

Perbandingan Algoritma NBC dan *Decision Tree* pada Sentimen Analisis Mengenai Vaksinasi Covid-19 di Indonesia

¹Bellin Halima Agtira, ²Hanny Hikmayanti Handayani, ³Anis Fitri Nur Masruriyah
^{1,2,3} Universitas Buana Perjuangan Karawang
Karawang, Indonesia

¹if19.bellinagtira@mhs.ubpkarawang.ac.id, ²hanny.hikmayanti@ubpkarawang.ac.id,
³anis.masruriyah@ubpkarawang.ac.id

*Penulis Korespondensi

Diajukan : 21/01/2023

Diterima : 29/01/2023

Dipublikasi : 29/01/2023

ABSTRAK

Perkembangan Twitter di Indonesia sebagai platform media sosial yang digemari masyarakat, yang menunjukkan Indonesia negara pengguna Twitter terbanyak kelima di dunia pada Januari 2022. Seiring berkembangnya tren saat ini, memaksa pemerintah Indonesia untuk melakukan berbagai cara lebih banyak berinteraksi dengan masyarakat, salah satunya dengan penggunaan Twitter. Memanfaatkan media *social* Twitter, pemerintah akan mendapatkan berbagai informasi yang sedang kontroversial di kalangan masyarakat. Seperti dilaksanakannya kegiatan vaksinasi oleh pemerintah Indonesia, banyak masyarakat menilai kontroversial dalam kegiatan vaksinasi ini dan banyak kalangan di masyarakat yang memberikan berbagai pendapatnya di media sosial. Melalui Twitter salah satunya, masyarakat berbagi postingan opini masing-masing tentang vaksinasi ini. Tahapan analisis pada pengujian ini akan membandingkan dua algoritma klasifikasi yakni *Naive Bayes Classifier (NBC)* dengan *Decision Tree* dengan menggunakan perhitungan *Confusion Matrix* untuk mengukur kinerja pada kedua algoritma tersebut. Hasil tes pengujian akhir disajikan dalam bentuk nilai akurasi, presisi dan recall. Perbandingan kedua algoritma klasifikasi *Naive Bayes* dengan *Decision Tree* hasil yang diperoleh memiliki perbedaan tingkat akurasi yang berbeda. Algoritma *Naive Bayes Classifier* mendapat hasil akurasi 93.96 % dan nilai presisi 91% dengan nilai recall 98%, sedangkan pada algoritma *Decision Tree* menghasilkan nilai akurasi sebesar 88.64 %, dengan nilai presisi 91% dan nilai recall 94%. Hasil ini didasarkan pada perhitungan *Confusion Matrix*. Kemudian bisa disimpulkan jika algoritma *NBC* penelitian ini lebih akurat dibanding algoritma *Decision Tree*.

Kata Kunci: Analisis Sentimen, *Decision Tree*, *Naive Bayes Classifier*, Twitter, Vaksinasi

I. PENDAHULUAN

Perkembangan Twitter di Indonesia sebagai platform media sosial yang digemari masyarakat, yang menunjukkan Indonesia negara pengguna Twitter terbanyak kelima di dunia pada Januari 2022. Twitter adalah media sosial sejenis microblogging yang dimana

para pengguna nya difasilitasi untuk menulis dan membagikan kegiatan ataupun opini. Dengan twitter, pengguna dapat membagikan keseharian dengan memposting foto atau video dan dapat menyampaikan berbagai pendapat mengenai sesuatu apa saja yang *Trending Topic* di Twitter yang dimana menjadi sebuah perbincangan hangat (Monavia Ayu, 2022).

Seiring berkembangnya tren saat ini, memaksa pemerintah Indonesia untuk melakukan berbagai cara lebih banyak berinteraksi dengan masyarakat, salah satunya dengan penggunaan Twitter. Memanfaatkan media social Twitter, pemerintah akan mendapatkan berbagai informasi yang sedang kontroversial di kalangan masyarakat. Salah satu halnya dengan adanya vaksinasi disaat terjadinya COVID-19 di Indonesia. Sejak maret 2020 penyebaran COVID-19, menyebabkan timbulnya dampak yang signifikan kepada aktivitas masyarakat. Maka pemerintah Indonesia melakukan upaya untuk mengurangi menybarnya COVID-19 dengan melakukan vaksinasi kepada masyarakat pada tahun 2021. Dilaksanakannya kegiatan vaksinasi ini oleh pemerintah Indonesia, banyak masyarakat menilai kontroversial dalam kegiatan vaksinasi ini dan banyak kalangan di masyarakat yang memberikan berbagai pendapatnya melalui media social, termasuk Twitter, masyarakat saling berbagi pendapat tentang vaksinasi ini (Muhadjir, 2021).

Analisis sentiment adalah metode yang digunakan untuk menganalisis bagaimana opini, pandangan, perilaku, dan perasaan seseorang ke dalam bahasa tulisan. Analisis sentiment digunakan dalam bentuk teks atau dokumen untuk menunjukkan suatu objek atau topik. Sehingga dengan penggunaan sentiment ini akan mengetahui opini seseorang secara tekstual dan setelah dilakukan evaluasi terhadap opini tersebut yaitu positif, negatif, dan netral (Fransiska Vina Sari, 2019)

Penelitian ini bertujuan untuk menganalisa sentimen pada *tweet* dengan mengelompokkan dua sentiment yaitu positif dan negatif dengan melakukan perbandingan kedua metode klasifikasi yaitu Algoritma *Naïve Bayes Classifier* dan Decision Tree, dengan tujuan menghasilkan nilai akhir berupa akurasi, presisi dan recall untuk kedua algoritma tersebut. Untuk mengetahui algoritma mana yang lebih akurat melakukan klasifikasi pada sentiment vaksinasi covid-19. Metode *Naïve Bayes Classifier (NBC)* ini merupakan metode klasifikasi biner yang menerapkan probabilitas statistik sederhana dengan aturan Bayesian menggunakan perkiraan independen yang kuat. Metode Bayes sendiri adalah metode analisis berdasarkan data sampel dan informasi sebelumnya. Kombinasi dari dua informasi ini disebut probabilitas posterior. Penerapan pada metode *NBC* ini dengan menggunakan data training untuk dilakukan pengujian pada data testing (Rahman, 2021). Sedangkan Decision tree disisi lain adalah bagan alir dengan bentuk pohon (*tree*) dimana setiap node (internal node) mewakili cabang dan atributnya menunjukkan hasil pengujian atau nilai atribut, sedangkan pada daunnya digunakan untuk mewakili kelas atau distribusi kelas. Decision tree juga biasanya digunakan untuk melakukan pengenalan pada pola statistic (Sarimuddin, 2020)

Pada penelitian yang sudah diteliti oleh (Ristasari Dwi Septiana, 2021) melakukan perbandingan klasifikasi pada vaksinasi covid-19 menggunakan metode *NBC* dengan *Chi-Square* dan *Particle Swarm Optimization*, kemudian peneliti tertarik untuk melakukan penelitian dengan topik tersebut dalam tugas akhir dengan topik "Perbandingan Algoritma *NBC* dan Decision Tree Pada Model Analisis Sentimen Mengenai Vaksinasi Covid-19 Di Indonesia".

II. STUDI LITERATUR

Penelitian terdahulu

Pada penelitian sebelumnya tentang sentimen analisis dilakukan oleh (Ainun Nisa, 2019) menggunakan Algoritma NBC dan Chi-Square pada penyedia di layanan Telekomunikasi. Hasil yang telah didapatkan menunjukkan nilai rata-rata performansi yang tertinggi dengan hasil naïve bayes dan *chi-square* mendapatkan nilai signifikan 0,01 dengan nilai presisi 83%, pada nilai akurasi sebesar 85,5%, nilai recall sebesar 86% dan f1- score sebesar 84%.

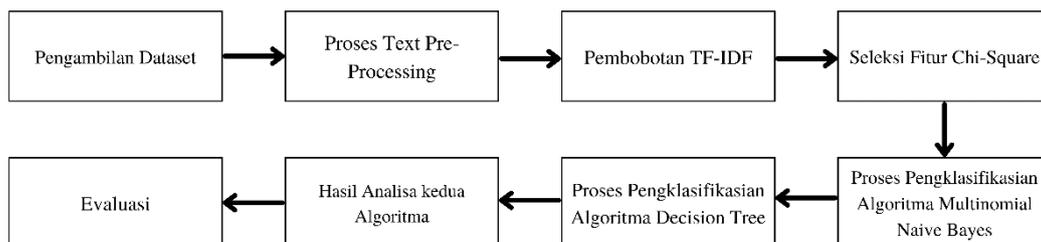
Penelitian selanjutnya yaitu melakukan sentimen analisis terhadap vaksinasi covid-19 dengan membandingkan Algoritma NBC dan Chi-Square, lalu membandingkan NBC dan Particle Swarm Optimization yang telah diteliti oleh (Ristasari Dwi Septiana, 2021). Hasil dari penelitian tersebut menunjukkan bahwa klasifikasi NBC dengan *Chi-Square* diperoleh akurasi 69,13%, sedangkan NBC dengan *Particle Swarm Optimization* diperoleh akurasi sebesar 66,02%.

Lalu pada penelitian terakhir ini melakukan analisis sentiment untuk mereview pada konsumen terhadap lokasi untuk makanan tradisional yang direkomendasi dengan metode Naïve Bayes dan *Chi-Square* yang diteliti oleh (Novan Dimas Pratama, 2018) pada seleksi fitur 25% memperoleh tingkat akurasi 81% dan pada seleksi fitur 50% memperoleh tingkat akurasi sebesar 80%.

III. METODE

Prosedur Penelitian

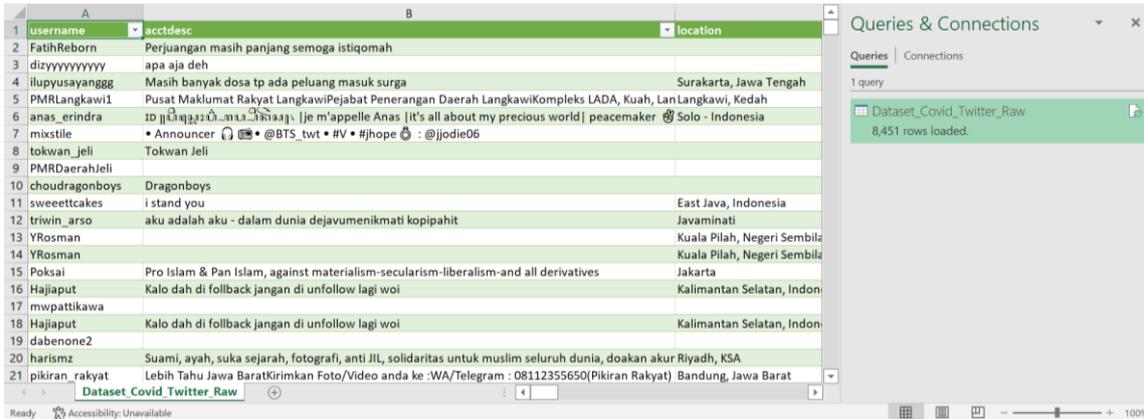
Prosedur penelitian yang ditunjukkan pada Gambar 1 dimulai dengan mengambil data dari (Kurniadi Ahmad Wijaya, 2022). Kemudian dilakukan proses text *preprocessing*, lalu berikutnya dilakukan proses pembobotan TF-IDF dan seleksi fitur *Chi-Square*. Setelah proses perhitungan bobot TF-IDF dan *Chi-square*. Proses berikutnya melakukan pengklasifikasian algoritma dengan *Naïve Bayes Classifier* dan *Decision Tree* dan akan menghasilkan Analisa pada kedua algoritma tersebut. Skema prosedur pada penelitian ditunjukkan dengan gambar dibawah ini :



Gambar 1 Flowchart prosedur penelitian

Karakter Data

Dataset yang akan digunakan untuk penelitian ini yaitu *Dataset_Covid_Twitter_Raw* meliputi 8451 data dan 10 *column*. Data yang diambil masih berupa data mentah yang berisi banyak data yang tidak penting dan atribut yang tidak mendukung.

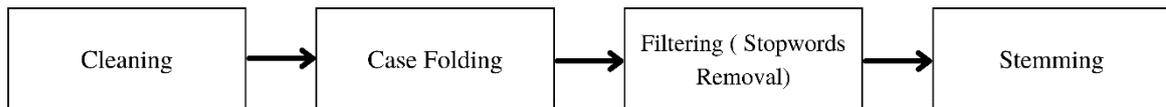


Gambar 2 Dataset

Text Pre-Processing

Text preprocessing adalah langkah pertama dalam mengubah data yang awalnya tidak berstruktur menjadi data yang terstruktur. ini dilakukan dengan tujuan untuk menemukan, memproses dan mencari, mengolah dan menyusun informasi dan menganalisa data teks yang terstruktur ataupun tidak yang akan memudahkan untuk proses pengolahan data pada tahapan proses data selanjutnya. (Arinda Ayu Puspitasari, 2018).

Penelitian ini memiliki beberapa metode yang digunakan, yaitu:



Gambar 3 Text Pre-Processing

1. *Cleaning*

Cleaning merupakan proses dimana semua karakter yang bersifat non alphabet dihilangkan bertujuan untuk mengurangi *noise* seperti URL, email, simbol, angka dan karakter khusus akan dihilangkan.

Tabel 1 tabel perbandingan hasil cleaning

Sebelum dilakukan <i>cleaning</i>	Setelah dilakukan <i>cleaning</i>
Pemerintah menjamin bahwa vaksin covid- 19 aman dan halal https://t.co/qqmpEbK5Xi #VaksinUntukNegeri #VaksinCOVID19 #vaksincorona #vaksingratisutkrakyat	Pemerintah menjamin bahwa vaksin covid aman dan halal VaksinUntukNegeri VaksinCOVID19 vaksincorona vaksingratisutkrakyat

2. *Case Folding*

Case folding adalah proses *preprocessing* yang merubah seluruh abjad pada teks *tweet*, dan *case folding* tidak akan memproses karakter apapun selain huruf a-z yang diterima.

Tabel 2 tabel perbandingan hasil case folding

Sebelum dilakukan <i>case folding</i>	Setelah dilakukan <i>case folding</i>
---------------------------------------	---------------------------------------

Pemerintah menjamin bahwa vaksin covid-19 aman dan halal https://t.co/qqmpEbk5Xi #VaksinUntukNegeri #VaksinCOVID19 #vaksincorona #vaksingratisutkrakyat	pemerintah menjamin bahwa vaksin covid aman dan halal vaksinuntuknegeri vaksincovid vaksincorona vaksingratisutkrakyat
---	--

3. *Filtering (Stopwords Removal)*

Filtering atau *Stopwords Removal* merupakan *text preprocessing* dimana memproses penghapusan kata yang tidak memiliki makna berupa yang, di, kata ke, angka.

Tabel 3 tabel perbandingan hasil filtering

Sebelum dilakukan <i>filtering</i>	Setelah dilakukan <i>filtering</i>
Pemerintah menjamin bahwa vaksin covid-19 aman dan halal https://t.co/qqmpEbk5Xi #VaksinUntukNegeri #VaksinCOVID19 #vaksincorona #vaksingratisutkrakyat	pemerintah menjamin bahwa vaksin covid aman halal vaksinuntuknegeri vaksincovid vaksincorona vaksingratisutkrakyat

4. *Stemming*

Stemming ialah hasil dari proses filtering untuk mencari kata dasar. Pada tahapan *stemming* ini dilakukan pemulihan bentuk kata menjadi representasi sama, pada proses *stemming* terdapat dua aturan yaitu menggunakan pendekatan aturan dan pendekatan kamus.

Tabel 4 tabel perbandingan hasil Stemming

Sebelum dilakukan <i>stemming</i>	Setelah dilakukan <i>stemming</i>
Pemerintah menjamin bahwa vaksin covid-19 aman dan halal https://t.co/qqmpEbk5Xi #VaksinUntukNegeri #VaksinCOVID19 #vaksincorona #vaksingratisutkrakyat	perintah menjamin bahwa vaksin covid aman halal vaksinuntuknegeri vaksincovid vaksincorona vaksingratisutkrakyat

Pembobotan TF-IDF

Pembobotan (TF) *Term Frequency* – (IDF) *Inverse Documen Frequency* (TF-IDF) ini ialah proses perhitungan bobot setiap kata yang seringkali dipergunakan untuk layanan informasi. Metode ini lebih efisien dan menghasilkan hasil yang lebih akurat. TF-IDF ini melakukan perhitungan (TF) dan (IDF), dimana dalam proses digunakan untuk mencari seberapa erat keterkaitan suatu term dengan suatu dokumen dengan cara membagi bobot setiap kata dan kedua konsep tersebut digabungkan, yaitu seberapa sering kemunculan kata pada dokumen atau sebaliknya, kemudian mengetahui seberapa jauh hubungan term pada dokumen tersebut dengan membagikan bobot di setiap kata, dan menggabungkan dua konsep yang mengandung kata tersebut. (Herwijayanti, 2018).

Perhitungan yang digunakan pada perhitungan TF-IDF:

$$idf = \log \log \left(\frac{N}{df} \right)$$

$$W = tf * idf$$

Keterangan :

Teks/dokumen yaitu D, lalu t merupakan huruf pada teks, W adalah bobot dokumen pada d untuk kata ket, kemudian tf adalah banyak kata pada dokumen, Ft,d ialah banyaknya dari kata d, berikutnya IDF adalah kebalikan frekuensi dokumen, lalu dft adalah jumlah kata yang mengandung kata I dan N adalah jumlah total dokumen.

Chi-Square

Chi-Square ialah proses dimana memilih sebuah fitur relevan yang sesuai kriteria untuk menghasilkan hasil fitur optimal. Chi square adalah seleksi fitur yang menguji ketergantungan istilah dengan kategori, menggunakan beberapa kondisi uji chi-square. Jika tabel kondisi berbentuk 2x2, maka tidak mungkin ada sel dengan jumlah kurang dari 5 pada frekuensi harapan (fh). Dan jika tabel kondisi berbentuk seperti 2 x 3, berarti jumlah sel tersebut tidak boleh melebihi nilai 20 %.

Naive Bayes Classifier

Algoritma *Naive Bayes Classifier* yaitu melakukan pengelompokkan yang memprediksi tingkat dalam suatu situasi. Metode klasifikasi NBC ini menggunakan metod static dengan probabilitas, ini di kemukakan oleh Thomas Bayes yaitu seorang ilmuwan Inggris, yang memprediksikan sebuah peluang kejadian pada masa mendatang yang berdasar pada kejadian pada masa yang lalu yang dikenal dengan teorema bayes. Atribut pengklasifikasian dari *Naive Bayes Classifier* adalah prediksi naif tentang independensi di setiap peristiwa. Algoritma *Naive bayes* ini memprediksi apakah atribut objek bersifat independen. Kinerja sistem klasifikasi dengan *Naive Bayes Classifier* ini tergantung kepada data yang ada dan akan digunakan sebagai data latih. Jika data tersebut bisa digunakan sebagai data latih, maka dari itu, sistem klasifikasi yang digunakan akan menghasilkan kinerja yang baik, dan sistem akan dapat digunakan untuk melakukan klasifikasi dengan jumlah data yang lebih besar. Model *machine learning* yang akan dilakukan pada dataset dengan menggunakan teorema Bayes dengan rumus sebagai berikut:

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

Keterangan :

H sebagai Hipotesis Data, X sebagai Data class yang dikenali, lalu P(H) sebagai perhitungan pada Hipotesis H, lalu P(H|X) sebagai Probabilitas pada Hipotesis H pada kondisi X, P(X) adalah nilai Probabilitas pada X, dan P(X|H) adalah Probabilitas pada X untuk Hipotesis H.

Decision Tree

Decision tree ialah flowchart dengan berbentuk pohon pada setiap nodenya (node internal) memiliki atribut dan cabang-cabangnya dengan menunjukkan hasil tes atau nilai atribut. Sedangkan pada daun digunakan untuk mewakili kelas atau distribusi kelas. Decision tree juga biasanya digunakan untuk melakukan pengenalan pada pola statistic

Perbandingan Algoritma Naïve Bayes dan Decision Tree

Setelah mengetahui *Word cloud* nya, lalu ke proses pembobotan kata oleh TF-IDF, dan setelah menemukan hasil, lalu diklasifikasikan dengan membandingkan kedua algoritma yaitu menggunakan algoritma *Naïve Bayes Classifier* dan *Decision tree* untuk mengevaluasi pengujiannya menggunakan *confusion matrix*.

Kinerja model yang diuji selanjutnya adalah *precision* dan *recall*. *Precision* merupakan perhitungan kinerja yang memberikan sebuah informasi dari prediksi data positif yang sebenarnya positif. Kemudian *Recall* merupakan perhitungan kinerja yang memberikan sebuah informasi mengenai prediksi data positif yang diprediksi sebagai data negatif.

Tabel 6 Hasil perbandingan Algoritma NBC dan Decision Tree

	Jumlah Prediksi Benar	Jumlah Prediksi Salah	Precision (%)	Recall (%)	Accuracy (%)
NBC	1589	102	91%	98%	93.96 %
Decision Tree	726	93	91%	94%	88.64%

Berdasarkan hasil pengujian yang dilakukan dengan membandingkan kedua algoritma klasifikasi *NBC* dan *Decision Tree*. Hasil penelitian ini didapatkan tingkat akurasi yang berbeda. Pada algoritma *NBC* memperoleh tingkat akurasi sebesar 93.96 %, dengan hasil presisi 91% dan hasil *recall* 98%, sedangkan pada algoritma *Decision Tree* memperoleh tingkat akurasi sebesar 88.64 %, dengan hasil presisi 91% dan nilai *recall* 94%.

IV. KESIMPULAN

Berdasarkan hasil uji pada perbandingan kedua algoritma *Naïve Bayes* dan *Decision tree*, maka dapat disimpulkan bahwa:

Algoritma *Naïve Bayes Classifier* pada penelitian ini lebih akurat karena mendapatkan hasil lebih besar. Dengan ditunjukkannya tingkat akurasi yang didapat sebesar 0.93 atau 93.96%. Untuk mendapatkan hasil perbandingan kedua algoritma pada penelitian ini, dilakukan uji dengan menggunakan algoritma *Decision tree* dan mendapatkan tingkat akurasi sebesar 0.88 atau 88.64%.

VI. REFERENSI

- Gading, T. S (2021). Analisis Sentimen Pada Tweet Dengan Tagar #Bpjsrasarentenir Menggunakan Metode Support. Pekanbaru: Universitas Islam Riau.
- Farahdiva, A. A. (2021). Analisis Sentimen Pada Tweet Dengan Tagar #Yangcuranggaktenang Mennggunakan Metode Decision Tree C4.5. Pekanbaru : Universitas Islam Riau.
- Blidex, J. S. (2021). Analisis Sentimen Klasifikasi Tweet Vaksin Covid 19 Dengan Naïve Bayes. *Jurnal Mahajana Informasi*, 8.
- Frizka, F., Ema, U., & Hanif, A. F. (2021). Analisis Sentiment Opini Terhadap Vaksin Covid-19 Pada Media Social Twitter Menggunakan Support Vector Machine Dan Naïve Bayes. *Jurnal Komtika (Komputasi Dan Informatika)*, 1-7.
- Harijiatno, S. D. (2019). Analisis Sentimen Pada Twitter Menggunakan Multinomial Naive Bayes. Yogyakarta: Universitas Sanata Dharma.

- Imron, A. (2019). Analisis Sentimen Terhadap Tempat Wisata Di Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier. *Yogyakarta: Universitas Islam Indonesia*.
- Merinda, L., Abdurrahim, A., & Lailis, S. (2019). Analisis Sentimen Tweet Vaksin Covid-19 Menggunakan Recurrent Neural Network Dan Naive Bayes. *Jurnal Resti(Rekayasa Sistem Dan Teknologi Informasi)*, 802-808.
- Miranti, A. Z. L., Nurul, A. S. W., Muhammad, S. R., & Galuh, W. S. (2022). Penerapan Metode K-Means Clustering Dalam Menganalisis Sentimen Masyarakat Terhadap K-Popers Pada Twitter. *Jurnal Ilmiah Komputer*, 10.
- Puji, N., Endang, S., & Tri, L. (2020). Analisis Sentimen Terhadap Penggemar K-Popdi Media Sosial Twitter menggunakan Naive Bayes (Studi Kasus Penggemar Grup Bts). *Journal Information Engineering And Educational Technology*, 4.
- Ristasari, D. S., Agung, B. S., & Tukiyyat, T. (2021). Analisis Sentimen Vaksinasi Covid-19 Pada Twitter Menggunakan Naive Bayes Classifier Dengan Feature Selection Chi-Squared Statistic Dan Particle Swarm Optimization. *Pamulang: Jurnal Sistem Komputer Dan Kecerdasan Buatan*.
- Zulfikar, F., & Nila, F. P. (2021). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid19 Berdasarkan Opini Pada Twitter Menggunakan Algoritma Naive Bayes. *Jurnal Teknik Informatika*, 8.
- Alifia, S. R., & Abdul, R. K. (2022). Analisis Sentimen Pengguna Twitter Terhadap Vaksin Booster Menggunakan Metode Naive Bayes Dan K-Nearest Neighbor. *SAINTEK (Jurnal Sains dan Teknologi)*. 1-5
- Bagus, M. A., Ahmad, T. A., & Rochmat, H. (2021). Analisis Sentimen dan Emosi Vaksin Sinovac pada Twitter menggunakan Naive Bayes dan Valence Shifter. *Jurnal Teknologi Terpadu*, 10.
- Rani, M., Dian, E. R., & Bayu, R. (2022). Analisis Sentimen pengguna Twitter terhadap Vaksinasi Sinovac dan AstraZeneca menggunakan Algoritma CART. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 8.
- Slamet, H. R., & Muhammad, I. W. (2022). Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naive Bayes dan K-NN. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, 1-9.
- Ainun, N., Eko, D., & Ibnu, A (2019). Analisis Sentimen Menggunakan Naive Bayes Classifier Dengan Chi-square Feature Selection Terhadap Penyedia Layanan Telekomunikasi. *eProceedings of Engineering*, 10.
- Fransiska, V. S., & Arief, W. (2019). Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naive Bayes Classifier Menggunakan Metode Naive Bayes Classifier. *Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 6.
- Novan, D. P., Yuita, A. S., & Putra, P. A. (2018). Analisis Sentimen pada Review Konsumen Menggunakan Metode Naive Bayes dengan Seleksi Fitur Chi Square untuk Rekomendasi Lokasi Makanan Tradisional. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 7.
- Sarimuddin., Jayanti, Y. S., Muh, M., Muh, A. M., Reski, S. A., & (2020). Klasifikasi Data Aging Tunggakan Nasabah Menggunakan Metode Decision Tree Pada ULaMMUnit Kolaka. *Informatics Journal*, 7.
- Arinda, A. P., Edy, S., & Indriati. (2018). Klasifikasi Dokumen Tumbuhan Obat Menggunakan Metode Improved K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 7.