

ANALISIS SENTIMEN ULASAN APLIKASI MOLA PADA GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

SENTIMENT ANALYSIS OF MOLA APPLICATION REVIEWS ON GOOGLE PLAY STORE USING SUPPORT VECTOR MACHINE ALGORITHM

Muhammad Diki Hendriyanto¹, Azhari Ali Ridha², Ultach Enri³

^{1,2,3}Universitas Singaperbangsa Karawang
muhammad.diki18199@student.unsika.ac.id

ABSTRACT

MOLA is one of the video streaming platform applications on the google play store. The mola application has been downloaded 5 million times but only has a 3.2 rating. On the Google Play Store app rating is followed by user reviews of the app. There are quite a lot of reviews that are unstructured and contain opinions from users about their satisfaction with the application so that it is often taken into consideration by potential users to choose the application used. Based on this, sentiment analysis was carried out using the Support Vector Machine algorithm to find out how the sentiments of users towards the MOLA application on the google play store were carried out. This study uses the Knowledge Discovery in Database (KDD) method. The data used is a review of the MOLA application with as many 520 data consisting of 312 positive reviews and 208 negative reviews. The best results are obtained in scenario 1 (90:10) using the RBF (Radial Basis Function) kernel which produces 92.31% accuracy, 96.3% precision, 89.66% recall, and 92.86% f1-score.

Keywords: Sentiment Analysis, Support Vector Machine, MOLA

ABSTRAK

MOLA adalah salah satu aplikasi platform *streaming* video di *google play store*. Aplikasi mola sudah diunduh sebanyak 5 juta kali namun hanya memiliki rating 3,2. Pada *google play store* pemberian rating aplikasi diikuti dengan ulasan dari para pengguna terhadap aplikasi. Ulasan yang ada cukup banyak dan tidak terstruktur serta mengandung opini dari para pengguna tentang kepuasan mereka terhadap aplikasi tersebut sehingga sering dijadikan pertimbangan oleh calon pengguna untuk memilih aplikasi yang digunakan. Berdasarkan hal tersebut, maka dilakukan analisis sentimen menggunakan algoritma *Support Vector Machine* untuk mengetahui bagaimana sentimen para pengguna *terhadap* aplikasi MOLA di *google play store*. Penelitian ini menggunakan metode *Knowledge Discovery in Database (KDD)*. Data yang digunakan merupakan ulasan aplikasi MOLA sebanyak 520 data yang terdiri dari 312 ulasan positif dan 208 ulasan negatif. Hasil terbaik diperoleh pada skenario 1 (90:10) menggunakan kernel RBF (*Radial Basis Function*) yang menghasilkan *accuracy* 92,31%, *precision* 96,3%, *recall* 89,66%, dan *f1-score* 92,86%.

Kata Kunci: Analisis Sentimen, *Support Vector Machine*, MOLA

PENDAHULUAN

Banyaknya aplikasi yang telah tersedia saat ini membuktikan bahwa telah terjadi perkembangan teknologi dan internet yang sangat cepat. Berdasarkan data dari Asosiasi Penyelenggara Jasa Internet Indonesia pada tahun 2020, pengguna internet di Indonesia hingga akhir tahun 2020 diperkirakan sudah mencapai 196,7 juta pengguna. Aplikasi yang sudah tersedia dapat dengan mudah diunduh melalui

Google Play Store pada perangkat *mobile* berbasis android. Menurut databoks.katadata.co.id, unduhan aplikasi global pada kuartal pertama tahun 2021 saja meningkat sebesar 8,7% dibandingkan periode yang sama tahun 2020, dengan *Google Play Store* menjadi penyumbang unduhan terbesar sebesar 3,8 miliar (Rizki et al., 2020; Sarbaini et al., 2021).

MOLA merupakan salah satu aplikasi yang terdapat di *Google Play*

Store. MOLA merupakan aplikasi *platform streaming video* yang menyajikan program hiburan dan olahraga. Aplikasi MOLA dirilis di Google Play Store pada tanggal 19 Juli 2019. Aplikasi MOLA sangat cocok digunakan saat sedang berpergian karena hanya membutuhkan koneksi internet.

Kini pengguna berbayar MOLA per Agustus 2021 sudah mencapai lebih dari 1 juta dengan tingkat pertumbuhan tahunan sebesar 100% dalam jumlah pengguna di Indonesia (Perwitasari, 2021). Pada Google Play Store hingga bulan Januari 2021 aplikasi MOLA telah diunduh sebanyak 5 juta kali namun hanya memiliki rating aplikasi sebesar 3,2.

Pemberian rating aplikasi di *Google Play Store* diikuti dengan ulasan dari para pengguna terhadap aplikasi tersebut. Ulasan tersebut mengandung opini dari para pengguna mengenai aplikasi tersebut dan calon pengguna melihat ulasan dari sebuah aplikasi sebagai pertimbangan sebelum memutuskan untuk menggunakan aplikasi tersebut (Saputra et al., 2019). Ulasan di *Google Play Store* ini cukup banyak dan tidak terstruktur sehingga diperlukan suatu teknik untuk mengetahui bagaimana ulasan pengguna terhadap aplikasi tersebut (Fransiska & Irham Gufroni, 2020). Oleh karena itu, diperlukan analisis sentimen pada data ulasan. Analisis sentimen adalah teknik untuk mengolah data tekstual untuk memperoleh informasi pada teks (Wahyudi & Kusumawardana, 2021).

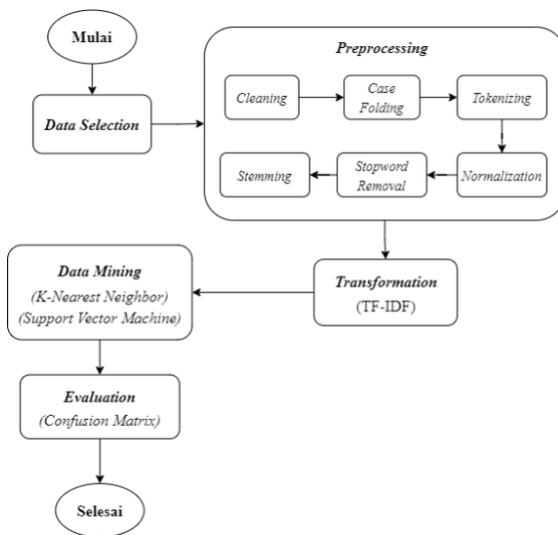
Terdapat beberapa algoritma yang dapat diterapkan dalam analisis sentimen seperti *K-Nearest Neighbor (KNN)*, *Naïve Bayes*, dan *Support Vector Machine (SVM)*. Berdasarkan beberapa referensi, SVM menghasilkan akurasi terbaik dalam melakukan analisis sentimen dibandingkan algoritma klasifikasi yang lain. Pada penelitian

sebelumnya SVM memiliki tingkat akurasi yang lebih tinggi sebesar 81,46% dibandingkan *Naïve Bayes* sebesar 75,41% saat diterapkan untuk analisis sentimen ulasan Bahasa Indonesia di *Google Play Store* (Ilmawan & Mude, 2020). Pada penelitian lain yang membandingkan SVM, *Naïve Bayes*, dan KNN untuk analisis sentimen *gadget*, SVM menghasilkan rata-rata akurasi terbaik sebesar 96,43% (Iskandar & Nataliani, 2021). SVM mempunyai fungsi yang bisa mentransformasikan data ke ruang dimensi yang lebih tinggi yaitu ruang kernel yang disebut dengan fungsi kernel *trick* sehingga data dapat dipisahkan dengan lebih baik dibandingkan algoritma klasifikasi lainnya (Mukarramah et al., 2021). Beberapa kernel pada SVM yaitu kernel *Linear*, *Radial Basis Function (RBF)*, *Sigmoid*, dan *Polynomial* (Zuriel & Fahrurozi, 2021). Pada penelitian analisis sentimen ulasan aplikasi Ruangguru dengan membandingkan 3 kernel algoritma SVM diperoleh hasil bahwa kernel linear merupakan kernel terbaik dengan akurasi hingga 89,7% (Irfani et al., 2020). Sedangkan dalam penelitian lainnya akurasi terbaik diperoleh oleh kernel RBF dalam analisis sentimen pada maskapai penerbangan sebesar 84,37% (Husada & Paramita, 2021).

Berdasarkan penjelasan di atas, karena belum ada hasil pasti mengenai akurasi kernel terbaik dalam melakukan analisis sentimen, maka pada penelitian ini akan dilakukan perbandingan akurasi antara kernel *Linear*, kernel RBF, kernel *Sigmoid*, dan kernel *Polynomial* pada algoritma SVM dalam melakukan analisis sentimen ulasan aplikasi MOLA untuk mengetahui bagaimana sentimen para pengguna terhadap aplikasi MOLA di *Google Play Store* dan akurasi kernel terbaik.

METODE

Penelitian ini menggunakan metode KDD (*Knowledge Discovery in Database*) karena memiliki keunggulan proses pengidentifikasian pola yang terorganisir dari sekumpulan data yang kompleks, sehingga datanya mudah untuk dipahami (Ramos et al., 2021; Rizki et al., 2021). Gambar 1 menunjukkan alur penelitian yang menerapkan metode KDD.



Gambar 1. Alur Penelitian

Alur penelitian pada gambar 1 dijelaskan sebagai berikut.

1. Data Selection

Data selection merupakan tahapan pertama dalam KDD, pada tahap ini akan dilakukan pengumpulan, seleksi, dan pelabelan data. Data dikumpulkan dari situs *Google Play Store* pada tanggal 1 Desember 2021 sampai 31 Januari 2022.

2. Preprocessing

Pada tahap preprocessing ini, data mentah yang telah terkumpul akan diolah menjadi data yang dapat digunakan pada tahap selanjutnya. Tahap *preprocessing* terdiri dari 6 proses yaitu:

a. *Cleaning*, merupakan proses untuk menghilangkan atribut yang tidak

berpengaruh terhadap klasifikasi yaitu tanda baca, karakter kosong, dan emoji.

- b. *Case Folding*, merupakan proses untuk mengubah semua huruf yang ada pada dokumen menjadi huruf kecil.
- c. *Tokenizing*, merupakan proses untuk memisahkan kalimat menjadi beberapa bagian kata.
- d. *Normalization*, merupakan proses untuk memperbaiki kesalahan yang ada pada kata seperti ejaan yang salah agar kata yang memiliki makna sama menjadi setara.
- e. *Stopword Removal*, merupakan proses untuk menghilangkan kata-kata yang dianggap tidak berpengaruh terhadap kalimat.
- f. *Stemming*, merupakan proses untuk mengubah kata-kata yang ada menjadi bentuk kata dasar.

3. Transformation

Tahap transformation adalah tahap mengubah data menjadi bentuk yang dapat diolah pada tahapan data mining. Pada tahap ini akan dilakukan pembagian data ke dalam 3 skenario terlebih dahulu yaitu skenario 1 (90% data *training* dan 10% data testing), skenario 2 (80% data *training* dan 10% data testing), skenario 3 (90% data *training* dan 10% data testing). Setelah itu akan dilakukan pembobotan kata dengan TF-IDF yang berguna untuk mengubah data berupa teks menjadi vektor bobot. Perhitungan metode TF-IDF dinyatakan dalam persamaan sebagai berikut.

$$w_t = TF_{t,d} \times IDF_t = TF_{t,d} \times \log \frac{N}{DF_t}$$

4. Data Mining

Pada tahap data mining akan dilakukan proses klasifikasi sentimen pada data ulasan dengan menggunakan empat kernel algoritma *Support Vector*

Machine yaitu Kernel *Linear*, Kernel RBF, Kernel *Sigmoid*, dan Kernel *Polynomial*. Berikut merupakan persamaan dari setiap kernel:

- a. Kernel *Linear*

$$K(x_i, x) = x_i x$$

- b. Kernel RBF

$$K(x_i, x) = \exp\left(\frac{-\|x_i - x\|^2}{2\sigma^2}\right)$$

- c. Kernel *Polynomial*

$$K(x_i, x) = (x_i x)^d$$

- d. Kernel *Sigmoid*

$$K(x_i, x) = \tanh(\sigma(x_i x) + c)$$

e. **Evaluation**

Tahap evaluasi akan menggunakan *confusion matrix* untuk mengetahui performansi dari setiap kernel algoritma SVM dengan hasil evaluasi berupa nilai *accuracy*, *precision*, *recall*, dan *f1-score*.

HASIL DAN PEMBAHASAN

1. Data Selection

Pengumpulan data ulasan aplikasi MOLA yang dilakukan dengan teknik scraping pada tanggal 1 Desember 2021 sampai 31 Januari 2022 menggunakan *library google play scraper* menghasilkan sebanyak 520 data. Data yang telah terkumpul memiliki 5 atribut yaitu *reviewId*, *userName*, *at*, *content*, dan *score*. Selanjutnya dilakukan penyeleksian atribut yang digunakan yaitu hanya *content*. Kemudian dilakukan pelabelan data ulasan secara manual dan divalidasi oleh ahli Bahasa dan Sastra Indonesia. Pelabelan data tersebut menghasilkan 312 ulasan positif dan 208 ulasan negatif. Hasil pada tahap *data selection* dapat dilihat pada gambar 2.

		Ulasan	Label
0	Awalnya bintang 4. Skrang 1. Knpa... Waktu per...		Negatif
1	Jika bkn karena tayangan liga inggris sy males...		Negatif
2	Terus diperbarui agar makin mantap		Positif
3	Alhamdulillah gak memegang hak siar EPL lagi, ma...		Negatif
4		Mpoz	Negatif
...	
515	Mau tanya dong min Kok akun punya saya gak bis...		Negatif
516	Waduh baru mau langganan ternyata banyak perma...		Negatif
517		Mantap	Positif
518		Ok	Positif
519		Siap aktif	Positif

520 rows × 2 columns

Gambar 2. Hasil Data Selection

2. Preprocessing

Pada tahap ini dilakukan untuk mengubah data agar sesuai dan dapat diolah pada tahapan selanjutnya. Contoh hasil dari 6 proses pada tahap *preprocessing* yang telah dilakukan dapat dilihat pada tabel 1.

Tabel 1. Contoh Hasil Preprocessing

Proses	Hasil
Data Ulasan	Aplikasi anti kritik, yang error terus tapi ga ada permintaan maaf atau ganti rugi
<i>Cleaning</i>	Aplikasi anti kritik yang error terus tapi ga ada permintaan maaf atau ganti rugi
<i>Case Folding</i>	aplikasi anti kritik yang error terus tapi ga ada permintaan maaf atau ganti rugi
<i>Tokenizing</i>	['aplikasi', 'anti', 'kritik', 'yang', 'error', 'terus', 'tapi', 'ga', 'ada', 'permintaan', 'maaf', 'atau', 'ganti', 'rugi']
<i>Normalization</i>	['aplikasi', 'anti', 'kritik', 'yang', 'eror', 'terus', 'tapi', 'tidak', 'ada', 'permintaan', 'maaf', 'atau', 'ganti', 'rugi']
<i>Stopword Removal</i>	['aplikasi', 'anti', 'kritik', 'eror', 'permintaan', 'maaf', 'ganti', 'rugi']
<i>Stemming</i>	['aplikasi', 'anti', 'kritik', 'eror', 'minta', 'maaf', 'ganti', 'rugi']

3. Transformation

Tahap *transformation* ini diawali dengan melakukan pembagian data menjadi data *training* dan data testing dalam 3 skenario. Pembagian data ini menggunakan fungsi *train test split*

dengan parameter *random_state*=42. Tabel 2 menunjukkan hasil pembagian data dalam 3 skenario.

Tabel 2. Hasil Pembagian Data

Skenario Rasio Perbandingan	Data Training	Data Testing
90:10	468	52
80:20	416	104
70:30	364	156

Setelah data terbagi menjadi data training dan data testing, selanjutnya dilakukan pembobotan kata dengan metode TF-IDF. Gambar 3 merupakan contoh hasil pembobotan kata dengan TF-IDF pada data *training* skenario 1 (90:10).

468 rows x 578 columns

Gambar 3. Contoh Hasil TF-IDF

4. Data Mining

Pada tahap ini dilakukan klasifikasi sentimen dengan tiga skenario pembagian data yang telah dilakukan menggunakan empat kernel pada algoritma *Support Vector Machine*. Hasil klasifikasi dipaparkan pada tabel 3.

Tabel 3. Hasil Klasifikasi SVM

Skenario	Kernel	Accuracy
90:10	<i>Linear</i>	88,46%
	RBF	92,31%
	<i>Polynomial</i>	88,46%
	<i>Sigmoid</i>	90,38%
80:20	<i>Linear</i>	90,38%
	RBF	91,35%
	<i>Polynomial</i>	90,38%
70:30	<i>Sigmoid</i>	89,42%
	<i>Linear</i>	89,74%

RBF	90,38%
<i>Polynomial</i>	89,74%
<i>Sigmoid</i>	90,38%

5. Evaluation

Tahap ini dilakukan evaluasi dengan *confusion matrix*. Tabel 4 memaparkan hasil perbandingan dari seluruh evaluasi menggunakan *confusion matrix* terhadap setiap skenario dari tahap data mining dengan 4 kernel algoritma *Support Vector Machine*.

Tabel 4. Perbandingan Hasil Evaluasi Skenario 1 (90:10)

Skenario	Kernel	Accuracy	Precision	Recall	F1-score
90:10	<i>Linear</i>	88,46%	89,66%	89,66%	89,66%
	RBF	92,31	96,3%	89,66%	92,86%
	<i>Polynomial</i>	88,46%	89,66%	89,66%	89,66%
	<i>Sigmoid</i>	90,38%	90%	93,1%	91,53%
80:20	<i>Linear</i>	90,38%	93,33%	90,32%	91,8%
	RBF	91,35%	94,92%	90,32%	92,56%
	<i>Polynomial</i>	90,38%	93,33%	90,32%	91,8%
	<i>Sigmoid</i>	89,42%	91,8%	90,32%	91,06%
70:30	<i>Linear</i>	89,74%	92,13%	90,11%	91,11%
	RBF	90,38%	94,19%	89,01%	91,53%
	<i>Polynomial</i>	89,74%	92,13%	90,11%	91,11%
	<i>Sigmoid</i>	90,38%	93,18%	90,11%	91,62%

Tabel 4 menunjukkan bahwa hasil evaluasi terbaik diperoleh pada skenario 1 dengan menggunakan kernel RBF yang menghasilkan akurasi sebesar 92,31%, *precision* 96,3%, *recall* 89,66%, dan *f1-score* 92,86%.

Kata-kata yang sering muncul pada data ulasan divisualisasikan dalam bentuk word cloud. Gambar 4 menunjukkan hasil visualisasi ulasan positif, sedangkan Gambar 5 menunjukkan hasil visualisasi ulasan negatif.



Gambar 4. Word Cloud Ulasan Positif



Gambar 5. Word Cloud Ulasan Negatif

Berdasarkan gambar 4 dan 5, kata-kata yang sering muncul pada ulasan positif yaitu “bagus”, “mola”, “mantap”, dan “gratis”. Sedangkan kata-kata yang sering muncul pada ulasan negatif yaitu “aplikasi”, “mola”, “error”, “macet”.

SIMPULAN

Berdasarkan penelitian yang telah dilakukan, maka didapat kesimpulan sebagai berikut. Sentimen pengguna terhadap aplikasi MOLA menghasilkan 312 ulasan positif dan 208 ulasan negatif. Pada hasil visualisasi kata-kata yang sering muncul pada ulasan positif yaitu “bagus”, “mola”, “mantap”, dan “gratis”. Sedangkan kata-kata yang sering muncul pada ulasan negatif yaitu “aplikasi”, “mola”, “error”, “macet”. Pada hasil evaluasi kinerja empat kernel algoritma *Support Vector Machine* dalam menganalisis sentimen ulasan aplikasi MOLA dengan tiga skenario *split data* diperoleh hasil terbaik pada skenario 1 dengan rasio perbandingan

data *training* 90% dan data *testing* 10% dengan hasil *accuracy* 92,31%, *precision* 96,3%, *recall* 89,66%, dan *f1-score* 92,86%

Adapun saran untuk penelitian selanjutnya yaitu:

1. Menggunakan *dataset* yang lebih besar agar hasil akurasi yang didapat semakin akurat.
2. Melakukan skenario yang dapat mengatasi *imbalanced data*.
3. Dapat menerapkan seleksi fitur seperti *Information Gain*, *Chi-Square*, dan sebagainya.

DAFTAR PUSTAKA

- Fransiska, S., & Irham Gufroni, A. (2020). Sentiment Analysis Provider by.U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method. *Scientific Journal of Informatics*, 7(2), 2407–7658. <http://journal.unnes.ac.id/nju/index.php/sji>
- Husada, H. C., & Paramita, A. S. (2021). Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM). *Teknika*, 10(1), 18–26. <https://doi.org/10.34148/teknika.v10i1.311>
- Ilmawan, L. B., & Mude, M. A. (2020). Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store. *ILKOM Jurnal Ilmiah*, 12(2), 154–161. <https://doi.org/10.33096/ilkom.v12i2.597.154-161>
- Irfani, F. F., Triyanto, M., Hartanto, A. D., & Kusnawi. (2020). Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector

- Machine. *JBMI (Jurnal Bisnis, Manajemen, Dan Informatika)*, 16(3), 258–266. <https://doi.org/10.26487/jbmi.v16i3.8607>
- Iskandar, J. W., & Nataliani, Y. (2021). Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek. *JURNAL RESTI*, 5(158), 1120–1126.
- Mukarramah, R., Atmajaya, D., & Ilmawan, L. B. (2021). Performance comparison of support vector machine (SVM) with linear kernel and polynomial kernel for multiclass sentiment analysis on twitter. *ILKOM Jurnal Ilmiah*, 13(2), 168–174. <https://doi.org/10.33096/ilkom.v13i2.851.168-174>
- Perwitasari, A. S. (2021). *Tumbuh 100%, Mola kini punya lebih dari 1 juta pelanggan berbayar di Indonesia*. Kontan.Co.Id. <https://industri.kontan.co.id/news/tumbuh-100-mola-kini-punya-lebih-dari-1-juta-pelanggan-berbayar-di-indonesia>
- Ramos, S., Soares, J., Cembranel, S. S., Tavares, I., Foroozandeh, Z., Vale, Z., & Fernandes, R. (2021). Data mining techniques for electricity customer characterization. *Procedia Computer Science*, 186, 475–488. <https://doi.org/10.1016/j.procs.2021.04.168>
- Rizki, M., Umam, M. I. H., & Hamzah, M. L. (2020). Aplikasi Data Mining Dengan Metode CHAID Dalam Menentukan Status Kredit. *SITEKIN: Jurnal Sains, Teknologi dan Industri*, 18(1), 29-33.
- Rinaldi, A., Rahmadani, N., Papilo, P., Silvia, S., & Rizki, M. (2021). Analisa Pengambilan Keputusan Pemilihan Bahan Dalam Pembuatan Kemeja Menggunakan Metode TOPSIS. *SITEKIN: Jurnal Sains, Teknologi dan Industri*, 18(2), 163-172.
- Sarbaini, S., Cynthia, E. P., & Arifandy, M. I. (2021). Pengelompokan Diabetic Macular Edema Berbasis Citra Retina Mata Menggunakan Fuzzy Learning Vector Quantization (FLVQ). *SITEKIN: Jurnal Sains, Teknologi dan Industri*, 19(1), 75-80.
- Saputra, S. A., Rosiyadi, D., Gata, W., & Husain, S. M. (2019). Sentiment Analysis Analysis of E-Wallet Sentiments on Google Play Using the Naive Bayes Algorithm Based on Particle Swarm Optimization. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(3), 377–382. <https://doi.org/10.29207/resti.v3i3.1118>
- Wahyudi, R., & Kusumawardana, G. (2021). Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine. *Jurnal Informatika*, 8(2), 200–207. <https://doi.org/10.31294/ji.v8i2.9681>
- Zuriel, H. P. P., & Fahrurrozi, A. (2021). Implementasi Algoritma Klasifikasi Support Vector Machine Untuk Analisa Sentimen Pengguna Twitter Terhadap Kebijakan Psbb. *Jurnal Ilmiah Informatika Komputer*, 26(2), 149–162. <https://doi.org/10.35760/ik.2021.v26i2.4289>