

Analisis Sentimen Pola Pikir Masyarakat Indonesia Terkait Virus Covid-19 Dalam Media Sosial Twitter Menggunakan Metode Rule Based Leksikon

Zuraida Malini Cantika Riskiyanti^{a1}, I Ketut Gede Darma Putra^{a2}, AA.Kt.Agung Cahyawan Wiranatha^{a3}

^aProgram Studi Teknologi Informasi, Fakultas Teknik, Universitas Udayana
Bukit Jimbaran, Bali, Indonesia, telp. (0361) 701806
e-mail: ¹zuraidamalini@student.unud.ac.id, ²ikgdarmaputra@unud.ac.id,
³agung.cahyawan@unud.ac.id

Abstrak

Virus Covid-19 telah berhasil membuat kekacauan dari berbagai sektor di seluruh penjuru dunia. Tidak hanya memengaruhi kondisi fisik seseorang, tetapi juga memengaruhi kondisi psikologis pola pikir masyarakat. Disisi lain *pandemic* Covid-19 juga berdampak pada tingginya penggunaan media sosial dikalangan masyarakat Indonesia. Opini-opini individu akan berbagai hal dituangkan di web dan dapat dikumpulkan menjadi data yang besar untuk diolah. Twitter API merupakan aplikasi yang diciptakan oleh pihak Twitter agar mempermudah pihak developer lain untuk mengakses Informasi web Twitter seperti data berjumlah besar yang digunakan pada penelitian ini. *Rule Based* Leksikon adalah metode klasifikasi yang memanfaatkan aturan-aturan (*rule*) untuk membedakan kelas yang satu dengan kelas yang lain dengan kelas kata sentimen positif, negatif, dan netral. Penelitian ini membahas hasil analisis sentimen pola pikir masyarakat Indonesia selama pandemi Covid-19 menggunakan metode klasifikasi *Rule Based* Leksikon dengan data yang bersumber dari media sosial Twitter sebanyak 4.068.464 data *tweet* yang memperoleh nilai akurasi rata-rata sebesar 81%, nilai *precision* 93 %, nilai *recall* 95 %, dan nilai *f1score* 83%. Hasil sentimen pola pikir terbanyak keseluruhan terdapat pada bulan Juli 2021 dengan 10.011 (23,53%) data *tweet* berklasifikasi sentimen positif bertopik PPKM di Provinsi Bangka Belitung, 8.629 (19,42%) berklasifikasi sentimen netral bertopik PPKM di Provinsi Gorontalo, serta 4.216 (8,61%) berklasifikasi sentimen negatif di Provinsi Bengkulu yang mengarah pada pola pikir *konformitas toxic positivity*.

Kata kunci: Analisis Sentimen, Covid-19, Psikologi Pola Pikir, Rule Based Leksikon, Twitter

Abstract

The Covid-19 virus has managed to wreak havoc in various sectors around the world. Not only does it affect a person's physical condition, but also affects the psychological condition of the people's mindset. On the other hand, the Covid-19 pandemic has also had an impact on the high use of social media among the Indonesian people. Individual opinions on various matters are expressed on the web and can be collected into big data for processing. Twitter API is an application created by Twitter to make it easier for other developers to access Twitter web information such as the large amount of data used in this research. Rule Based Lexicon is a classification method that utilizes rules to distinguish one class from another with positive, negative, and neutral sentiment word classes. This study discusses the results of the sentiment analysis of the mindset of the Indonesian people during the Covid-19 pandemic using the Rule Based Lexicon classification method with data sourced from Twitter social media as much as 4,068,464 tweet data which obtained an average accuracy value of 81%, a precision value of 93%, 95% recall value, and 83% f1 score. The results of the highest overall mindset sentiment were in July 2021 with 10,011 (23.53%) tweet data classified as positive sentiment on the topic of PPKM in Bangka Belitung Province, 8,629 (19.42%) classified as neutral sentiment on PPKM topic in Gorontalo Province, and 4,216 (8.61%) classified as negative sentiment in Bengkulu Province which leads to a toxic positivity conformity mindset.

Keywords: Sentiment Analysis, Covid-19, Mindset Psychology, Rule Based-Lexicon, Twitter

1. Pendahuluan

Wabah dalam aspek kesehatan adalah hal yang sulit untuk dihindari dalam hidup ini. Virus merupakan penyakit yang sulit ditangani menggunakan obat, tapi dapat perlahan sembuh seiring dengan daya tahan tubuh yang kuat. Virus Covid-19 telah berhasil membuat kekacauan dari berbagai sektor di seluruh penjuru dunia [9]. Perekonomian yang kacau dengan diberlakukannya *lockdown* di berbagai negara berhasil membuat pemerintahan dari setiap negara kewalahan [10]. Berbagai sektor lainnya pun ikut terkena dampak dari virus yang sangat sulit terdeteksi penyebarannya ini. Kombinasi manusia dan analisis komputerisasi otomatis membuat pemerintah atau lembaga bisa secara efektif mendeteksi siapapun. *Microblogging* adalah bagian media sosial yang memungkinkan penggunaanya untuk menulis dan berbagi pesan singkat (280 karakter di Twitter) yang berisi opini, informasi, pertanyaan dan juga diskus. Status media sosial dan dokumen teks lainnya seperti blog, situs berita, dan email adalah teks bahasa alami [26].

Jumlah opini yang beragam akan menyulitkan dan memerlukan waktu untuk membaca secara keseluruhan. Pertukaran data secara elektronis seharusnya bersifat *lest investment*, dimana pelaku bisnis tidak perlu membeli peralatan baru sebagai infrastruktur untuk melakukan pertukaran data, dengan kata lain tetap menggunakan peralatan yang sudah tersedia [25]. Pengolahan pendapat/opini belakangan ini dikenal dengan *text mining* atau analisis sentimen. *Sentiment Analysis* merupakan cabang dari *text classification* yang bertujuan untuk mengklasifikasikan sentimen (opini) dari sebuah teks secara otomatis. Apakah teks mengandung opini negatif (negatif sentimen), opini positif (positif sentimen) atau netral (netral sentimen) [3]. Pendekatan statistik sudah banyak digunakan untuk menyelesaikan problem *sentiment analysis* saat ini. Metode yang sering digunakan pada *sentiment analysis* diantaranya *Naïve Bayes*, *Support Vector Machine (SVM)*, dan *Maximum Entropy*. Pendekatan *Rule Based method* menggunakan bantuan *human expert* berupa aturan (*rule*) yang digunakan untuk memisahkan data sesuai kelas masing-masing [3].

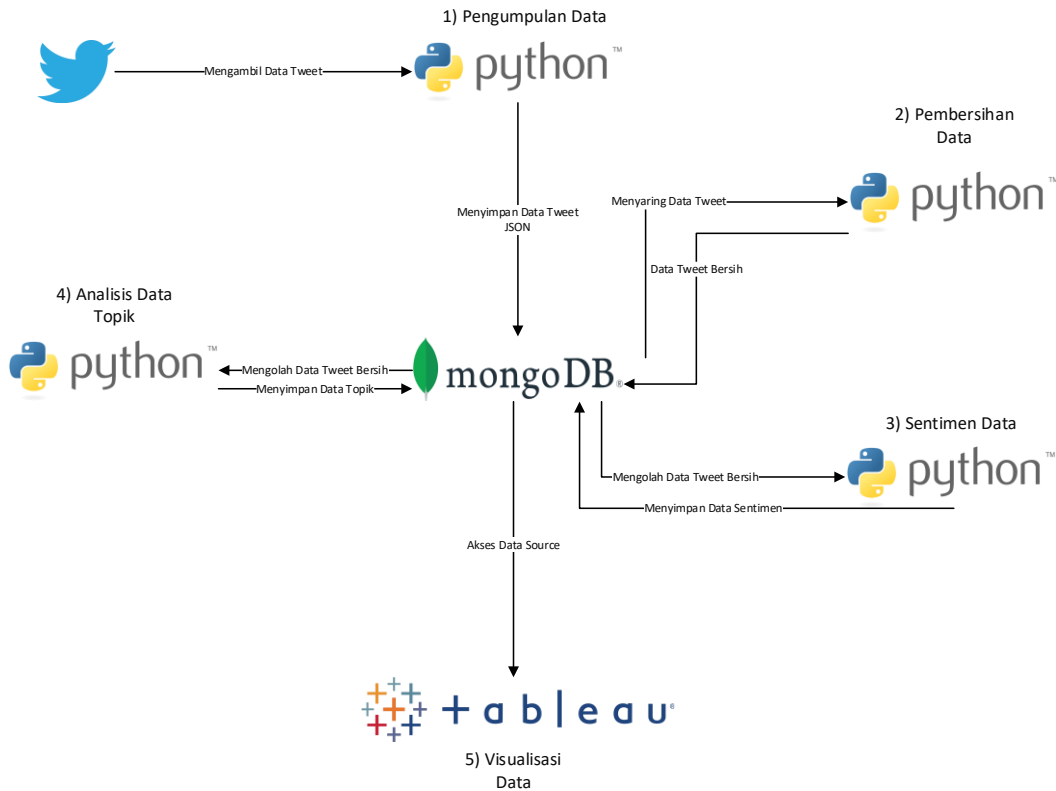
Semua kesulitan diberbagai sektor memicu adanya pemikiran negatif dari tiap individu. Tidak hanya memengaruhi kondisi fisik seseorang, tetapi juga memengaruhi kondisi psikologis pola pikir masyarakat [12]. Pemikiran-pemikiran negatif dari masyarakat ini diperparah dengan pemerintah yang kurang dapat menangani kesejahteraan masyarakat sehingga timbul teori-teori baru sesuai pola pikir masyarakat Indonesia. Pemikiran ini akan menjadi pedang bermata dua bagi sebagian masyarakat yang percaya dengan konspirasi berdasarkan nalar pola pikir mereka saat pandemi untuk menguak kebenaran dari apa yang mereka yakini. Disisi lain *pandemic* Covid-19 juga berdampak pada tingginya penggunaan media sosial dikalangan masyarakat Indonesia [14]. Sumber data Twitter berbahasa Indonesia dipilih karena penelitian dalam bahasa Indonesia belum banyak dilakukan [5]. Penelitian ini membantu menguak pemikiran masyarakat Indonesia selama *pandemic* Covid-19 dari kacamata analisis big data.

2. Metode Penelitian

Penelitian menggunakan metode *Rule Based Leksikon* yang dimulai dengan tahapan analisis sentimen secara umum seperti tahap pengambilan data, pembersihan, proses data, dan visualisasi data.

2.1. Gambaran Umum Penelitian

Implementasi Big Data dalam penulisan ini membutuhkan suatu gambaran umum yang dapat merepresentasikan data yang didapatkan untuk dianalisa secara umum. Gambaran umum Analisis Sentimen Pola Pikir Masyarakat Indonesia Terkait Virus Covid-19 Dalam Media Sosial Twitter Menggunakan Metode *Rule Based Leksikon* divisualisasikan pada Gambar 1.

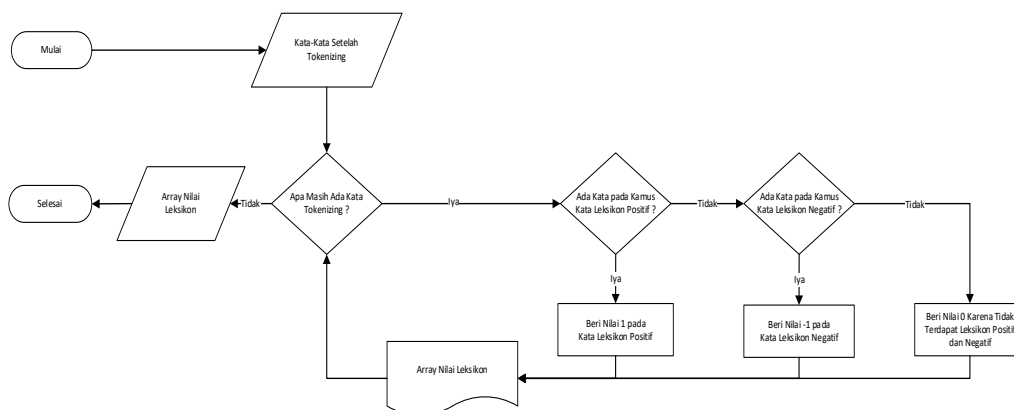


Gambar 1. Gambaran Umum Penelitian

Gambaran umum penelitian memperlihatkan alur pengambilan data melalui sosial media Twitter, data yang telah diambil akan disimpan dalam MongoDB datastore, data yang disimpan berikutnya diproses ke tahap *cleaning* dan *filtering* serta analisis menggunakan kode Bahasa pemrograman *Python*. Data bersih yang telah mengalami proses *cleaning* dan analisis akan divisualisasi dalam bentuk beberapa grafik menggunakan Tableau.

2.2. Diagram Metode Rule Based Leksikon

Diagram Metode *Rule Based Leksikon* merupakan diagram proses yang digunakan pada tahap processing data sehingga didapatkan hasil sesuai perhitungan. Diagram Metode *Rule Based Leksikon* dapat dilihat pada Gambar 2.



Gambar 2. Diagram Metode Rule Based Leksikon

Diagram proses metode *Rule Based Leksikon* dimulai dari input data *tweet* kata-kata setelah *tokenizing*. Selanjutnya dilakukan pengambilan keputusan dengan pertanyaan apakah masih ada kata yang akan di-*tokenize*, jika iya maka dilanjutkan ke pengambilan keputusan berikutnya jika tidak maka proses akan berhenti pada *array* nilai leksikon. Kondisi bila ada kata

yang belum di-*tokenize* adalah; jika ada kata bersentimen positif pada kamus positif leksikon maka diberikan nilai 1, jika tidak ada kata bersentimen positif & negatif pada kamus positif & negatif leksikon maka diberikan nilai 0, jika ada kata bersentimen negatif pada kamus negatif leksikon maka diberikan nilai -1.

3. Studi Literatur

Studi Literatur atau Kajian Pustaka meliputi semua sumber yang dijadikan sebagai bahan rujukan dalam penelitian ini, beserta tahap perancangan penelitian yang dijabarkan sebagai berikut.

3.1. Perancangan Data

Perancangan Data merupakan tahap perancangan proses pengumpulan kamus data *list stopwords* Bahasa Indonesia, *emoticon* parameter dan kamus data leksikon sebagai parameter kata klasifikasi emosi serta data *tweet*. Penelitian ini menggunakan *list stopwords* Bahasa Indonesia yang bersumber dari sebuah penelitian tesis yang menyediakan *stoplist* untuk Bahasa Indonesia. *Stoplist* yang diusulkan berasal dari hasil analisis frekuensi kata dalam Bahasa Indonesia [1]. Penelitian ini menggunakan *emoticon parameter* yang bersumber dari sebuah penelitian yang menyediakan *emoticon parameter*, Kamus emosikon ini juga diberi bobot yang akan menentukan perubahan bobot sentimen pada sebuah kalimat [2]. Kamus data leksikon menggunakan penelitian yang disusun menggunakan kumpulan kata-kata dari *tweet* bahasa Indonesia. InSet dibangun dengan memberi bobot secara manual pada setiap kata dan ditingkatkan dengan menambahkan kumpulan stemming dan sinonim yang menghasilkan sumber penelitian 3.609 kata positif dan 6.609 kata negatif dengan skor berkisar antara -5 dan +5 [4].

3.2. Rule Based-Lexicon

Metode *Lexicon-Rule Based* merupakan kombinasi dari kedua metode yang bertujuan dari untuk meningkatkan hasil *recall* dari kedua metode [5]. Pendekatan berbasis leksikon merupakan salah satu metode pada sentimen analisis yang memanfaatkan kamus yang berisi daftar kata yang mengandung polaritas kata di dalamnya. *Rule Based method* merupakan metode klasifikasi yang memanfaatkan aturan-aturan (*rule*) untuk membedakan kelas yang satu dengan kelas yang lain. *Rule* dibuat berdasarkan karakter dari masing-masing kelas dan dinotasikan dalam bentuk "*IF* (kondisi) *THEN* (solusi)". Dimana "*IF*" merupakan kondisi prasyarat (*rule antecedant*) yang terdiri dari satu atau lebih atribut tes, dimana tesnya bersifat logika. Sedangkan "*THEN*" merupakan konsekuen (*rule consequent*) yang berisi hasil prediksi kelas. Jika di dalam kalimat tersebut terdapat kata yang berupa opini, maka kalimat tersebut dianggap kalimat opini [5]. Misalnya untuk memisahkan data ke dalam kelas positif, negatif dan netral rule yang bisa digunakan adalah [3]:

- Jika jumlah frekuensi kemunculan kata bersentimen positif lebih banyak dari kata bersentimen negatif, maka data digolongkan sebagai kelas positif.
- Jika jumlah frekuensi kemunculan kata bersentimen negatif lebih banyak dari kata bersentimen positif, maka data digolongkan sebagai kelas negatif.
- Jika jumlah frekuensi kemunculan kata bersentimen positif sama dengan kata bersentimen negatif, maka data digolongkan sebagai kelas netral.
- Jika tidak ditemukan kata bersentimen positif maupun kata bersentimen negatif pada data, maka digolongkan sebagai kelas netral.

3.3. Pengumpulan Data Tweet

Pengumpulan data yang digunakan sebagai data uji analisis sentimen adalah tahapan dari Pengumpulan data *tweet*. Data *tweet* ini berupa *tweet* dikumpulkan sosial media Twitter menggunakan *filter* bahasa Indonesia dan koordinat wilayah Indonesia. Struktur data *tweet* ini berupa format JSON yang kemudian disimpan pada database MongoDB.

3.4. Pre-processing Data

Tahapan ini merupakan proses ekstraksi teks, dari data tidak terstruktur diubah menjadi data yang terstruktur agar bisa diolah lebih lanjut untuk proses klasifikasi. Tujuan dilakukannya pre-processing untuk mempersiapkan dokumen teks menjadi data yang siap diolah pada tahapan selanjutnya dan untuk mengurangi noise. Tahapan yang ada pada pre-processing adalah tokenizing (memotong kata), formalization (mengubah ke bentuk standar sesuai kamus KBBI), filtering (membuat stopword), stemming (mengubah ke bentuk dasar). Pada tahapan ini dilakukan proses pembersihan data dari noise dan mengubah data ke bentuk dasarnya (kata dasar). Sesuai tujuan dari tahapan preprocessing yakni menormalisasi data dan mereduksi dimensi dari data, maka pada tahapan ini dilakukan proses cleansing, formalization dengan menggunakan kamus lokal, serta menghapus nama jalan dengan harapan dapat meningkatkan akurasi [3]. Penelitian ini menggunakan 5 tahapan pembersihan; *data cleaning, lowering case, emoticon conversion, stemming, and stopwords data*.

3.5. Processing Data

Tahapan Processing dimulai dari input data *tweet after pre-processing* kemudian *tokenizing* kata *unigram* (1 suku kata) dan *bigram* (2 suku kata) dari data *tweet* yang diinputkan. Berikutnya dilakukan perhitungan peluang kata dan peluang kelas berdasarkan data *tweet* menggunakan metode *Rule Based* Leksikon, dan yang terakhir menampilkan hasil keluaran klasifikasi emosi berdasarkan nilai probabilitas tertinggi dari perhitungan probabilitas menggunakan metode *Rule Based* Leksikon. Setelah *tokenizing*, selanjutnya dilakukan pengambilan keputusan dengan pertanyaan apakah masih ada kata yang akan di-*tokenize*, jika iya maka dilanjutkan ke pengambilan keputusan berikutnya jika tidak maka proses akan berhenti pada *array* nilai leksikon. Kondisi bila ada kata yang belum di-*tokenize* adalah; jika ada kata bersentimen positif pada kamus positif leksikon maka diberikan nilai 1, jika tidak ada kata bersentimen positif & negatif pada kamus positif & negatif leksikon maka diberikan nilai 0, jika ada kata bersentimen negatif pada kamus negatif leksikon maka diberikan nilai -1.

3.6. Visualisasi Data

Perancangan visualisasi data merupakan serangkaian tahapan dalam menampilkan hasil analisis data, yang mendeskripsikan mengenai diagram proses visualisasi data menggunakan *tools* Tableau yang dimulai dari import data dari database MongoDB, kemudian dilanjutkan dengan proses menambahkan *measure latitude* dan *longitude* untuk menentukan titik koordinatnya, selanjutnya memilih *filter dimensions* berupa *emotions text, place name, & country code*, dan yang terakhir menampilkan hasil klasifikasi emosi pada sebaran wilayah Indonesia.

3.7. Pengujian dan Validasi

Pengukuran kinerja dari sistem/ metode klasifikasi dilakukan dengan menghitung nilai *accuracy, recall, precision, dan f1score*. Perhitungan yang digunakan untuk menghitung nilai *accuracy, recall, precision* dan *f1score* ditunjukkan oleh rumus-rumus berikut.

Accuracy diperoleh dengan membandingkan jumlah data hasil klasifikasi (prediksi) yang sesuai dengan jumlah keseluruhan data. Semakin tinggi nilai akurasi yang diperoleh, maka hasil klasifikasi semakin baik [3,6].

$$\text{Accuracy} = \frac{\text{Correctly predicted data}}{\text{Total of all data}} \quad (1)$$

Perhitungan persamaan *Precision* diperoleh dengan membandingkan jumlah data hasil klasifikasi yang relevan dan total jumlah data yang ditemukan pada kelas tertentu [3,6].

$$\text{Precision} = \frac{\text{Amount of predicted data class c correct}}{\text{Amount of data predicted as c}} \quad (2)$$

Perhitungan persamaan *Recall* diperoleh dengan membandingkan jumlah data hasil klasifikasi yang relevan dan total data yang dianggap relevan [3,6].

$$\text{Recall} = \frac{\text{The number of correct predictions of class c data}}{\text{The actual number of data as c}} \quad (3)$$

F1score menjelaskan mengenai perhitungan yang bertujuan untuk menghitung mean atau rata-rata dari *Precision* dan *Recall* [7,8].

$$F1score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \quad (4)$$

Confusion Matrix adalah sebuah tabel yang dimana digunakan untuk menghitung performa model klasifikasi data uji [7,8]. *Confusion Matrix* dapat membantu analisis data dan membantu memetakan masing-masing kelas yang memiliki *Precision* dan *Recall* rendah. *Confusion Matrix* punya empat istilah:

1. True Negative (TN): Model memprediksi data ada di kelas Negatif dan yang sebenarnya data memang ada di kelas Negatif.
2. True Positive (TP): Model memprediksi data ada di kelas Positif dan yang sebenarnya data memang ada di kelas Positif.
3. False Negative (FN): Model memprediksi data ada di kelas Negatif, namun yang sebenarnya data ada di kelas Positif.
4. False Positive (FP): Model memprediksi data ada di kelas Positif, namun yang sebenarnya data ada di kelas Negatif.

4. Hasil dan Pembahasan

Bagian hasil dan pembahasan ini berisi hasil analisis sentimen yang telah dilakukan berdasarkan hasil data yang telah diolah, divisualisasikan, dan diuji dengan beberapa metode validasi.

4.1. Dataset

Data *Tweet* atau Data *Testing* diambil menggunakan *Tweepy* API, kemudian data disimpan dengan format JSON pada MongoDB *database*. Penelitian ini menggunakan tweet data yang diambil pada akhir Mei 2021 sampai pertengahan Desember 2021 selama pandemi Covid-19.

```

_id: ObjectId("60b623025875dee052838e45")
created_at: "Tue Jun 01 12:07:25 +0000 2021"
id: 1399699102321971202
id_str: "1399699102321971202"
full_text: "aduhh bsk ujian 😊"
truncated: false

_id: ObjectId("61572ffce50ee0e850c3ac06")
id: 1399699102321971200
created_at: "Tue Jun 01 12:07:25 +0000 2021"
full_text: "aduhh bsk uji uji hidup wkwkwk negatif"
place: Object
lang: "in"
    
```

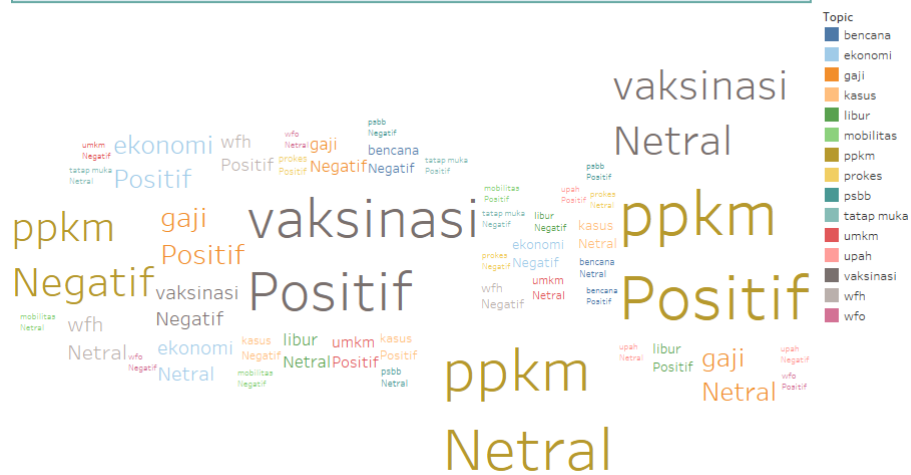
Gambar 2. Sampel Data *Tweets* Sebelum & Sesudah *Preprocessing* dan *Processing*.

Gambar 2 menggambarkan sampel data tweet tanggal 1 Juni 2021 dengan data *collection* berlokasi Indonesia. Disisi kiri gambar menunjukkan data mentah yang belum dibersihkan. Disisi kanan gambar menunjukkan data yang telah melalui *preprocessing* dan *processing* Ketika kata telah dipilah sehingga menghasilkan hasil sentimen leksikon negatif.

4.2. Visualisasi Data

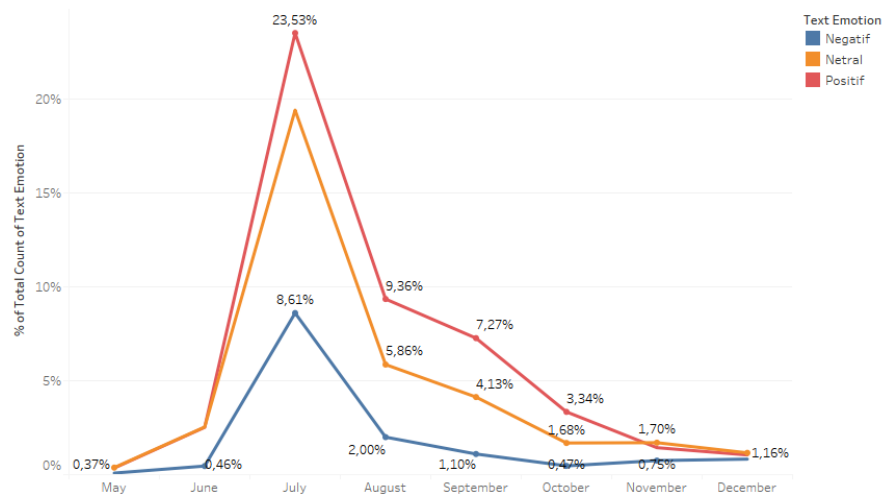
Hasil visualisasi data menggunakan bentuk *Text Bubbles (Wordcloud)*, *Lines*, *Maps*, *Stacked Bars*, *TreeMaps* berdasarkan hasil data mulai dari Bulan Mei 2021 sampai Desember 2021 selama Pandemi Covid-19.

Pola Pikir Masyarakat Indonesia Terkait Covid-19 Bulan Mei s/d Desember 2021



(a)

Pola Pikir Masyarakat Indonesia Terkait Covid-19 Bulan Mei s/d Desember 2021

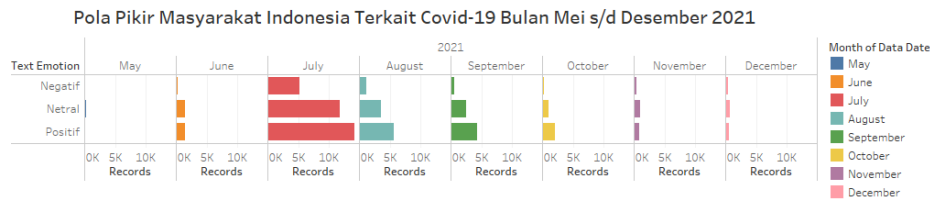


(b)

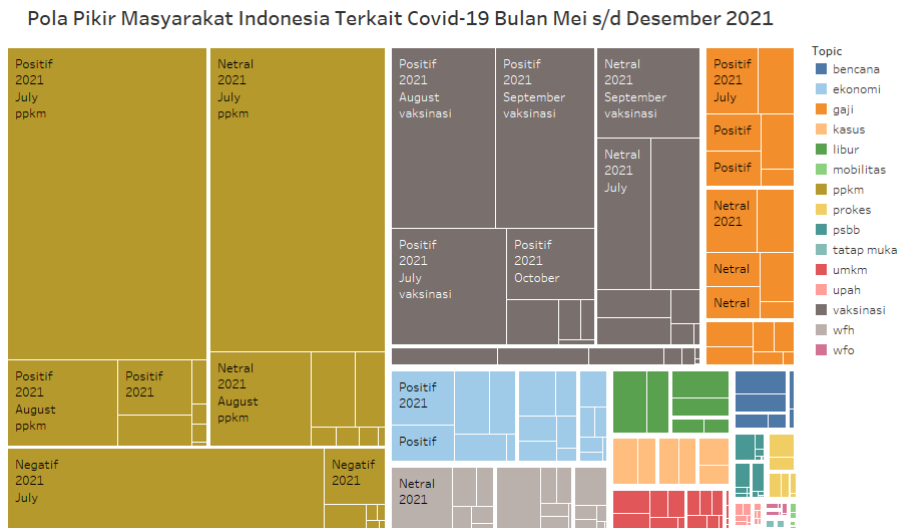
Pola Pikir Masyarakat Indonesia Terkait Covid-19 Bulan Mei s/d Desember 2021



(c)



(d)



(e)

Gambar 3. Visualisasi Data Tweet Sentimen Poal Pikir Masyarakat Indonesia.

Text Bubbles (Wordcloud) pada gambar 3 (a) menunjukkan topik yang paling sering dibahas dalam tweet terkait virus Covid-19. Pembahasan topik terbanyak ditandai dengan kata ukuran terbesar, kemudian diciutkan ke terkecil. PPKM mendominasi topik yang paling dipikirkan masyarakat Indonesia dengan tweet terbanyak pada Juli 2021 yang dilansir di laman berita kontan.co.id bahwa “Pemberlakuan Pembatasan Kegiatan Masyarakat Darurat (PPKM)” yang dimulai pada 3 Juli 2021, awalnya akan berakhir pada 20 Juli 2021[17]. Berikut diakhir pendataan, penulis masih menemukan bahwa peraturan PPKM terus mengalami perubahan hingga Desember 2021 yang diberitakan di laman berita cnnindonesia.com bahwa pemerintah memutuskan untuk tidak menerapkan PPKM level 3 pada periode Nataru di semua daerah. Selain itu, masih ada masyarakat Indonesia yang membahas peraturan PSBB yang telah diganti dengan PPKM. Dalam pembahasan PSBB, publik mencari perbedaan dengan PPKM. Sektor ekonomi selalu menjadi perbincangan hangat dari data yang diperoleh karena pada (28/7) katadata.co.id juga melaporkan bahwa IMF memperkirakan ekonomi Indonesia dan India akan paling menderita tahun ini di antara negara-negara G20. karena vaksinasi tertinggal ditulis oleh Said [24]. Rangkuman artikel ilmiah tersebut menjelaskan bahwa selama pandemi COVID-19 telah memicu krisis ekonomi global yang sangat mendadak yang memicu resesi ekonomi. Topik terkait ekonomi lainnya termasuk gaji, upah, dan ekonomi. Dilansir dari berbagai artikel laman berita (16/11) "Kenaikan Upah Minimum 1,09 Persen, Menteri Tenaga Kerja Klaim Klaim Upah Indonesia Tertinggi di Dunia" [18]. Vaksinasi merupakan tweet terbanyak kedua oleh masyarakat Indonesia, berdasarkan hasil visualisasi data penelitian ini. Dikutip dari laman repulika.co.id, vaksinasi COVID-19 Indonesia menduduki peringkat kelima dunia (4/08) [20]. Perubahan kebiasaan kerja seperti; wfh, wfo/ tatap muka, mobilitas juga menjadi perbincangan hangat di kalangan masyarakat Indonesia seperti dilansir berbagai laman berita; (29/6) “Pilihan PPKM Darurat: WFH 100% untuk sektor yang tidak penting, dilarang makan di luar” [17], (16/11) “WFO Wajib! Ini peraturan kerja terbaru PNS selama PPKM”, dan (13/12) “Hindari Lonjakan Covid-19, Pemerintah Atur Mobilitas”. Selain itu, pola pikir masyarakat Indonesia juga terkena dampak bencana yang mengiringi pandemi Covid-19.

Gambar 3 (b) menunjukkan visualisasi dari keseluruhan data topik sentimen dalam bentuk garis persentase. Persentase keseluruhan tweet topik sentimen mengungguli Juli 2021 dengan persentase 4,196%. Gambar Peta 3 (c) menggambarkan perolehan data topik sentimen terbanyak di Provinsi Gorontalo, Bengkulu, Bangka Belitung, Aceh, Riau di Negara Indonesia. *Stacked Bars* pada gambar 3 (d) menjelaskan bahwa perolehan tweet paling sentimental turun pada Juli 2021 dengan lebih dari 10 ribu tweet sentimen positif. Selain itu, *Tree Maps* 3 (e) memberikan penggambaran topik pphk pada Juli 2021 yang merupakan topik terbanyak secara keseluruhan dengan visual bagan bagian terbesar. Topik sentimen secara keseluruhan paling banyak terjadi pada Juli 2021 dengan 10.011 tweet pada klasifikasi sentimen positif, 8.629 pada klasifikasi sentimen netral, dan 4.216 pada klasifikasi sentimen negatif.

Data menunjukkan bahwa dalam kurun waktu 8 bulan dari Mei hingga Desember, masyarakat Indonesia cenderung lebih banyak menggunakan pilihan kata dan emotikon positif yang mengarah pada pola pikir tidak sehat "*Toxic Positivity Conformity*" dalam menggunakan media sosial selama masa pandemi Covid-19. Dikutip dari laman kumparan.com yang ditulis Salsabila, mengutip dari beberapa sumber buku, konformitas merupakan salah satu jenis pengaruh sosial yang mengubah perilaku dan sikap individu agar sesuai dengan norma sosial yang berlaku. Beberapa alasan yang melatarbelakangi seseorang melakukan konformitas adalah; karena keinginan untuk disukai, ketakutan akan penolakan, dan keinginan untuk merasa benar [15]. Menurut Nurul Kusuma Hidayati, M.Psi., Psikolog [16], *toxic positivity* adalah ketika sikap positif kemudian digeneralisasikan ke semua situasi dan mengabaikan perasaan dan emosi negatif. Tidak terasa, didengar, bahkan tidak diakui", dalam kuliah *online* tentang *toxic positivity* dan kesehatan mental, Fakultas Psikologi UGM (26/3). Hal ini didukung oleh hasil survei pada tahun 2020, 69% masyarakat mengalami gangguan psikologis. masalah selama Covid-19 yang ditulis di halaman mediaindonesia.com (29/9) "Survei: 69% Orang Mengalami Masalah Psikologis Selama Covid-19" [22]. Tulisan dari halaman sumber berita bbc.com ditulis oleh Brown, J., peneliti untuk *Microsoft* mensurvei 476 orang dan menganalisis profil Twitter mereka untuk kata-kata depresi, gaya bicara, hubungan, dan emosi. Selanjutnya, mereka mengembangkan klasifikasi yang secara akurat memprediksi depresi sebelum menyebabkan gejala pada tujuh dari 10 kasus [23]. Penelitian analisis sentimen sebelumnya oleh Lailiyah, dkk [3] juga menunjukkan bahwa sebagian besar masyarakat Indonesia menyampaikan keluhan dengan menggunakan bahasa yang santun meskipun sebenarnya merasa kesal dan kecewa yang membuat kalimat tersebut mengandung lebih dari satu sentimen, atau ada sentimen ambigu di mana hal ini dapat mengubah sentimen kalimat. Sentimen kalimat negatif menjadi positif. Skripsi ini [3] juga memberikan informasi bahwa banyak ditemukan sarkasme dalam data pengaduan dari Twitter.

4.2. Visualisasi Data

Data *Testing* dikumpulkan menggunakan *Tweepy* API, kemudian data disimpan dalam format JSON di *database* MongoDB. Pengujian data juga berisi data tweet yang telah melewati tahap *preprocessing* dan *processing*.

Tabel 1. Pengujian data *tweet* Twitter.

Kelas	Data Test	TPos	TNet	TNeg	Prediksi
Skenario 1	302	96	76	70	Positif
Skenario 2	602	189	147	145	Positif
Skenario 3	902	283	223	227	Positif
Skenario 4	1202	378	294	303	Positif
Skenario 5	1502	474	365	370	Positif

Hasil pengujian dengan menggunakan tabel *confusion matrix* menunjukkan bahwa beberapa kelas prediksi pengujian skenario 1 sampai 5 benar menunjukkan klasifikasi sentimen positif.

Tabel 2. Hasil pengujian data tweet Twitter.

	Skenario 1	Skenario 2	Skenario 3	Skenario 4	Skenario 5
<i>Accuracy</i>	80,1 %	80 %	81,26 %	81,1 %	80,4 %
<i>Precision</i>	93 %	92,9 %	93 %	93,2 %	92,5 %
<i>Recall</i>	96 %	94 %	94,3 %	94,5 %	92,5 %
<i>F1score</i>	82%	81,41%	83,28%	83,45%	82,35%

Berdasarkan Tabel 2, nilai akurasi yang diperoleh oleh sentimen leksikal bahasa Indonesia mencapai 81% untuk data dari media sosial Twitter. Hal ini membuktikan bahwa Metode *Rule Based Lexicon* cukup baik dalam hal penghitungan data tweet berbahasa Indonesia di media sosial Twitter. Hasil penelitian ini tentunya lebih baik dari penelitian sebelumnya oleh Lailiyah, dkk [3] yang diteliti menggunakan metode *Rule Based* dengan hasil tingkat sentimen leksikal bahasa Indonesia mencapai 65,4% untuk data dari Twitter. Hal yang membuat nilai akurasi lebih tinggi karena penelitian ini menggunakan kamus data penelitian leksikon [4] yang memiliki jumlah kamus data lebih banyak mencapai lebih dari sepuluh ribu kamus data leksikon sehingga lebih baik untuk proses pengklasifikasian data.

5. Kesimpulan

Penelitian ini menggunakan data yang bersumber dari media sosial Twitter sebanyak 4.068.464 data *tweet*. *Tweet* tersebut selanjutnya diolah lalu diproses pengklasifikasian menggunakan metode Rule Based-Leksikon, maka hasil penelitian ini dibagi menjadi beberapa poin;

1. Hasil sentimen pola pikir terbanyak keseluruhan ada pada bulan Juli 2021 dengan 10.011 (23,53%) data *tweet* pada klasifikasi sentimen positif bertopik PPKM di Provinsi Bangka Belitung, 8.629 (19,42%) pada klasifikasi sentimen netral bertopik PPKM di Provinsi Gorontalo, dan 4.216 (8,61%) pada klasifikasi sentimen negatif di Provinsi Bengkulu.
2. Hasil visualisasi data topik *tweet sentiment* yang telah dilakukan menunjukkan hasil klasifikasi sentimen pada penelitian ini didominasi oleh sentimen tentang kebijakan PPKM dan Vaksinasi yang bernilai positif, netral, lalu kemudian negatif.
3. Data menunjukkan dalam kurun waktu 8 bulan terhitung bulan Mei sampai dengan Desember 2021, masyarakat Indonesia cenderung lebih banyak menggunakan pilihan kata dan emotikon positif sehingga menunjukkan hasil masyarakat Indonesia memiliki pola pikir tidak sehat "*Konformitas Toxic Positivity*" yang menyebabkan mereka cenderung menggunakan pilihan kata dan emotikon positif untuk dinilai benar, disukai, dan diakui telah berhasil menekan emosi negatif dalam media sosial Twitter terkait Covid-19.
4. Pengujian data *testing* menghasilkan nilai rata-rata nilai *accuracy* sebesar 81 %, nilai *precision* sebesar 93 %, nilai *recall* sebesar 95 %, dan nilai *f1score* sebesar 83%. Nilai akurasi lebih tinggi karena penelitian ini menggunakan kamus data leksikon lebih dari sepuluh ribu kamus data leksikon sehingga lebih baik untuk proses mengklasifikasi datanya.

References

- [1] Tala, F. (2003). A study of stemming effects on information retrieval in Bahasa Indonesia.
- [2] Wahid, D. H., & Azhari, S. N. (2016). Peringkasan sentimen ekstrakatif di twitter menggunakan hybrid TF-IDF dan cosine similarity. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 10(2), 207-218.
- [3] Lailiyah, M. (2017). *Sentiment Analysis Menggunakan Rule Based Method Pada Data Pengaduan Publik Berbasis Lexical Resources* (Doctoral dissertation, Institut Teknologi Sepuluh Nopember).

- [4] Koto, F., & Rahmaningtyas, G. Y. (2017, December). Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs. In *2017 International Conference on Asian Language Processing (IALP)* (pp. 391-394). IEEE.
 - [5] Widyanto, Lutfi (2017) *DETEKSI SUBJEKTIFITAS TEKS BERBAHASA INDONESIA MENGGUNAKAN METODE LEXICON - RULE BASED*. Undergraduate (S1) thesis, University of Muhammadiyah Malang.
 - [6] Winarko, E. (2017). Sentimen analisis *tweet* berbahasa Indonesia dengan deep belief network. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 11(2), 187-198.
 - [7] Kusumasari, D., & Rafizan, O. (2018). Studi Implementasi Sistem Big Data untuk Mendukung Kebijakan Komunikasi dan Informatika. *Masyarakat Telematika Dan Informasi: Jurnal Penelitian Teknologi Informasi dan Komunikasi*, 8(2), 81-96.
 - [8] Wulandari, D. A. P., Sudarma, M., & Paramaita, N. (2019). Pemanfaatan Big Data Media Sosial Dalam Menganalisa Kemenangan Pilkada. *Majalah Ilmiah Teknologi Elektro*, 18(1), 101-104.
 - [9] Wang, C. J., Ng, C. Y., & Brook, R. H. (2020). Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive *testing*. *Jama*, 323(14), 1341-1342.
 - [10] Hamid, A. R. A. H. (2020). Social responsibility of medical journal: a concern for COVID-19 pandemic. *Medical Journal of Indonesia*, 29(1), 1-3.
 - [11] Gupta, R. K., Vishwanath, A., & Yang, Y. (2020). COVID-19 Twitter dataset with latent topics, sentiments and emotions attributes. *arXiv e-prints*, arXiv-2007.
 - [12] Herlambang, H., Saputra, N. E., Supian, S., Iranda, A., & Rahman, M. A. (2021). Studi Deskriptif Tentang Dampak Covid-19 Terhadap Psikologis Pada Masyarakat Jambi. *PSIKODIMENSIA*, 20(1), 10-21.
 - [13] Sumitro, P. A. (2021). Analisis Sentimen Terhadap Vaksin Covid-19 di Indonesia pada Twitter Menggunakan Metode Lexicon Based. *J-ICOM-Jurnal Informatika dan Teknologi Komputer*, 2(2), 50-56.
 - [14] Universitas Islam Indonesia, "Www.Uii.Ac.Id.", 27 Juli 2021. [Online]. Available: <https://www.uii.ac.id/media-sosial-membawa-dampak-positif-dan-negatif/>. Diakses pada 05 Desember 2021
 - [15] Salsabila, D., "Kumparan.Com", 11 Desember 2021. [Online]. Available: <https://kumparan.com/diva-salsabila-2021/mengenal-konformitas-ikut-ikutan-atau-penyesuaian-1x5lu2adncV/full>. Diakses pada 12 Desember 2021
 - [16] Afifah D., "Psikologi.Ugm.Ac.Id.", 26 Maret 2021. [Online]. Available: <https://psikologi.ugm.ac.id/kuliah-online-toxic-positivity-dan-kesehatan-mental/>. Diakses pada 06 December 2021
 - [17] Hidayat, K., "Kontan.Co.Id", 29 Juni 2021. [Online]. Available: <https://newssetup.kontan.co.id/news/opsi-ppkm-darurat-wfh-100-untuk-sektor-tak-penting-makan-di-tempat-dilarang>. Diakses pada 06 Desember 2021
 - [18] Wahyudi, A.N., "Bisnis.Com", 16 November 2021. [Online]. Available: <https://ekonomi.bisnis.com/read/20211116/12/1466774/kenaikan-upah-minimum-109-persen-menaker-klaim-upah-di-ri-tertinggi-di-dunia>. Diakses pada 06 Desember 2021
 - [19] Kementerian Keuangan Republik Indonesia. "Www.Kemenkeu.Go.Id", 27 September 2021. [Online]. Available: <https://www.kemenkeu.go.id/publikasi/berita/pemerintah-terus-perkuat-umkm-melalui-berbagai-bentuk-bantuan/>. Diakses pada 08 Desember 2021
 - [20] Azizah, N., "Republika.Co.Id", 04 Oktober 2021. [Online]. Available: <https://www.republika.co.id/berita/r0gj51463/vaksinasi-covid19-indonesia-peringkat-kelima-dunia>. Diakses pada 07 Desember 2021
 - [21] Antara., "Jpnn.Com", 11 Agustus 2021. [Online]. Available: <https://www.jpnn.com/news/pemda-mulai-mengizinkan-pembelajaran-tatap-muka-dan-wfo-ini-syaratnya>. Diakses pada 05 Desember 2021
 - [22] Puspa, A., "Mediaindonesia.Com", 29 September 2021. [Online]. Available: <https://mediaindonesia.com/humaniora/436215/survei-69-masyarakat-alami-masalah-psikologis-selama-covid-19>. Diakses pada 06 Desember 2021
 - [23] Brown, J., "Bbc.Com", 16 Januari 2018. [Online]. Available: <https://www.bbc.com/indonesia/vert-fut-42679432>. Diakses pada 08 Desember 2021
-

- [24] Said, A., A., "Katadata.Co.Id", 28 Juli 2021. [Online]. Available: <https://katadata.co.id/agustiyanti/finansial/610150ede3554/imf-ramal-ekonomi-ri-paling-menderita-di-antara-negara-g20>. Diakses pada 09 Desember 2021
 - [25] Hanafi, A., Sukarsa, I. M., & Wiranatha, A. A. K. A. C. (2017). Pertukaran data antar database dengan menggunakan teknologi API. *Lontar Komputer*, 8(1), 22-30.
 - [26] Ermawati, M., & Buliali, J. L. (2018). Text Based Approach For Similar Traffic Incident Detection from Twitter. *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, 9(2), 63.
-