

IMPLEMENTASI METODE SUPPORT VECTOR MACHINE PADA OPINION MINING MASYARAKAT TERKAIT CHATGPT

Muhammad Muadin¹, Junadhi², Rahmiati³, Hadi Asnal⁴

¹⁻⁴ Prodi Teknik Informatika, STMIK Amik Riau

Email: m.muadinsar@gmail.com, junadhi@sar.ac.id, rahmiati@sar.ac.id, hadiasnal@sar.ad.id

Abstract

ChatGPT is able to provide benefits to lecturers and students because they can immediately get accurate information. Conversely, ChatGPT also allows students to cheat during exams and ask ChatGPT to make essays on coursework. Many opinions or opinions were conveyed via social media Twitter regarding the presence of ChatGPT. Some people consider ChatGPT a threat. Others actually see ChatGPT as an opportunity. Sentiment analysis is a fairly popular research field, because it can provide benefits for various aspects. This study uses the Support Vector Machine (SVM) algorithm and uses 189919 datasets. The data splitting process uses three models, namely 60:40 (60% is Training 40% is testing), 70:30 (70% is Training 30% is testing), and 80:20 (80% is Training 20% is testing). The results of the study obtained test data and training data with a 60:40 combination resulting in an accuracy rate of 68%. Then the combination of training data and test data 70:30 produces an accuracy rate of 69% and with a combination of 80:20 produces an accuracy value of 70%. Therefore, it can be concluded that community opinion mining on ChatGPT using the SVM method produces the best accuracy of 70% with a ratio of 80:20.

Keywords: ChatGPT, Support Vector Machine, Opinion Mining, Twitter

Abstrak

ChatGPT mampu memberikan manfaat kepada para dosen dan mahasiswa karena mereka langsung bisa mendapatkan informasi dengan akurat. Sebaliknya, ChatGPT juga memungkinkan para mahasiswa untuk menyontek saat ujian berlangsung dan meminta ChatGPT untuk membuat essai soal mata kuliah. Berbagai pendapat atau opini banyak disampaikan melalui media sosial twitter terkait kehadiran ChatGPT. Sebagian masyarakat menganggap ChatGPT sebagai ancaman. Sebagian lainnya justru memandang ChatGPT sebagai peluang. Analisis sentimen merupakan bidang penelitian yang cukup populer, karena dapat memberikan keuntungan untuk berbagai aspek. Penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) dan menggunakan 189919 dataset. Proses splitting data menggunakan tiga model yaitu 60:40 (60% adalah Training 40% adalah testing), 70:30 (70% adalah Training 30% adalah testing), dan 80:20 (80% adalah Training 20% adalah testing). Hasil penelitian memperoleh data uji dan data latih dengan kombinasi 60:40 menghasilkan tingkat akurasi sebesar 68%. Kemudian kombinasi data latih dan data uji 70:30 menghasilkan tingkat akurasi sebesar 69% dan dengan kombinasi 80:20 menghasilkan nilai akurasi sebesar 70%. Oleh karena itu, dapat disimpulkan bahwa opinion mining masyarakat terhadap ChatGPT menggunakan metode SVM menghasilkan akurasi yang terbaik adalah 70% dengan rasio perbandingan 80:20.

Keywords: ChatGPT, Support Vector Machine, Opinion Mining, Twitter

1. PENDAHULUAN

Transformasi teknologi saat ini melahirkan berbagai teknologi baru. Salah satunya adalah ChatGPT (*Generative Pre-training Transformer*) sebuah chatbot berbasis AI (Artificial Intelligence). ChatGPT merupakan perangkat lunak berupa model bahasa besar

yang dikembangkan oleh perusahaan teknologi OpenAI dan diluncurkan pada November 2022 lalu (Fais Sya' bani et al., 2022). Kehadiran ChatGPT pada akhir tahun lalu ini mampu mengguncang dunia teknologi karena fitur dan kemampuannya yang canggih.

Kehadiran ChatGPT mampu merubah sistem belajar mengajar dunia pendidikan. Satu sisi, ChatGPT mampu memberikan manfaat kepada para dosen dan mahasiswa karena mereka langsung bisa mendapatkan informasi dengan akurat (Zen Munawar et al., 2023). Sebaliknya, ChatGPT juga memungkinkan para mahasiswa untuk menyontek saat ujian berlangsung dan meminta ChatGPT untuk membuatkan esai soal mata kuliah. Berbagai pendapat atau opini banyak disampaikan melalui media sosial twitter terkait kehadiran ChatGPT. Sebagian masyarakat menganggap ChatGPT sebagai ancaman. Sebagian lainnya justru memandang ChatGPT sebagai peluang (Lund & Wang, 2023).

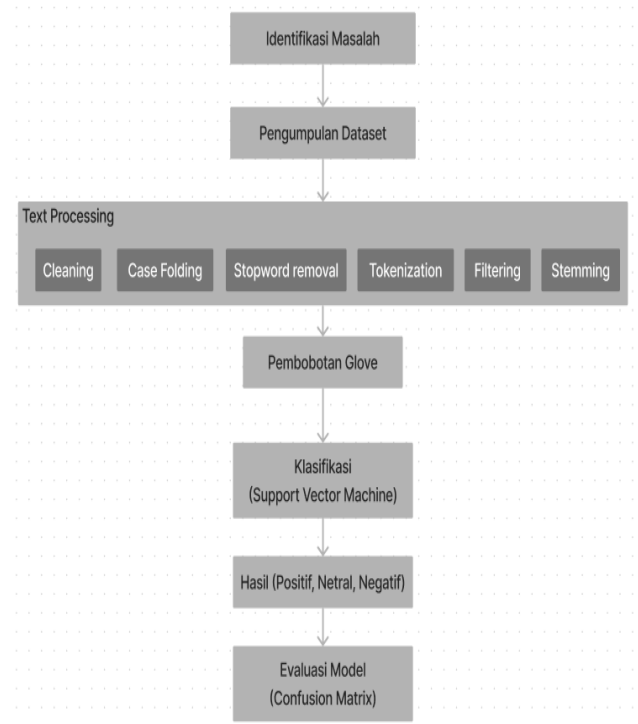
Mining opini adalah kombinasi mining teks dan pengolahan bahasa alam. Support Vector Machine (SVM) adalah salah satu teknik text mining yang dapat menyelesaikan masalah opinion mining. SVM dapat digunakan untuk mengkategorikan opini ke dalam opini positif dan negatif. Penelitian tentang kemampuan SVM untuk mengklasifikasikan teks telah dilakukan oleh (Alhaq et al., 2021). Dari proses pengujian secara kualitatif disebutkan bahwa teks bisa diklasifikasikan dengan akurasi yang tinggi.

Penelitian di bidang opinion mining mulai marak pada tahun 2002. (Alhaq et al., 2021) melakukan penelitian tentang opinion mining dengan menggunakan data berupa data review konsumen atas suatu marketplace. Metode yang digunakan adalah Support Vector Machine (SVM). Hasil terbaik yang dicapai adalah 93% akurasi terhadap data review marketplace bukalapak. Penelitian lain yang dilakukan oleh (Fitriyana et al., 2023) dengan metode SVM tentang ulasan aplikasi Jamsostek Mobile menghasilkan akurasi 96%.

Berdasarkan pernyataan di atas, peneliti tertarik untuk melakukan penelitian analisis sentimen menggunakan metode SVM dengan tujuan menganalisis opini masyarakat terhadap kehadiran ChatGPT, serta mengklasifikasikan ulasan atau opini ke dalam kategori positif, negatif dan netral sehingga dapat menjadi alat untuk bahan pertimbangan bagi masyarakat dalam mengambil keputusan, dan dapat mempermudah masyarakat dalam melihat kekurangan dari ChatGPT.

2. METODE PENELITIAN

Proses melakukan sebuah penelitian data dan informasi yang bersifat objektif yang akan digunakan sebagai titik acuan dalam penelitian, dengan adanya data-data tersebut diharapkan penelitian yang dihasilkan adalah penelitian yang berkualitas (Junadhi et al., 2022). Proses dalam melakukan penelitian ini digambarkan sebagai berikut:



Gambar 1. Alur Penelitian

Tahapan yang dilakukan dapat dijelaskan sebagai berikut:

2.1 IDENTIFIKASI MASALAH

Pembuatan sistem ini diawali dengan menentukan masalah, dalam penelitian ini adalah mengklasifikasikan opini masyarakat terhadap kehadiran chatgpt di dunia pendidikan sehingga dapat diketahui pola-pola tertentu yang dijadikan pengetahuan baru.

2.2 PENGUMPULAN DATASET

Dataset yang digunakan pada penelitian ini diambil mulai dari November 2022 hingga 26 Maret 2023 yang diperoleh dari media sosial

Twitter. Berikut adalah bentuk tampilan data tweet yang diperoleh.



Gambar 2. Data Tweet ChatGPT

Pada gambar 2 diatas terlihat bahwa dataset yang diperoleh masih belum dilakukan proses text processing. Tweet masih mengandung simbol-simbol kata imbuhan dan emoticon.

2.3 TEXT PROCESSING

Tahap *Text Processing* diartikan sebagai tahap awal untuk untuk membersihkan kata-kata yang tidak perlu atau kata-kata yang tidak memiliki makna. Proses keseluruhan dilakukan menggunakan program python3, sehingga dilakukan secara otomatis (Firdaus & Firdaus, 2021). Preprocessing yang akan dilakukan mencakup berbagai aktivitas, diantaranya adalah:

1. *Cleaning*
Menghilangkan karakter yang tidak diinginkan seperti tanda baca, url, emoji, angka, atau karakter khusus dari teks (Fridom Mailo et al., 2019).
2. *Casefolding*
Mengubah teks menjadi huruf kecil atau huruf besar yang bertujuan untuk menghilangkan perbedaan yang ditimbulkan oleh perbedaan penulisan huruf besar dan kecil dalam dokumen teks (Fridom Mailo et al., 2019).
3. *Tokenization*
Memecah teks menjadi kata-kata atau frase yang lebih kecil (Rifaldi & Fadlil, 2023).
4. *Stopword Removal*
Menghilangkan kata-kata yang tidak memiliki makna yang signifikan seperti “apakah”, “ke”, “luar”, ”biasanya”, dll (Olive et al., 2020).
5. *Filtering*
Menghilangkan kata-kata yang tidak sesuai dengan topik yang dianalisis, hal ini akan membuat data lebih bersih dan akurat sebelum digunakan dalam proses analisis (Almira et al., 2021).

6. *Stemming*
Menghilangkan imbuhan dari kata-kata untuk mengurangi variasi kata yang sama (Olive et al., 2020).

2.4 PEMBOBOTAN KATA (GLOVE)

Setelah data sudah dibersihkan dan selesai dari tahapan preprocessing maka selanjutnya semua datadikonversikan kedalam bentuk angka agar memiliki nilai bobotnya sehingga dapat di proses oleh algoritma klasifikasi (Made Dwi Dharma Sreya, 2022). Nilai bobot ini ditentukan berdasarkan konteks atau makna dari kata tersebut dalam teks. Pembobotan kata pada penelitian ini menggunakan metode GloVe.

2.5 KLASIFIKASI DATA

Sebelum melakukan klasifikasi data, perlu dilakukannya beberapa tahapan seperti *splitting data*, pembangunan algoritma SVM dan melatih algoritma tersebut untuk melakukan evaluasi.

Splitting data digunakan membagi data menjadi dua bagian, yaitu data latih (*training*) dan data uji (*testing*). Pada penelitian ini, penulis melakukan empat percobaan.

1. Percobaan pertama, 60% data digunakan sebagai data latih dan 40% sebagai data uji.
2. Percobaan kedua, 70% data sebagai data latih dan 30% sebagai data uji.
3. Percobaan ketiga, 80% data sebagai data latih dan 20% sebagai data uji.
4. Percobaan keempat, 90% data sebagai data latih dan 10% sebagai data uji.

Setelah SVM dilatih, tahap selanjutnya adalah melakukan evaluasi terhadap performa dari algoritma SVM, performa akan diuji berdasarkan metode *confusion matrix* dengan menggunakan data uji.

2.6 HASIL

Setelah data diproses dengan algoritma SVM maka akan di peroleh hasil berupa data yang menunjukkan kategori positif, negatif, dan netral. Pada saat klasifikasi, algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan. Selanjutnya data divisualisasikan dalam bentuk diagram

agar menghasilkan grafik dari masing-masing kategori.

2.7 EVALUASI MODEL CONFUSION MATRIX

Cara mengetahui performa dari metode SVM, maka dilakukan pengujian terhadap model. Hasil klasifikasi akan ditampilkan dalam bentuk *confusion matrix*. Nilai akurasi yang didapatkan dihitung berdasarkan jumlah nilai dari diagonal *confusion matrix* dibagi dengan jumlah seluruh data. Selanjutnya evaluasi dilakukan untuk mengetahui kinerja model (Dwi Cahyo et al., 2019). Evaluasi model dilakukan dengan cara melihat tingkat akurasi metode melalui *confusion matrix* dan tabel akurasi serta presisi untuk tiap model. Setelah data uji diujikan, maka akan menghasilkan daftar kelas-kelas dari data uji, sebut saja prediksi kelas. Kemudian prediksi kelas dibandingkan dengan kelas yang sebenarnya dari data uji yang disembunyikan sebelumnya. Sehingga dapat dilihat dan dihitung nilai *accuracy*, *precision*, *recall*, dan *f1-score* (Jatikusumo & Derajad Wijaya, 2021).

Setelah diketahui hasil, selanjutnya untuk performa metode klasifikasi dari setiap kelasnya dapat dilihat melalui nilai presisi, *recall*, dan *f1-score* pada setiap kelasnya. Hasil nilai presisi, *recall*, dan *f1-score* memiliki nilai sebesar 0-1. Semakin tinggi nilainya maka semakin baik. Gambaran hasil dari proses evaluasi model dapat dilihat pada Tabel 1.

Tabel 1. Nilai Presisi, *Recall*, *F1-Score*

Jenis Klasifikasi	Presisi	<i>recall</i>	<i>f1-score</i>
Positif	?	?	?
Netral	?	?	?
Negatif	?	?	?

Sehingga disimpulkan hasil dari evaluasi model dapat dilihat nilai presisi, *recall*, dan *f1-score* di setiap kelasnya.

3. HASIL DAN PEMBAHASAN

Langkah pertama yang dilakukan penulis mengimport library yang diperlukan, dapat dilihat pada gambar dibawah ini.

```
#masukan modul yang dibutuhkan
import pandas as pd
import numpy as np
import nltk
import string
import re
```

Gambar 3. Library Python

Berikutnya dilakukan proses load data yang menggunakan library pandas untuk membaca data dari file CSV dengan nama "ChatGPT_Tweet_Labeled.csv" dan memuat 189919 data tersebut ke dalam variabel "tweets". Setelah itu, program tersebut menampilkan 5 baris pertama dari data dengan menggunakan method head() pada variabel "tweets".

```
#masukan data kembali yang akan diproses
def load_data():
    data_tweets = pd.read_csv('ChatGPT_Tweet_Labeled.csv')
    return data_tweets

tweets = load_data()
tweets = tweets.drop('Unnamed: 0', axis=1)
tweets.head() #tampilkan bagian atas saja
```

	tweets	labels
0	ChatGPT: Optimizing Language Models for Dialog...	neutral
1	Try talking with ChatGPT, our new AI system wh...	good
2	ChatGPT: Optimizing Language Models for Dialog...	neutral
3	THRILLED to share that ChatGPT, our new model ...	good
4	As of 2 minutes ago, @OpenAI released their ne...	bad

Gambar 4. Data Tweet Hasil Load

Proses berikutnya dilakukan cleaning data untuk membersihkan tweet dari karakter-karakter yang tidak relevan atau tidak diinginkan.

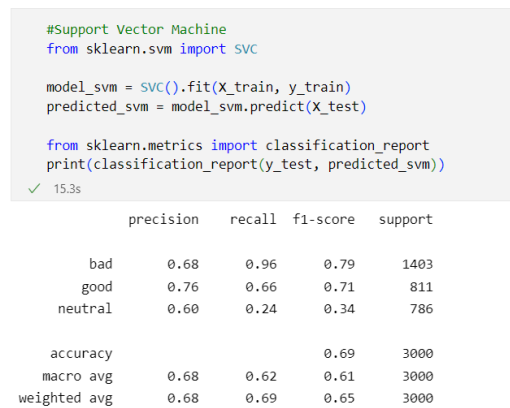
```
#cleaning
def cleaning(Text):
    Text = re.sub(r'\$\w*', '', Text)
    Text = re.sub(r'^rt[\s]+', '', Text)
    Text = re.sub('((www\.[^\s]+)|(https?://[^\s]+))', '', Text)
    Text = re.sub('&quot;', '', Text)
    Text = re.sub(r'^d+', '', str(Text))
    Text = re.sub(r'^b[a-zA-Z]\b', '', str(Text))
    Text = re.sub(r'^\w$', '', str(Text))
    Text = re.sub(r'(\.|\+)', r'\1', Text)
    Text = re.sub(r'\s+', '', str(Text))
    Text = re.sub(r'#', '', Text)
    Text = re.sub(r'^a-zA-Z0-9]', '', str(Text))
    Text = re.sub(r'\b[w[1,2]\b', '', Text)
    Text = re.sub(r'\s\s+', '', Text)
    Text = re.sub(r'^RT[\s]+', '', Text)
    Text = re.sub(r'^b[\s]+', '', Text)
    Text = re.sub(r'^link[\s]+', '', Text)
    return Text

tweets['cleaning'] = tweets['remove_user'].apply(cleaning)
```

Gambar 5. Kode Cleaning Data

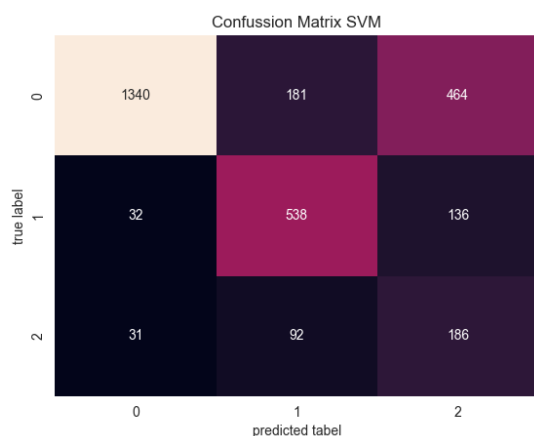
Proses klasifikasi dengan perbandingan 70:30 memperoleh hasil accuracy sebesar 69%. Hasilnya dapat dilihat pada gambar dibawah ini.

Klasifikasi 70 : 30



Gambar 10. Klasifikasi Data 70:30

Selanjutnya ditampilkan visualisasi dari confusion matrix klasifikasi data perbandingan 70:30.

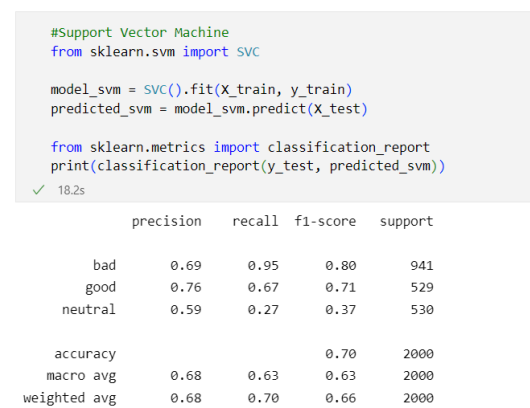


Gambar 11. Confussion Matrix 70:30

3. Klasifikasi 80:20

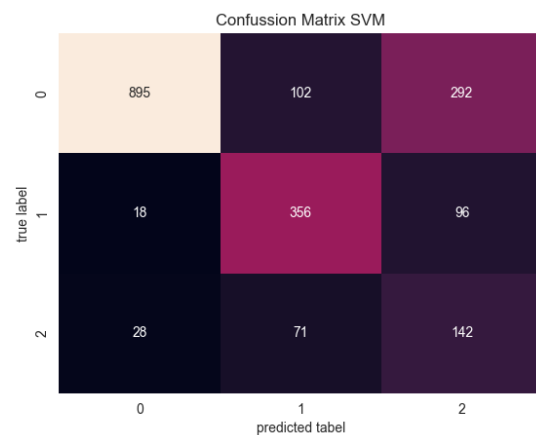
Proses klasifikasi dengan perbandingan 80:20 memperoleh hasil accuracy sebesar 70%. Hasilnya dapat dilihat pada gambar dibawah ini.

Klasifikasi 80 : 20



Gambar 12. Klasifikasi Data 80:20

Berikut ini visualisasi dari confusion matrix klasifikasi data perbandingan 80:20.



Gambar 13. Confussion Matrix 80:20

Output dari kode tersebut adalah laporan klasifikasi yang memuat nilai presisi (*precision*), *recall*, *f1-score*, dan nilai support untuk setiap kelas. Laporan ini memberikan gambaran tentang seberapa baik model dapat mengklasifikasikan setiap kelas. Semakin tinggi nilai *precision*, *recall*, dan *f1-score*, semakin baik model dalam mengklasifikasikan kelas tersebut. Sedangkan nilai support menunjukkan jumlah sampel pada kelas tersebut dalam data uji.

Berdasarkan hasil evaluasi yang didapatkan, nilai-nilai evaluasi tersebut di sajikan dalam bentuk tabel sebagai berikut.

Tabel 2. Perbandingan Nilai Evaluasi Support Vector Machine Terhadap Percobaan Splitting Data.

Splitting Data	60 : 40	70 : 30	80 : 20
<i>Accuracy</i>	0.68	0.69	0.70
<i>Precision</i>	0.68	0.68	0.68
<i>Recall</i>	0.68	0.69	0.70
<i>F1 Score</i>	0.64	0.65	0.66

4. SIMPULAN

Penelitian sentimen analisis terhadap penggunaan chatgpt di kalangan masyarakat cukup beragam. Berdasarkan hasil penelitian sentimen analisis dengan label positif berjumlah 3000, label netral berjumlah 2500 dan label negatif cukup tinggi yaitu diatas 4000. Selanjutnya penelitian ini diproses menggunakan algoritma SVM dengan hasil penelitian memperoleh data uji dan data latih dengan kombinasi 60:40 menghasilkan tingkat akurasi sebesar 68%. Kemudian kombinasi data latih dan data uji 70:30 menghasilkan tingkat akurasi sebesar 69% dan dengan kombinasi 80:20 menghasilkan nilai akurasi sebesar 70%. Oleh karena itu, dapat disimpulkan bahwa opinion mining masyarakat terhadap ChatGPT menggunakan metode SVM menghasilkan akurasi yang terbaik adalah 70% dengan rasio perbandingan 80:20.

5. DAFTAR PUSTAKA

- Alhaq, Z., Mustopa, A., & Santoso, J. D. (2021). *Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter*.
- Almira, A., Suendri, & Ikhwan, A. (2021). Implementasi Data Mining Menggunakan Algoritma Fp-Growth pada Analisis Pola Pencurian Daya Listrik. *Jurnal Informatika Universitas Pamulang*, 6(2), 442–448.
<http://www.openjournal.unpam.ac.id/index.php/informatika/article/view/12278>
- Dwi Cahyo, M. P., Widodo, & Prasetya Adhi, B. (2019). Kinerja Algoritma Support Vector Machine dalam Menentukan Kebenaran Informasi Banjir di Twitter. *PINTER : Jurnal Pendidikan Teknik Informatika Dan Komputer*, 3(2), 116–121.
<https://doi.org/10.21009/pinter.3.2.5>
- Fais Sya' bani, M. R., Enri, U., & Padilah, T. N. (2022). Analisis Sentimen Terhadap Bakal Calon Presiden 2024 Dengan Algoritme Naïve Bayes. *JURIKOM (Jurnal Riset Komputer)*, 9(2), 265.
<https://doi.org/10.30865/jurikom.v9i2.3989>
- Firdaus, A., & Firdaus, W. I. (2021). Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan). In *Jurnal JUPITER* (Vol. 13, Issue 1).
- Fitriyana, V., Hakim, L., Candra Rini Novitasari, D., Hanif Asyhar, A., Studi Matematika, P., Sains Dan Teknologi, F., Sunan Ampel Surabaya, U., & Timur, J. (2023). Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine. In *Jurnal Buana Informatika* (Vol. 14, Issue 1).
- Fridom Mailo, F., Lazuardi, L., Manajemen dan kebijakan Kesehatan Fakultas Kedokteran, D., Masyarakat dan Keperawatan Universitas Gadjah Mada, K., Sistem Informasi Manajemen Kesehatan Fakultas Kedokteran, D., Masyarakat dan Keperawatan, K., & Gadjah Mada, U. (2019). Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia. In *Jurnal Sistem Informasi Kesehatan Masyarakat Journal of Information Systems for Public Health* (Vol. 4, Issue 1).
- Jatikusumo, D., & Derajad Wijaya, H. (2021). IJCIT (Indonesian Journal on Computer and Information Technology) Pendeteksi Lokasi Kejadian Covid-19 Menggunakan Social Media dengan Kombinasi Algoritma Stemming Bahasa Indonesia. In *IJCIT (Indonesian*

- Journal on Computer and Information Technology*) (Vol. 6, Issue 1).
- Junadhi, Agustin, Rifqi, M., & Anam, M. K. (2022). Sentiment Analysis of Online Lectures using K-Nearest Neighbors based on Feature Selection. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 11(3), 216–225. <https://doi.org/10.23887/janapati.v11i3.51531>
- Lund, B. D., & Wang, T. (2023). *Chatting about ChatGPT: How may AI and GPT impact academia and libraries?* <https://ssrn.com/abstract=4333415>
- Made Dwi Dharma Sreya, E. B. S. (2022). *Penggunaan Metode GloVe untuk Ekspansi Fitur pada Analisis Sentimen Twitter dengan Naïve Bayes dan Support Vector Machine.*
- Olive, I., Putra, D., Rega Prilianti, K., Lucky, P., & Irawan, T. (2020). *Implementasi Text Mining Untuk Analisis Opini Masyarakat Terhadap Kinerja Layanan Transportasi Online Dengan Analisis Faktor.* 8(2).
- Rifaldi, D., & Fadlil, A. (2023). *DECODE: Jurnal Pendidikan Teknologi Informasi Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet “Mental Health.”* 3(2), 161–171. <https://doi.org/10.51454/decode.v3i2.131>
- Zen Munawar, Herru Soerjono, Novianti Indah Putri, Hernawati, & Andina Dwijayanti. (2023). Manfaat Kecerdasan Buatan ChatGPT Untuk Membantu Penulisan Ilmiah. *TEMATIK*, 10(1), 54–60. <https://doi.org/10.38204/tematik.v10i1.1291>