

ANALISIS SENTIMEN PADA SOSIAL MEDIA TENTANG IMPLEMENTASI KEBIJAKAN PSE KOMINFO MENGGUNAKAN ALGORITME *LEXICON* BASED

Ahmad Syaifuddin¹⁾, Mohammad Muslimin²⁾

1) Program Studi Teknik Informatika Universitas Islam Majapahit

2) Program Studi Teknik Industri Universitas Islam Majapahit

E-mail: syaifuddin@unim.ac.id¹⁾, muslimin@unim.ac.id²⁾

Abstrak

Twitter merupakan salah satu media sosial yang memungkinkan penggunanya untuk bebas membuat tweet apa yang sedang terjadi, termasuk kritik dan masukan bagi instansi pemerintah. Masyarakat di media sosial banyak yang menunjukkan reaksi terhadap kebijakan PSE Kominfo dengan menuliskan opini pada twitter. Opini yang ada pada twitter dapat diambil informasi unik yang dapat menjadi masukan bagi pemerintah dan bagi pengguna twitter lain. Pemanfaatan algoritme Lexicon Based dapat memberikan hasil klasifikasi pada tweet berbahasa Indonesia berupa sentimen positif, negatif atau netral. Data diambil menggunakan library Twint didapatkan data sebanyak 3300 tweet, kemudian dilakukan preprocessing data (stemming, stopword, removing duplicate tweets) selanjutnya dihitung polaritas tweet menggunakan metode Lexicon Based. Hasil penelitian didapatkan opini masyarakat dengan sentimen negatif sebesar 80,1%, sehingga dapat disimpulkan bahwa opini masyarakat di twitter yang masih negatif atau kontra dengan implementasi kebijakan PSE.

Kata kunci: *Twitter, PSE Kominfo, Lexicon Based*

Pendahuluan

Kebijakan PSE (Penyelenggara Sistem Elektronik) yang diterapkan oleh Kementerian Komunikasi dan Informatika RI saat ini menjadi bahan perbincangan hangat di Indonesia, terutama di media sosial. Kebijakan pendaftaran PSE sendiri menuai pro dan kontra, masyarakat yang kontra dengan aturan dari Kemenkominfo menilai dengan memblokir situs Steam, Epic Games dan Paypal merupakan tindakan sewenang-wenang, melanggar hukum dan merugikan masyarakat. Pemblokiran dilakukan dengan dalih aplikasi atau situs PSE tidak terdaftar di web PSE-Kominfo berdasarkan Peraturan Menteri Komunikasi dan Informatika Nomor 5 Tahun 2020. Masyarakat yang pro terhadap kebijakan PSE juga datang dari berbagai sektor, yang menilai kebijakan PSE merupakan langkah awal untuk menjaga kedaulatan digital di Indonesia [1].

Penyelenggara Sistem Elektronik (PSE) adalah individu, penyelenggara pemerintah, bisnis, dan warga negara yang menyediakan dan/atau mengoperasikan sistem elektronik kepada pengguna untuk keuntungan pribadi atau keuntungan bersama. PSE harus terdaftar di Web PSE yang dikelola Kementerian Komunikasi dan Informatika Republik Indonesia [2].

Perbincangan mengenai kebijakan PSE paling ramai dibahas pada jejaring internet berupa media sosial. Pengguna internet di Indonesia telah mencapai 150 juta, dengan tingkat penetrasi 56% yang tersebar di seluruh wilayah Indonesia. Angka ini hanya sedikit berbeda dengan jumlah pengguna internet seluler yang mencapai 142,8 juta dengan tingkat penetrasi 53%. Berdasarkan kontribusi netizen menurut provinsi, Sumatera Utara, Kalimantan Barat, Jawa Barat, dan Sulawesi Selatan adalah yang terbesar. Bengkulu memiliki tingkat penetrasi tertinggi sebesar 85%, diikuti oleh DKI Jakarta, Kalimantan Barat, Sulawesi Tenggara, Papua, dan Nusa Tenggara Barat [3].

Twitter telah menjadi media sosial yang paling populer di kalangan masyarakat umum, dimana pengguna dapat menulis tentang berbagai topik dan mendiskusikan hal terkini [4]. Pengguna Twitter dapat mengekspresikan pendapat mereka melalui Tweet. Setiap Tweet dapat berupa opini, saran, atau kritik tentang topik tertentu. Salah satu trending topik adalah PSE Kominfo.

Banyak penelitian telah dilakukan tentang analisis sentimen, sebagian besar menggunakan data komentar dari Twitter, Facebook, berita online, blog, dan lainnya, namun tidak banyak yang belum fokus pada pengelolaan teks berbahasa Indonesia. Karena informasi tentang PSE Kominfo masih baru, penelitian terkait analisis sentimen pada sosial media tentang implementasi kebijakan PSE Kominfo ini dapat membantu pemerintah dalam hal ini Kemenkominfo RI untuk pengambilan

kebijakan kedepan terutama terkait implementasi Penyelenggara Sistem Elektronik (PSE) di Indonesia.

Berdasarkan permasalahan diatas, penelitian ini menawarkan solusi untuk diterapkan analisis sentimen pada sosial media tentang implementasi kebijakan PSE Kominfo menggunakan algoritme *Lexicon Based*. Analisis sentimen digunakan untuk mengetahui klasifikasi tweet atau komentar masyarakat Indonesia, sehingga memudahkan dalam mengetahui kecenderungan tweet atau opini tentang PSE Kominfo. Sebagai bagian dari pemrograman NLP, langkah-langkah yang dilaksanakan umumnya mengikuti tahapan penelitian bidang text mining dengan melakukan preprocessing.

Kontribusi paper ini adalah:

1. Analisis sentimen pada sosial media berbahasa Indonesia tentang implementasi kebijakan PSE Kominfo menggunakan algoritme *Lexicon Based*.
2. Evaluasi polaritas setiap kata memakai kamus data lexicon InSet (*Indonesian Sentiment Lexicon*) yang dibuat [5].
3. Penggunaan library Twint untuk crawling tweet tanpa memerlukan Twitter API.
4. Visualisasi data dan implementasi *wordcloud* untuk *positive word* dan *negative word*.

Studi Pustaka

Yanda Nooryuda Prasetya [6] melakukan penelitian dengan menerapkan analisis sentimen *Lexicon Based* to Twitter pada topik Covid-19, mengklasifikasikan opini publik menjadi tiga klasifikasi positif, negatif, dan netral. Penelitian ini juga memakai kamus InSet. Penggunaan library *GetOldTweet* untuk mengambil data melalui proses crawling melalui Twitter API dan memprosesnya menggunakan bahasa pemrograman Python. Data Twitter yang dipakai adalah data dari bulan Maret hingga Agustus 2020 dengan proses crawling dikumpulkan data sebanyak 151.618 data tweet. Penelitian ditemukan bahwa opini publik Twitter percaya Covid-19 itu nyata, terhitung 58,08%.

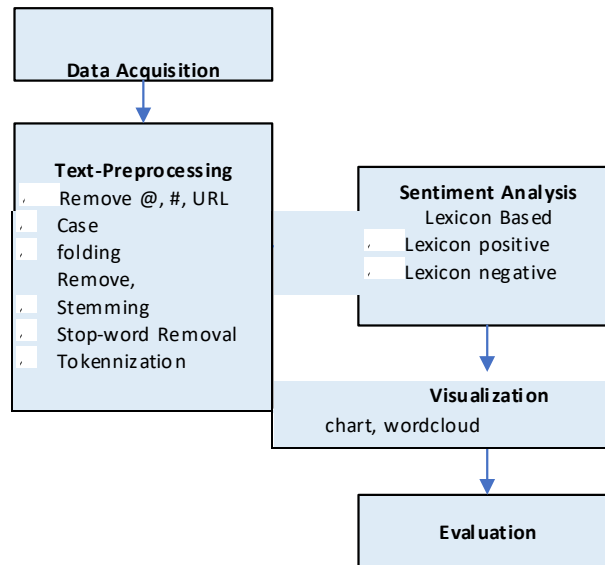
Putu Pasek Okta Mahawardana [7] melakukan analisis sentimen di sosial media Twitter dengan topik figure pemimpin kandidat calon presiden 2024. Pengambilan data memanfaatkan library *tweepy* dengan Twitter API, didapatkan sebanyak 500 tweet. Algoritme analisis sentimen yang dipakai yaitu *TextBlob* dengan klasifikasi positif, negatif dan netral. Hasil penelitian didapatkan sebesar 78% masyarakat cenderung netral terhadap calon presiden 2024.

Delima Ayu Wulandari [8] dalam penelitiannya menggunakan klasifikasi algoritme *naive bayes* untuk analisis sentimen Twitter tentang respon netizen pada RUU cipta kerja. Dataset didapatkan dengan *crawling* data Twitter menggunakan *python* didapatkan sebanyak 1667 data yang relevan, kemudian dataset akan dibagi menjadi data training dan data testing dengan rasio 75:25. Hasil penelitian didapatkan reaksi masyarakat pada RUU cipta kerja cenderung positif dengan nilai precision 92%, recall 84%, dan f1-score 86%.

Erwin Yudi Hidayat [9] menggunakan algoritme Deep Neural Network (DNN) dalam analisis sentimen Twitter untuk menilai opini publik. Crawling data dengan Twitter API yang didapatkan hingga 5504 record. Proses pelabelan dilakukan pada 3902 record menggunakan aplikasi *Sentiment Strength Detection*. Pelatihan model menggunakan algoritma DNN dengan nilai yang bervariasi untuk jumlah hidden layer, urutan node, dan *learning rate*. Bereksperimenlah dengan rasio data pelatihan 90:10 untuk data pengujian. Model terdiri dari 3 hidden layer, dengan penempatan node 128, 256, dan 128 untuk setiap layer dalam model, dan menggunakan *learning rate* 0,005, model dapat menghasilkan nilai akurasi 88,72%

Metodologi Penelitian

Langkah prosedur penelitian ini menggunakan standar untuk analisis *text mining*, yaitu standar *preprocessing* dengan tambahan penghapusan *at (@)*, *hashtag (#)*, *HTTP*, *stemming*, *stop word removal* dan analisa sentimen dengan algoritme *Lexicon Based*, ditunjukkan pada Gambar 1.

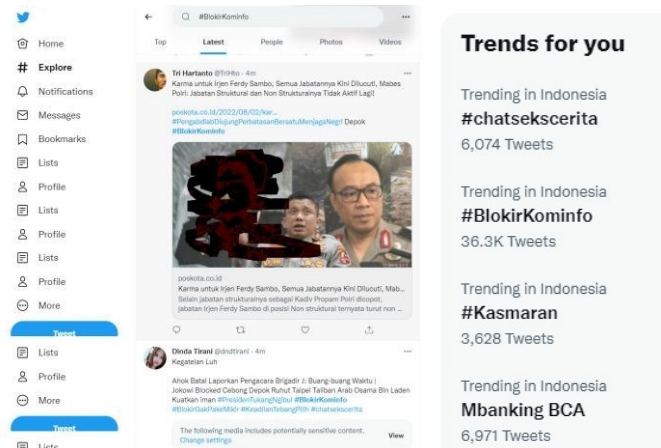


Gambar 1. Alur Sistem

Langkah pertama yaitu melakukan akuisisi data yaitu dengan crawling tweet menggunakan library Twint pada python, kemudian dilakukan *text-preprocessing*, selanjutnya dilakukan perhitungan algoritme *Lexicon Based* dan divisualisasi menggunakan chart dan wordcloud. Langkah terakhir yaitu evaluasi.

1. Pengumpulan Data

Pengumpulan data dengan cara *crawling* tweet menggunakan library Twint pada python. Library Twint dipilih karena dapat mengambil data tweet pada twitter tanpa memerlukan Twitter API. Penggunaan kata kunci “PSE Kominfo” dianggap paling relevan terhadap topik penelitian dibandingkan dengan “#BlokirKominfo” yang sedang tranding di Twitter. Karena banyak tweet yang bukan berisi informasi terkait PSE Kominfo namun memakai #BlokirKominfo.



Gambar 2. #Blokir Kominfo

Pengambilan data dibatasi pada tweet berbahasa Indonesia, data yang diambil yaitu tweet sampai tanggal 31 Juli 2022 dengan tidak dibatasi jumlah tweet, hal ini memungkinkan untuk mendapatkan banyak tweet yang berhubungan dengan PSE Kominfo. Data tweet yang didapatkan sebanyak 3300 tweet dengan waktu proses 1 menit 48 detik.

```

1 # Scraping tweets and then store to csv file using twint (https://github.com/twintproject/twint)
2
3 c = twint.Config()
4 c.Search = 'PSE Kominfo'
5 c.Lang = "in"
6 c.Hide_output = True
7 c.Debug = True
8 c.Pandas = True
9 twint.run.Search(c)

```

Gambar 3. Scraping tweet menggunakan library twint

2. Preprocessing

Langkah preprocessing data meliputi beberapa langkah utama pengerjaan yakni pembersihan data, *stemming* dan *stop word removal* [10].

- Pembersihan data dilakukan untuk menghilangkan penulisan huruf besar hingga huruf kecil dan menghilangkan karakter yang tidak perlu. Selanjutnya data yang diolah berasal dari pesan media sosial dan memerlukan penanganan khusus seperti menghapus @, #, dan URL, menghapus media gambar, video dan menghapus bilangan angka.
- Stemming* dilakukan untuk mengembalikan kata infleksi dalam Bahasa Indonesia menjadi bentuk aslinya. Alat untuk melakukan *stemming* menggunakan library *sastrawi*.
- Stop word removal* mengacu pada susunan *stop word* yang memanfaatkan *stop word removal indonesian* dari library NLTK.
- Menghapus dokumen ganda. Karena data tweet yang diolah, sehingga banyak retweet dari Dokumen ganda tersebut harus diambil salah satu.

Preprocessing didapatkan tweet sebanyak 3093, berkurang bila dibandingkan dengan dataset sebelum preprocessing, hal ini disebabkan karena banyak tweet ganda atau hasil retweet, sehingga data akan diambil salah satu tweet saja. Setelah melalui tahap preproses selanjutnya menghitung nilai sentimen polaritas menggunakan algoritme lexicon based.

3. Lexicon Based

Tweet yang mengandung kata yang terdapat pada opini atau tweet akan dilabeli sebagai tweet dengan polaritas tinggi (positif), polaritas rendah (negatif) atau polaritas netral. Tahap untuk menentukan klasifikasi berdasarkan nilai polaritas, diperlukan kamus data yang menentukan apakah kata tersebut lexicon positif atau lexicon negatif. Kamus kata lexicon yang digunakan yaitu InSet (*Indonesia Sentiment Lexicon*) yang dikembangkan [5] untuk menilai polaritas masing-masing kata. Kamus InSet Lexicon terdapat 3.609 kata positif beserta nilai polaritas dan 6.609 kata negatif beserta nilai polaritasnya dengan bobot nilai -5 sampai dengan +5.

Hasil dan Pembahasan

Analisis sentimen dilakukan dengan mengimplementasikan algoritme *Lexicon Based*, sehingga didapatkan kelompok tweet yang tergolong dalam polaritas positif, negatif dan netral. Uji coba *Lexicon Based* menggunakan python dan visualisasi hasil polaritas menggunakan *word cloud*.

1. Hasil Polaritas Sentimen Tweet

Hasil klasifikasi sentimen dihitung menggunakan algoritme *Lexicon Based* dilakukan perhitungan bobot setiap kata pada tweet tentang topik PSE Kominfo dari mulai ramai diperbincangkan sampai dengan tanggal 31 Juli 2022. Dari total data yang telah melalui tahap preprocessing yaitu sebanyak 3093 data tweet diperoleh hasil seperti yang terlihat pada tabel dibawah.

Tabel 1. Hasil Klasifikasi Sentimen		
Polaritas Sentimen	Jumlah Tweet	Prosentase (%)
Positif	409	13,2
Negatif	2477	80,1
Netral	207	6,7
Jumlah	3093	100



Visualisasi *wordcloud* berdasarkan kata yang paling banyak muncul terlihat bahwa kata kominfo, PSE dan daftar merupakan kata yang paling banyak muncul dengan frekuensi kemunculan tertinggi. Sesuai dengan topik yang diangkat yaitu PSE Kominfo, masyarakat banyak yang menggunakan tagar #kominfo dan kata PSE pada tweet mereka.

Sentimen positif pada tweet terkait PSE Kominfo diketahui dengan menghitung nilai dari polaritas berdasarkan kamus lexicon positif InSet. Hasil dari 10 tweet positif terkait PSE Kominfo dapat dilihat pada gambar dengan visualisasi wordcloud dibawah ini.

Gambar 6. Top 10 Tweet Positif Tentang PSE Kominfo

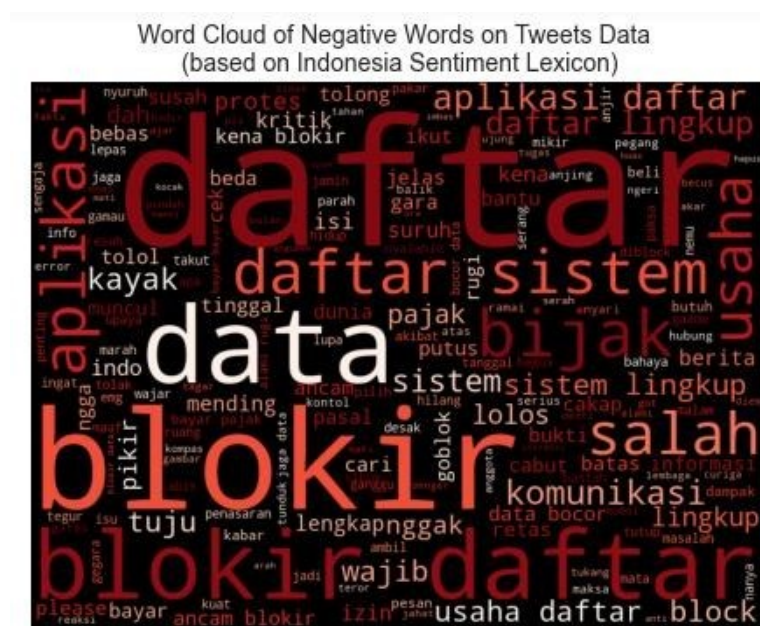


12

Sentimen negatif pada tweet terkait PSE Kominfo diketahui dengan menghitung nilai dari polaritas berdasarkan kamus lexicon negatif InSet terlihat pada gambar dengan visualisasi wordcloud dibawah ini.

	text_clean	polarity_score	polarity
1	kominfo goblok udah beli game ori dan kena ppn pulak malah di blok storenya suruh daftar pse lah itu pajak pembelian buat apa ajg kalo di blok juga hadehhh goblok goblok goblok mau nyari duit korupsi gini amat	-43	negative
2	jadi penasaran kominfo maksa buat daftar pse ini prosedurnya gmn sih ngrim woro lewat broadcast email yg berujung masuk folder spam atau hubungin satu tp proposal presentasi jelek makanya mereka ogah daftar penasaran sm cara kallon meyakinkan mereka spy mau daftar	-43	negative
3	kemungkinan template argumen membela pse dan blokir bbrp platform kominfo ingin menjaga data warga kita agar tidak dicuri asing caranya gimana salah aplikasi itu dong kenapa gamau daftar pse kalau bentrok dgn privacy policy isi buku tamu doang kok gamau	-42	negative
4	batas antara kritik dan menghina itu tipis banget dan subjektif makanya dengan peraturan kominfo tentang pse ini digabung sama uu ite sama kuhap pejabat dikritik merasa dirinya dihina tinggal lapor trus pemerintah pun berhak minta data akun anonim yg kritik dan setelah	-39	negative
5	minta maaf tapi pse gak dihapus sama juga bohong diserang warganet malah minta tolong ke kita padahal kita pun menderita karena ulah kominfo menolong kominfo itu bagaimana melepas anjing galak yang terjepit viralkan terus tagar ini	-38	negative
6	kita saat ini dalam kondisi dimana developer kalau ga daftar pse aplikasinya bakal kena blokir kalaupun daftar data privasi kita terancam di obrak abrik kominfo curangnya lagi pemerintah sedang membuat pasal yang bikin kita ga bisa bilang kominfo goblok	-35	negative
7	dampak dari aturan pse kominfo privasi tidak akan aman kebebasan pendapat dan kritikan bakal jadi susah aplikasi gak daftar pse bakal kena blokir terjadinya kebobolan data user akan menyulitkan para konten creator sma freelance dan masih banyak lagi	-34	negative
8	mungkin kominfo terlalu meremehkan sistem pertahanannya steam paypal sama epic games ya wkww gw takutnya kalau daftar pse malah tambah gampang kena bobol jir kemaren bpjs aja bisa kebobolan web kominfo juga baru ini kena hack	-31	negative
9	thefck saya tidak membahas masalah pemblokiran steam dkk yang membuat sebagian masyarakat dirugikan yang terpenting di aturan pse ini sebenarnya tiga pasal permenkominfo no th pasal ayat dan pasal ayat dan pasal yang bersifat karet kalau kominfo sewenang wenang	-31	negative
10	satyo sep mereka udh bayar pajak kominfo gk ada urusanya ama pajak tpi yah pse ini ngelanggar privasi alias krena pse pemerintah bisa korek privasi luyi wa kita semua yg isinya juga gk ada bahayanya plingan mentok share link bokep bahayanya juga bisa diketahui ama negara	-31	negative

Gambar 8. Top 10 Tweet Negatif Tentang PSE Kominfo



Gambar 7. Wordcloud Kata dari Tweet Positif yang Sering Muncul

Tweet negatif dengan nilai polaritas paling tinggi dengan skor -43, berdasarkan visualisasi wordcloud kata dari tweet negatif yang sering muncul terlihat bahwa kata blokir, sistem, daftar dan datamenjadi kata dominan dengan frekuensi tinggi. Hal tersebut menunjukkan opini netizen pada klasifikasi sentimen negatif yang memang waktu awal dirilis, sistem PSE Kominfo banyak PSE yang mengalami kesulitan pada waktu mendaftar karena muncul pesan kesalahan pada pengguna/user. Selain itu masyarakat juga masih belum percaya dan khawatir apabila data pribadi yang dimasukkan pada PSE akan diketahui oleh Kominfo. Seperti halnya Lembaga SAFEnet yang menyatakan banyak pasal yang mengancam hak atas privasi dan kebebasan berekspresi, dengan PSE mendaftar, berarti sudah memberikan data penggunanya. SAFEnet menjelaskan tiga poin penting terkait pasal itu, yaitu: 1. Konten pengguna dihapus setiap saat jika dianggap bertentangan dengan kebijakan publik 2. Sistem dan data pengguna

diakses oleh lembaga pemerintah dan lembaga penegak hukum 3. Komunikasi pribadi dibuka oleh aparat penegak hukum [1].

Kesimpulan

Berdasarkan hasil yang dilakukan analisis sentimen pada sosial media Twitter terhadap kebijakan PSE Kominfo di Indonesia menggunakan algoritme *Lexicon Based* dengan melakukan perhitungan bobot setiap kata, dari 3300 data twitter menghasilkan opini masyarakat dengan sentimen kategori negatif sebesar 80,1%, sehingga dapat disimpulkan bahwa opini publik di media sosial Twitter yang masih negatif atau menentang implementasi kebijakan PSE.

Daftar Pustaka

- [1] L. Septiani, "Ramai Tagar Blokir Kominfo, Ini Pro dan Kontra Aturan PSE - Teknologi Katadata.co.id," *katadata.co.id*, 2021.
<https://katadata.co.id/yulawati/digital/62e7c0d407df1/ramai-tagar-blokir-kominfo-ini-pro-dan-kontra-aturan-pse> (accessed Aug. 03, 2022).
- [2] P. Kominfo, "Penyelenggaraan Sistem Elektronik Lingkup Privat." Menteri Komunikasi dan Informatika Republik Indonesia, Jakarta, pp. 1–40, 2020.
- [3] L. Rizkinaswara, "Penggunaan Internet di Indonesia – Ditjen Aptika," 2019.
<https://aptika.kominfo.go.id/2019/08/penggunaan-internet-di-indonesia/> (accessed Aug. 03, 2022).
- [4] H. Hamdan, P. Bellot, and F. Bechet, "Lsislif: Feature Extraction and Label Weighting for Sentiment Analysis in Twitter," *SemEval 2015 - 9th Int. Work. Semant. Eval. co-located with 2015 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. NAACL-HLT 2015 - Proc.*, pp. 568–573, 2015, doi: 10.18653/V1/S15-2095.
- [5] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," *Proc. 2017 Int. Conf. Asian Lang. Process. IALP 2017*, vol. 2018-January, pp. 391–394, Feb. 2018, doi: 10.1109/IALP.2017.8300625.
- [6] Y. N. Yudha, D. Winarso, and syahril, "Analisis Sentimen Masyarakat Pada Twiter Terhadap Isu Covid-19 Menggunakan Metode Lexicon Based," *J. FASILKOM (teknologi Inf. dan Ilmu KOMputer)*, vol. 11, no. 2, pp. 97–103, Aug. 2021, doi: 10.37859/JF.V11I2.2772.
- [7] P. P. O. Mahawardana, G. A. Sasmita, and I. P. A. E. Pratama, "Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap 'Figure Pemimpin' Menggunakan Python | JITTER: Jurnal Ilmiah Teknologi dan Komputer," *JITTER J. Ilm. Teknol. dan Komput.*, vol. 3, no. 1, pp. 810–820, 2022, Accessed: Aug. 03, 2022. [Online]. Available: <https://ojs.unud.ac.id/index.php/jitter/article/view/82975>
- [8] D. Wulandari, R. Saedudin, and R. Andreswari, "Analisis Sentimen Media Sosial Twitter Terhadap Reaksi Masyarakat Pada Ruu Cipta Kerja Menggunakan Metode Klasifikasi Algoritma Naive Bayes | Wulandari | eProceedings of Engineering," in *eProceedings of Engineering*, 2021, pp. 9007–9016. Accessed: Aug. 03, 2022. [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15883>
- [9] E. Y. Hidayat, R. W. Hardiansyah, and A. Affandy, "Analisis Sentimen Twitter untuk Menilai Opini Terhadap Perusahaan Publik Menggunakan Algoritma Deep Neural Network," *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 2, pp. 108–118, Sep. 2021, doi: 10.25077/TEKNOSI.V7I2.2021.108-118.
- [10] A. Syaifuddin, R. A. Harianto, and J. Santoso, "Analisis Trending Topik untuk Percakapan Media Sosial dengan Menggunakan Topic Modelling Berbasis Algoritme LDA," *INSYST J. Intell. Syst. Comput.*, vol. 2, no. 1, pp. 12–19, Apr. 2020, doi: 10.52985/INSYST.V2I1.150.