



Sentiment Analysis Tweet KTT G-20 di Media Sosial Twitter Menggunakan Metode Naïve Bayes

Arta Tirtayasa¹, Alfian Listiyo Wibowo²

^{1,2}Program Studi Sistem Informasi, Fakultas Informatika, Institut Teknologi Telkom
Purwokerto

Jl. DI Panjaitan No.128, Kec. Purwokerto Selatan., Kabupaten Banyumas, Jawa Tengah
53147

e-mail: 19103147@ittelkom-pwt.ac.id¹, 19103032@ittelkom-pwt.ac.id²

Abstract

The G-20 or The Group of Twenty is a group consisting of 19 countries with major economies plus 1 European Union. This group was formed in 1999 as a systematic forum with the aim of discussing important issues or problems related to the world economy. The countries included in the G-20 include Australia, Canada, Saudi Arabia, United States, India, Russia, South Africa, Turkey, Argentina, Brazil, Mexico, France, Germany, Italy, United Kingdom, China, India, Japan, and South Korea. From these data it can be concluded that the G-20 Summit is a forum capable of improving the standard of living of many people because of its ability to produce international policies, laws and regulations. Indonesia was once in the world's spotlight because in November 2022, Indonesia will host the G-20 Summit in Nusa Dua, Bali, to be precise. Ordinary people use Twitter to express emotions related to something, both negative and positive emotions. With the implementation of sentiment analysis data from tweets from 500 data tweets using the Naive Bayes algorithm, the result is an accuracy of 69%. The accuracy value with class precision for positive predictions is 78%, while the class precision accuracy value for negative predictions is 36%. The positive class recall accuracy value is 81%, while the negative class recall accuracy value is 32%.

Keywords: G-20, Sentiment Analysis, Twitter, Naïve Bayes

1. Pendahuluan

G-20 atau *The Group of Twenty* adalah kelompok yang terdiri dari 19 negara dengan perekonomian besar ditambah dengan 1 Uni Eropa. Kelompok ini dibentuk pada tahun 1999 sebagai forum sistematis yang bertujuan untuk membahas isu-isu atau permasalahan penting terkait dengan perekonomian dunia (Santi et al., 2022). Negara-negara yang termasuk ke dalam G-20 antara lain: Australia, Kanada, Arab Saudi, Amerika Serikat, India, Rusia, Afrika Selatan, Turki, Argentina, Brazil, Meksiko, Perancis, Jerman, Italia, Britania Raya, Cina, Indonesia, Jepang, dan Korea Selatan. G-20 didirikan guna menyediakan mekanisme baru dialog informal untuk memperluas diskusi terkait isu-isu kebijakan ekonomi dan keuangan secara sistematis dan mempromosikan kerja sama dunia yang stabil dan berkelanjutan (Hajnal, 2019).



Menurut University of Toronto, sejak KTT (Konferensi Tingkat Tinggi) G-20 pertama diadakan, setidaknya terdapat 23 isu yang telah dibahas, 194 komitmen yang terukur, dan tingkat kepatuhan rata-rata sebesar 71%. Dari data tersebut, dapat disimpulkan bahwa KTT G-20 merupakan forum yang mampu meningkatkan taraf kehidupan banyak orang atas kemampuannya dalam menghasilkan kebijakan, hukum, dan peraturan internasional (Astuti, 2020). Indonesia sempat menjadi sorotan dunia karena pada bulan November 2022, Indonesia menjadi tuan rumah KTT G-20, tepatnya di Nusa Dua, Bali. Indonesia merupakan satu-satunya negara di Asia Tenggara yang menjadi anggota G-20. Topik terkait KTT G-20 di Bali pun sempat menjadi *trending* di salah satu sosial media, yaitu Twitter.

Twitter menjadi salah satu sosial media yang paling banyak menyebarkan informasi terkait dengan KTT G-20. Masyarakat biasa menggunakan Twitter untuk mengungkapkan emosi terkait suatu hal, baik emosi negatif maupun positif. Terdapat lima kategori emosi dasar yang dimiliki oleh manusia, yaitu cinta, senang, sedih, marah, dan takut. Emosi-emosi tersebut dapat dikenali dengan menggunakan *sentiment analysis*. Selain itu, *sentiment analysis* dapat digunakan untuk melihat kecenderungan opini terhadap suatu masalah atau peristiwa yang terjadi di lingkungan masyarakat, baik opini negatif maupun positif (Gafatia & Hadinata, 2021).

2. Tinjau Pustaka

Penelitian ini memiliki referensi dari penelitian I Wayan Desta Gafatia dan Novri Hadinata yang berjudul “Analisis Pro Kontra Vaksin Covid 19 Menggunakan *Sentiment Analysis* Sumber Media Sosial Twitter” perbedaannya dengan penelitian kali ini adalah menggunakan data yang telah diambil dari API twitter dan topik yang diambil adalah covid 19 yang trending pada tahun 2021. Penelitian ini juga menggunakan metode naïve bayes dan menggunakan 100 data set dengan hasil 8 %, positive sebanyak 21 %, *weakly positive*, 4 % *strongly positive*, 61 % netral, 1% *negative*, 4% *weakly negative*, 1% *strongly negative* (Gafatia & Hadinata, 2021).

Penelitian yang dilakukan Rahmanda Wahyu Adinegoro dkk yang berjudul “Aplikasi Website Sentiment Analysis Ulasan Tokopedia” penelitian ini sama halnya menggunakan sentiment analysis menggunakan naïve bayes tetapi menggunakan 4 algoritma naïve bayes yang berbeda agar mengetahui algoritma dengan performa yang terbaik dan menghasilkan complement dimana hasil tingkat akurasi prediksi sebesar 83,66% dengan akurasi sebesar 84,37% pada pengujian sistem pada data ulasan sebanyak 32 data (Adinegoro et al., 2020).

Penelitian yang dilakukan oleh Valonia Inge Santoso dkk yang berjudul Penerapan “Sentiment Analysis Hasil Evaluasi Dosen dengan Metode Support Vector Machine” hasil dari penelitian ini adalah sentiment analysis dengan metode SVM terhadap hasil evaluasi dosen FTI UKDW Program Studi Teknik Informatika tahun ajaran 2014/2015 semester gasal. Dengan metode SVM, sistem dapat melakukan sentiment analysis dengan menggunakan 3 kelas sentimen. Akurasi tertinggi SVM pada sistem ini yaitu 67,83%. Akurasi tertinggi dicapai pada sistem yang tidak menerapkan perubahan bobot pada dokumen uji dan menggunakan range > 0 untuk kelas sentimen positif, < 0 untuk kelas sentimen negatif, dan 0 untuk kelas sentimen netral, serta pelatihan menggunakan set data latih ke 6 (Santoso¹ et al., 2017).

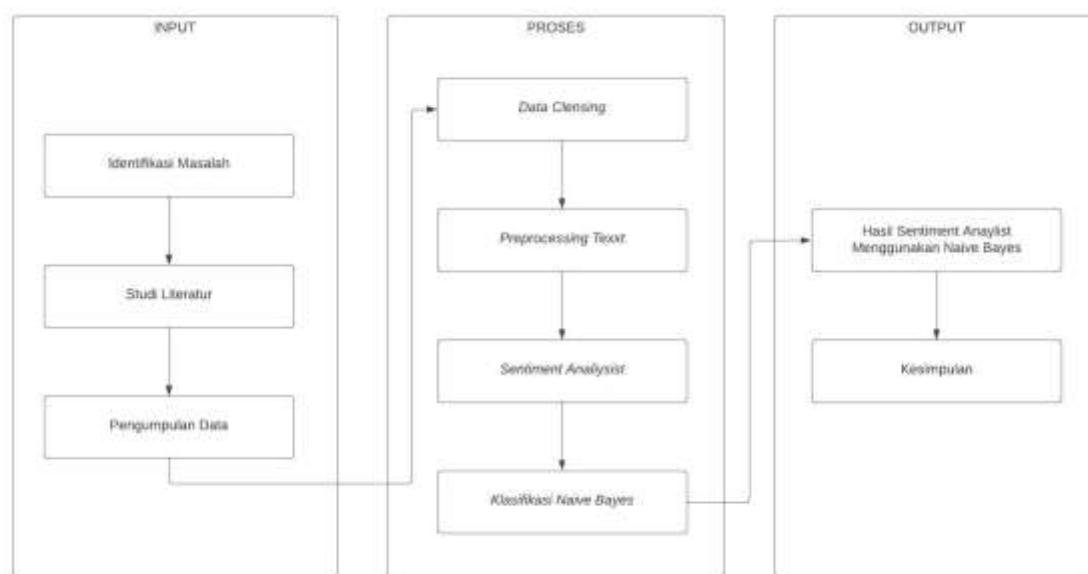
Penelitian yang dilakukan oleh Ganjar Tata Pangestu dan Miftahurrahma Rosyda yang berjudul “Sentiment Analisis Tweet Pilkada 2020 saat Pandemi COVID-19 di Media Sosial Twitter Menggunakan Metode 1D Convolutional Neural Network” hasil dari penelitian ini hasil akurasi sebesar 72.50% dengan epoch yang digunakan sebesar 25 epoch. Dari peningkatan epoch mengalami kenaikan akurasi sebesar 7-12% dari variasi epoch sebelumnya (Pangestu & Rosyda, 2022).

3. Metode Penelitian

Metode penelitian ini menggunakan Teknik *Sentiment Analysis* yang bersumber dari media sosial yaitu Twitter. Skema penelitian yang dilakukan oleh penulis yang pertama adalah mendaftarkan akun ke twitter developer yang beralamat di situs <https://developer.twitter.com/> yang telah disediakan oleh pihak Twitter, setelah mendaftarkan Twitter *developer* kemudian penulis bisa mendapatkan akses keys inilah yang akan digunakan untuk aplikasi meneliti tweets dari sebuah akun Twitter. Langkah selanjutnya adalah menentukan klasifikasi *naive bayes* sesuai yang telah ditetapkan oleh penulis berdasarkan topik berdasarkan kata kunci atau hashtag KTT G20 yang merupakan topik yang sedang trending di bulan november 2022 yaitu KTT G20 di Indonesia. Setelah menentukan kata kunci berupa hashtag kemudian penulis mendapatkan hasil sentimen analysis yang berupa pro kontra meliputi *positive*, netral maupun *negative* dan kemudian didokumentasi untuk keperluan penelitian.

3.1 Diagram Alir Penelitian

Tahapan diagram alir penelitian yang akan dilakukan adalah berdasarkan metode deksriptif untuk mendapatkan *trends* yang terjadi di ruang lingkup masyarakat dalam bentuk kata, melalui *Application Programming Interface* (API) adalah suatu *library* yang berbentuk *JavaScript* yang telah di sediakan oleh Twitter metode tersebut merupakan metode penelitian yang berlandaskan yang berlandaskan paspositivisme yaitu memandang realita, fenomena yang diklarifikasikan tetap dan terukur. Adapun berikut merupakan diagram alir penelitian seperti gambar 1.



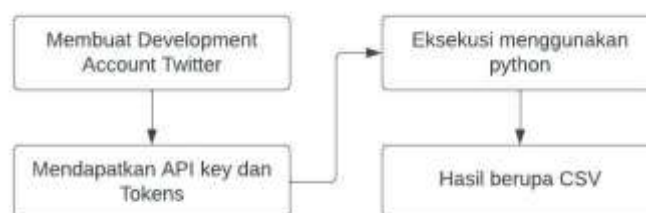
Gambar 1. Tahapan Penelitian

3.2 Identifikasi Masalah

Penelitian ini menggunakan data yang telah di crawling menggunakan tool jupyter di software anaconda menggunakan python, data yang diambil adalah sekitar 1000 data yang telah diambil di API twitter dengan kata kunci KTT G20. Data yang sudah diambil dikonfersikan menjadi sebuah file CSV dan dilanjutkan ke proses preprocessing test yang dimana proses pembersihan simbol pada teks penghilangan stopword dan tokenisasi setelah data selesai akan disimpan menjadi bentuk csv.

3.3 Pengumpulan Data

Crawling Data merupakan suatu proses dimana melakukan penambangan data atau penarikan data dari suatu *platform* yang ingin diambil datanya, dalam penelitian ini *crawling data* disebut *platform media social* Twitter. Adapun tahapan *Crawling data* sebagai berikut:



Gambar 2. Proses *Crawling Data*

Berikut adalah penjelasan dari gambar 2 yakni:

1. Membuat *Development Account* Twitter
2. Setelah membuat *Development Account Twitter* bisa mengakses atau mengambil *API key* dan *token* yang telah di sediakan oleh Tweiiter.
3. Ekseksusi menggunakan *Python* untuk *Crawiling Data*
4. Setelah data mentah berhasil diambil hasil akan berupa file CSV

3.4 Data Cleansing

Data Cleansing merupakan sebuah tahapan atau metode untuk membersihkan data dan memilah data apa saja yang dibutuhkan seperti data yang tidak relevan, tidak akurat, data duplikat dan juga data yang tidak terformat yang akan diproses atau diolah, Proses *Data Cleansing* ini dilakukan karena data yang telah diambil adalah data yang diperoleh dari Twitter yang terbilang data yang cukup kotor. Proses ini bertujuan untuk meningkatkan kualitas data yang telah dikumpulkan (Darwis et al., 2021).

3.5 Proses pemodelan sentiment analisis menggunakan naïve bayes

Data yang sudah bersih setelah proses *preprocessing text* akan lanjutkan ke tahap selajutnya yaitu melakukan sentiment analisis yang dimana mengetahui sebuah kata masuk kedalam kategori positif netral ataupun negatif. Proses terakhir dimana setelah melakukan sentiment analisis yaitu adalah kasifikasi menggunakan naïve bayes untuk mengetahui emsional dari kata yang diproses seperti *anger*, *disgust*, *joy*, *surprise* dan *negative*.

3.6 Naïve Bayes Classifier

Naïve bayes classifier didasarkan pada teorema Bayes untuk probabilitas bersyarat. Teorema ini mengkuantifikasi probabilitas bersyarat dari variabel acak (variabel kelas), mengingat pengamatan yang diketahui tentang nilai set variabel acak lainnya (variabel fitur). Adapun persamaan teorema *Naïve Bayes Classifier* yaitu:

$$P(A | B) = \frac{P(B | A).P(A)}{P(A)}$$

Keterangan: Dimana, $P(A|B)$ merupakan peluang tweet yang memiliki atribut X akan diklasifikasikan sebagai kelas A. $P(A)$ merupakan probabilitas kelas A dalam data latih, sedangkan $P(B)$ merupakan peluang atribut B pada seluruh data latih dan $P(B|A)$ merupakan Peluang atribut B terjadi dalam kelas A (Bhatia et al., 2018).

3.7 Preprocessing Text

Proses pengubahan bentuk menjadi data yang terstruktur sesuai kebutuhannya untuk proses dalam data mining, yang biasanya akan menjadi nilai-nilai numerik. Proses ini sering disebut *Text Preprocessing* (Siregar et al., 2019). Setelah data mentah yang telah di *Crawling* akan dieksekusi untuk dibersihkan dikarenakan data yang penelitian ini diambil dari data Twitter, data yang terbilang kotor. Beberapa proses yang dilakukan seperti berikut:



Gambar 3. *Preprocessing Text*

Berikut adalah penjelasan dari gambar 3 yakni:

1. *Input data mentah*
Data yang telah diambil pada proses *crawling data* dimasukan dan dieksekusi menggunakan Python
2. *Lower Case*
Perintah ini akan mengubah seluruh data yang berupa teks akan diubah menjadi huruf kecil semua (Feldman & Sanger, 2006).
3. *Tokenizing*
Proses penguraian deskripsi yang semula berupa kalimat-kalimat menjadi kata-kata dan menghilangkan delimiter-delimiter seperti tanda titik (.), koma (,), tanda kutip ("), tanda kurung (()), spasi dan karakter angka yang ada pada kata tersebut (Weiss, 2005).


```

In [1]: def remove_ap(text):
    # remove link, addid 2000
    text = text.replace('http://', '').replace('https://', '').replace('www.', '')
    # remove non ASCII
    text = text.encode('ascii', 'replace').decode('ascii')
    # remove hashtag, line, hashtag
    text = re.sub('#[a-zA-Z0-9]+', '', text)
    # remove line
    return text.replace('\n', '').replace('https://', '').replace('http://', '')

In [2]: # remove non ASCII
def remove_non_ascii(text):
    return re.sub(r'[^a-zA-Z0-9]', '', text)

# remove hashtag
ap['teks'] = ap['teks'].apply(remove_ap)

# remove non ASCII
def remove_non_ascii(text):
    return text.translate(str.maketrans('', '', string.punctuation))

ap['teks'] = ap['teks'].apply(remove_non_ascii)

# remove hashtag, addid 2000 & hashtag
def remove_hashtag_text(text):
    return re.sub(r'#[a-zA-Z0-9]+', '', text)

ap['teks'] = ap['teks'].apply(remove_hashtag_text)

# remove multiple hashtag, hashtag, hashtag, hashtag
def remove_hashtag_multiple(text):
    return re.sub(r'#[a-zA-Z0-9]+', '', text)

ap['teks'] = ap['teks'].apply(remove_hashtag_multiple)

# remove single char
def remove_single_char(text):
    return re.sub(r'[^a-zA-Z0-9]', '', text)

ap['teks'] = ap['teks'].apply(remove_single_char)

```

Gambar 5. Proses Data Cleansing

4.3 Preprocessing Text

Setelah proses *crawling data* langkah selanjutnya adalah *preprocessing text*. Adapun gambar proses *preprocessing text* sebagai berikut:

1. Lower Case

Proses *Lower Case* merupakan mengubah kalimat atau paragraf yang isinya mempunyai huruf kapital dari setiap awalan *tweet* maupun isi *tweet* seperti nama negara, nama kota, nama seseorang, lembaga, organisasi dll akan di konfersikan menjadi sebuah kalimat atau paragraf dengan menggunakan huruf kecil (*lower case*) ini adalah satu dari proses yang harus dijalankan dari *preprocessing text*. Berikut merupakan contoh dan hasil dari proses *lower case* pada gambar 6.

```

In [5]: ap['teks'] = ap['teks'].str.lower()

print('Case Folding Result :')
print(ap['teks'].head(10))

Case Folding Result :
0      keren keys ktt g20
1  semangat ggn memperluas g20 ke seluruh sektor...
2  semoga para pemimpin negara yang hadir merasa...
3      https://t.co/8rcvtd11j
4      https://t.co/8rcvtd11j
5      https://t.co/8rcvtd11j
6  kalo ga salah ini acara udh diundur, dari b...
7  itu strategi team pemungut mereka, daftar...
8  mengapresiasi atas dukungan publik pada peny...
9  bisa nonton rapat ktt g20 kak 8y'm
Name: teks, dtype: object

```

Gambar 6. Proses Lower Case

2. Tokenizing

Proses *Tokenizing* merupakan mengubah kalimat atau paragraf menjadi kata-kata yang terpisah, proses ini diterapkan dari proses sebelumnya yaitu *lower case* dari isi tweet dari pengguna Twitter. Tujuannya adalah mendapatkan kata yang nantinya akan diproses ke tahap selanjutnya. Berikut merupakan contoh eksekusi dan hasil dari isi tweet yang telah diproses *tokenizing* seperti gambar 7.

```
In [20]: nltk.download('punkt')
tokenized_sents = [word_tokenize(i) for i in datasets.sinkt20]
for i in tokenized_sents:
    print(i)

[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\artat\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!

['keren', 'kaya', 'ktt']
['semangat', 'pgn', 'memperluas', 'gas', 'ke', 'seluruh', 'sektor', 'untuk', 'mendukung', 'transisi', 'energi', 'gas', 'bumi', 'kita', 'dalam', 'ktt']
['semoga', 'para', 'pemimpin', 'negara', 'yang', 'badin', 'merasa', 'guas', 'selama', 'penyelenggaraan', 'ktt']
[]
['kalo', 'ga', 'salah', 'ini', 'acara', 'aduh', 'diundur', 'dari', 'bulan', 'kemarin', 'karna', 'ada', 'gala']
['itu', 'strategi', 'team', 'pemenangan', 'merekadaster', 'sendiri', 'diadakan', 'ev']
['mengapresiasi', 'atas', 'dukungan', 'publik', 'pada', 'penyelenggaraan', 'ktt', 'Indonesia', 'di', 'hal']
['bisa', 'nonton', 'rapat', 'ktt', 'kak']
['inilah', 'akibatnya', 'kalo', 'ktt', 'dipake', 'acara', 'pegang', 'mangrove', 'segala', 'bukannya', 'mikirin', 'ekonomi']
['in', 'nyinyir', 'terus', 'kerjanya', 'bak', 'jokowi', 'itu', 'aduh', 'salikota', 'gubernur', 'presiden', 'dan', 'ketua', 'ktt']
['pagi', 'pa', 'prabs', 'mau', 'bertanya', 'sepatu', 'hitam', 'yg', 'dipakai', 'waktu', 'ktt', 'kemarin', 'merknya', 'apa', 'ya']
['membuka', 'pelaksanaan', 'persidangan', 'kongres', 'penda', 'seasia', 'timur', 'di', 'trano', 'luxury', 'hotel', 'kota', 'b
```

[illegible]

Page 8

Gambar 9. Proses *Stemming*

Hasil dari data Twitter setelah melewati beberapa proses yakni *data cleansing* dan *preprocessing text* dari 500 data yang telah diambil dari Twitter dapat diketahui bahwa *sentiment* dari setiap *tweet* oleh pengguna Twitter terdapat 383 *tweet* dengan respon positif dan 116 *tweet* dengan respon negatif. Berikut merupakan gambaran *sentiment* dari 500 data *tweet* seperti pada gambar 9.



	precision	recall	f1-score	support
negatif	0.36	0.32	0.34	25
positif	0.78	0.81	0.80	75
accuracy			0.69	100
macro avg	0.57	0.57	0.57	100
weighted avg	0.68	0.69	0.68	100

Text(50, 72222222222214, 0.5, 'Label sebenarnya')

Confusion matrix

	Label prediksi negatif	Label prediksi positif
Label sebenarnya negatif	8	17
Label sebenarnya positif	14	61

Page 10



Apa yang telah tertera pada gambar 11 hasil dari perhitungan *confusion matrix* guna mengetahui akurasi dari algoritma *naïve bayes* dari 500 data Twitter yakni mendapatkan hasil akurasi sebesar 69%, pada gambar 11 diketahui untuk jumlah *True Negatif* (TN) adalah 8, *True Positif* (TP) adalah 61, *False Negatif* (FN) adalah 17, *False Positif* (FP) 14, *class precision* positif sebesar 78%, *class precision* negatif sebesar 36%, *recall* positif sebesar 81%, *recall* negatif sebesar 32%, *f1-score* positif 80% dan *f1-score* negatif sebesar 34%.

5. Kesimpulan

Hasil dari klasifikasi *sentiment analysis* oleh pengguna Twitter dengan kata kunci “KTT G-20” yang terjadi pada bulan November tahun 2022 menunjukkan bahwa hasil dari *tweets* yang telah diambil sebanyak 500 data mempunyai hasil sebanyak 383 *tweets* positif, sedangkan *tweets* dengan sentimen negatif mempunyai hasil sebanyak 116 *tweets*. Hal ini menunjukkan bahwa respon terhadap terselenggaranya Konferensi Tingkat Tinggi *Group of Twenty* (KTT G-20) yang telah diselenggarakan di Bali menunjukkan hasil yang positif mengenai KTT G-20 di Bali. Dalam pengolahan data menggunakan algoritma *naïve bayes* menghasilkan sebuah akurasi dengan nilai 69%. Nilai akurasi dengan *class precision* untuk prediksi positif adalah 78%, sedangkan nilai akurasi *class precision* untuk prediksi negatif adalah 36%. Nilai akurasi *class recall* positif adalah 81%, sedangkan nilai akurasi *class recall* negatif adalah 32%.

Referensi

- Adinegoro, R. W., Dyar, E., 2}, W., & Arifiyanti, A. A. (2020). Aplikasi Website Sentiment Analysis Ulasan Tokopedia. *Jurnal Informatika Dan Sistem Informasi (JIFoSI)*, 1(3), 963–971. <http://jifosi.upnjatim.ac.id/index.php/jifosi/article/view/193>
- Astuti, W. R. D. (2020). Kerja Sama G20 dalam Pemulihan Ekonomi Global dari COVID-19. *Andalas Journal of International Studies (AJIS)*, 9(2), 131. <https://doi.org/10.25077/ajis.9.2.131-148.2020>
- Bhatia, S., Sharma, M., & Bhatia, K. K. (2018). Sentiment Analysis and Mining of Opinions. *Studies in Big Data*, 30(May), 503–523. https://doi.org/10.1007/978-3-319-60435-0_20
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, 15(1), 131. <https://doi.org/10.33365/jtk.v15i1.744>
- Feldman, R., & Sanger, J. (2006). The Text Mining Handbook. In *The Text Mining Handbook*. <https://doi.org/10.1017/cbo9780511546914>
- Gafatia, I. W. D., & Hadinata, N. (2021). Analisis Pro Kontra Vaksin Covid 19 Menggunakan Sentiment Analysis Sumber Media Sosial Twitter. *Jurnal Pengembangan Sistem Informasi Dan Informatika*, 2(1), 34–42. <https://doi.org/10.47747/jpsii.v2i1.544>
- Hajnal, P. I. (2019). *THE G20 Evolution, Interrelationships, Documentation* (J. J. Krinton (ed.); Second Edi). Routledge.
- Pangestu, G. T., & Rosyda, M. (2022). Sentiment Analysis Tweet Pilkada 2020 Saat Pandemi COVID-19 di Media Sosial Twitter Menggunakan Metode 1D



- Convolutional Neural Network. *Jurnal Media Informatika Budidarma*, 6(2), 1017. <https://doi.org/10.30865/mib.v6i2.3765>
- Santi, P. N. P., Ardani, W., & Putri, I. A. S. (2022). Presidensi G20 sebagai Sarana Marketing dan Branding Pariwisata Indonesia serta Pengaruhnya terhadap Peningkatan Kunjungan Wisatawan pada Era Pandemi Covid-19 (Studi Kasus di Hotel Melia Bali). *Lensa Ilmiah: Jurnal Manajemen Dan Sumberdaya*, 1(1), 15–20. <https://doi.org/10.54371/jms.v1i1.167>
- Santoso¹, V. I., Virginia², G., & Lukito³, Y. (2017). Penerapan Sentimen Analisis Pada Hasil Evaluasi Dosen Dengan Metode SVM. *Jurnal Transformatika*, 14(2), 72.
- Siregar, R. R. A., Siregar, Z. U., & Arianto, R. (2019). Klasifikasi Sentiment Analysis Pada Komentar Peserta Diklat Menggunakan Metode K-Nearest Neighbor. *Kilat*, 8(1), 81–92. <https://doi.org/10.33322/kilat.v8i1.421>
- Weiss, S. M. (2005). *Text mining: predictive methods for analyzing unstructured information* (S. M. Weiss (ed.)) [Book]. Springer. <https://doi.org/10.1007/978-0-387-34555-0>.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>)