

# Analysis of the Relationship between Public Sentiment on Social Media and Indonesian Covid-19 Dynamics

<sup>1</sup>Nana Mulyana Maghfur, <sup>2</sup>Fakhri Muhammad, <sup>3</sup>Apriade Voutama

<sup>1,2</sup>Program Studi Teknik Informatika, Universitas Singaperbangsa Karawang

<sup>3</sup>Program Studi Sistem Informasi, Universitas Singaperbangsa Karawang

Email: nana.mulyana18171@student.unsika.ac.id

## Abstract

*The bad side of the open access nature of social media is that it frees anyone to have an opinion as they please and is often accompanied by other agendas such as spreading panic, false information, fake news, hate speech and even distorting public opinion. This condition can be fatal in the pandemic era where public opinion can worsen the pandemic situation. Therefore, it is important to know whether it is true that changes in public sentiment in response to news on social media can affect the dynamics of the spread of COVID-19. We use sentiment analysis using machine learning methods to extract daily sentiment data and test its correlation with daily Covid-19 case data in Indonesia. The results of the associative hypothesis test with a Pearson correlation value of 0.151 show that public sentiment on social media towards the news of the COVID-19 variant is positively correlated with the dynamics of the daily Covid-19 cases. Therefore, the author invites all social media users, including the author himself, to be more vigilant and careful in giving opinions and accepting other people's opinions on social media.*

**Keywords:** Covid-19, Sentiment Analysis, SVM, Twitter, Hypothesis Test

## 1. INTRODUCTION

Pandemi covid-19 yang telah berlangsung sekitar 2 tahun kebelakang sudah merenggut setidaknya 5.411.759 jiwa manusia [1] serta menyebabkan pukulan berarti bagi ekonomi dunia [2], [3], Indonesia sendiri sempat mengalami dua kali gelombang puncak kasus harian yang cukup buruk sampai dikatakan sebagai episentrum covid-19 di Asia pada gelombang kedua saat varian delta menyerang [4]. Selain itu pandemi ini juga telah merubah cara masyarakat bersosialisasi dengan adanya ketentuan untuk *social distancing* dan *lockdown* di berbagai negara. Oleh sebab itu, media sosial yang sebelumnya sudah berkembang secara pesat menjadi semakin dibutuhkan masyarakat guna melampiasikan kebutuhan bersosialisasi saat adanya banyak keterbatasan di dunia nyata. Peranan media sosial menjadi penting di era pandemi sebagai sarana efektif menyebarkan informasi, berita dan anjuran pihak berwenang dengan tujuan untuk membangun kesadaran dalam diri dan lingkungan masyarakat Indonesia agar selalu mematuhi protokol kesehatan dimana tujuan akhirnya tentu agar penyebaran virus ini bisa diatasi dan keadaan pandemi bisa cepat berakhir [5].

Namun sifat media sosial yang mempunyai akses terbuka bagaikan pedang bermata dua, disatu sisi hal ini memberikan kemudahan akses bagi siapa saja dan dimana saja, disisi lain hal ini membebaskan siapa saja untuk berpendapat sesuka mereka dan kerap kali disertai agenda lain seperti

menyebarkan kepanikan, informasi palsu, berita bohong, ujaran kebencian bahkan pengalangan opini publik [6]. Kondisi tersebut bisa berakibat fatal di era pandemi, karena jika terdapat suatu berita berisikan fakta anjuran kesehatan yang oleh oknum tertentu ingin dibelokan kebenarannya, mereka bisa dengan mudah memanfaatkan keterbukaan akses di media sosial ini [7]. Karena sentimen masyarakat terhadap berita berisi fakta menjadi berubah, orang-orang akan cenderung tidak mepercayai fakta tersebut dan lebih memilih mengabaikannya yang mana tentu saja akan memperparah kondisi pandemi. Oleh karena itu penting untuk diketahui apakah benar perubahan sentimen masyarakat menanggapi suatu berita di media sosial dapat mempengaruhi dinamika persebaran covid-19.

Artikel ini ditulis untuk membuktikan keterkaitan sentimen masyarakat di platform media sosial *twitter* dengan dinamika covid-19 yang mana dalam prosesnya memanfaatkan metode *data mining* dan *natural language processing* (NLP) untuk mengekstrak sentimen teks *reply* yang terdapat pada *tweet* berisi judul berita untuk selanjutnya sentimen harian akan disandingkan dengan data kasus harian covid-19 di Indonesia dan diuji keterkaitannya menggunakan metode uji hipotesis statistik asosiatif. Hasil Analisa dalam penelitian ini dapat memberikan informasi pada masyarakat apakah polaritas opini di sosial media dapat berdampak secara nyata di era pandemi yang jika benar adanya dapat mengajak masyarakat agar lebih berhati-hati dalam bersosial media agar tidak dengan mudah termakan upaya fabrikasi opini oleh oknum-oknum yang tidak bertanggung jawab.

Analisis sentimen pada teks *tweet* yang dikumpulkan dari platform media sosial *twitter* sudah sering dilakukan di banyak bahasa dan metodenya pun beragam dari yang berbasis *lexicon* kata positif dan negatif sampai metode *machine learning* dan *deep learning* baik secara *supervised* maupun *unsupervised* [8]. Untuk analisis sentimen pada bahasa Indonesia khususnya untuk data teks *tweet*, pendekatan *supervised learning* berupa klasifikasi teks menjadi tiga kategori yaitu positif, negatif dan netral biasanya lebih banyak dipilih dibandingkan pendekatan lainnya dikarenakan kemudahan akses algoritma serta proses *fine tuning* [9].

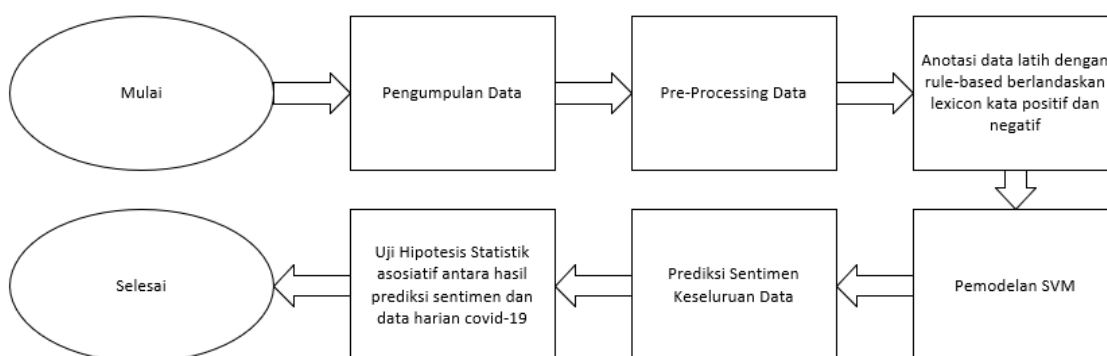
Seperti pada artikel [10] yang menggunakan tiga algoritma *supervised learning* yaitu *support vector machine* (SVM), Naïve Bayes, dan *k-nearest neighbor* (KNN) untuk mengekstraksi sentimen dari *tweet* seputar covid-19 secara keseluruhan untuk mengetahui tanggapan masyarakat mengenai adanya pandemi covid-19 dengan hasil yang menunjukkan bahwa performa algoritma SVM paling baik diantara ketiga algoritma yang digunakan dengan nilai akurasi sebesar 90,1% dibanding Naïve Bayes yang mendapat akurasi 79,2% serta KNN sebesar 62,1%. Lalu pada artikel [11] yang berusaha melakukan analisis sentimen untuk mengetahui opini masyarakat mengenai pembelajaran tatap muka melalui komentar pada media sosial youtube menggunakan implementasi *deep learning* berupa LSTM dan LDA untuk arsitektur *neural network* menghasilkan model dengan nilai akurasi nilai ROC AUC sebesar 84%. Selanjutnya pada artikel [12], penulis artikel tersebut berusaha menganalisa

sentimen masyarakat pada satu topik viral menggunakan pendekatan *rule-based* berlandaskan *lexicon* kata positif dan kata negatif bahasa Indonesia dimana performa *rule-based* ini tidak diukur dengan *metric* sehingga tidak diketahui seberapa baik metode ini dibandingkan dengan metode lain namun metode ini cepat dan mudah dioperasikan karena tidak memerlukan data dengan label seperti pendekatan *supervised learning* dan tidak memerlukan proses pelatihan model terlebih dahulu.

Dari beberapa metode yang telah dijabarkan berdasarkan artikel-artikel sebelumnya yang terkait dengan penelitian ini, SVM merupakan metode dengan performa tertinggi dilihat dari nilai akurasi. Namun kualitas dari metode *machine learning* dan *deep learning* sangat bergantung pada kualitas data latih yang digunakan untuk pemodelan [13], apalagi jika pendekatannya menggunakan *supervised learning* dimana proses pelabelan data menjadi sangat krusial bagi kesesuaian model dalam memprediksi nantinya [14]. Hal ini menyebabkan proses pelabelan data haruslah dilakukan manual oleh ahli atau menggunakan metode lain yang tidak memerlukan pelabelan [15]. Opsi yang tersedia untuk pelabelan menggunakan metode lain yaitu dengan pendekatan *unsupervised* atau *rule-based*, namun pendekatan *unsupervised* sendiri sudah selayaknya suatu penelitian terpisah dimana kompleksitasnya lebih besar dibanding pendekatan *supervised* sendiri. Maka pelabelan dengan pendekatan *rule-based* berlandaskan *lexicon* dirasa paling mudah dijangkau agar penelitian lebih efisien. Maka dari itu dalam penelitian ini kami menggabungkan proses pengambilan sentimen menggunakan pendekatan *rule-based* sebagai langkah awal untuk mendapatkan sentimen sebagian data untuk kemudian melatih model berbasis SVM guna memprediksi sentimen dari data lainnya seperti yang dilakukan pada artikel [16].

## 2. METHODS

Penelitian mengikuti alur yang dijabarkan pada Gambar 1, dimana keseluruhan proses dilakukan menggunakan bahasa pemrograman *python* secara lokal.



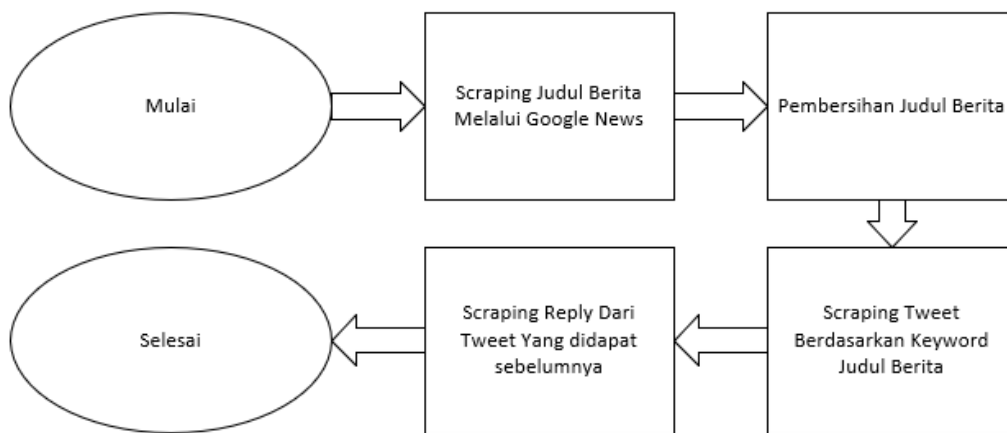
Gambar 1 Alur Penelitian

### 2.1 Ruang Lingkup

Dalam penelitian ini kami teks *tweet* yang kami gunakan bukanlah mengenai covid-19 secara umum, melainkan kami fokuskan topik bahasan seputar varian covid-19. Kami menetapkan ini karena topik mengenai varian covid-19 adalah salah satu bahasan yang hangat tahun ini yang disebabkan varian baru hasil mutasi virus sampai saat ini belum berhenti bermunculan dan sifat varian baru yang cenderung semakin mematikan serta berpotensi menimbulkan gelombang baru persebaran covid-19. Jangka waktu yang kami ambil yaitu antara 1 Januari 2021 sampai 31 Oktober 2021 karena varian covid-19 merambah ke Indonesia dimulai dari awal 2021 [17].

### 2.2 Pengumpulan Data

Data yang kami gunakan dikumpulkan melalui proses *scraping* media sosial twitter menggunakan tool *snsrape* dan *tweepy* yang terdapat dalam bahasa pemrograman *python*. Alur lengkap proses pengumpulan data dapat dilihat pada Gambar 2 dimana proses dimulai dengan terlebih dahulu melakukan *scraping* judul berita berbahasa Indonesia dari *Google News* dengan keyword “varian covid”. Hasil *scraping* judul berita kemudian dibersihkan dari adanya format url, nama situs berita dan juga tanda baca. Contoh judul berita hasil *scraping* dapat dilihat pada Tabel 1. Langkah selanjutnya setiap judul berita dijadikan *keyword* untuk pencarian *tweet* saat melakukan *scraping* di *twitter*. Namun yang kami tetapkan sebagai dataset bukanlah *tweet* utama yang cocok dengan *keyword* melainkan teks *reply* dari *tweet* utama yang berisikan *keyword* judul berita.



Gambar 2 Alur Pengumpulan Data

Tabel 1 Contoh Judul Berita Hasil *Scraping*

<b>Teks Judul Berita</b>	<b>Teks Judul Berita Setelah Proses <i>Cleaning</i></b>
Beijing Lockdown! Virus Covid-19 Impor Dari Indonesia - CNBC Indonesia	Beijing Lockdown Virus Covid Impor Dari Indonesia
Varian Baru Virus Corona Jerman Masih Diselidiki - CNN Indonesia	Varian Baru Virus Corona Jerman Masih Diselidiki

### 2.3 *Pre-Processing Data*

Data hasil *scraping* tidak bisa langsung digunakan untuk analisis sentimen karena seperti yang umum diketahui bahwa *tweet* orang-orang kebanyakan berisikan banyak *noise* (format url, *email*, *mention*, *hashtag*, *stop words*), emotikon, tanda baca, kata-kata slang dan kesalahan pengejaan. Oleh karena itu penting dilakukan proses *cleaning* dan normalisasi teks agar informasi yang terdapat pada data teks saat pemodelan dapat terekstraksi dengan maksimal. Seluruh proses *cleaning* dilakukan dengan memanfaatkan *regular expression* (regex) dan untuk menormalisasikan kata-kata slang digunakan *lexicon* dari artikel [18]. Contoh teks *reply tweet* yang didapatkan dari proses pengumpulan data dan setelah melalui proses *cleaning* dapat dilihat pada Tabel 2.

Tabel 2 Contoh Teks *Tweet* Sebelum dan Sesudah *Cleaning*

<b>Teks Hasil <i>Scraping</i></b>	<b>Teks Setelah <i>Cleaning</i></b>
@cnbcindonesia akhir nya kita swasembada ☐ Swasembada Covid 19 ☐	akhir nya kita swasembada swasembada covid 19
@basukane @KompasTV Kek makanan aja ade varian-nya. Rasa coklat & bluberi. ☐☐☐☐☐☐	kayak makanan saja ada varian-nya rasa coklat ama bluberi

### 2.4 *Pemodelan Analisis Sentimen*

Dalam proses pemodelan, seperti yang disebutkan sebelumnya bahwa akan digunakan kombinasi pendekatan *rule-based* dengan metode *supervised machine learning* SVM dimana *rule-based* berfungsi sebagai annotator berbasis mesin untuk membuat data latih yang dapat digunakan untuk melatih model SVM. Dengan berlandaskan pada *lexicon* [19] berisi daftar kata positif dan negatif dalam bahasa Indonesia, setiap kalimat *tweet* pada data yang dijadikan data latih akan dicocokkan berapa jumlah kata positif dan berapa jumlah kata negatif didalamnya. Kemudian dihitunglah polaritas dari setiap *tweet* dengan rumus (1) [20]. Contoh hasil dari penggunaan *rule-based* dapat dilihat pada Tabel 3.

$$Skor\ Polaritas = \frac{Jumlah\ Kata\ Positif - Jumlah\ Kata\ Negatif}{Jumlah\ Total\ Kata\ Positif\ Dan\ Negatif} \quad (1)$$

Dengan Catatan:

- Jika skor > 0, maka sentimen kalimat secara keseluruhan dianggap “positif”
- Jika skor < 0, maka sentimen kalimat secara keseluruhan dianggap “negatif”
- Jika skor = 0, maka sentimen kalimat secara keseluruhan dianggap “netral”

Tabel 3 Contoh Hasil Pelabelan Sentimen Menggunakan *Rule-Based*

Teks	Skor Polaritas	Sentimen
sudah benar ini dikembalikan lagi ke yang punya	1.0	Positif
bagaimana cara nya	0.0	Netral
kok ujung <sup>2</sup> x malah promo produk jadi jangan salahkan kami kalo bilang beberapa nakes itu sales	-1.0	Negatif

Setelah data latih selesai diproses menggunakan *rule-based* dan sentimen serta polaritas setiap *tweet* dalam data latih diketahui, selanjutnya pemodelan SVM dapat dilakukan. Karena diketahui sentimen yang merupakan tiga kelas klasifikasi juga diketahui nilai polaritas yang merupakan data numerik, maka pemodelan SVM bisa dilakukan secara klasifikasi maupun regresi. Kami akhirnya memutuskan untuk melakukan keduanya guna melihat perbandingan performa keduanya.

### 2.5 Uji Hipotesis Asosiatif

Langkah terakhir untuk mengetahui keterkaitan antara sentimen masyarakat terhadap berita seputar varian covid-19 di media sosial dengan dinamika kasus harian covid-19 yaitu dengan melakukan uji hipotesis statistik asosiatif menggunakan *pearson product moment test* dan data hasil prediksi model analisis sentimen baik yang dilakukan secara klasifikasi maupun regresi dengan data harian covid-19 di Indonesia dalam *range* waktu yang sama dengan *range* waktu *scraping* judul berita dan *tweet* yang dibahas sebelumnya.

## 3. RESULTS AND DISCUSSION

Setelah semua prosedur penelitian dilakukan, secara garis besar didapatkan dua *output* utama. Pertama, model analisis sentimen berbasis SVM baik secara klasifikasi maupun regresi yang telah dilatih menggunakan data latih yang dihasilkan pendekatan *rule-based* menggunakan *lexicon* kata positif dan negatif bahasa Indonesia. Kedua, hasil uji hipotesis yang dapat memenuhi tujuan utama

penelitian ini untuk membuktikan apakah benar sentimen publik di media sosial terhadap berita seputar varian covid-19 berhubungan dengan dinamika persebaran covid-19 itu sendiri.

### 3.1 Hasil Pemodelan

Kedua model yang dihasilkan melalui pelatihan SVM secara klasifikasi dan regresi kami bandingkan performanya dengan menggunakan data *test* yang dianotasikan manual dan menghasilkan prediksi dengan performa yang dapat diukur melalui skor R2, *mean absolute error* (MAE), *mean squared error* (MSE), *root mean squared error* (RMSE) yang dapat dilihat pada Tabel 4. Dari sini diketahui bahwa model klasifikasi lebih unggul berdasarkan pada skor R2 yang lebih tinggi dan MAE yang lebih rendah, namun untuk skor MSE dan RMSE performa model klasifikasi sedikit lebih rendah disbanding model regresi meski perbedaannya hanya 0,0007 untuk MSE dan 0,001 untuk RMSE.

Tabel 4 Perbandingan Performa Model SVM Klasifikasi dan Regresi

Teks	Regresi	Klasifikasi
R2	0.6727367264013685	0.7120246115229778
MAE	0.23409784677691967	0.12063344201210992
MSE	0.13117028444554185	0.13181183046110853
RMSE	0.36217438402728297	0.3630589903322992

Selain itu untuk model klasifikasi kami ujikan pula menggunakan metric akurasi, presisi, *recall*, dan *F1-score* dengan hasil yang dapat dilihat pada Gambar 3 dan dilengkapi penjabaran *confusion matrix* pada Gambar 4 dimana hasilnya akurasinya yang sebesar 88,5% sudah diatas beberapa model lain pada studi literatur dan hanya sedikit lebih rendah dari model SVM lain yang mencapai 90,1%.

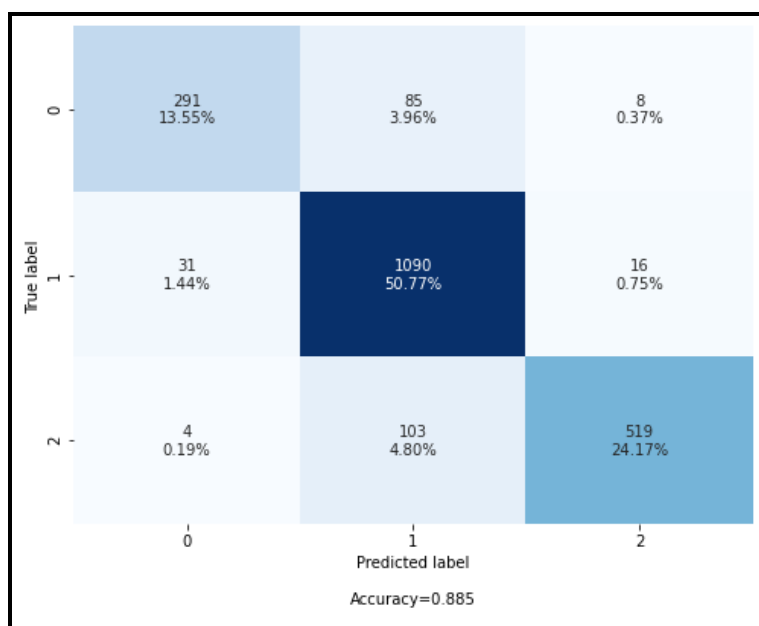
```

Classification report :
              precision    recall  f1-score   support

     0       0.89         0.76         0.82         384
     1       0.85         0.96         0.90        1137
     2       0.96         0.83         0.89         626

 accuracy                   0.88         2147
 macro avg       0.90         0.85         0.87         2147
 weighted avg    0.89         0.88         0.88         2147
    
```

Gambar 3 Report Hasil Klasifikasi



Gambar 4 *Confusion Matrux*

### 3.2 Hasil Pengujian Hipotesis Asosiatif

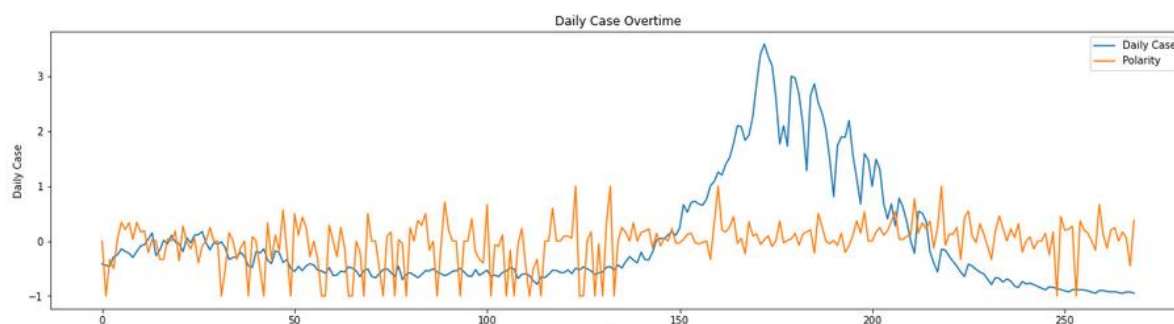
Dalam pengujian hipotesis asosiatif ini kami menggunakan tool *statistical package for the social sciences* (SPSS) dari IBM yang dapat memudahkan kami melakukan pengujian dibandingkan harus menghitung manual. Hasil Pengujian dapat dilihat pada Gambar 5 dimana dapat diketahui dengan nilai *pearson correlation* sebesar 0,151 menunjukkan bahwa sentimen publik di media sosial terhadap berita varian covid-19 memiliki korelasi positif dengan dinamika kasus harian covid-19 meski nilainya tidaklah besar. Dari hasil uji ini juga diketahui tingkat signifikansi hubungan dua variabel yaitu sebesar 0,013 yang jika dibandingkan dengan nilai kritis pada tingkat galat 0,05 menunjukkan bahwa hubungan kedua variabel signifikan pada taraf galat tersebut. Grafik kedua nilai variabel perharinya yang telah distandarisasi telah kami visualisasikan sebagaimana terlihat pada Gambar 6.

Correlations			
		Polarity	V4
V4	Pearson Correlation	1	.151 <sup>*</sup>
	Sig. (2-tailed)		.013
	N	269	269
Polarity	Pearson Correlation	.151 <sup>*</sup>	1
	Sig. (2-tailed)	.013	
	N	269	269

\*. Correlation is significant at the 0.05 level (2-tailed).

Gambar 5 Hasil Uji Korelasi Dengan SPSS





Gambar 6 Grafik Perbandingan Nilai Kedua Variabel Setiap Harinya

#### 4. CONCLUSION

Dari hasil penelitian dan juga pembahasannya, kesimpulan yang dapat ditarik yaitu selain model analisis sentimen yang dihasilkan mendapat performa cukup mumpuni namun juga telah membuktikan hipotesis awal yang menyatakan bahwa sentimen publik di media sosial terhadap berita varian covid-19 memiliki hubungan dengan dinamika kasus harian covid-19 telah dibuktikan benar adanya dan dalam taraf galat 0,05 hubungan keduanya signifikan. Informasi yang didapatkan dari penelitian ini yaitu bahwa dinamika opini publik di media sosial pada era pandemi ini bukanlah sesuatu yang bisa dianggap remeh karena dapat berimplikasi nyata terhadap kondisi pandemi itu sendiri. Maka dari itu penulis mengajak semua pengguna media sosial termasuk penulis sendiri agar lebih waspada dan berhati-hati dalam beropini serta menerima opini orang lain di media sosial.

#### REFERENCES

- [1] WHO, "WHO Coronavirus (COVID-19) Dashboard," *Geneva World Health Organization*, 2020.
- [2] W. McKibbin and R. Fernando, "The economic impact of COVID-19," *Econ. Time COVID-19*, vol. 45, no. 10.1162, 2020.
- [3] S. Maital and E. Barzani, "The global economic impact of COVID-19: A summary of research," *Samuel Neaman Inst. Natl. Policy Res.*, vol. 2020, pp. 1–12, 2020.
- [4] O. Dyer, "Covid-19: Indonesia becomes Asia's new pandemic epicentre as delta variant spreads," *BMJ*, vol. 374, p. n1815, Jul. 2021.
- [5] A. Wong, S. Ho, O. Olusanya, M. V. Antonini, and D. Lyness, "The use of social media and online communications in times of pandemic COVID-19," *J. Intensive Care Soc.*, vol. 22, no. 3, pp. 255–260, Aug. 2021.
- [6] H. Sahni and H. Sharma, "Role of social media during the COVID-19 pandemic: Beneficial, destructive, or reconstructive?," *Int. J. Acad. Med.*, vol. 6, no. 2, pp. 70–75, Apr. 2020.
- [7] A. Clark, "COVID-19-related Misinformation: Fabricated and Unverified Content on Social Media," *Anal. Metaphys.*, no. 19, pp. 87–93, 2020.
- [8] K. Chakraborty, S. Bhattacharyya, and R. Bag, "A Survey of Sentiment Analysis from Social Media Data," *IEEE Trans. Comput. Soc. Syst.*, vol. 7, no. 2, pp. 450–464, Apr. 2020.
- [9] E. Miranda, M. Aryuni, R. Hariyanto, and E. S. Surya, "Sentiment Analysis using Sentiwordnet and Machine Learning Approach (Indonesia general election opinion from the twitter content),"

- in *2019 International Conference on Information Management and Technology (ICIMTech)*, 2019, vol. 1, pp. 62–67.
- [10] F. S. Pamungkas and I. Kharisudin, “Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter,” *PRISMA Pros. Semin. Nas. Mat.*, vol. 4, pp. 628–634, Feb. 2021.
- [11] F. Z. Ahmad, M. F. S. Arifandy, M. R. Caesarardhi, and N. A. Rakhmawati, “Bagaimana Masyarakat Menyikapi Pembelajaran Tatap Muka: Analisis Komentar Masyarakat pada Media Sosial Youtube Menggunakan Algoritma Deep Learning Sekuensial dan LDA,” *J. Linguist. Komputasional*, vol. 4, no. 2, pp. 40–46, 2021.
- [12] Rifiana Arief and Karel Imanuel, “Analisis Sentimen Topik Viral Desa Penari Pada Media Sosial Twitter Dengan Metode Lexicon Based,” *J. Ilm. Matrik*, vol. 21, no. 3, Dec. 2019.
- [13] B. Vidgen and L. Derczynski, “Directions in abusive language training data, a systematic review: Garbage in, garbage out,” *PLOS ONE*, vol. 15, no. 12, p. e0243300, Des 2020.
- [14] K. Mishev, A. Gjorgjevikj, I. Vodenska, L. T. Chitkushev, and D. Trajanov, “Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers,” *IEEE Access*, vol. 8, pp. 131662–131682, 2020.
- [15] V. Prabhakaran, A. M. Davani, and M. Díaz, “On Releasing Annotator-Level Labels and Information in Datasets,” *ArXiv211005699 Cs*, Oct. 2021.
- [16] P. Chikersal, S. Poria, and E. Cambria, “SeNTU: sentiment analysis of tweets by combining a rule-based classifier with supervised learning,” presented at the Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), 2015, pp. 647–651.
- [17] I. S. Joyosemito and N. M. Nasir, “Gelombang Kedua Pandemi Menuju Endemi Covid-19: Analisis Kebijakan Vaksinasi dan Pembatasan Kegiatan Masyarakat di Indonesia,” *J. Sains Teknol. Dalam Pemberdaya. Masy.*, vol. 2, no. 1, 2021.
- [18] N. Aliyah Salsabila, Y. Ardhito Winatmoko, A. Akbar Septiandri, and A. Jamal, “Colloquial Indonesian Lexicon,” in *2018 International Conference on Asian Language Processing (IALP)*, 2018, pp. 226–229.
- [19] D. H. Wahid and A. Sn, “Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity,” *IJCCS Indones. J. Comput. Cybern. Syst.*, vol. 10, no. 2, pp. 207–218, Jul. 2016.
- [20] W. Budiharto and M. Meiliana, “Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis,” *J. Big Data*, vol. 5, no. 1, p. 51, Dec. 2018.