

Analisis Topik Tagar Covidindonesia pada Instagram Menggunakan *Latent Dirichlet Allocation*

Kevin Rafi Adjie Putra Santoso ⁽¹⁾, Asmaul Husna ⁽²⁾, Nadia Widyawati Putri ⁽³⁾, Nur Aini Rakhmawati ^{(4)*}

Sistem Informasi, Fakultas Teknologi Elektro dan Informatika Cerdas, Institut Teknologi Sepuluh Nopember, Surabaya

e-mail : {kevin.18052,asmaulhusna.18052,nadiaputri.180521}@mhs.its.ac.id,
nur.aini@is.its.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 11 April 2021, direvisi 12 Juli 2021, diterima 13 Juli 2021, dan dipublikasikan 25 Januari 2022.

Abstract

In this era, technology is increasingly sophisticated, this is evidenced by the number of people using the internet via cell phones, laptops, and other communication tools. One of the developments of this technology is social media such as Instagram. Along with technological developments, Instagram users can upload and share photos and videos using hashtags (#) so that other users can find the results of their posts. Instagram has now become one of the social media used by more than 1 billion people in the world. In this study, the authors wanted to know the dominant topics discussed through the hashtag covidindonesia. This research was conducted using the Latent Dirichlet Allocation (LDA) method. The analysis was carried out after doing text mining on 84 captions from various users on Instagram. To determine the optimal number of topics, by looking at the value of perplexity and topic coherence. The results obtained are the top 5 topics that are the content material in the uploaded video. These topics include covidindonesia, covid_19, pandemics in Indonesia, and discussion of covid-19 virus mutations.

Keywords: *Data Crawling, Instagram, Latent Dirichlet Allocation, Covid Indonesia, Topic Modeling*

Abstrak

Pada era sekarang, teknologi semakin canggih, hal ini dibuktikan dengan banyaknya orang yang menggunakan internet melalui telepon genggam, laptop, dan alat komunikasi lainnya, Salah satu perkembangan dari teknologi ini ialah media sosial seperti Instagram. Seiring dengan perkembangan teknologi, pengguna Instagram dapat mengunggah dan membagikan foto dan juga video dengan menggunakan *hashtag* (#) agar pengguna lain dapat menemukan hasil unggahan mereka. Instagram pun kini menjadi salah satu media sosial yang digunakan lebih dari 1 miliar orang di dunia. Pada penelitian ini penulis ingin mengetahui topik dominan yang dibahas melalui *hashtag* covidindonesia. Penelitian ini dilakukan dengan menggunakan metode *Latent Dirichlet Allocation* (LDA). Analisis dilakukan setelah melakukan *text mining* pada 84 *caption* dari berbagai user yang ada di Instagram. Untuk menentukan jumlah topik yang optimal, dengan melihat *nilai perplexity* dan *topic coherence*. Hasil yang didapatkan adalah 5 topik teratas yang menjadi bahan konten dalam video yang diunggah. Topik tersebut antara lain covidindonesia, covid_19, *pandemic* di Indonesia, dan pembahasan mutasi virus covid-19.

Kata Kunci: *Crawling Data, Instagram, Latent Dirichlet Allocation, Covid Indonesia, Pemodelan Topik*

1. PENDAHULUAN

Pada era sekarang, teknologi semakin canggih, hal ini dibuktikan dengan banyaknya orang yang menggunakan internet melalui telepon genggam, laptop, dan alat komunikasi lainnya, dengan masifnya perkembangan teknologi, masyarakat dapat dengan mudah memenuhi apa yang dibutuhkannya dalam waktu yang relatif cepat serta mudah dalam hal penggunaannya. Dengan keberadaan internet, secara tidak langsung menghasilkan sebuah generasi yang baru, yaitu generasi $ne(x)t$. Generasi ini dipandang menjadi sebuah generasi masa depan yang diasuh dan



dibesarkan dalam lingkungan budaya baru media digital yang interaktif, yang berwatak menyendiri, kemudian berkomunikasi secara personal, meleak komputer atau meleak teknologi, dan dibesarkan dengan *videogram* (Ibrahim, 2018). Salah satu perkembangan dari teknologi ialah media sosial seperti Instagram yang muncul pada 6 Oktober 2010. Instagram merupakan salah satu jenis media sosial yang kehadirannya cukup fenomenal, dalam waktu sembilan bulan saja, unggah foto dalam Instagram mencapai angka 150 juta foto di San Francisco. Mengalahkan situs *media-sharing* sejenis Flickr dan situs jejaring Facebook yang fenomenal (Nugraha & Akbar, 2019). Instagram sebagai sebuah media sosial dibangun berdasarkan teknologi Web 2.0 yang membuat penggunanya dapat menyediakan dan berbagi konten (Abdul Talib & Mat Saat, 2017). Pengguna Instagram dapat membagikan foto atau video yang mereka unggah kepada teman dan pengikut mereka. Selain itu, pengguna juga dapat saling berinteraksi dengan melihat, menyukai, dan mengomentari unggahan yang dibagikan di Instagram. Seiring dengan perkembangan teknologi, pengguna Instagram dapat mengunggah dan membagikan foto dan juga video dengan menggunakan *hashtag* (#) agar pengguna lain dapat menemukan hasil unggahan mereka. Instagram pun kini menjadi salah satu media sosial yang digunakan lebih dari 1 miliar orang di dunia (Carman, 2018).

Salah satu topik yang marak dibahas sekarang ialah pandemi covid-19 di Indonesia. Pada awalnya virus ini diduga akibat paparan pasar grosir makanan laut yang banyak menjual banyak spesies hewan hidup. Penyakit ini dengan cepat menyebar di dalam negeri ke bagian lain China. Munculnya 2019-nCoV telah menarik perhatian global tak terkecuali di Indonesia. Pasien pertama yang terkonfirmasi covid-19 di Indonesia berawal dari suatu acara di Jakarta di mana penderita kontak dengan seseorang warga Negara asing (WNA) asal Jepang yang tinggal di Malaysia. Setelah pertemuan tersebut penderita mengeluh demam, batuk dan sesak nafas (Azizah, 2020). Karena penyebaran virus covid-19 yang tidak henti-henti ini akan berdampak secara sosial dan ekonomi. Dalam hal ini Indonesia harus bersiap siaga dalam menghadapinya terutama dalam hal sistem kesehatan yang ada. Melihat pandemi covid-19 yang tak kunjung usai, masyarakat pun banyak memberikan tanggapan mereka melalui Instagram dengan menggunakan *hashtag* *covidindonesia*.

Berdasarkan uraian di atas, penelitian ini menawarkan solusi untuk melakukan analisis *topic modelling* terhadap topik covid-19 yang dibicarakan di Instagram. *Topic modelling* merupakan metode *non-hierarchical clustering* yang secara otomatis mengklusterkan ke dalam topik yang muncul dari pemodelan sehingga didapatkan topik *cluster* yang sesuai (Alfanzar et al., 2020). Penelitian ini akan mengambil data dari Instagram dengan menggunakan *hashtag* *covidindonesia*.

Penelitian mengenai Covid-19 sudah pernah dilakukan oleh (Boer et al., 2020) dalam penelitiannya yang berjudul "Analisis Pendapat Siswa Tentang Pembelajaran Berbasis Media Televisi Selama Pandemi Covid-19". Penelitian tersebut bertujuan untuk mengidentifikasi pemberitaan mengenai penanganan virus Covid-19 yang ada di media *online*. Metode yang digunakan adalah metode *framing* di mana para peneliti akan menganalisis suatu fenomena dengan cara memilah kembali dan menyederhanakan berbagai aspek yang menarik perhatian sebagian besar orang. Penelitian ini merupakan penelitian deskriptif kualitatif dengan batasan hanya tiga situs media saja yang dianalisis, yaitu CNNIndonesia.com, Liputan6.com, dan Kompas.com. Sedangkan batasan pada penelitian yang kami lakukan adalah karena kami terbatas pada *caption* Instagram, sehingga terdapat perbedaan karakteristik data yang bersumber dari berbagai media sosial, baik dari panjang karakter per dokumen dan sifat kebahasaan yang diterapkan, mengingat *caption* Instagram biasanya menggunakan bahasa informal.

Metode LDA ini juga pernah digunakan oleh (Xue et al., 2020) dalam penelitiannya yang berjudul "Public discourse and sentiment during the COVID 19 pandemic: Using latent dirichlet allocation for topic modeling on Twitter". Di mana penelitian dilakukan dengan tujuan untuk memahami wacana dan reaksi psikologis dari pengguna Twitter terhadap Covid-19. Jumlah *tweet* yang dianalisis adalah sekitar 1,9 juta *tweets* berbahasa Inggris yang berkaitan dengan virus corona.



Tweet yang dikumpulkan dimulai dari tanggal 23 Januari hingga 7 Maret 2020. Dalam mengolah dan menganalisis datanya, ada tiga tahapan yang para peneliti lakukan yaitu (1) pengambilan sampel, (2) pengumpulan data, dan (3) *pre-processing* data. Ada beberapa batasan yang terdapat pada penelitian tersebut, yang pertama adalah peneliti hanya mengambil 19 sampel tren *hashtag* dalam pencarian *tweet* untuk penelitian ini, yang kedua bahwa pengguna Twitter tidak merepresentasikan seluruh populasi manusia dan hanya merepresentasikan opini dari orang-orang yang aktif bermain media sosial, dan yang terakhir *tweet* yang menggunakan bahasa selain Bahasa Inggris akan secara otomatis tidak diikutsertakan ke dalam data penelitian.

Metode yang digunakan pada penelitian (Rahmawati et al., 2021) yang berjudul “Analisis topik konten *channel* YouTube K-pop Indonesia menggunakan Latent Dirichlet Allocation”. Penelitian dilakukan dengan cara melakukan *crawling data* untuk mengumpulkan data kemudian dilanjutkan dengan Latent Dirichlet Allocation (LDA) dalam klasifikasi datanya. Dalam penelitian itu perhitungan *perplexity* cukup rendah, di mana model LDA yang baik itu dilihat dari nilai *perplexity* yang rendah. *Perplexity* merupakan nilai ukuran statistik tentang seberapa baik model probabilitas memprediksi sampel (Computing for the Social Science, 2021). Batasan yang ada pada penelitian ini adalah sampel hanya diambil dari 10 *channel* Youtube milik warga Indonesia yang berhubungan dengan K-Pop, di mana sampel ini kurang merepresentasikan keseluruhan konten dan media yang ada di Youtube.

Melihat hasil dari penelitian sebelumnya, penulis mengambil kesimpulan bahwa dengan menggunakan metode LDA, peneliti akan menghasilkan akurasi yang sama baiknya dengan penelitian sebelumnya. Dengan menggunakan salah satu metode *topic modeling* yaitu metode LDA, dapat membantu menentukan topik yang dibicarakan pada *hashtag* covidindonesia. LDA mampu meringkas, mengklusterkan, menghubungkan, dan memproses *data text* yang sangat besar, ditambah pula metode yang paling populer dalam *topic modelling* adalah metode LDA.

Penelitian berbasis metode LDA memang sudah banyak dilakukan pada beberapa *platform* media sosial, khususnya aplikasi Twitter. Penerapan metode LDA ini paling banyak ditemukan dan diimplementasikan pada aplikasi media sosial Twitter. Tetapi implementasinya pada *platform* Instagram masih jarang dilakukan, khususnya di Indonesia. Oleh sebab itu, peneliti memilih media sosial Instagram sebagai *platform* yang akan dianalisis mengenai unggahan pengguna yang menggunakan *hashtag* Covid-19. Peneliti memilih Instagram karena aplikasi ini memiliki keunggulan di mana dalam sekali unggahan, pengguna dapat mengunggah *caption*, lokasi, dan juga media berupa foto atau video. Diharapkan penelitian ini dapat memberikan informasi mengenai topik covid-19 yang dibicarakan di Instagram. Selain itu, praktisi akademik dapat menjadikan karya tulis ini sebagai rujukan dari penelitian yang berkaitan.

1.1 Tinjauan Pustaka

1.1.1 *Crawling Data*

Dalam melakukan analisis terhadap sebuah data, tentu diperlukan pengambilan data dari sumber yang diperlukan. Media sosial seperti Instagram dan Twitter kerap menjadi sumber dari data yang ingin dianalisis. Pengambilan data inilah yang disebut dengan *crawling* (Eka Sembodo et al., 2016). Proses dari *crawling* sendiri memerlukan waktu yang tidak sedikit, karena mesin *crawler* harus bisa menjalankan besarnya volume data yang ada dari target *platform* yang diinginkan (Zuliarso & Mustofa, 2009).

1.1.2 *Text Mining*

Text mining merupakan salah satu metode yang digunakan untuk melakukan pengelompokan atau pengklasifikasian data-data yang berupa tulisan, dokumen, atau teks. *Text mining* termasuk salah satu metode *data mining* yang fokusnya memang untuk mengolah data, teks, atau dokumen dalam jumlah yang besar. Sebelum melakukan *text mining*, perlu dilakukan pemrosesan awal data, baru kemudian data tersebut diolah dengan menggunakan algoritma yang tepat dan sesuai dengan luaran yang diinginkan (Harjanta, 2015).



1.1.3 Topic Modelling

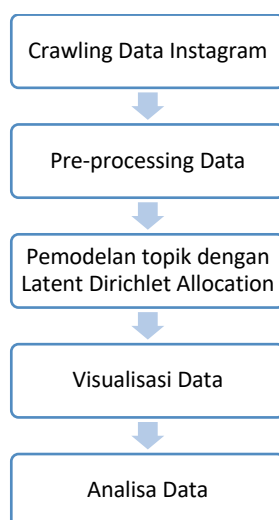
Topic modelling merupakan salah satu metode yang digunakan untuk memodelkan suatu topik dan termasuk salah satu dari *Natural Language Processing* dan memiliki fungsi utama untuk menganalisis suatu teks yang berupa algoritma. Nantinya permodelan ini akan digunakan untuk dapat mengidentifikasi pola-pola yang tidak tampak secara langsung dari serangkaian teks dengan menggunakan metode pendistribusian kata-kata yang ada di dalam teks atau dokumen tersebut. Luaran dari *topic modelling* ini adalah sekumpulan topik yang terdiri dari beberapa kelompok kata berdasarkan pola tertentu yang akan muncul secara bersamaan di dalam teks atau dokumen tersebut. Ada beberapa jenis *topic modelling*, seperti *Latent Semantic Allocation* (LSA), *Probabilistic Latent Semantic Analysis* (PLSA), dan *Latent Dirichlet Allocation* (LDA) (Nauri et al., 2021). Jenis yang paling populer saat ini adalah *Latent Dirichlet Allocation* (LDA) karena metode ini cukup mudah dengan melakukan klusterisasi dan meringkas data yang ada, selain itu metode *Latend Dirichlet Allocation* (LDA) ini juga cocok digunakan untuk memproses data yang banyak atau besar (Zulhanif et al., 2017).

1.1.4 Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) merupakan salah satu metode untuk mendeteksi topik dalam sekumpulan data serta melakukan pemodelan topik. *Latent Dirichlet Allocation* (LDA) merupakan metode *topic modeling* dan topik analisis yang paling populer saat ini dan banyak digunakan dalam melakukan analisis pada dokumen yang berukuran sangat besar (Rahmawati et al., 2021). *Latent Dirichlet Allocation* (LDA) saat ini sering digunakan karena dapat melakukan klusterisasi, melakukan peringkasan, menghubungkan, dan dapat memproses data dengan memberikan bobot pada masing-masing dokumen yang nantinya menghasilkan daftar topik. Ide dasar dari *Latent Dirichlet Allocation* (LDA) ini menganggap bahwa dokumen yang kita ujikan dapat direpresentasikan sebagai sebuah model yang dicampur dari berbagai topik yang dibutuhkan, oleh sebab itu teknik ini disebut sebagai laten (Zulhanif et al., 2017).

2. METODE PENELITIAN

Penelitian ini dimulai dengan mengumpulkan informasi mengenai penelitian serupa yang pernah dilakukan dan studi literatur. Studi literatur didapatkan melalui referensi terpercaya yaitu paper ataupun jurnal, artikel ilmiah, dan *website* resmi yang membahas topik covid indonesia, *topic modelling*, dan *Latent Dirichlet Allocation*. Kemudian dilakukan *crawling data* atau pengumpulan data pada Instagram yakni terkait tagar covidindonesia. Data dikumpulkan dengan menggunakan *command* *insta-crawl hashtag* yang selanjutnya dilakukan pemodelan topik serta analisa. Alur metodologi penelitian disajikan pada Gambar 1.



Gambar 1 Alur Metodologi Penelitian



2.1 *Crawling* Data Instagram

Selanjutnya, dilakukan pengumpulan data atau *crawling* yang digunakan untuk mengumpulkan informasi pada sebuah *website* yaitu Instagram. Pencarian pada Instagram dengan menggunakan *hashtag* covidindonesia akan menjadi dasar pencarian semua *caption* yang kemudian akan dilakukan *indexing* untuk mencari data yang ada pada setiap *post* yang telah diunggah. Data yang dicari ialah data yang terkait dengan *hashtag* covidindonesia, melalui data ini nantinya akan dicari topik apa yang dibahas. Kemudian, akan dilakukan penambangan data berupa *username*, *caption*, dan waktu unggahan *caption* yang akan dihimpun untuk proses penelitian berikutnya.

2.2 *Pre-processing* Data

Data yang telah dihimpun kemudian dilakukan proses *cleaning*, yang mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (Jasmir, 2016). Dari sekumpulan data yang telah didapatkan kemudian dilakukan proses tokenisasi yang bertujuan untuk memisahkan teks ke dalam unit-unit kecil yang biasa disebut dengan token (Rangkuti, 2020). Setelah melakukan tokenisasi, diperlukan pembuangan *stopwords*. Tidak sedikit dalam data teks kita menemukan kata-kata yang sebetulnya tidak diperlukan dalam penelitian. Kata-kata seperti yang, untuk, dari, dan sebagainya tentu bukanlah kata yang berdampak serius pada penelitian bahkan dapat dikatakan tidak dibutuhkan sama sekali.

2.3 Pemodelan Topik dengan *Latent Dirichlet Allocation*

Data yang telah dipisah ke dalam unit-unit kecil kemudian diterapkan ke LDA. Penerapan LDA berfungsi untuk pengenalan pola pada perhitungan statistika dengan cara menemukan proyeksi linear dari data yang akan memaksimalkan jarak antar kelas dan meminimalkan jarak data yang memiliki kesamaan. Tahap pertama dalam LDA ialah mengisi parameter yang akan digunakan sebagai batas data yang digunakan. Kemudian, menerapkan *semi random distribution*, metode ini didasari atas metode distribusi *Dirichlet*. Selanjutnya dilakukan iterasi untuk melihat parameter yang dapat menentukan distribusi dari topik dan kata dari dokumen.

Untuk menentukan banyaknya topik dapat dilakukan dengan menggunakan perhitungan *perplexity*. Nilai *perplexity* didapatkan dari perhitungan dengan menentukan kemungkinan dari log teks dokumen yang tidak terlihat, jika hasil yang didapatkan ialah nilai *perplexity* rendah, dapat dipastikan model tersebut adalah baik. Sementara itu, alur pemodelan topik dengan LDA dimulai dari memuat data yang sudah disimpan ke dalam variabel-variabel tertentu yang sesuai dengan topik penelitian ini yaitu Covid-19, kemudian setelah data-datanya dimuat akan dilanjutkan dengan membentuk model topik dari parameter yang sudah ditentukan, misalnya jumlah kata, jumlah topik, jumlah *hashtag*, dan lain-lain. Kemudian dilakukan pendokumentasian *logging* untuk memantau dan mengawasi aktivitas-aktivitas yang terjadi selama proses pembentukan model dari topik itu sendiri (Putra & Kusumawardani, 2017). Pemodelan topik dengan LDA ini sesuai digunakan untuk menganalisis topik campuran yang memuat beberapa kata sehingga pemodelan topik dengan LDA ini diimplementasikan pada penelitian ini (Nurlayli & Nasichuddin, 2019).

2.4 Visualisasi Data

Data yang didapatkan dari hasil LDA akan menjadi data mentah yang kemudian akan dilakukan visualisasi dengan menggunakan grafik tipe *bar-chart* untuk mempermudah proses analisis data. Dalam visualisasi ini menampilkan perbandingan persebaran topik yang satu dengan yang lainnya.

2.5 Analisa Data

Input dalam proses Analisa data ini ialah data mentah yang didapatkan dari penerapan LDA beserta visualisasi dari hasil LDA yang telah dilakukan pada tahap sebelumnya. Pada proses



analisis ini, data dianalisis menggunakan metode *topic modelling* yang dapat mengelompokkan topik-topik tertentu berdasarkan kemiripan kata kunci atau pola. Hasil visualisasi data yang sebelumnya berbentuk grafik bertipe *bar-chart* akan dianalisis terkait perbandingan topik yang telah ditemukan. Hasil dari analisis ini akan terus dievaluasi melalui kode program pada zenodo (Husna et al., 2020) untuk menemukan topik-topik yang relevan dan topik yang dibahas dari *hashtag covidindonesia*. *Source code* yang diimplementasikan untuk menghitung nilai *perplexity* dan nilai *topic coherence* ditampilkan pada Gambar 2.

```

corpus = [dictionary.doc2bow(text) for text in texts]
for k in range(3, 4):
    ldamodel = gensim.models.ldamodel.LdaModel(corpus, num_topics=k, id2word = dictionary,
        passes=20, iterations=100, alpha=[0.01]*k, eta=[0.01]*len(dictionary.keys()))
    coherence_model_lda = CoherenceModel(model=ldamodel, texts=texts,
        dictionary=dictionary, coherence='c_v')
    coherence_lda = coherence_model_lda.get_coherence()
    print(k,ldamodel.log_perplexity(corpus),coherence_lda)
    
```

Gambar 2 Rumus Perhitungan Nilai *Perplexity* dan Nilai *Topic Coherence*

3. HASIL DAN PEMBAHASAN

3.1 Hasil *Crawling*

Tabel 1 Contoh *Caption* *Hashtag* Covidindonesia

No.	<i>Caption</i> Instagram
1	Selama sepekan terakhir kasus kematian harian di Indonesia terus meningkat,Bahkan se Asia Indonesia peringkat ke 2 setelah India# <i>covid19</i> # <i>covidindonesia</i> # <i>pandemic</i> # <i>coronaindonesia2w</i>
2	Badan Pemerika Keuangan (BPK) mencatat total anggaran penanganan Covid-19 mencapai Rp 1.035,2 triliun. Auditor Utama Keuangan Negara III BPK Bambang Pamungkas mengatakan, anggaran penanganan Covid-19 itu berasal dari anggaran pendapatan dan belanja negara (APBN) sebesar Rp 937,42 triliun.â€œKemudian dari anggaran pendapatan dan belanja daerah (APBD) sebesar Rp 86,36 triliun dan dari sektor moneter sebesar Rp 6,50 triliun,â€ katanya dalam konferensi pers via daring, Selasa (29/12).Selain itu, anggaran tersebut juga berasal dari badan usaha milik negara (BUMN) dengan total anggaran sebesar Rp 4,02 triliun. Adapun dari badan usaha milik daerah (BUMD) sekitar Rp 320 miliar dan berasal dari dana hibah dan masyarakat sebesar Rp 625 miliar.@ <i>momovele.id</i> # <i>momovele</i> # <i>anggaran</i> # <i>covid19indonesia</i> # <i>covidindonesia</i> # <i>covid_19</i> # <i>covid19news</i> # <i>ppkm</i> # <i>ppkmdarurat</i> # <i>ppkmjawabali</i> # <i>ppkmdaruratjawabali1h</i>
3	Corona makin menggila.. Semoga ini bermanfaat yahSejumlah cara lain untuk mengaktifkan sistem imun tubuh. Salah satunya menjaga pola makan dengan gizi seimbang.Selain itu, minum air putih sedikitnya 6 gelas/hari, olahraga setidaknya 3 kali dalam seminggu minimal 30 menit, dan menjaga kebersihan tubuh secara keseluruhan dengan mandi setiap hari.Tidak kalah penting, mencuci tangan dengan sabun atau handsanitizer setiap kali akan makan atau minum dan keluar dari kamar mandi, serta istirahat atau tidur yang cukup 6-8 jam/hari.Bagikan tips ini ke keluarga/teman/kerabat kalian.dapatkanMadu Murni dan Rempah Herbaldi @ <i>kampusmadu</i> # <i>jsrsehat</i> # <i>imuntubuh</i> # <i>covid</i> # <i>coronavirus</i> # <i>covidindonesia</i> # <i>immunitysistem</i> # <i>jagaimunitas</i> # <i>imunitastubuh</i> # <i>kampusmadu</i> # <i>jsrherbal</i> # <i>resepherbal</i> # <i>herbalcovid</i> # <i>obatcovid</i> # <i>infokesehatan</i> # <i>informasiupdate</i> # <i>updateinfo</i> # <i>infolengkapjsr</i> # <i>zaidulakbar</i> # <i>jsrresep</i> # <i>resepjsr</i> # <i>covidjsr</i> # <i>jintenhitam</i> # <i>ladahitamorganik</i> # <i>maduasli</i> # <i>manfaatmadu</i> # <i>manfaatjinten</i> # <i>produkAllah1w</i>

Pada penelitian ini, *crawling* data dilakukan dengan menggunakan *hashtag covidindonesia* pada Instagram. Data yang didapatkan dari pencarian di Instagram dengan *filter* unggahan terpopuler dan terbaru dari beragam pengguna Instagram. Berdasarkan hasil *crawling*, diperoleh 150 *caption* yang memiliki *hashtag covidindonesia*. Tabel 1 merupakan contoh *caption* pada



instagram yang diambil dari tanggal 02/03/2021 sampai 11/07/2021 dengan batasan menggunakan *hashtag* covidindonesia.

3.2 Hasil Pre-processing Data

Setelah mengumpulkan data *caption* Instagram dengan proses *crawling*, maka data tersebut tidak dapat langsung digunakan dalam proses analisis. Data yang telah diperoleh harus diproses terlebih dahulu untuk mendapatkan hasil analisis yang maksimal (Husna et al., 2020). Tahap *pre-processing* data ini mencakup membersihkan data dan menghilangkan *stopwords*. Proses tokenisasi juga dilakukan pada data yaitu dengan menggunakan spasi untuk mengganti jarak enter yang ada pada *caption*. Selanjutnya, dilakukan penghapusan *stopwords* atau kata-kata yang tidak memberikan informasi penting untuk penelitian ini. Kata-kata seperti yang, untuk, dari, dan sebagainya dimasukkan menjadi *stopwords* secara manual pada kode dan dihilangkan dari token. Untuk contoh sampel data yang sudah dilakukan proses *preprocessing* dapat dilihat dalam Tabel 2, di mana pada tabel tersebut terdapat 10 sampel data teratas yang telah dilakukan *pre-processing*. Pada Tabel 2 tersebut, penulis memasukkan 3 informasi yaitu *dominant topic*, jumlah *dominant topic*, dan *keywords* yang digunakan pada unggahan di Instagram. Fungsi *len()* juga digunakan dalam kode untuk memisahkan tanda baca yang tidak diperlukan serta selanjutnya dapat dilanjutkan ke proses LDA.

Tabel 2 Sampel Hasil Data Pre-Processing

<i>Dominant Topic</i>	Jumlah <i>Dominant Topic</i>	<i>Keywords</i>
0,0	7	covidindonesia, covid_19, kerumunan, covid19
1,0	2	covidindonesia, kesehatan, ppkm, covid19
2,0	1	covidindonesia, covid_19, kesehatan, covid19indonesia

3.3 Hasil Penerapan LDA

Pemodelan LDA dilakukan dengan mengubah data menjadi *dictionary* dan *corpus*. Dokumen token yang telah didapatkan melalui tahap *pre-processing*, selanjutnya diubah menjadi sebuah id atau kata kunci dari *dictionary*. Lalu, dokumen token akan diubah menjadi matriks *document-term* untuk selanjutnya dapat membuat model LDA. *Packages* LdaModel digunakan untuk membuat model LDA dan kemudian akan dilakukan perhitungan skor koherensi dari topik menggunakan *CoherenceModel*. Perhitungan skor koherensi topik juga dibantu oleh *packages* Gensim pada Python.

3.4 Hasil Analisis

Berdasarkan penerapan LDA yang telah dilakukan pada data, dapat diketahui jumlah topik yang ada pada *caption* Instagram yang mengandung *hashtag* covidindonesia serta nilai *perplexity* dan nilai *topic coherence* yang menunjukkan seberapa baik model yang dihasilkan. Semakin kecil nilai *perplexity* maka semakin bagus model yang dihasilkan. Model yang baik juga memiliki nilai *coherence* yang tinggi. Dengan menggunakan variasi 3, 4, dan 5 topik didapatkan nilai *perplexity* dan nilai *topic coherence* yang berbeda-beda seperti ditunjukkan pada Tabel 3.

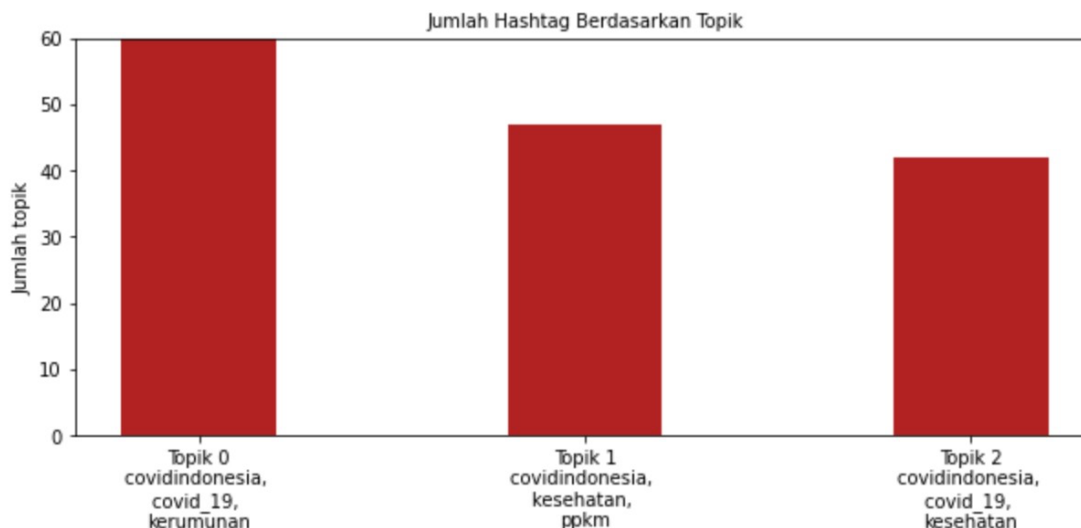
Tabel 3 Nilai *Perplexity* dan *Topic Coherence*

Jumlah Topik	Nilai <i>Perplexity</i>	Nilai <i>Topic Coherence</i>
3	-8.704555042489837	0.6555015739681846
4	-8.587164083739	0.6227750210917353
5	-8.501029065460557	0.5541026680947877

Berdasarkan hasil pada Tabel 3, diketahui bahwa nilai *perplexity* dan nilai *coherence* dari model menggunakan 3 topik mengindikasikan bahwa model tersebut lebih baik dibandingkan dengan 4 dan 5 topik. Dengan 3 topik, model memiliki nilai *perplexity* yang sangat kecil dan nilai *topic*



coherence yang besar. Hasil distribusi topik pembicaraan pada *caption* dengan *hashtag* dapat dilihat pada Gambar 3. Grafik pada Gambar 3 menunjukkan persebaran 3 topik yang ada pada *caption* Instagram yang mengandung *hashtag* covidindonesia beserta jumlahnya. Daftar topik yang dominan juga didapatkan dari hasil penerapan LDA seperti disajikan pada tabel Tabel 4.



Gambar 3 Jumlah *Hashtag* Berdasarkan Topik

Tabel 4 Daftar Topik Dominan

No.	Dominan Topic	Keywords	Jumlah
1	0,0	covidindonesia, covid_19, kerumunan	66
2	1,0	covidindonesia, kesehatan, ppkm	47
3	2,0	covidindonesia, covid_19, kesehatan	42

Berdasarkan Tabel 4, didapatkan bahwa setiap topik memiliki 3 kata kunci yang berkaitan dengan *hashtag* covidindonesia. Topik 0 memiliki kata kunci *covidindonesia*, *covid_19*, dan *kerumunan*. Topik 0 memiliki jumlah sebanyak 66. Topik 1 dibentuk dari 47 *caption* yang memiliki kata kunci *covidindonesia*, *kesehatan*, dan *ppkm*. Pada topik 2 membahas mengenai *covidindonesia*, *covid_19*, serta *kesehatan*. Topik ini memiliki jumlah yaitu 42.

4. KESIMPULAN

Berdasarkan analisis yang dilakukan dengan menggunakan metode *crawling data* dan *Latent Dirichlet Allocation (LDA)* pada platform Instagram, didapatkan bahwa model LDA dengan jumlah topik sebanyak 3 topik merupakan model dengan nilai terbaik. Tiga topik teratas yang berkaitan dengan tagar #covidindonesia di antaranya berupa *hashtag* atau pun suatu kata tertentu. *Hashtag* yang ditemukan di antaranya adalah *covidindonesia*, *covid_19*, *kerumunan*, *ppkm*, dan *kesehatan*. Apabila diurutkan dari topik yang paling dominan, posisi pertama didominasi oleh kata kunci *covidindonesia*, *covid_19*, dan *kerumunan*. Di posisi kedua didominasi dengan kata kunci *covidindonesia*, *kesehatan*, dan *ppkm*. Sementara untuk posisi ketiga didominasi dengan kata kunci *covidindonesia*, *covid_19* dan *kesehatan*. Dari hasil analisis dan penelitian yang sudah dilakukan, topik-topik dan kata kunci di atas memang topik dan kata kunci yang relevan dengan tagar *covidindonesia*. Topik yang telah didapatkan melalui metode pemodelan topik *Latent Dirichlet Allocation (LDA)* menunjukkan topik utama yang sedang ramai diperbincangkan. Hal ini dapat bermanfaat bagi pembaca khususnya pengguna media sosial Instagram di Indonesia terkait perkembangan pandemi terkini.

Data yang digunakan dalam penelitian ini bersumber dari *caption* status di Instagram yang memiliki karakteristik bahasa informal sehari-hari. Berdasarkan pengamatan yang dilakukan



penulis, didapatkan dugaan sementara bahwa terdapat perbedaan karakteristik data yang bersumber dari berbagai media sosial, baik dalam hal karakteristik kebahasaan yang diterapkan maupun panjang karakter per dokumen, sehingga hal ini dapat menjadi masukan bagi penelitian selanjutnya.

DAFTAR PUSTAKA

- Abdul Talib, Y. Y., & Mat Saat, R. (2017). Social proof in social media shopping: An experimental design research. *SHS Web of Conferences*, 34, 02005. <https://doi.org/10.1051/shsconf/20173402005>
- Alfanzar, A. I., Khalid, K., & Rozas, I. S. (2020). Topic Modelling Skripsi Menggunakan Metode Latent Dirichlet Allocation. *JSil (Jurnal Sistem Informasi)*, 7(1), 7. <https://doi.org/10.30656/jsii.v7i1.2036>
- Azizah, K. N. (2020). *Kronologi 2 Pasien Pertama Virus Corona COVID-19 di Indonesia*. DetikHealth. <https://health.detik.com/berita-detikhealth/d-4922758/kronologi-2-pasien-pertama-virus-corona-covid-19-di-indonesia>
- Boer, K. M., Pratiwi, M. R., & Muna, N. (2020). Analisis Framing Pemberitaan Generasi Milenial dan Pemerintah Terkait Covid-19 di Media Online. *Communicatus: Jurnal Ilmu Komunikasi*, 4(1), 85–104. <https://doi.org/10.15575/cjik.v4i1.8277>
- Carman, A. (2018). *Instagram now has 1 billion users worldwide*. The Verge. <https://www.theverge.com/2018/6/20/17484420/instagram-users-one-billion-count>
- Computing for the Social Science. (2021). *Topic modeling*. Computing for the Social Science. <https://cfss.uchicago.edu/notes/topic-modeling/>
- Eka Sembodo, J., Budi Setiawan, E., & Abdurahman Baizal, Z. (2016). Data Crawling Otomatis pada Twitter. *INDOSC 2016, October*, 11–16. <https://doi.org/10.21108/INDOSC.2016.111>
- Harjanta, A. T. J. (2015). Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining. *Jurnal Informatika UPGRIS*, 1(Juni), 1–9.
- Husna, A., Santoso, K. R. A. P., Putri, N. W., & Rakhmawati, N. A. (2020). *asmaul-husna/covidindonesia-LDA: Final | Zenodo*. <https://doi.org/10.5281/zenodo.4679443>
- Ibrahim, I. S. (2018). *Kritik Budaya Komunikasi, Budaya, Media, dan Gaya Hidup Dalam Proses Demokratisasi di Indonesia* (S. O. Pavitasari (ed.)). Jelasutra.
- Jasmir. (2016). Implementasi Teknik Data Cleaning dan Teknik Roughset pada Data Tidak Lengkap dalam Data Mining. *Seminar Nasional APTIKOM (SEMNASTIKOM)*, 1(1), 99–106.
- Naury, C., Fudholi, D. H., & Hidayatullah, A. F. (2021). Topic Modelling pada Sentimen Terhadap Headline Berita Online Berbahasa Indonesia Menggunakan LDA dan LSTM. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(1), 24. <https://doi.org/10.30865/mib.v5i1.2556>
- Nugraha, B., & Akbar, M. F. (2019). Perilaku Komunikasi Pengguna Aktif Instagram. *Jurnal Manajemen Komunikasi*, 2(2), 95. <https://doi.org/10.24198/jmk.v2i2.21330>
- Nurlayli, A., & Nasichuddin, M. A. (2019). Topik Modeling Penelitian Dosen JPTEI UNY pada Google Scholar Menggunakan Latent Dirichlet Allocation. *Elinvo (Electronics, Informatics, and Vocational Education)*, 4(2), 154–161. <https://doi.org/10.21831/elinvo.v4i2.28254>
- Putra, K. B., & Kusumawardani, R. P. (2017). Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA). *Jurnal Teknik ITS*, 6(2). <https://doi.org/10.12962/j23373539.v6i2.23205>
- Rakhmawati, A., Nikmah, N. L., Perwira, R. D. A., & Rakhmawati, N. A. (2021). Analisis topik konten channel YouTube K-pop Indonesia menggunakan Latent Dirichlet Allocation. *Teknologi*, 11(1), 16–25. <https://doi.org/10.26594/teknologi.v11i1.2155>
- Rangkuti, M. D. E. (2020). *Analisis topik komentar video beberapa akun youtube e-commerce indonesia menggunakan metode latent dirichlet allocation*. UIN Syarif Hidayatullah.
- Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., & Zhu, T. (2020). Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. *PLOS ONE*, 15(9), e0239441. <https://doi.org/10.1371/journal.pone.0239441>
- Zulhanif, Sudartianto, Tantular, B., & Jaya, I. G. N. M. (2017). Aplikasi Latent Dirichlet Allocation (LDA) pada Clustering Data Teks. *Jurnal Logika*, 7(1), 46–51.
- Zuliarso, E., & Mustofa, K. (2009). Crawling Web berdasarkan Ontology. *Jurnal Teknologi Informasi DINAMIK*, XIV(2), 105–112. <https://doi.org/10.35315/dinamik.v14i2.97>

