

Analisis Sentimen Kalimat Depresi Pada Pengguna Twitter Dengan Naive Bayes, Support Vector Machine, Random Forest

Mohammad Fachriza, Munawar

Fakultas Ilmu Komputer, Universitas Esa Unggul
¹m.fachriza.20@gmail.com, ²moenawar@gmail.com

History Artikel

Diterima : 11 Agustus 2023 Disetujui : 06 September 2023 Dipublikasikan : 20 Oktober 2023

Abstract

Pengguna aktif media sosial terus meningkat, dengan Twitter menjadi salah satu platform yang populer. Media sosial, termasuk Twitter, menjadi tempat bagi pengguna untuk mengekspresikan pendapat, termasuk mengenai masalah kesehatan mental seperti depresi. Penelitian ini bertujuan untuk menganalisis sentimen tweet pengguna Twitter terkait depresi menggunakan metode klasifikasi seperti Naïve Bayes Classifier, Support Vector Machine (SVM), dan Random Forest. Pengumpulan data tweet menggunakan metode crawling menggunakan API yang disediakan oleh Twitter dengan kata kunci yang berhubungan dengan depresi. Data tweet yang digunakan sebanyak 1502 tweet, yang selanjutnya dibersihkan pada tahap preprocessing dan diberi label dengan validasi oleh pakar terkait depresi, data yang sudah diberi label sebagai data untuk pengujian pada algoritma yang digunakan. Dari hasil pengujian performa pada algoritma yang diuji dapat disimpulkan bahwa algoritma Random Forest memiliki hasil performa yang lebih tinggi dibandingkan dengan Naive Bayes Clasifier dan Support Vector Machine dengan Hasil akurasi: 83.33%, presisi: 83.04%, recall: 83.33%, dan f1-scores: 82.62%. Penelitian ini Juga memberikan Word cloud untuk memberikan gambaran visual tentang kata-kata yang paling sering muncul dalam tweet. Kata-kata yang dominan dapat memberikan indikasi tentang topik atau isu yang paling mendominasi terkait kesehatan mental seseorang.

Keywords: *Analisis Sentimen, Twitter, Naive Bayes Classifier, Random Forest, Depresi. Wordcloud*

Abstrak

Active social media users continue to increase, with Twitter being one of the most popular platforms. Social media, including Twitter, is a place for users to express opinions, including regarding mental health issues such as depression. This study aims to analyze the sentiments of Twitter users' tweets related to depression using classification methods such as the Naïve Bayes Classifier, Support Vector Machine (SVM), and Random Forest. Tweet data collection uses the crawling method using the API provided by Twitter with keywords related to depression. The tweet data used was 1502 tweets, which were then cleaned at the preprocessing stage and labeled with validation by experts related to depression, data that had been labeled as data for testing the algorithm used. From the results of performance testing on the algorithms tested, it can be concluded that the Random Forest algorithm has higher performance results compared to the Naive Bayes Classifier and Support Vector Machine with results of accuracy: 83.33%, precision: 83.04%, recall: 83.33%, and f1-scores: 82.62%. This research also provides a word cloud to provide a visual overview of the words that appear most often in tweets. Dominant words can give an indication of the most dominant topic or issue regarding a person's mental health.

Kata Kunci: *Sentiment Analysis, Twitter, Naive Bayes Classifier, Random Forest, Depression*

How to Cite: Mohammad Fachriza (2023). Analisis Sentimen Kalimat Depresi Pada Pengguna Twitter Dengan Naive Bayes, Support Vector Machine, Random Forest. *KOMPUTEK : Jurnal Teknik Universitas Muhammadiyah Ponorogo*, Vol 7 (2): Halaman 49-58

© 2023 Universitas Muhammadiyah Ponorogo. All rights reserved

ISSN 2614-0985 (Print)

ISSN 2614-0977 (Online)

PENDAHULUAN

Perkembangan teknologi, terutama internet, telah mengubah cara komunikasi dengan pesat. Banyak platform, termasuk media sosial, telah dikembangkan untuk berkomunikasi secara online. Indonesia memiliki 191 juta pengguna aktif media sosial pada Januari 2022, meningkat 12,35% dari tahun sebelumnya. Twitter menjadi salah satu platform populer di Indonesia dengan 18,45 juta pengguna pada tahun 2022. Melalui media sosial, pengguna dapat mengekspresikan pendapat mereka dengan bebas.

Twitter adalah salah satu platform media sosial yang sangat populer dan digunakan secara luas untuk beropini. Didirikan oleh Jack Dorsey, Twitter memungkinkan pengguna untuk mengirim dan menyebarkan pesan dalam bentuk tweet dan retweet (Krisdiyanto, 2021).

Penelitian ini menggunakan data opini yang diambil dari media sosial Twitter. Twitter dipilih karena banyaknya pengguna yang aktif menyampaikan cerita, opini, dan kegiatan mereka di platform tersebut (Primadhani Tirtopangarsa & Maharani, 2021).

Depresi adalah kondisi medis yang menyebabkan perasaan sedih yang dapat memiliki dampak negatif pada pikiran, tindakan, perasaan, dan kesehatan mental seseorang. Ini bisa menjadi reaksi normal terhadap peristiwa kehidupan seperti kehilangan orang yang dicintai atau menghadapi kegagalan. Mendiagnosis orang lain dengan masalah kesehatan mental berdasarkan pandangan pribadi tanpa kemampuan untuk beradaptasi juga dapat menjadi faktor yang mempengaruhi depresi (Hadi et al., 2017).

Penelitian ini bertujuan untuk menganalisis sentimen tweet pengguna media sosial Twitter yang menggunakan kata kunci "Depresi" untuk mengetahui apakah tweet-tweet tersebut mengandung unsur depresi atau tidak. Metode klasifikasi yang digunakan dalam penelitian ini adalah Naïve Bayes Classifier, Support Vector Machine (SVM), dan Random Forest. Karena metode tersebut

memiliki kelebihan dan kekurangan masing-masing.

Beberapa peneliti sebelumnya yang telah melakukan penelitian tentang analisis sentimen. Diantaranya penelitian yang dilakukan oleh (Maulana et al., 2020) mengenai analisa sentimen cyberbullying di jejaring sosial twitter dengan algoritma naïve bayes. Dalam penelitian tersebut menghasilkan tingkat akurasi 76%, precision 76,09%, recall 97,22% dan specificity sebesar 21,4%.

Penelitian yang dilakukan oleh (Adhitya Karel Maulaya & Junadhi, 2022) mengenai analisis sentimen menggunakan support vector machine masyarakat indonesia di twitter terkait bjorka. Dalam penelitian tersebut menghasilkan tingkat akurasi sebesar 62,33%. Penelitian yang dilakukan oleh (Syarifuddin, 2020) mengenai analisis sentimen opini publik terhadap efek psbb pada twitter dengan algoritma decision tree-knn-naïve bayes. Dalam penelitian tersebut algoritma Decision Tree menghasilkan nilai accuracy 83,3%, precision 79% dan recall 87,17%.

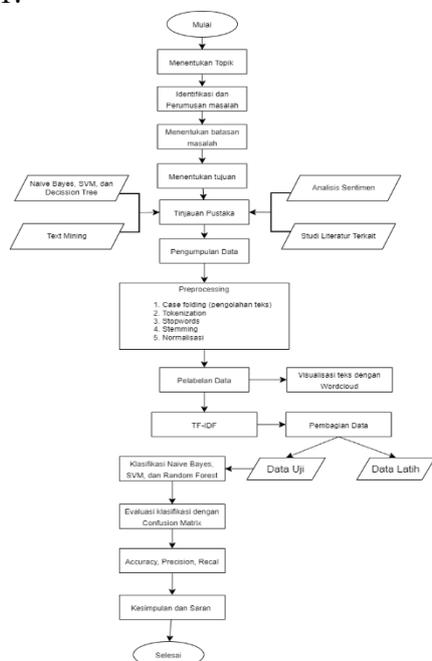
Penelitian yang dilakukan oleh (Nugraha, 2021) mengenai analisis sentimen berbasis emoticon pada komentar instagram bahasa indonesia menggunakan naïve bayes. Dalam penelitian ini menghasilkan nilai akurasi yang cukup baik pada metode Naïve Bayes Classifier dengan nilai 96.3%. penelitian yang dilakukan oleh (Hutagalung et al., 2021) mengenai analisis sentimen terhadap aplikasi ruangguru menggunakan algoritma naïve bayes, random forest dan support vector machine. Dalam penelitian ini algoritma random forest menghasilkan nilai akurasi sebesar 97,16% serta nilai AUC 0,996.

Berdasarkan penelitian sebelumnya, maka dalam penelitian ini menggunakan metode Naive Bayes Classifier, dan Random Forest untuk mengetahui nilai akurasi, presisi, recall, dan f1_scores dan algoritma yang terbaik untuk klasifikasi mengenai tweet depresi.

METODOLOGI PENELITIAN

Pendekatan kuantitatif dengan metode crawling digunakan dalam penelitian ini

dengan objek penelitian berupa tweet dengan kata kunci "depresi", "kesehatanmental", "gangguankejiwaan" pada periode bulan Januari – Maret Tahun 2022. Penelitian ini melewati beberapa tahapan seperti pada gambar 1.



Gambar 1. Tahap Penelitian

Kode python digunakan dalam penelitian ini dengan platform editor jupyter notebook.

Pengumpulan Data

Pengumpulan data dilakukan dengan crawling pada Twitter menggunakan API yang tersedia, dengan bantuan software Orange3 dan Jupyter Notebook. Peneliti menggunakan kata kunci "depresi", "kesehatan mental", dan "gangguan kejiwaan" pada periode Januari - Maret 2022, dan berhasil mengumpulkan 5000 data. Namun, terdapat data ganda atau sama, sehingga peneliti menghapus data yang berulang tersebut. Setelah proses tersebut, didapatkan 1502 data yang siap untuk tahap preprocessing.

Teks Preprocessing

Data yang digunakan dalam penelitian ini adalah tweet dari platform Twitter yang berbentuk teks yang tidak terstruktur dan

mengandung banyak noise. Sebelum melakukan klasifikasi, data perlu diubah menjadi lebih terstruktur melalui tahap preprocessing. Proses preprocessing ini akan menggunakan metode text mining dengan bantuan Jupyter Notebook dan bahasa pemrograman Python, serta library NLTK (Natural Language Toolkit) yang menyediakan banyak fitur untuk pemrosesan bahasa alami dalam Python. Tahapan-tahapan dalam preprocessing meliputi case folding, tokenizing, filtering, stemming, dan normalisasi.

1. Case folding (pengolahan teks) pada tahap ini, dilakukan pembersihan data dengan mengubah huruf kapital menjadi huruf kecil, serta menghilangkan noise seperti tanda baca, username/mention, URL, angka, emoticon, hashtag, karakter khusus, dan lainnya.
2. Tokenizing pada tahap ini proses memecah kalimat atau teks menjadi kata-kata individual yang disebut token. Dengan melakukan tokenizing lebih mudah menghitung frekuensi kemunculan setiap kata dalam teks tersebut.
3. Stopwords Removal pada tahap ini, dilakukan penyaringan dan penghapusan kata-kata yang tidak bermakna menggunakan kamus stopwords dari library Sastrawi. Selain itu, juga ada tambahan kata-kata yang tidak terdapat dalam library stopwords.
4. Stemming pada tahap ini mengubah kata-kata yang memiliki imbuhan menjadi kata dasar. Tujuannya adalah untuk menggambarkan kata tersebut dengan bentuk yang sama dengan kata lain yang tidak memiliki imbuhan. Proses ini menggunakan library Sastrawi.
5. Normalisasi kata tahap untuk mengubah bentuk kata ke dalam bentuk baku sesuai dengan kata yang terdapat dalam Kamus Besar Bahasa Indonesia (KBBI). Proses ini

menggunakan kamus data dari kamus data online. Tujuannya adalah untuk mempermudah proses selanjutnya, yaitu pelabelan data.

Berikut ini merupakan contoh hasil preprocessing, dapat dilihat pada tabel

Tabel 1. Contoh Hasil Preprocessing

Sebelum	Sesudah
['fase', 'frustasi', 'bentar', 'depresi']	['fase', 'frustasi', 'bentar', 'depresi']

Pelabelan Data

Setelah tahap preprocessing selesai dan data sudah dibersihkan, langkah selanjutnya adalah pelabelan data untuk mengetahui informasi sentimen pada data tweet tersebut. Pelabelan ini akan diproses dengan menggunakan library VaderSentiment dan juga melalui validasi oleh pakar.

Contoh hasil yang diperoleh dari pelabelan dapat dilihat pada tabel 2.

Tabel 2. Contoh Hasil Pelabelan

Tweet	Label Pakar
['fase', 'frustasi', 'bentar', 'depresi']	Negatif

Pembobotan Kata (TF-IDF)

Setelah tahap preprocessing dan pelabelan data selesai, langkah selanjutnya adalah melakukan ekstraksi fitur pada data. Proses ini bertujuan untuk mengubah data menjadi vektor matriks agar lebih mudah diolah oleh sistem. Ekstraksi fitur ini membantu mengurangi dimensi data, meningkatkan kualitas data, dan memudahkan pemahaman oleh algoritma pembelajaran mesin. Pada penelitian ini, proses ekstraksi fitur dilakukan dengan metode TF-IDF (Term Frequency Inverse Document Frequency). Metode ini melibatkan perhitungan pembobotan kata dari beberapa sampel data yang ada dalam penelitian.

D1 = “fase frustasi bentar depresi” D2 = “orang bilang anak sd depresi alami runding” D3 = “untung tidak ambil depresi”

Tabel 3. Contoh Perhitungan tf-idf

N O	Term	TF			d f	Idf $\log(\frac{N}{df})$	W = df x idf		
		D 1	D 2	D 3			D1	D2	D3
1	fase	1	0	0	1	0,477	0,477	0	0
2	frustasi	1	0	0	1	0,477	0,477	0	0
3	bentar	1	0	0	1	0,477	0,477	0	0
4	depresi	1	1	1	3	0	0	0	0
5	bilang	0	1	0	1	0,477	0	0,477	0
6	anak	0	1	0	1	0,477	0	0,477	0
7	sd	0	1	0	1	0,477	0	0,477	0
8	alami	0	1	0	1	0,477	0	0,477	0
9	runding	0	1	0	1	0,477	0	0,477	0
10	untung	0	0	1	1	0,477	0	0	0,477
11	tidak	0	0	1	1	0,477	0	0	0,477
12	ambil	0	0	1	1	0,477	0	0	0,477

Dari Tabel 3 dapat dilihat simulasi dari operasi perhitungan TF-IDF dari dokumen/teks masukan. Ekstraksi ini bertujuan untuk mengubah dokumen teks dalam bentuk numerik agar mudah dijadikan sebagai vektor inputan saat pelatihan data dan pengklasifikasian data menggunakan algoritma Naïve Bayes, Support Vector Machine dan Random Forest. Pada tahapan selanjutnya perhitungan TF-IDF menggunakan library sklearn menggunakan jupyter notebook.

Pembagian Data (Splitting Data)

Kemudian dataset hasil dari tf-idf dibagi menjadi dua bagian, yaitu data latih dan data uji. Dengan proporsi data latih sebanyak 80% dan data uji sebanyak 20% dari total dataset dengan nilai random state sebesar 42. Sehingga total dari masing-masing dataset menjadi 1,120 baris data latih, dan 282 baris data uji.

Klasifikasi (Pembentukan Model)

Dalam tahap pemodelan ini memproses klasifikasi data menggunakan tiga metode, yaitu Naive Bayes, Support Vector Machine, dan Random Forest yang akan dihitung menggunakan library sklearn dengan pemrograman python dan menggunakan jupyter notebook.

- **Naive Bayes**

Naive Bayes Classifier adalah metode klasifikasi yang digunakan dalam text mining untuk melakukan analisis sentimen. Metode ini memiliki potensi yang baik untuk mengklasifikasi dalam hal akurasi dan perhitungan data. Naive Bayes Classifier menghitung probabilitas jika kategori keputusan adalah benar karena merupakan vektor informasi objek, yang mengamsusikan bahwa atribut objek ini adalah independent (Khadafi et al., 2022). Pada persamaan (1) merupakan rumus untuk menghitung nilai probabilitas:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

$P(H|X)$ adalah Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob). $P(X|H)$ adalah probabilitas X berdasarkan kondisi tersebut. $P(H)$ adalah probabilitas hipotesis H (prior prob). $P(X)$ adalah probabilitas dari X

- **Support Vector Machine**

Support Vector Machine adalah metode yang membandingkan pemilihan parameter standar nilai diskrit dengan nilai tetap disebut himpunan kandidat untuk mengklasifikasikan akurasi. *Support Vector Machine* (SVM) diperkenalkan oleh Vapnik, Boser dan Guyon pada tahun 1992 (Riadi et al., 2019). Support Vector Machine melakukan training agar menghasilkan bobot, dan

bias. Di mana dalam melakukan training akan melakukan update nilai bobot dan bias seperti pada rumus (2). Hasil dari training adalah suatu persamaan garis seperti pada persamaan (3) (Nafi'iyah, 2020)

$$w^T x_i + b \geq 1 \quad (2)$$

$$w^T x_i + b \leq -1$$

Dimana persamaan x_i adalah variabel input, dari persamaan diatas hasil training berupa persamaan (3)

$$f_{w,b}(x) = \text{sgn}(w^T x + b) \quad (3)$$

- **Random Forest**

Metode Random Forest adalah metode pelatihan yang berbasis ensemble learning, yang menggunakan algoritma Decision Tree. Prosesnya melibatkan pembuatan beberapa model Decision Tree dengan menggunakan data uji yang sama untuk setiap model. Selanjutnya, hasil prediksi dari setiap model Decision Tree digabungkan melalui proses majority voting dengan menggunakan metode modus. Hasil akhir dari proses ini menjadi kelas prediksi (Sejati et al., 2022). Pada persamaan (4) merupakan rumus untuk menghitung Entropy, dan persamaan (5) merupakan rumus untuk menghitung Gain pada metode Random Forest.

$$Entropy(E) = -\sum_{i=1}^n p_i \log_2 p_i \quad (4)$$

dimana n adalah jumlah kelas yang mungkin, dan p_i adalah probabilitas dari kelas i .

$$Gain = Entropy_{parent} - Entropy_{children} \quad (5)$$

$Entropy_{parent}$ adalah entropi dari node induk dan $Entropy_{children}$ adalah entropi rata-rata dari node anak.

Evaluasi Klasifikasi (Confusion Matrix)

Evaluasi merupakan langkah terakhir dalam setiap data mining dan klasifikasi teks. Pada tahap ini akan mengevaluasi hasil percobaan, membandingkan dan menganalisis terhadap kinerja klasifikasi, umumnya untuk mengevaluasi hasil implementasi pada sentimen analisis menggunakan confusion matrix (Rizki et al., 2021). Confusion matrix adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah (Normawati & Prayogi, 2021).

Tabel 4 Confusion Matrix

Predict Class	True Class	
	Positif	Negatif
Positif	True Positif (TP)	False Positif (FP)
Negatif	False Negatif (FN)	True Negatif (TN)

Dari tabel 4 dapat digunakan untuk menghitung nilai akurasi, presisi, recall dan f1-scores dari setiap algoritma.

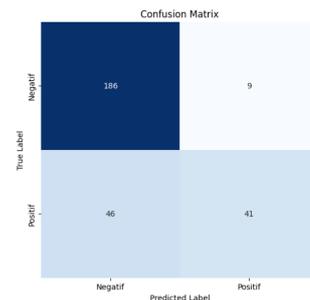
HASIL DAN PEMBAHASAN

Pada bagian ini merupakan pembahasan mengenai metode Naive Bayes. Support Vector Machine, dan Random Forest yang digunakan untuk analisis sentimen terhadap tweet depresi yang diperoleh dari Twitter.

Dengan bantuan program python. Pengujian metode menggunakan library scikit-learn (sklearn) yang sering digunakan dalam melakukan analisis data dan pembuatan model machine learning pada JupyterLab dengan pemrograman Python.

Naive Bayes

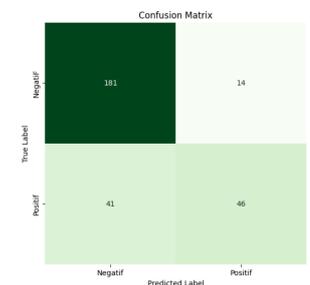
Pemodelan menggunakan algoritma Naive Bayes pada kode “naive_bayes = MultinomialNB()” merupakan objek dari kelas MultinomialNB, yang merupakan implementasi dari algoritma Naive Bayes dengan metode Multinomial. Objek ini akan digunakan untuk membangun dan melatih algoritma Naive Bayes. Dari hasil tersebut menghasilkan confusion matrix dapat dilihat pada gambar 2



Gambar 2. Hasil Confusion Matrix Naive Bayes

Support Vector Machine

Pemodelan menggunakan algoritma Support Vector Machine dengan kode “svm_classifier” adalah objek untuk SVM dengan menentukan kernel yang digunakan. Dalam penelitian ini, peneliti menggunakan “kernel='linear'”, yang artinya peneliti menggunakan kernel linear untuk melakukan pemetaan fitur. Dari hasil tersebut menghasilkan confusion matrix dapat dilihat pada gambar 3

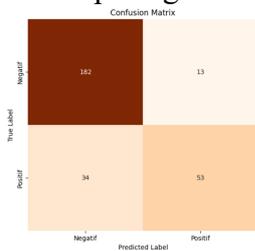


Gambar 3. Hasil Confusion Matrix SVM

Random Forest

pengujian algoritma Random Forrest dengan menginisialisasi algoritma menggunakan “RandomForestClassifier(random_state=42)”.

Dari hasil tersebut menghasilkan confusion matrix dapat dilihat pada gambar 4.



Gambar 4. Hasil Confusion Matrix Random Forest

Perbandingan Algoritma

Setelah mendapatkan hasil confusion matrix setiap algoritma, selanjutnya penelitian ini melakukan perbandingan terhadap performa tiga algoritma klasifikasi: Naïve Bayes Classifier, Support Vector Machine, dan Random Forest dalam mengklasifikasikan tweet yang mengandung unsur depresi dan tidak mengandung unsur depresi di media sosial Twitter. Hasil pengujian performa dari ketiga algoritma tersebut dapat dilihat pada tabel berikut:

Tabel 5 Perbandingan Performa Algoritma

	Naive Bayes Classifier	Support Vector Machine	Random Forrest
Akurasi	80.5%	80.4%	83.33%
Presisi	80.74%	80.03%	83.04%
Recall	80.5%	80.49%	83.33%
F1-score	78.71%	79.33%	82.62%

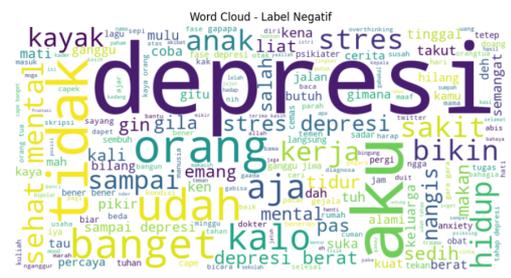
Berdasarkan hasil pengujian algoritma pengklasifikasi pada tweet yang mengandung depresi, Random Forest Classifier menunjukkan performa terbaik dengan akurasi 83.33%, presisi 83.04%, recall 83.33%, dan F1-score 82.62%. Naive Bayes memiliki akurasi 80.5%, presisi 80.74%, recall 80.5%, dan F1-score 78.71%. Sedangkan SVM memiliki akurasi 80.4%, presisi 80.03%, recall 80.49%, dan F1-score 79.33%.

Visualisasi

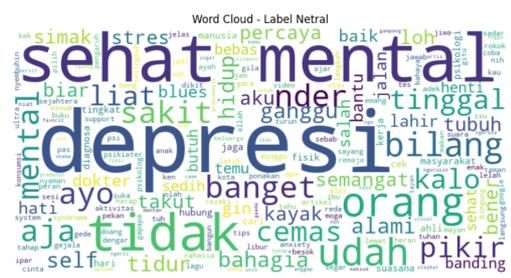
Setelah mendapatkan label atau sentimen pada data dan menguji algoritma klasifikasi, langkah selanjutnya dalam penelitian ini adalah memvisualisasikan kata-kata yang sering muncul dengan membuat wordcloud. Wordcloud adalah representasi visual yang menarik dari frekuensi kemunculan kata-kata dalam teks, di mana kata-kata yang paling sering muncul ditampilkan dengan ukuran yang lebih besar. Pada penelitian ini visualisasi wordcloud dilakukan menggunakan library matplotlib dan wordcloud pada bahasa pemrograman Python.



Gambar 5. Wordcloud Label Positif



Gambar 6. Wordcloud Label Negatif



Gambar 7. Wordcloud Label Netral

Dari wordcloud, dapat disimpulkan bahwa kata-kata "sehat mental" dan "depresi" adalah yang paling dominan dan sering muncul dalam tweet yang dianalisis. Hal ini menunjukkan bahwa banyak pengguna media sosial Twitter berbicara atau mengekspresikan perhatian terhadap kesehatan mental, terutama mengenai isu depresi. Selain itu, kata "depresi" juga

muncul pada label positif dan netral, yang mengindikasikan bahwa ada sentimen positif yang terkait dengan kata tersebut dalam tweet-tweet tersebut, seperti pengalaman positif setelah mengatasi depresi.

KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan, pengumpulan data tweet mengenai depresi dari periode Januari - Maret 2022 dengan kata kunci terkait menggunakan metode crawling dengan API Twitter, menghasilkan 5000 data. Setelah menghapus data yang ganda atau sama, tersisa 1502 data yang siap untuk digunakan. Melalui validasi pelabelan oleh pakar, 937 tweet diklasifikasikan sebagai sentimen negatif (mengandung depresi), 465 tweet diklasifikasikan sebagai sentimen positif (tidak mengandung depresi), dan 93 tweet diklasifikasikan sebagai sentimen netral.

Hasil pengujian algoritma pengklasifikasi Random Forest Classifier menunjukkan performa terbaik dengan akurasi 83.33%, presisi 83.04%, recall 83.33%, dan F1-score 82.62%. Naive Bayes memiliki akurasi 80.5%, presisi 80.74%, recall 80.5%, dan F1-score 78.71%. Sedangkan SVM memiliki akurasi 80.4%, presisi 80.03%, recall 80.49%, dan F1-score 79.33%. Dari hasil ini, dapat disimpulkan bahwa Random Forest memiliki performa yang paling baik dalam mengklasifikasikan tweet yang mengandung depresi

Berdasarkan penelitian dan analisis data, ditemukan bahwa jumlah data tweet yang digunakan tidak terlalu banyak, dan terdapat ketidakseimbangan antara kelas positif dan negatif. Untuk penelitian selanjutnya, disarankan untuk meningkatkan jumlah data tweet dan menggunakan pendekatan ensemble learning seperti SMOTE (Synthetic Minority Over-sampling Technique). Metode ini menggabungkan oversampling dan undersampling untuk menangani ketidakseimbangan kelas pada dataset, khususnya pada kelas minoritas yang jumlahnya lebih sedikit. Dengan pendekatan ini diharapkan dapat meningkatkan performa model klasifikasi secara keseluruhan.

DAFTAR PUSTAKA

- Adhitya Karel Maulaya, & Junadhi. (2022). Analisis sentimen menggunakan support vector machine masyarakat indonesia di twitter terkait bjorka. *Jurnal CoSciTech (Computer Science and Information Technology)*, 3(3), 495–500. <https://doi.org/10.37859/coscitech.v3i3.4358>
- Hadi, I., Fitriwijayati, Usman, R. D., & Rosyanti, L. (2017). Gangguan Depresi Mayor. *Hijp : Health Information Jurnal Penelitian*, 9(1), 16. <https://myjurnal.poltekkes-kdi.ac.id/index.php/HIJP>
- Hutagalung, A. S., Negara, A. B. P., & Pratama, E. E. (2021). Aplikasi Pendeteksi Cyberbullying Terhadap Komentar Postingan Media Sosial Instagram dengan Metode Naïve Bayes Classifier Berbasis Website. *Jurnal Sistem Dan Teknologi Informasi (Justin)*, 9(3), 364. <https://doi.org/10.26418/justin.v9i3.44843>
- Khadafi, M. Al, Kartika, K. P., Febrinita, F., Informasi, F. T., Balitar, U. I., Blitar, K., Based, L., Sentimen, A., Widodo, P. J., Program, P., & Kesehatan, J. (2022). Penerapan Metode Naïve Bayes Classifier Dan Lexicon Based Untuk Analisis Sentimen Cyberbullying Pada Bpjs. *Jurnal Mahasiswa Teknik Informatika*, 6(2), 725–733.
- Krisdiyanto, T. (2021). Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 7(1), 32. <https://doi.org/10.24014/coreit.v7i1.12945>
- Maulana, F. A., Ernawati, I., Labu, P., & Selatan, J. (2020). Analisa sentimen cyberbullying di jejaring sosial twitter

- dengan algoritma naïve bayes. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, 529–538. <https://conference.upnvj.ac.id/index.php/senamika/article/view/619>
- Nafi'iyah, N. (2020). Algoritma SVM untuk Memprediksi Pengunjung Wisata Musium di Jakarta. *KERNEL: Jurnal Riset Inovasi Bidang Informatika Dan Pendidikan Informatika*, 1(1), 33–41. <https://doi.org/10.31284/j.kernel.2020.v1i1.1156>
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 5(2), 697–711.
- Nugraha, K. A. (2021). Analisis Sentimen Berbasis Emoticon pada Komentar Instagram Bahasa Indonesia Menggunakan Naïve Bayes. *Jurnal Teknik Informatika Dan Sistem Informasi*, 7(3), 715–721. <https://doi.org/10.28932/jutisi.v7i3.4094>
- Primadhani Tirtopangarsa, A., & Maharani, W. (2021). Sentiment Analysis of Depression Detection on Twitter Social Media Users Using the K-Nearest Neighbor Method Analisis Sentimen Deteksi Depresi pada Pengguna Media Sosial Twitter dengan Menggunakan Metode K-Nearest Neighbor. *Semnasif*, 1(1), 247–258.
- Riadi, I., Umar, R., & Aini, F. D. (2019). Analisis Perbandingan Detection Traffic Anomaly Dengan Metode Naive Bayes Dan Support Vector Machine (Svm). *ILKOM Jurnal Ilmiah*, 11(1), 17–24. <https://doi.org/10.33096/ilkom.v11i1.361.17-24>
- Rizki, M. F., Auliasari, K., & Primaswara Prasetya, R. (2021). Analisis Sentiment Cyberbullying Pada Sosial Media Twitter Menggunakan Metode Support Vector Machine. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 5(2), 548–556. <https://doi.org/10.36040/jati.v5i2.3808>
- Sejati, P., Pilliang, M., Akbar, H., Unggul, U. E., Barat, J., Korespondensi, P., & Neighbor, K. (2022). STUDI KOMPARASI NAIVE BAYES, K-NEAREST NEIGHBOR, DAN RANDOM FOREST UNTUK PREDIKSI CALON MAHASISWA YANG DITERIMA ATAU MUNDUR. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 9(7), 1341–1348. <https://doi.org/10.25126/jtiik.202296737>
- Syarifuddin, M. (2020). ANALISIS SENTIMEN OPINI PUBLIK TERHADAP EFEK PSBB PADA TWITTER DENGAN ALGORITMA DECISION TREE-KNN-NAÏVE BAYES. *INTI Nusa Mandiri*, 15(1), 87–94. <https://doi.org/10.33480/inti.v15i1.1433>