

ANALISA SENTIMEN TERHADAP TAGAR #dirumahaja MELALUI TWITTER DI INDONESIA

Andreyestha¹, Adhi Dharma Suriyanto², Witriana Endah Pangesti³

^{1,2}Universitas Bina Sarana Informatika, ³STMIK Nusa Mandiri, Indonesia
Email : andreyes2505@bsi.ac.id¹, adhi.ais@bsi.ac.id², witriana.weg@nusamandiri.ac.id³

ABSTRAK

Kebiasaan baru atau New Normal kini mau tidak mau diterapkan diberbagai bidang. Selama pandemi Covid-19, masyarakat dituntut untuk melakukan berbagai hal di rumah mereka masing-masing. Media sosial menjadi semakin pesat pemanfaatannya oleh masyarakat, salah satunya twiiter memiliki banyak fungsi dan tujuan. Twitter dapat menjadi penyebar informasi berupa himbauan, seperti tagar dirumahaja yang digaungkan oleh pemerintah dan masyarakat Indonesia. Penggunaan tagar dirumahaja tersebut memiliki berbagai macam opini, seperti opini positif maupun negatif. Penelitian ini mencoba menganalisa kicauan berbahasa Indonesia yang menggunakan tagar dirumahaja untuk mengetahui bagaimana pandangan masyarakat di Indonesia mengenai tagar tersebut. Metode yang digunakan yaitu *scraping* data dari Twitter dengan jumlah data 200 kicauan kemudian dianalisa menggunakan metode text mining berbasis lexicon, hasil sentimen yang didapat dari kicauan menggunakan metode *Bing Vector*.

Hasil analisa sentimen pada penelitian ini mendapat skor positif berjumlah 49, dan skor negatif berjumlah 48. Skor negatif hampir mengimbangi skor positif disebabkan tingginya persentase emosi *Anticipation* sebesar 40% yang menggambarkan kekhawatiran warganet pengguna twitter terhadap pandemi virus covid-19 yang mereka tuangkan melalui kicauan bertagar dirumahaja.

Kata kunci: *Text Mining*, Analisa Sentimen, dirumahaja

ABSTRACT

New Normal now inevitably apply in various fields. During the Covid-19 pandemic, people were required to do various things in their own homes. Social media is increasingly being used by the community, one of which is that Twitter has many functions and purposes. Twitter can disseminate information in the form of appeals, such as dirumahaja hashtags echoed by the government and the people of Indonesia. The use of these dirumahaja hashtags has various kinds of opinions, such as positive and negative opinions. This study tries to analyze Indonesian language tweets that use dirumahaja hashtags to find out how people in Indonesia view these hashtags. The method used is scraping data from Twitter with a total of 200 data tweets then analyzed using the lexicon-based text mining method, the sentiment results obtained from tweets using the Bing Vector method.

The results of the sentiment analysis in this study got a positive score of 49, and a negative score of 48. The negative score almost offset the positive score due to the high percentage of Anticipation emotions of 40% which illustrates the concern of Twitter netizens about the COVID-19 pandemic which they pour through dirumahaja tagged tweets.

Keywords: *Text Mining, Sentiment Analysis, dirumahaja*

PENDAHULUAN

Selama pandemi Covid-19, masyarakat dituntut untuk melakukan berbagai hal di rumah mereka masing-masing. Bekerja dari rumah, sekolah dari rumah, larangan berpergian dan berkumpul serta kegiatan lain yang biasanya dilakukan di luar, seolah menjadi halangan bagi semua kalangan.

Kebiasaan baru atau New Normal kini mau tidak mau diterapkan diberbagai bidang, meski hampir semua sektor industri banyak yang mengalami penurunan, tidak sedikit juga dari mereka yang mengambil kesempatan untuk tetap maju dan berkembang. Media sosial contohnya, semakin pesat pemanfaatannya oleh masyarakat.

Media sosial merupakan perpaduan sosiologi dan teknologi yang memberikan kesempatan orang untuk saling terhubung secara online baik dalam hubungan personal maupun bisnis (Khasanah, Samsiana, & Handayanto, 2020). Media sosial menjadi solusi ampuh bagi setiap kalangan untuk menyebarkan informasi secara cepat, tepat, dimana dan kapan saja. Media sosial seperti twiiter memiliki banyak fungsi dan tujuan (Singh, Dwivedi, Singh, & Pathania, 2019) mulai dari alat komunikasi, media berwirausaha atau berjualan (Ajimat, Denok Sunarso, 2020), bahkan penyebaran informasi berupa himbuan. Seperti tagar dirumahaja yang digaungkan diberbagai media sosial salahsatunya yang populer yaitu Twitter (Alom, Carminati, & Ferrari, 2020).

Selama pandemi, organisasi publik harus memainkan peran utama dalam mengelola sumber daya yang tersedia, mengembangkan dan melaksanakan strategi pencegahan dan tindakan, dan berkoordinasi dengan media berita untuk memberikan informasi yang terbukti dan berguna kepada warga (García-díaz, Cánovas-garcía, & Valencia-garcía, 2020). Pemerintah dan masyarakat Indonesia menjadikan Twitter sebagai sarana untuk menyebarkan informasi dengan ciri khas tagar dirumahaja, agar masyarakat di Indonesia mematuhi protocol kesehatan untuk menjaga diri mereka dan lingkungan dari Virus Covid-19.

Penggunaan tagar dirumahaja tersebut memiliki berbagai macam opini, seperti opini positif maupun negatif. Beberapa menggunakan tagar dirumahaja sebagai himbuan, peringatan, motivasi, atau dalam bentuk informasi lainnya. Teknologi memberikan peluang yang sangat besar untuk memerangi wabah penyakit menular dan memiliki peran yang luar biasa, terutama dalam analisis sentimen untuk media sosial. Pentingnya hal ini karena perannya yang luar biasa dalam menganalisis sentimen publik (Alamoodi et al., 2020).

Maka penelitian ini mencoba menganalisa kicauan berbahasa Indonesia yang menggunakan tagar dirumahaja untuk mengetahui bagaimana pandangan masyarakat di Indonesia mengenai tagar tersebut. Dalam penelitian ini yaitu sentimen terhadap kicauan bertagar dirumahaja, opini-opini tersebut akan dikelompokkan kedalam opini positif, negative, atau netral.

METODE

Metode yang digunakan pada penelitian ini menggunakan scraping data dari Twitter kemudian dianalisa menggunakan metode text mining berbasis lexicon dengan menggunakan API dari Twitter dan Pemograman berbasis R. Hasil sentimen yang didapat dari kicauan menggunakan metode Bing Vector, dimana pada Bing Vector menggunakan skala -1 sebagai sentimen negatif dan +1 sebagai sentimen positif.

Pengumpulan Data

Data diperoleh melalui twitter dengan kata kunci “dirumahaja”, data yang digunakan berjumlah 200 kicauan.

Text Mining

Text Mining adalah "penemuan dan ekstraksi pengetahuan yang menarik dari teks bebas atau tidak terstruktur" (Hartog, 2017). Pengetahuan berasal dari pola dan hubungan dan dapat digunakan untuk mengungkapkan fakta, atau tren. *Text Mining* telah digunakan untuk berbagai tujuan, seperti menyelidiki persepsi konsumen terhadap hotel berdasarkan informasi tekstual pelanggan online, pemanfaatan posting media sosial untuk melakukan analisis kompetitif, penggunaan posting media sosial untuk perencanaan produk dan analisis data besar di sektor keuangan (Pejic-bach, Bertonce, Meško, & Krstić, 2019).

Secara umum, algoritma ini dibagi menjadi klasifikasi atau *supervised learning*, dan pengelompokan atau *unsupervised learning*. Klasifikasi mengkategorikan item berdasarkan fitur-fiturnya dalam kumpulan kategori yang telah ditentukan sebelumnya, dan pengelompokan mengkategorikan item berdasarkan kesamaan di antara mereka (Máverick, Anderson, Evandro, & Romero, 2019).

Preprocessing Teks

Pre-processing data dapat membantu mengubah data agar mudah digunakan sebelum tugas analisa sentimen dapat dilakukan. Untuk bisa mengolah dokumen teks supaya bisa menghasilkan luaran yang diinginkan maka harus dilakukan pengolahan teks terlebih dahulu (Rahutomo, Puspitasari, & Sulistyoningrum, 2020), tahapan preprocessing teks meliputi menghapus angka, tanda baca, stopword dan mengubah teks menjadi huruf kecil. Pre-processing data dapat dijabarkan berdasarkan tahapan berikut (Palkar, Gala, Shah, & Shah, 2016).

a. Remove numbers

Menghapus angka karena tidak memainkan peran vital dalam menentukan orientasi teks dan karenanya, mereka dihapus dari data.

b. Remove punctuation

Dalam tahapan ini, tanda baca dihapus dari teks. Sehingga hanya kata yang tersisa dari kicauan yang didapat.

c. Remove stop words

Kata yang sering muncul tetapi tidak memiliki informasi penting contohnya adalah kata sambung seperti "untuk", "di", "atau", "dari", dll dihapus. Stop word removal dapat membantu mengurangi kebutuhan memori saat mengklasifikasikan kicauan, fungsi utamanya dari pembuangan stopwords adalah mencegah agar stopwords tidak mempengaruhi hasil pada proses selanjutnya (Budiman, 2020).

d. Mengubah Teks Menjadi Huruf Kecil

Kalimat tersebut dikonversi menjadi huruf kecil untuk menjaga konsistensi dan memetakan kata-kata terlepas dari kasusnya.

Analisa Sentimen

Analisa sentimen merupakan sistem komputasi yang mempelajari pendapat, sikap dan emosi terhadap suatu entitas yang dapat mewakili individu, peristiwa atau topik dari sekumpulan teks (Soekarno, No, Lowokwaru, Malang, & Timur, 2019). Pada Analisa Sentimen terdapat kumpulan studi komputasi opini, sentimen, emosi, penilaian, dan sikap orang terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atributnya (Zhang, Wang, & Liu, 2018). Analisa Sentimen berkembang pesat bertepatan dengan media sosial di Web yang semakin banyak digunakan, seperti, blog, forum diskusi, Twitter, dan jejaring sosial lainnya. Lexicon dapat dikembangkan secara manual, Untuk memprediksi sentimen keseluruhan dari suatu dokumen, rumus atau algoritma diperlukan untuk menggabungkan nilai sentimen dari setiap kata untuk menghasilkan skor sentiment (Khoo, 2017).

Dengan menjamurnya situs ulasan produk, penilaian pengguna dan penggunaan media sosial yang masif, ada banyak sekali dokumen yang dijelaskan di Internet. Ketika *Corpus* pelatihan yang cukup besar tidak tersedia, *lexicon* sentimen yang ada diperlukan untuk kategorisasi sentimen

DISKUSI

Penelitian ini menggunakan data yang diperoleh dari twiiter berjumlah 200 kicauan, *scraping* data menggunakan RStudio berbasis pemrograman R dengan memanfaatkan TwitterAPI yang sudah dibuat. Data yang didapat berjumlah 200 dokumen dengan mengambil kicauan dari region Indonesia saja, berikut kicauan berbahasa Indonesia pada penelitian ini mengenai tagar dirumahaja.

Tabel 1. Kicauan dirumahaja

No	Kicauan
1	Semenjak kita diharuskan untuk tetap #DiRumahAja, pasti kamu berandai-andai untuk pergi liburan. Ya ga sih?
2	Tetap semangat untuk para ibu yang menjalani kehamilan di masa pandemi #COVID19 ini! Kurangi risiko tertular dengan... https://t.co/qT7vBT3JLL "
3	Baru kali ini elliptical nyampe 5 km. Kalo 46 menit dibuat gowes udah nyampe pantai kuta nih. Krn hampir tiap pagi... https://t.co/twxwEBNqC0
4	Hanya om @Partono_ADjem yang bisa menurunkan berat badan selama libur panjang, karena #dirumahaja
5	Cegah penyebaran virus covid-19 dengan 3M
6	RT @KemenkesRI: Update situasi terkini perkembangan #COVID19 di Indonesia #BersatuLawanCovid19 #dirumahaja #JagaJara..."
7	Sama kyk tahun sebelumnya, tahun ini dimulai dengan #dirumahaja. untungnya mulai dari tgl 31-3 Januari kmrin gue di... https://t.co/KgmM4CWvIT

Sumber: Twitter (2020)

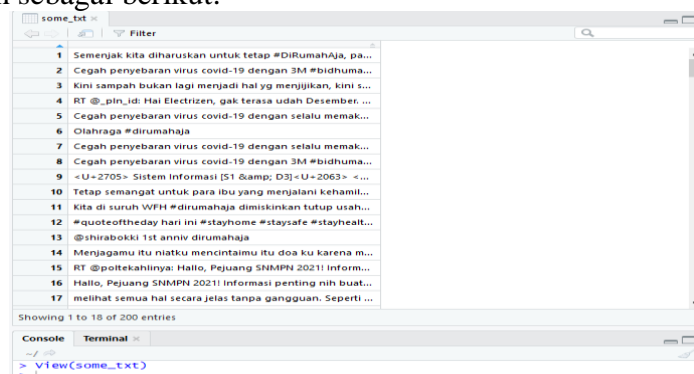
Setelah kicauan didapat langkah selanjutnya adalah mengolah kicauan tersebut dengan ekstraksi fitur, dan membuat *word cloud* dari kumpulan kicauan yang didapat. Sebelum melakukan ekstraksi fitur, langkah yang dapat dilakukan sebagai berikut.

Menentukan Direktori

Menentukan direktori penyimpanan project pada RStudio.

Membuat Corpus

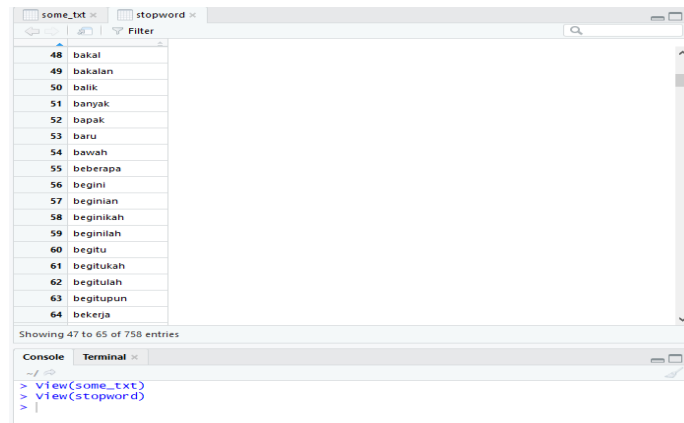
Corpus merupakan kumpulan teks yang berisi kicauan yang telah didapat, corpus tersebut akan digunakan sebagai bahan penelitian. Kicauan akan diambil teksnya saja sehingga konten seperti video atau gambar dihilangkan, dari teks tersebut dibuat corpus. Corpus ditampilkan sebagai berikut.



Gambar 1. Corpus

Memuat Stopword

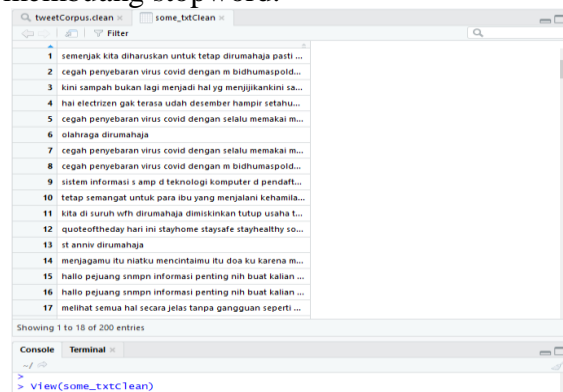
Stopword adalah kumpulan teks untuk menghilangkan kata-kata yang tidak diperlukan. Stopword yang digunakan pada penelitian ini berbahasa Indonesia diupload oleh akun masdevid di Github, melalui link <https://github.com/masdevid/ID-Stopwords>.



Gambar 2. Stopword

Proses Ekstraksi Fitur

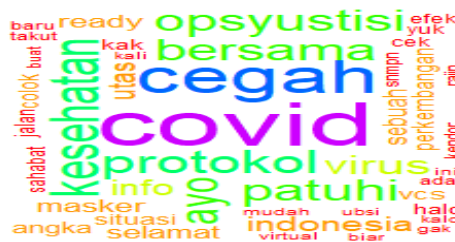
Pada proses ekstraksi fitur, kicauan yang telah dikumpulkan akan diubah formatnya yang sebelumnya tidak terstruktur menjadi terstruktur. Tahapan ini meliputi mengubah huruf kapital menjadi huruf kecil, pembuangan URL, menghilangkan punctuation atau tanda baca, pembuangan angka dan membuang stopwords.



Gambar 3. Hasil Ekstraksi Fitur

Membuat Word cloud

Word cloud salah satu metode untuk menampilkan data teks secara visual. Grafik ini populer dalam text mining karena mudah dipahami. Dengan menggunakan word cloud, gambaran frekuensi kata-kata dapat ditampilkan dalam bentuk yang menarik namun tetap informatif. Semakin sering satu kata digunakan, maka semakin besar pula ukuran kata tersebut ditampilkan dalam word cloud.

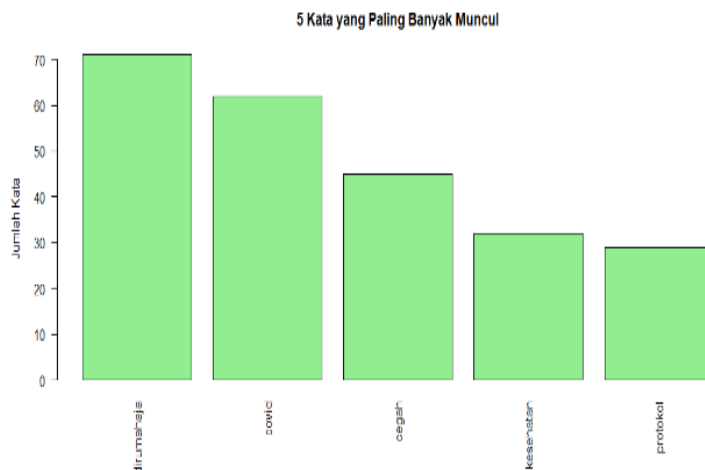


Gambar 4. Word Cloud

Word cloud berhasil dibuat, covid ditampilkan ditengah menunjukkan kata tersebut sering muncul. Selain kata covid, beberapa kata seperti cegah, bersama, protocol, dan kesehatan juga tampil dominan dan muncul relative tinggi karena banyak himbauan maupun info kenaikan jumlah pasien positif covid yang disebar melalui twitter, seperti pada kicauan “@KemenkesRI: Update situasi terkini perkembangan #COVID19 di Indonesia (3/1) #BersatuLawanCovid19 #dirumahaja”. di disusul kata virus, masker, info, dan patuhi.

Membuat Barplot

Barplot merupakan grafik yang akan menampilkan kata dengan frekuensi terbanyak.

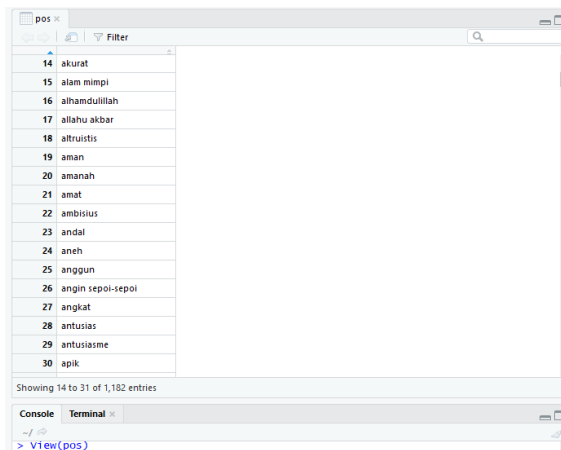


Gambar 5. Barplot

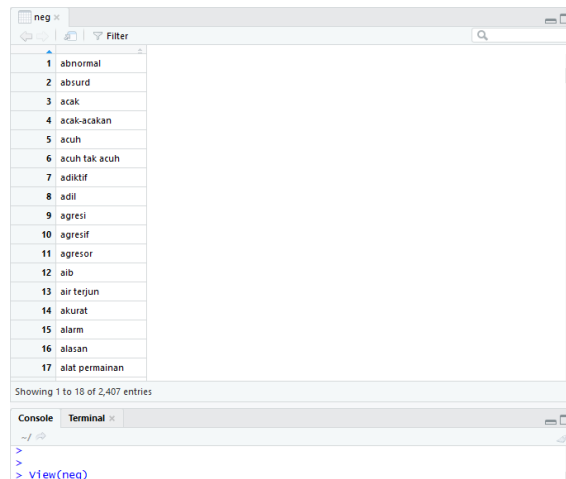
Pada barplot tersebut ditampilkan 5 kata yang sering muncul pada kicauan dengan tagar dirumahaja, yaitu kata dirumahaja, covid, cegah, kesehatan, dan protokol.

Menghitung Score Sentimen

Setelah berhasil membuat *word cloud* dan barplot, selanjutnya yaitu menghitung score sentimen atau nilai yang akan menghitung jumlah dari kicauan positif, negatif, ataupun netral dengan terlebih dahulu memuat kumpulan kata opini positif dan negatif. Pada penelitian ini menggunakan kumpulan kata opini positif dan negatif berbahasa Indonesia dari akun ghitub masdevid pada link. <https://github.com/masdevid/ID-OpinionWords>.



Gambar 6. Kata Opini Positif



Gambar 7. Kata Opini Negatif

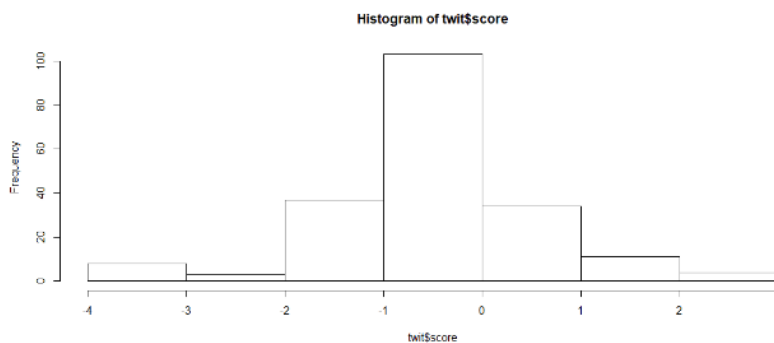
Metode untuk menghitung skor sentimen pada penelitian ini menggunakan *Bing Vector*, dimana metode tersebut menampilkan skala biner dengan -1 menunjukkan negatif dan +1 menunjukkan sentimen positif. Pada kicauan dengan tagar dirumahaja dapat dilihat hasil skornya sebagai berikut.

```
> twit$score
[1] 1 -1 -3 1 -1 0 -1 -1 1 0 0 0 0 0 1 1 -2 0 -2 0 0 -1 1 0 1
[26] -2 1 -1 0 -3 1 1 0 1 -1 0 0 -1 -1 1 -1 -1 -1 -1 -1 0 -3 -1
[51] -1 -1 -1 0 0 0 0 0 0 -1 0 0 1 0 0 0 1 0 0 0 2 0 0 1 1
[76] 2 1 0 -1 1 -1 0 1 1 -1 1 0 0 0 0 0 -1 -1 -1 -1 0 0 -1 0 0
[101] 2 0 1 -1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -1 1 0 0 0 -3
```

Pada skor diatas angka yang berada di dalam kurung [1, 26, 51, 76, 101], merupakan index kicauan. Sedangkan angka disampingnya merupakan skor sentimen kicauan tersebut. Lalu membuat dan menampilkan table hasil perhitungan skor sentimen sebagai berikut.

```
> table(twit$score)
-4 -3 -2 -1 0 1 2 3
1 7 3 37 103 34 11 4
```

Selanjutnya membuat dan menampilkan histogram dari skor sentimen.



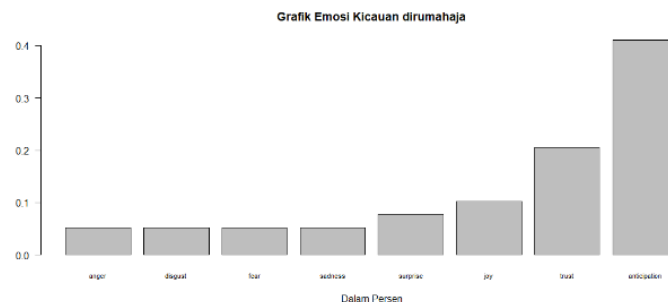
Gambar 7. Histogram twit\$score

Angka positif yang berada di sisi kanan histogram menunjukkan opini positif terhadap kicauan dengan tagar dirumahaja, angka nol menunjukkan opini netral, sedangkan angka

negatif yang terdapat dibagian kiri histogram menunjukkan opini negatif terhadap kicauan dengan tagar dirumahaja. Hasil penelitian yang didapat pengguna twitter lebih banyak memberikan opini positif dibanding opini negatif terhadap tagar dirumahaja meskipun dengan hasil tipis hanya berbeda satu kicauan saja, dari 200 kicauan terdapat 49 kicauan yang terdeteksi sebagai kicauan positif, 103 kicauan netral, dan 48 kicauan yang terdeteksi sebagai kicauan negatif.

Membuat Grafik Emosi

Hasil skor sentimen telah didapat, selanjutnya adalah membuat grafik emosi untuk mengetahui emosi dari kicauan bertagar dirumahaja.



Pada grafik di atas terlihat *Anticipation*, *Trust* dan *Joy* menjadi 3 emosi terbanyak. Grafik ini juga memperjelas bagaimana kicauan negatif hampir mengimbangi kicauan positif, dimana tingginya persentase emosi *Anticipation* sebesar 40% yang menggambarkan kekhawatiran warganet pengguna twitter terhadap pandemi virus covid-19 yang mereka tuangkan melalui kicauan bertagar dirumahaja.

KESIMPULAN

Hasil analisa sentimen pada penelitian ini, dapat dikatakan banyak warganet pengguna twitter yang memberikan opini positif melalui kicauan bertagar dirumahaja untuk merespon pandemi covid-19 di Indonesia yang ditunjukkan dengan skor positif berjumlah 49 di sisi kanan tabel dan histogram. Kebanyakan warganet bersikap netral dengan tingginya nilai nol yang ditunjukkan dalam tabel dan histogram *twit\$score*, meskipun skor dari kicauan negatif hampir mengimbangi kicauan positif dengan skor 48.

Tingginya persentase emosi *Ancipation* sebesar 40% juga mempengaruhi kicauan negatif yang hampir mengimbangi kicauan positif, dimana tingginya emosi tersebut pada grafik menggambarkan kekhawatiran warganet pengguna twitter terhadap pandemi virus covid-19 yang mereka tuangkan melalui kicauan bertagar dirumahaja.

DAFTAR PUSTAKA

- Ajimat, Denok Sunarso, F. S. P. (2020). BERWIRAUSAHA MEMANFAATKAN MEDIA SOSIAL PADA DAERAH SEPATAN, *1*(1), 69–76.
- Alamoodi, A., Zaidan, B., Zaidan, A., Albahri, O., Malik, R., Almahdi, E., ... Alaa, M. (2020). Sentiment Analysis and Its Applications in Fighting COVID-19 and Infectious Diseases: A Systematic Review. *Expert Systems With Applications*, 114155. <https://doi.org/10.1016/j.eswa.2020.114155>
- Alom, Z., Carminati, B., & Ferrari, E. (2020). A deep learning model for Twitter spam detection. *Online Social Networks and Media*, *18*, 100079.

- <https://doi.org/10.1016/j.osnem.2020.100079>
Budiman, A. E. (2020). Analisis Pengaruh Teks Preprocessing Terhadap Deteksi Plagiarisme Pada Dokumen Tugas Akhir, *6*, 475–488.
- García-díaz, J. A., Cánovas-garcía, M., & Valencia-garcía, R. (2020). Ontology-driven Aspect-based Sentiment Analysis classification: An Infodemiological case study regarding infectious diseases in Latin America. *Future Generation Computer Systems*. <https://doi.org/10.1016/j.future.2020.06.019>
- Hartog, D. N. Den. (2017). *Text Mining in Organizational Research*. <https://doi.org/10.1177/1094428117722619>
- Khasanah, F. N., Samsiana, S., & Handayanto, R. T. (2020). Pemanfaatan Media Sosial dan Ecommerce Sebagai Media Pemasaran Dalam Mendukung Peluang Usaha Mandiri Pada Masa Pandemi Covid 19, *1*(1), 51–62.
- Khoo, C. S. G. (2017). Lexicon-based sentiment analysis : Comparative evaluation of six sentiment lexicons. <https://doi.org/10.1177/0165551517703514>
- Máverick, R. F., Anderson, A., Evandro, P., & Romero, C. (2019). Text mining in education, (June). <https://doi.org/10.1002/widm.1332>
- Palkar, R. K., Gala, K. D., Shah, M. M., & Shah, J. N. (2016). Comparative Evaluation of Supervised Learning Algorithms for Sentiment Analysis of Movie Reviews. *International Journal of Computer Applications*, *142*(1), 975–8887. Retrieved from <http://www.ijcaonline.org/archives/volume142/number1/palkar-2016-ijca-909660.pdf>
- Pejic-bach, M., Bertonce, T., Meško, M., & Krstić, Ž. (2019). Text mining of industry 4 . 0 job advertisements. *International Journal of Information Management*, (July), 1–16. <https://doi.org/10.1016/j.ijinfomgt.2019.07.014>
- Rahutomo, F., Puspitasari, D., & Sulistyoningrum, T. E. (2020). Implementasi Single Pass Clustering pada Preprocessing Temu Kembali Koleksi Berita, *6*(1), 86–94.
- Singh, P., Dwivedi, Y. K., Singh, K., & Pathania, A. (2019). Can twitter analytics predict election outcome? An insight from 2017 Punjab assembly elections. *Government Information Quarterly*, (December), 101444. <https://doi.org/10.1016/j.giq.2019.101444>
- Soekarno, J., No, H., Lowokwaru, K., Malang, K., & Timur, J. (2019). IMPLEMENTASI SUPPORT VECTOR MACHINE PADA ANALISA, *10*, 83–88.
- Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis : A survey, (January), 1–25. <https://doi.org/10.1002/widm.1253>