

ANALISIS SENTIMEN TERHADAP REVIEW PENGGUNA INDRIVE DI GOOGLE PLAYSTORE MENGGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE*

Oleh:

Calvin Matulesy¹⁾

Adi Purnama²⁾

Universitas Widyatama, Bandung^{1,2)}

E-mail:

calvin.matulesy@widyatama.ac.id¹⁾

adi.purnama@widyatama.ac.id²⁾

ABSTRAK

Analisis sentimen merupakan suatu proses untuk menentukan status sebuah data berdasarkan sentimen yang di berikan. Sentimen yang akan di analisis biasanya sentimen positif dan negatif terhadap sebuah objek. Pada penelitian ini akan menganalisis sebuah sentimen dari pengguna aplikasi *Indrive* yang tengah beroperasi di Indonesia. Opini yang diberikan oleh setiap pengguna *Indrive* di Indonesia akan dianalisis berdasarkan tiga kelas sentimen, yaitu positif, netral dan negatif. Tujuan dari penelitian ini untuk menganalisis sentimen dari kepuasan pengguna dalam menggunakan *Indrive* dengan berbagai ulasan atas penggunaan aplikasi transportasi *online* dengan algoritma *Support Vector Machine*. Metodologi penelitian ini menggunakan tahapan pengumpulan data, pelabelan data, *pra*-proses, pembobotan kata, metode klasifikasi, evaluasi, dan hasil klasifikasi. Pengujian pada penelitian ini dilakukan dengan 3 (tiga) perbandingan *dataset* yaitu 9:1, 7:3, dan 5:5. Evaluasi dengan perhitungan *accuracy*, *precision*, *recall*, dan *f1-score* pada pengujian *Confusion Matrix*, di dapatkan hasil terbaik berdasarkan 3 kelas (positif, netral, dan negatif) pada perbandingan 90% data *training* dan 10% data *test*. Hasil *accuracy* sebesar 87,71% dengan rata-rata tiap kelas sentimen 87% *precision*, 88% *recall* dan 88% *f1-score*. Hasil akurasi akan menjadi lebih baik jika data *training* yang digunakan lebih besar dan data *test* lebih kecil.

Kata Kunci: Analisis Sentimen, *Indrive*, *Google Playstore*, *Support Vector Machine*

1. PENDAHULUAN

Perkembangan teknologi informasi beriringan dengan peningkatan inovasi dalam ilmu pengetahuan. Bahkan pengembangan teknologi saat ini menciptakan transportasi online yang menjadi tren transportasi umum saat ini di kalangan masyarakat. Sehingga banyak perusahaan berlomba-lomba untuk membentuk perusahaan transportasi online berbasis aplikasi online, salah satunya yaitu *Indrive*.

Dengan fenomena tersebut banyak masyarakat yang memberikan opininya tentang transportasi *online* di Indonesia saat ini melalui media sosial. Tidak hanya dengan media sosial, *Google Playstore* juga dapat menjadi tempat masyarakat

memberikan pendapat dan opininya. Penilaian sentimen dari pengguna aplikasi diambil dari dua atribut, yaitu nilai *rating* dan nilai komentar. Namun, untuk mengetahui tanggapan dari setiap konsumen terhadap kepuasan produk yang digunakan, dibutuhkan analisis data yang

relevan agar menjadi masukan bagi perusahaan *Indrive* itu sendiri.

Penelitian terdahulu yang menguji analisis sentimen terhadap penggunaan aplikasi gojek, (Irawan et al., 2022) membuktikan bahwa analisis sentiment pengguna aplikasi gojek dengan algoritma *K-Nearest Neighbor*. Data yang digunakan sebanyak 1409 berupa respons pengguna gojek di media sosial Twitter dengan menggunakan Twitter API. Hasil akurasi dan evaluasi menggunakan *Confusion Matrix*. Klasifikasi algoritma *K-Nearest Neighbor* ini terdiri dari perbandingan data latih dan data uji yaitu 8:2, dengan tingkat keakurasian 79,43% pada $k=15$. Hasil perhitungan dengan menggunakan parameter *precision* menunjukkan nilai positif 0.95, netral 0.72, dan negatif 0.44, parameter *recall* menunjukkan nilai positif 0.76, netral 0.94, dan negatif 0,17. Sedangkan hasil perhitungan dari parameter *f1-score* dari kedua perhitungan menunjukkan nilai positif 0,84, netral 0,81, dan negatif 0,25. Lebih lanjut penelitian lainnya yang menguji analisis sentimen terhadap *review customer* layanan ekspedisi JNE dan J&T Express, (Ika et al., 2021) penelitian ini menggunakan algoritma *Naïve Bayes*. Hasil performansi evaluasi analisis sentimen ini memiliki tingkat keakurasian yang paling baik dengan mengambil perbandingan data latih dan data uji 75:25.

Pada penelitian ini didapatkan tingkat keakurasian terhadap JNE sebesar 79%, dengan hasil evaluasi rata-rata dari *precision* 80%, *recall* 79%, dan *f1-score* 79%. Sedangkan hasil akurasi terhadap J&T Express sebesar 76%, dengan hasil rata-rata evaluasi parameter *precision* 76%, *recall* 76%, dan *f1-score* 76%.

Pada penelitian ini menggunakan algoritma *Support Vector Machine* dengan data yang lebih banyak agar memiliki nilai akurasi yang lebih baik. Algoritma SVM bertujuan dalam mengklasifikasi sebuah teks dengan pembobotan sebagai fitur. Penggunaan SVM sudah digunakan sejak tahun 1992 oleh Boser, Guyon, dan Vapnik. *Support Vector Machine* adalah sebuah metode yang digunakan dalam memprediksi permasalahan pengklasifikasian dan regresi (Athira Luqyana et al., 2018).

Berdasarkan uraian permasalahan di atas menunjukkan kebutuhan suatu *system* pengelolaan setiap respons pengguna aplikasi secara *rating* dan komentar. Dengan *system* ini dapat digunakan sebagai informasi penilaian sentimen terhadap aplikasi (Tri Anjasmos & Fitri Marisa, 2020). Analisis sentimen merupakan suatu proses untuk menentukan status sebuah data berdasarkan sentimen yang di berikan.

Tujuan dari penelitian ini untuk menganalisis sentimen kepuasan pengguna

Indrive dengan berbagai ulasan atas penggunaan aplikasi transportasi *online* dengan algoritma *Support Vector Machine*. Sehingga di harapkan penelitian dengan metode ini mendapatkan nilai akurasi yang lebih baik dan memberikan informasi kepada perusahaan *Indrive* atas evaluasi dari pengguna aplikasi *Indrive*.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen digunakan untuk mengumpulkan suatu polaritas dari ulasan atau pemikiran pada dokumen dengan nilai positif, negatif, dan netral.

2.2. Text Mining

Text mining merupakan sebuah teknik bagian dari data *mining* dengan proses dalam menemukan pola yang menarik dari kumpulan data teks yang jumlahnya banyak. (Sari, 2021). Proses utama dalam teknik ini untuk menemukan kata-kata dari dokumen harus melakukan analisis antar dokumen dengan menggunakan teknik statistik tertentu yaitu, analisis kelompok, klasifikasi dan juga asosiasi (Pramadani et al., n.d.).

2.3. Data Preprocessing

Data dari hasil *scrapping* tidak dapat diklasifikasi langsung dikarenakan belum dibersihkan, oleh sebab itu, diperlukan *pre-processing* data agar hasilnya terstruktur dan sudah bersih sehingga

dapat diklasifikasi (Mas Pintoko & Muslim, n.d.).

- **Cleaning**

Cleaning merupakan proses untuk menghilangkan karakter di dalam teks. Contohnya seperti '@', '#', links 'HTTP' dan karakter yang lain ~!@#\$\$%^&*()_+?<>.,:;{}[].

- **Casefolding**

Pada tahap *case folding*, input yang diterima pada sebuah kalimat hanya huruf 'a' sampai 'z'. Jika ada huruf lain selain huruf tersebut, maka akan dihilangkan

- **Tokenization**

Tokenization, yaitu proses memecah kalimat menjadi kata yang berdiri sendiri atau menjadi potongan-potongan.

- **Text Normalization**

Normalisasi teks dilakukan dengan memeriksa setiap kata terhadap kamus kata slang, kata abreviasi (singkatan) atau kamus kata tidak baku dari penelitian jbkalit.

- **Stopwords Removal**

Stopwords Removal, merupakan proses dalam menghilangkan sebuah kosakata. Contohnya kata penghubung, kata keterangan, dan kata yang tidak unik, misalnya “seorang”, “dan”, “atau” dan lainnya.

- **Stemming**

Stemming yaitu proses mengubah kata menjadi kata dasar atau bentuk asli dari kata tersebut.

2.4. Support Vector Machine

Proses *Support Vector Machine* (SVM) digunakan untuk memisahkan fungsi terbaik dari fungsi yang sudah ada dalam dua macam objek. Penyelesaian klasifikasi dua kelas dapat di selesaikan dengan persamaan berikut (Hendrastuty et al., 2021):

$$\min w^{ij} \frac{1}{2} (w^{ij})^T w^{ij} + c \sum_r t_r^{ij} \quad (1)$$

Subject to :

$$w^{ij} x_r + b^{ij} \geq t_r^{ij}, \text{ jika } y_r = i$$

$$w^{ij} x_r + b^{ij} \leq t_r^{ij}, \text{ jika } y_r \neq i$$

$$t_r^{ij} \geq 0,$$

2.5. Confusion Matriks

Confusion Matrix digunakan sebagai evaluasi untuk menghitung pencapaian sebuah model dalam memprediksi nilai positif sejati (TP), negatif sejati (TN), nilai positif palsu (FP), dan negatif palsu (FN) (Ilahi et al., 2022)

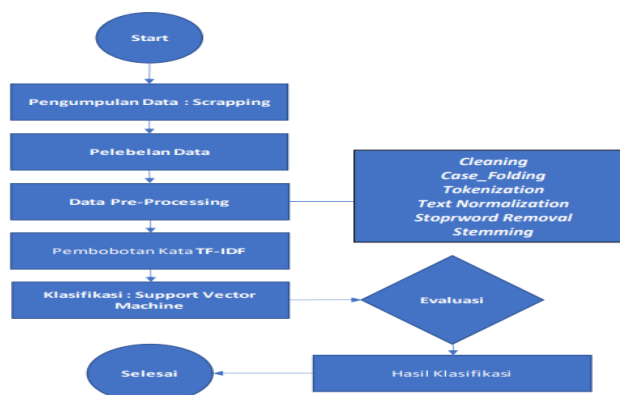
Confusion Matrix TF-IDF				
Prediction	Positif	True Positive	False Positive	False Positive
	Netral	False Neutral	True Neutral	False Neutral
	Negatif	False Negative	False Negative	True Negatif
		Positif	Netral	Negatif
		Actual		

Gambar 1. Contoh Confusion Matrix 3 x 3

3. METODE PENELITIAN

Metode penelitian ini berisi tentang langkah-langkah untuk memperoleh hasil analisis *sentimen* menggunakan metode

Support Vector Machine (Sholekha et al., 2022). Gambar di bawah ini merupakan tahapan melakukan *sentimen* analisis.



Gambar 2. Flow Chart Tahapan Analisis Sentimen

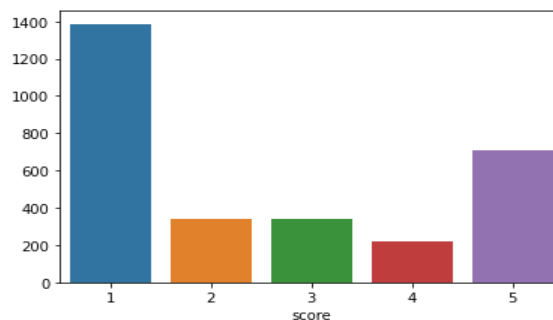
4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Mengambil data *sentimen* pengguna aplikasi *Indrve* pada *Google Playstore* dengan cara *scraping* data menggunakan *Google Colap*. Data yang di ambil merupakan data *sentiment* dengan *range* waktu dari tahun 2022 hingga 2023 saat ini. Keseluruhan data yang dalam

penelitian ini sebanyak 3000 data dari *sentimen* pengguna aplikasi *indrive* dan dilakukan pembagian *dataset* menjadi data *training* dan data testing.

Di bawah ini merupakan grafik data hasil *scraping* dengan *score/rating* yang diberikan sebagai penilaian saat menggunakan aplikasi *Indrive*.



Gambar 3. Data Review Pengguna Indrive Dengan Score

Penelitian ini menggunakan metode kuantitatif. Metode kuantitatif adalah penelitian terhadap masalah-masalah sosial yang didasarkan dengan uji teoritis dari nilai variabel, yakni dapat diukur dengan angka, dan di analisa menggunakan teknik statistik untuk menganalisis apakah penerapan dalam memprediksi teoritis itu benar (Sholekha et al., 2022).

• Teknik Pengumpulan Data

Adapun beberapa cara dalam mengumpulkan data sebagai berikut :

A. Observasi

Observasi merupakan tinjauan langsung terhadap sebuah objek yang diteliti dan merupakan kegiatan yang sedang berlangsung pada penelitian.

B. Penelitian Perpustakaan

Pengumpulan data ini merupakan tahap untuk mempelajari, meneliti dan mengkaji berbagai sumber referensi dari buku, jurnal , situs di internet, dan bacaan yang sesuai dengan penelitian yang di lakukan.

4.2. Pelabelan Data

Pelabelan data dilakukan dengan cara membubuhi tanda pada data yang

mencerminkan sentimen *positive*, *neutral*, dan *negative* terhadap sebuah ulasan data (Alrajak et al., 2020). Label data adalah hasil tahapan sebelumnya yang dilakukan perhitungan polaritas dari sentimen ulasan yang terambil, dan menghasilkan tiga

label, yaitu positif, netral, dan negatif (Herlinawati et al., 2020). Pada tahap ini akan di label sesuai dengan score atau *rating* yang akan diberikan oleh pengguna aplikasi *indrive*.

Tabel 1. Contoh Label Data

No	Data	Rating	Label
1	Aplikasinya bagus bgt, saya suka dengan fitur-fitur aplikasinya. Sangat memudahkan pengguna nya 😊😊.	5	Positif
2	Fitur nya bagus sih tapia gak kurang dikit ketika autobidnya di perbatuhi	3	Netral
3	Saya sangat susah untuk menggunakan aplikasinya dikarenakan fitur peesanannya kurang sesuai.	1	Negatif

Di bawah merupakan tabel pengguna *Indrive* dari jumlah label berdasarkan tahun dengan total sentimen pada tahun

2022 mencapai 1307 dan pada tahun 2023 mencapai 1693 :

Tabel 2. Total Jumlah Sentimen Berdasarkan Tahun

Label	Tahun		Total
	2022	2023	
Positif	435	488	923
Netral	153	188	341
Negatif	719	1017	1736

4.3. Data Preprocessing

Preprocessing data merupakan proses melakukan pembersihan dan penyaringan

data yang tidak terstruktur menjadi data yang terstruktur.

Tabel 3. Tahapan Pre-Processing

Contoh :

Aplikasinya bagus bgt, saya suka dengan fitur-fitur aplikasinya. Sangat memudahkan pengguna nya 😊😊.

Pre-Processing	Hasil
Cleaning	Aplikasinya bagus bgt saya suka dengan fitur fitur aplikasinya Sangat

	memudahkan pengguna nya
Case Folding	aplikasinya bagus bgt saya suka dengan fitur fitur aplikasinya sangat memudahkan pengguna nya
Tokenization	'aplikasinya', 'bagus', 'bgt', 'saya', 'suka', 'dengan', 'fitur', 'fitur', 'aplikasinya', 'sangat', 'memudahkan', 'penggunanya'.
Text Normalization	'aplikasinya', 'bagus', 'sekali', 'saya', 'suka', 'dengan', 'fitur', 'fitur', 'aplikasinya', 'sangat', 'memudahkan', 'penggunanya'.
Stopwords Removal	'aplikasinya', 'bagus', 'suka', 'fitur', 'fitur', 'aplikasinya', 'memudahkan'.
Stemming	'aplikasinya', 'bagus', 'suka', 'fitur', 'fitur', 'aplikasinya', 'mudah'.

4.4. Visualisasi

Hasil dari pra proses data ditunjukkan melalui visualisasi berupa *wordcloud*. Gambar 4 menyajikan visualisasi *wordcloud* untuk visual performa metode yang digunakan.

4.5. Pembobotan Kata TF-IDF

Proses ini dilakukan untuk memberikan bobot kata di tiap dokumen atau ulasan menggunakan TF-IDF agar memiliki nilai yang bisa membantu proses klasifikasi dengan perhitungan sesuai dengan rumus (3) (Alrajak et al., 2020).

$$TF = f(x) = \begin{cases} 1 + \log_{10}(tf_{t,d}), & \text{if } tf_{t,d} > 0 \\ 0, & \text{if } tf_{t,d} = 0 \end{cases} \quad (2)$$

$$IDF = \log \left(\frac{N}{df_t} \right) \quad (3)$$

Keterangan :

$tf_{t,d}$: jumlah kemunculan term (t) pada dokumen (d), jika tidak ada term atau $t=0$, maka TF menjadi 0.

N : jumlah dokumen teks.

df_t : jumlah dokumen yang mengandung term (t).

Perhitungan bobot kata TF-IDF dari perkalian TF dan IDF :

$$w_{t,d} = TF \times IDF \quad (4)$$

4.6. Proses Klasifikasi

Pada proses awal klasifikasi bertujuan untuk membagi dokumen menjadi dua bagian yaitu untuk data latih dan data uji. Data latih digunakan sebagai data yang dipelajari mesin dengan menggunakan metode tertentu. Analisis klasifikasi algoritma *Support Vector Machine* menggunakan beberapa perbandingan rasio data *training* dan data testing.

- **Hasil Akurasi**

Penelitian ini, melakukan pengujian akurasi dengan membagi data ke dalam

data training dan data testing dengan beberapa perbandingan. Berikut ini adalah

perbandingan pada data training dan data testing yang di tunjukan pada tabel.

Tabel 4. Hasil Akurasi Berdasarkan Perbandingan Dataset

Data Training	Data Testing	Hasil Akurasi
90	10	87,71%
70	30	87,14%
50	50	84,63%

4.7. Evaluasi

Evaluasi pada penelitian ini dengan *Confusion Matrix* dengan 3000 data menggunakan parameter *accuracy*, *precision*, *recall*, dan *fi-score*.

4.8. Klasifikasi Hasil Pengujian

Hasil klasifikasi berupa hasil sentimen positif, negatif, dan netral berdasarkan parameter yang digunakan pada *Confusion Matrix*. Hasil dari pengujian di atas memiliki nilai akurasi yang berbeda. Akurasi ini di hasilkan dari perbedaan data latih dan data uji. Pengujian pertama

memiliki hasil akurasi sebesar 87,71% menggunakan perbandingan 9:1 dengan data latih sebanyak 2700 dan data uji 300, pengujian kedua mendapatkan hasil akurasi sebesar 87,14% menggunakan perbandingan 7:3 dengan data latih 2100 dan data uji 900, sedangkan pengujian terakhir hasil akurasinya sebesar 84,63%. menggunakan perbandingan 5:5 dengan data latih dan data uji yang sama banyak sebesar 1500 data. Tabel 6 berikut meyajikan hasil klasifikasi dari beberapa tahap pengujian :

Tabel 5. Hasil Klasifikasi Menggunakan Support Vector Machine

Perbandingan Data		Jumlah Data		<i>Accuracy</i>	Label Sentimen	<i>Precision</i>	<i>Recall</i>	<i>F1-Scroe</i>
Data Training	Data Test	Data Training	Data Test					
90	10	2700	300	87,71%	Positif	0,80	0,87	0,83
					Netral	0,94	1	0,97
					Negatif	0,88	0,78	0,83
					Rata-Rata	87%	88%	88%
70	30	2100	900	87,14%	Positif	0,78	0,91	0,84
					Netral	0,91	1	0,95
					Negatif	0,93	0,75	0,83
					Rata-Rata	87%	89%	87%

50	50	1500	1500	84,63%	Positif	0,73	0,91	0,81
					Netral	0,87	0,99	0,92
					Negatif	0,93	0,72	0,81
					Rata-Rata	84%	87%	85%

5. SIMPULAN

Berdasarkan hasil penelitian dengan 3000 data yang telah diambil melalui Google Playstore digunakan sebagai bahan pengujian analisis sentimen dengan algoritma *Support Vector Machine* dapat disimpulkan, yaitu :

1. Sebanyak 3 (tiga) kali pengujian yang dilakukan dengan perbandingan data *training* dan data *test* yang berbeda, menghasilkan nilai akurasi yang tinggi pada perbandingan 9:1. Hasil akurasi sebesar 87,71%, dengan rata-rata kelas *sentiment precision* 87%, *recall* 88%, dan *f1-score* 88%.
2. Hasil akurasi akan menjadi lebih baik jika data *training* yang digunakan lebih besar dan data *test* lebih kecil.
3. TF-IDF membantu perhitungan Algoritma *Support Vector Machine* dalam proses membobotkan kata.
4. Penelitian ini memiliki kelebihan pada nilai parameternya yang baik, sehingga dapat digunakan oleh sistem.

5. Penelitian ini juga memiliki kekurangan pada saat *preprocessing* data, kata-kata yang berbahasa asing tidak di bersihkan sehingga membuat akurasi dalam penelitian ini kurang maksimal.

6. DAFTAR PUSTAKA

- Alrajak, M. S., Ernawati, I., Fakultas, I. N., & Komputer, I. (2020). *ANALISIS SENTIMEN TERHADAP PELAYANAN PT PLN DI JAKARTA PADA TWITTER DENGAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)*.
- Athira Luqyana, W., Cholissodin, I., & Perdana, R. S. (2018). *Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine* (Vol. 2, Issue 11). <http://j-ptiik.ub.ac.id>
- Hendrastuty, N., Rahman Isnain, A., & Yanti Rahmadhani, A. (2021). *Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine*. 6(3). <http://situs.com>

- Herlinawati, N., Yuliani, Y., Faizah, S., Gata, W., Komputer STMIK Nusa Mandiri Jl Damai No, I., Jati Barat, W., & Selatan, J. (2020). *ANALISIS SENTIMEN ZOOM CLOUD MEETINGS DI PLAY STORE MENGGUNAKAN NAÏVE BAYES DAN SUPPORT VECTOR MACHINE* (Vol. 5, Issue 2).
- Ika, N., Kalingara, P., Pratiwi, O. N., & Anggana, H. D. (2021). *ANALISIS SENTIMEN REVIEW CUSTOMER TERHADAP LAYANAN EKSPEDISI JNE DAN J&T EXPRESS MENGGUNAKAN METODE NAÏVE BAYES SENTIMENT ANALYSIS REVIEW CUSTOMER OF JNE AND J&T EXPRESS EXPEDITION SERVICES USING NAÏVE BAYES METHOD*. 8(5).
- Ilahi, M. W., Apriyani, C. N., Desiani, A., Gofar, N., Andriani, Y., & Halim, M. R. (2022). Classification of Geometric Batik Motif Typical of Indonesian Using Convolutional Neural Network. *JURNAL TEKNIK INFORMATIKA*, 15(1), 91–100. <https://doi.org/10.15408/jti.v15i1.24968>
- Irawan, F. R., Jazuli, A., & Khotimah, T. (2022). *ANALISIS SENTIMEN TERHADAP PENGGUNA GOJEK MENGGUNAKAN METODE K-NEAREST NEIGHBORS*. *Jurnal Informatika Dan Komputer) Akreditasi KEMENRISTEKDIKTI*, 5(1). <https://doi.org/10.33387/jiko>
- jbkalit github. (2018, May 15). *Sentimen Analisis*. Github. <https://github.com/jbkalit/Sentiment-Analysis-Final-Project/blob/master/normalisasi.csv>
- Kevin, V., Que, S., Analisis, :, Transportasi, S., Iriani, A., & Purnomo, H. D. (2020). Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization In *Jurnal Nasional Teknik Elektro dan Teknologi Informasi* | (Vol. 9, Issue 2). www.tripadvisor.com,
- Mas Pintoko, B., & Muslim, K. (n.d.). *Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier*.
- Pramadani, M., Putra, R., Rizky, K., & Wardani, N. (n.d.). *JUTIM (Jurnal Teknik Informatika Musirawas) PENERAPAN TEXT MINING DALAM MENGANALISIS KEPERIBADIAN PENGGUNA MEDIA SOSIAL*.
- Sari, T. W. D. (2021). Penerapan Text Mining Dengan Menggunakan Algoritma TF-IDF Untuk Klasifikasi Genre Novel. *Pelita Informatika: Informasi Dan Informatika*, 10.

Sholekha, I., Faqih, A., & Bahtiar, A.
(2022). Sentiment Analysis of Public
Opinion Covid-19 Vaccine Using
Naïve Bayes and Random Forest
Methods. *JURNAL TEKNIK
INFORMATIKA*, 15(1), 34–43.
<https://doi.org/10.15408/jti.v15i1.248>
47