominan, sedangkan topik terkait isu kesehatan dan psikologi menjadi dominan. Hasil ini relevan dengan kondisi di awal pandemi COVID-19 yang berdampak pada permasalahan kesehatan. Hal ini mempengaruhi inkonsistensi aktivitas masyarakat akibat masa pembatasan sosial, sehingga pembahasan terkait dampak topik tersebut meningkat signifikan.

Langkah selanjutnya adalah memetakan relasi perilaku pengguna dengan korelasi *Cramer's V*. Data yang diolah berfokus pada 121 pengguna aktif di kedua periode. Variabel perilaku yang dibandingkan adalah label topik bahasan dari hasil pemodelan topik dan label emosi dari hasil deteksi emosi. Perilaku pengguna ini dibandingkan per periode secara independen. Untuk mendapatkan hasil identifikasi perilaku yang relevan, perhitungan korelasi antar pengguna hanya dilakukan pada pengguna dengan jumlah data *tweet* di atas 100. Kemudian, setiap pengguna dalam jangka waktu tertentu, misalnya tiga bulan, biasanya memiliki intensitas *tweet* yang berbeda. Jika dua pengguna yang perilakunya dibandingkan memiliki jumlah *tweet* berbeda, maka dilakukan teknik *undersampling*. Hal ini dilakukan dengan mengakumulasi jumlah *like*, *mention*, dan *reply* pada pengguna yang memiliki jumlah *tweet* terbanyak. Data *tweet* tersebut selanjutnya direduksi dengan menyamakan jumlah *tweet* sesuai pengguna dengan jumlah *tweet* terendah. Hasil ini kemudian diukur similaritasnya dengan korelasi *Cramer's V*. Proses ini menghasilkan 98 relasi pengguna aktif pada periode sebelum COVID-19 dan 54 relasi pengguna aktif pada periode setelah COVID-19.

Hubungan antar pengguna dengan kesamaan perilaku yang kuat selanjutnya divisualisasikan ke jaringan perilaku sosial. Jaringan tersebut kemudian dievaluasi dengan tiga jenis metrik, yaitu *modularity*, *performance*, dan *covarege*. Evaluasi ketiga metrik ini digunakan untuk mengukur performa jaringan graf dalam memisahkan perilaku berdasarkan jarak dan

Tabel 5. Hasil Klusterisasi Perilaku Pengguna berdasarkan Similaritas Topik Bahasan dan Emosi

Periode	Klaster	Jumlah Pengguna	Inisialisasi Perilaku	Deskripsi Perilaku
Sebelum COVID-19	0	11	BF_0	Perilaku <i>tweet</i> yang didominasi oleh topik bahasan terkait politik dan pemerintahan dengan emosi marah
	1	13	BF_1	Perilaku <i>tweet</i> yang didominasi oleh topik bahasan terkait pekerjaan dengan emosi marah
	2	32	BF_2	Perilaku <i>tweet</i> yang didominasi oleh topik bahasan terkait aktivitas pribadi dengan emosi bahagia
	3	31	BF_3	Perilaku <i>tweet</i> yang didominasi oleh topik bahasan terkait aktivitas ekonomi sehari-hari dengan emosi marah
	4	11	BF_4	Perilaku <i>tweet</i> yang didominasi oleh topik bahasan terkait pekerjaan dengan emosi bahagia
Setelah COVID-19	0	11	AF_0	Perilaku <i>tweet</i> yang didominasi oleh topik bahasan terkait politik dan pemerintahan dengan emosi marah
	1	35	AF_1	Perilaku <i>tweet</i> yang didominasi oleh topik bahasan terkait masalah kesehatan dengan emosi marah
	2	8	AF_2	Perilaku <i>tweet</i> yang didominasi oleh topik bahasan terkait aktivitas pribadi dengan emosi bahagia

densitas antar *node*. Pada periode sebelum COVID-19, graf memiliki metrik *modularity*, *performance*, dan *covarege* masing-masing sebesar 0.56, 0.87 dan 0.83. Di sisi lain, graf periode setelah COVID-19 memiliki skor masing-masing sebesar 0.67, 0.97, dan 0.89. Hasil ini menunjukkan bahwa nilai modularitas kedua periode tersebut berada di atas nilai 0.5. Fragmentasi modularitas mendekati 1 menunjukkan pemisahan yang jelas antar komunitas, sedangkan nilai modularitas kurang dari 0.5 menunjukkan bahwa partisi komunitas cenderung tumpang tindih dan lebih rapat (Pascual-Ferrá et al., 2020). Selanjutnya, metrik *performance* membandingkan kedekatan antar *node* dalam satu komunitas jaringan. Di sisi lain, metrik *coverage* mengukur rasio jumlah hubungan *node* antar komunitas di seluruh jaringan. Linear dengan hasil tersebut, *performance* dan *coverage* memiliki skor rata-rata di atas 0.80. Hasil ini berarti jarak antar *node* dalam satu komunitas cukup dekat, sedangkan jarak antar komunitas cukup jauh dan renggang. Hal ini menunjukkan bahwa jaringan perilaku sosial ini memiliki pemisahan klaster yang cukup baik dalam membedakan kedekatan perilaku antar komunitas.

#### b. Deteksi Komunitas berbasis Perilaku Pengguna

Penentuan komunitas terdiri dari dua langkah. Pertama, memetakan kedekatan perilaku antar *node* pada jaringan perilaku pengguna dengan *DeepWalk*. Untuk mendapatkan hasil pemodelan yang optimal, penelitian ini melakukan serangkaian uji coba dengan kombinasi parameter: *walk length* ( $l$ ) = {20, 40, 60, 80, 100}, *walk per vertex* ( $\gamma$ ) = {13, 10, 30, 50, 90}, *embedding size* ( $d$ ) = {32, 64, 128, 192, 256}, dan *window size* ( $\omega$ ) = {5, 10, 15, 20, 25}. Dari hasil uji coba tersebut, parameter optimal dalam penelitian ini adalah  $l = 100$ ,  $\gamma = 90$ ,  $d = 64$ , dan  $\omega = 5$ .

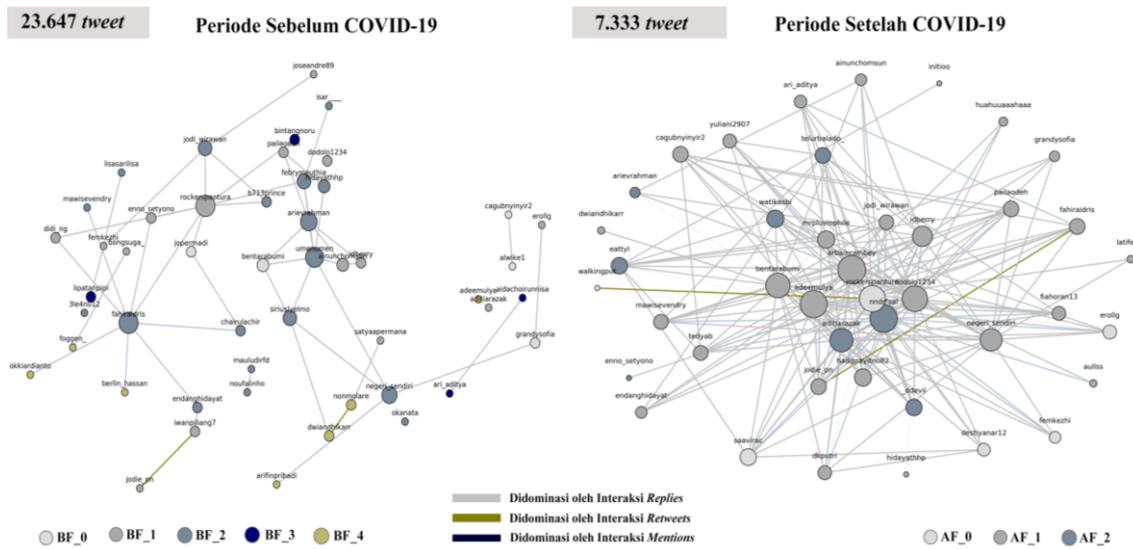
Langkah kedua adalah pengelompokkan vektor hasil representasi matriks kedekatan perilaku dengan algoritma DPCA. Untuk mendapatkan hasil yang optimal, pemodelan ini dievaluasi dengan tiga jenis metrik, yaitu SC, Calinski–Harabasz Index (CHI) dan Davies–Bouldin Index (DBI) (Bello-Organ et al., 2020). Pada periode sebelum COVID-19, hasil klusterisasi perilaku pengguna memiliki nilai evaluasi

maksimal pada klaster ke lima, dengan nilai SC, CHI, dan DBI masing-masing sebesar 0.36, 7.81 dan 1.25. Di sisi lain, klusterisasi perilaku pengguna pada periode setelah COVID-19 memiliki nilai evaluasi maksimal di klaster ke tiga, dengan nilai SC, CHI, dan DBI masing-masing sebesar 0.43, 28.40 dan 0.79.

Dari hasil evaluasi ketiga metrik tersebut, hal ini menunjukkan bahwa kedua periode memiliki nilai SC yang kurang baik. Hal ini dikarenakan algoritma ini mempertimbangkan pemisahan klaster berdasarkan densitas antar *node* pada jaringan. Sebaliknya, metrik SC melakukan partisi klaster berdasarkan jarak *intra-cluster* antar *node*. Hal ini berarti evaluasi SC dalam penelitian ini tidak memberikan hasil yang optimal. Di sisi lain, klusterisasi perilaku memiliki hasil yang cukup baik pada metrik CHI dan DBI. CHI memiliki nilai yang lebih baik jika nilainya semakin besar atau menjauhi 0, sedangkan DBI memiliki nilai yang lebih baik jika mendekati 0. Hasil ketiga metrik tersebut menunjukkan bahwa algoritma DPCA memiliki performa yang cukup baik dalam mengelompokkan kedekatan perilaku antar pengguna berdasarkan similaritas topik bahasan dan emosi.

#### c. Pendirian Pusat Komunitas

Representasi perilaku kolektif pengguna dipilih berdasarkan hasil komunitas perilaku yang diperoleh, yaitu periode sebelum COVID-19 memiliki lima klaster, sedangkan periode setelah COVID-19 hanya memiliki tiga klaster. Hasil klaster perilaku tersebut dapat dilihat pada Tabel 5. Hasil ini menunjukkan bahwa dominasi topik bahasan dan emosi pengguna terkait aktivitas pribadi, politik dan pemerintahan cenderung stabil pada kedua periode tersebut. Akan tetapi, terdapat perubahan perilaku, di mana masa pembatasan sosial dan pemberlakuan *work from home* telah mengurangi pembahasan *tweet* terkait aktivitas pekerjaan dan ekonomi. Hal ini membuat kedua topik bahasan tersebut tidak lagi menjadi topik utama, seperti terlihat pada agregasi topik di Tabel 4. Di sisi lain, pandemi COVID-19 juga mempengaruhi intensitas topik bahasan terkait isu kesehatan. Hal ini menyebabkan perilaku komunitas terkait topik ini meningkat, dengan didominasi oleh kemarahan



Gambar 2. Visualisasi Perilaku Komunitas Pengguna berdasarkan Similaritas Topik Bahasan dan Emosi

masyarakat terkait pengalaman distress publik. Transisi inilah yang menyebabkan perilaku *tweet* pengguna terkait isu kesehatan lebih mendominasi daripada perilaku *tweet* terkait aktivitas pekerjaan dan aktivitas ekonomi pada periode setelah COVID-19.

**4.3. Identifikasi Perilaku Komunitas**

Berdasarkan hasil analisis pada jaringan perilaku sosial, periode sebelum COVID-19 memiliki 62% kesamaan perilaku yang didominasi oleh relasi pengguna berdasarkan topik bahasan, sedangkan 38% sisanya berdasarkan emosi. Pada periode setelah COVID-19, 51% kesamaan perilaku didominasi oleh relasi pengguna berdasarkan topik bahasan, dan 49% sisanya berdasarkan emosi. Hasil ini menunjukkan bahwa adanya kenaikan kesamaan perilaku yang didasari oleh emosi pengguna pada periode setelah COVID-19. Hal ini berarti variabel topik bahasan dan emosi memiliki proporsi pengaruh yang cukup seimbang terhadap perubahan perilaku pengguna.

Berdasarkan hasil klusterisasi perilaku di Tabel 5, hal ini menunjukkan bahwa periode sebelum COVID-19 didominasi oleh 33% perilaku *tweet* terkait aktivitas pribadi dengan emosi senang. Di sisi lain, 65% perilaku pengguna pada periode setelah COVID-19 didominasi oleh komunitas dengan perilaku *tweet* terkait kesehatan dengan emosi marah. Hasil ini menunjukkan adanya perubahan dominasi perilaku yang cukup signifikan pada kedua periode tersebut, khususnya pada topik kesehatan. Hal ini dipengaruhi oleh tingginya diskusi dan penyebaran informasi terkait diseminasi penyakit, dampak sosial dan penanggulangan pandemi COVID-19, sehingga signifikansi topik tersebut meningkat.

Komunitas perilaku selanjutnya divisualisasi ke jaringan berbasis similaritas topik bahasan dan emosi berdasarkan aktivitas *tweet* temporal, seperti pada Gambar 2. Visualisasi tersebut menunjukkan bahwa periode sebelum COVID-19 memiliki pola jaringan yang terdistribusi, sehingga penyebaran informasi

antar pengguna terjadi secara terdesentralisasi. Hal ini menyebabkan aliran informasi menyebar dengan lambat ke beberapa pusat komunitas yang berbeda. Oleh karena itu, variasi perilaku pada periode ini cenderung lebih banyak dan memiliki potensi yang lama dalam mempengaruhi aktivitas atau perilaku pengguna lain. Di sisi lain, jaringan setelah COVID-19 membentuk jaringan berpola *clique graph*, dengan densitas antar interaksi pengguna yang tinggi. Hal ini menyebabkan aliran informasi menjadi tersentralisasi di pusat komunitas yang sama. Hal ini menyebabkan variasi perilaku dari jaringan ini cenderung lebih sedikit, tetapi memiliki potensi yang lebih tinggi dan cepat dalam mempengaruhi perilaku pengguna lain.

**5. KESIMPULAN**

Penelitian ini mengembangkan mekanisme identifikasi perilaku pengguna Twitter berdasarkan model ekstraksi perilaku kolektif pada aktivitas *tweet* temporal. Terdapat dua analisis perubahan perilaku dalam penelitian ini. Pertama, klusterisasi perilaku pengguna menunjukkan bahwa periode sebelum COVID-19 memiliki 98 pengguna representatif yang didominasi oleh 33% perilaku komunitas terkait aktivitas pribadi dengan emosi senang. Di sisi lain, periode setelah COVID-19 memiliki 54 pengguna representatif yang didominasi oleh 65% perilaku komunitas terkait kesehatan dengan emosi marah. Kedua, jaringan berbasis kesamaan perilaku pada periode sebelum COVID-19 memiliki pola jaringan yang terdistribusi. Hal ini menyebabkan variasi perilaku cenderung lebih banyak dan memiliki potensi yang rendah dalam mempengaruhi perubahan perilaku pengguna lain. Sebaliknya, periode setelah COVID-19 memiliki pola jaringan *clique graph*. Hal ini menyebabkan variasi perilaku cenderung lebih sedikit, tetapi memiliki potensi yang lebih tinggi dalam mempengaruhi perubahan perilaku pengguna lain. Kedua hasil tersebut dapat menjadi perhatian bagi Pemerintah DKI Jakarta, yang pada penelitian

ini dijadikan sebagai obyek lokasi pengambilan data, dalam membuat perumusan strategi untuk mengatasi situasi krisis mendatang. Misalnya, mengurangi potensi penyebaran perilaku negatif dengan memanfaatkan komunitas yang memiliki pengaruh perilaku positif di kalangan pengguna Twitter.

Penelitian ini memiliki tantangan utama, yaitu alur pertukaran informasi secara dinamis dapat mempengaruhi tingginya potensi perubahan emosi pengguna. Oleh karena itu, penelitian selanjutnya perlu menganalisis implikasi perubahan perilaku terhadap perubahan emosi pengguna Twitter pada struktur jaringan komunitas sosial.

## DAFTAR PUSTAKA

- APRIANTONI, A., THOORIQOH, H. A., FATICHAH, C., & PURWITASARI, D. 2021. Topic Detection in Sentiment Analysis of Twitter Texts for Understanding The COVID-19 Effect in Local Economic Activities. In *2021 13th International Conference on Information & Communication Technology and System (ICTS)* (pp. 354–359).
- AZER, J., BLASCO-ARCAS, L., & HARRIGAN, P. 2021. #COVID-19: Forms and drivers of social media users' engagement behavior toward a global crisis. *Journal of Business Research*, *135*(June), 99–111. <https://doi.org/10.1016/j.jbusres.2021.06.030>
- BASIRI, M. E., NEMATY, S., ABDAR, M., ASADI, S., & ACHARRYA, U. R. 2021. A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets. *Knowledge-Based Systems*, *228*, 107242. <https://doi.org/10.1016/j.knosys.2021.107242>
- BELLO-ORGAZ, G., MESAS, R. M., ZARCO, C., RODRIGUEZ, V., CORDÓN, O., & CAMACHO, D. 2020. Marketing analysis of wineries using social collective behavior from users' temporal activity on Twitter. *Information Processing and Management*, *57*(5), 102220. <https://doi.org/10.1016/j.ipm.2020.102220>
- COSTA, G., & ORTALE, R. 2021. Jointly modeling and simultaneously discovering topics and clusters in text corpora using word vectors. *Information Sciences*, *563*, 226–240. <https://doi.org/10.1016/j.ins.2021.01.019>
- CRESCI, S., DI PIETRO, R., PETROCCHI, M., SPOGNARDI, A., & TESCONI, M. 2020. Emergent properties, models, and laws of behavioral similarities within groups of twitter users. *Computer Communications*, *150*(October 2019), 47–61. <https://doi.org/10.1016/j.comcom.2019.10.019>
- GARCIA, K., & BERTON, L. 2021. Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. *Applied Soft Computing*, *101*, 107057. <https://doi.org/10.1016/j.asoc.2020.107057>
- GUTIERREZ, E., KARWOWSKI, W., FIOK, K., DAVAHLI, M. R., Liciaga, T., & Ahram, T. 2021. Analysis of human behavior by mining textual data: Current research topics and analytical techniques. *Symmetry*, *13*(7), 1–22. <https://doi.org/10.3390/sym13071276>
- HIDAYAT, R. A., KHASANAH, I. N., PUTRI, W. C., & MAHENDRA, R. 2021. Feature-Rich Classifiers for Recognizing Textual Entailment in Indonesian. *Procedia CIRP*, *189*(2019), 148–155. <https://doi.org/10.1016/j.procs.2021.05.094>
- JAIDKA, K., GUNTUKU, S. C., LEE, J. H., LUO, Z., BUFFONE, A., & UNGAR, L. H. 2021. The rural–urban stress divide: Obtaining geographical insights through Twitter. *Computers in Human Behavior*, *114*(April 2020), 106544. <https://doi.org/10.1016/j.chb.2020.106544>
- LI, L., ZHOU, C., HE, J., WANG, J., LI, X., & WU, X. 2018. Collective semantic behavior extraction in social networks. *Journal of Computational Science*. <https://doi.org/10.1016/j.jocs.2017.11.003>
- MOHAMMED, A., & FERRARIS, A. 2021. Factors influencing user participation in social media: Evidence from twitter usage during COVID-19 pandemic in Saudi Arabia. *Technology in Society*, *66*(December 2020), 101651. <https://doi.org/10.1016/j.techsoc.2021.101651>
- PASCUAL-FERRÁ, P., ALPERSTEIN, N., & BARNETT, D. J. 2020. Social Network Analysis of COVID-19 Public Discourse on Twitter: Implications for Risk Communication. *Disaster Medicine and Public Health Preparedness*, 1–9. <https://doi.org/10.1017/dmp.2020.347>
- PURWITASARI, D., RAHARJO, A. B., Korespondensi, P., Analisis, S. N., Topik, P., Pengguna, P., & Pengguna, P. 2021. Identifikasi Pengaruh Pandemi Covid-19 Terhadap Perilaku Pengguna Twitter Dengan Pendekatan Social Network Analysis, *8*(6), 1309–1318. <https://doi.org/10.25126/jtiik.202185213>
- RAHMANTI, A. R., NINGRUM, D. N. A., LAZUARDI, L., YANG, H. C., & LI, Y. C. 2021. Social Media Data Analytics for Outbreak Risk Communication: Public Attention on the “New Normal” During the COVID-19 Pandemic in Indonesia. *Computer Methods and Programs in Biomedicine*, *205*, 106083. <https://doi.org/10.1016/j.cmpb.2021.106083>
- SARSAM, S. M., AL-SAMARRAIE, H., ALZHRANI, A. I., ALNUMAY, W., & SMITH, A. P. 2021. A lexicon-based approach to detecting suicide-related messages on Twitter. *Biomedical Signal Processing and Control*, *65*(December 2020), 102355. <https://doi.org/10.1016/j.bspc.2020.102355>
- VALDEZ, D., TEN THIJ, M., BATHINA, K., RUTTER, L. A., & BOLLEN, J. 2020. Social media insights into US mental health during the COVID-19 pandemic: Longitudinal analysis of twitter data. *Journal of Medical Internet Research*, *22*(12). <https://doi.org/10.2196/21418>

WANDABWA, H. M., NAEEM, M. A., MIRZA, F., & PEARS, R. 2021. Multi-interest semantic changes over time in short-text microblogs[Formula presented]. *Knowledge-Based Systems*, 228, 107249.  
<https://doi.org/10.1016/j.knosys.2021.107249>

*Halaman ini sengaja dikosongkan*