

ANALISIS SENTIMEN TERHADAP RESESI EKONOMI GLOBAL DI INDONESIA MENGUNAKAN HYBRID LINEAR REGRESSION – NAIVE BAYES

SENTIMENT ANALYSIS OF GLOBAL ECONOMIC RECESSION IN INDONESIA USING HYBRID LINEAR REGRESSION – NAIVE BAYES

Divya Nabila Herisnan¹, Masyitah Elwinda², Rashid³, Rahmaddeni⁴

^{1,2,3,4}STMIK Amik Riau

divanabillapku123@gmail.com

ABSTRACT

This research aims to analyze Indonesian people's sentiment towards the global economic recession using a hybrid approach that combines Linear Regression and Naive Bayes algorithms. The data used was taken from the Kaggle dataset which contains 2,138 tweets related to the economic recession. The research stages include problem identification, dataset collection, preprocessing, word weighting using TF-IDF, data classification, sentiment distribution, and model evaluation. The results of the analysis show that public sentiment towards the global economic recession in Indonesia is dominated by neutral reactions. The hybrid Linear Regression and Naïve Bayes method with TF-IDF word weighting at 60:40 data splitting gave the highest accuracy results of 97.66%. Overall, the highest accuracy using Linear Regression is 98.01%. This research provides deeper insight into public perceptions of economic recessions, which can help policymakers and industry design more effective strategies to deal with the impact of the global economic crisis.

Keywords: *Global Economic Recession, Sentiment Analysis, Hybrid, Linear Regression, Naive Bayes.*

ABSTRAK

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia terhadap resesi ekonomi global menggunakan pendekatan hybrid yang menggabungkan algoritma Linear Regression dan Naive Bayes. Data yang digunakan diambil dari dataset Kaggle yang berisi 2.138 tweet terkait resesi ekonomi. Tahapan penelitian meliputi identifikasi masalah, pengumpulan dataset, preprocessing, pembobotan kata menggunakan TF-IDF, klasifikasi data, distribusi sentimen, dan evaluasi model. Hasil analisis menunjukkan bahwa sentimen publik terhadap resesi ekonomi global di Indonesia didominasi oleh reaksi netral. Metode hybrid Linear Regression dan Naïve Bayes dengan pembobotan kata TF-IDF pada splitting data 60:40 memberikan hasil akurasi tertinggi sebesar 97,66%. Secara keseluruhan, akurasi tertinggi dengan menggunakan Linear Regression yaitu sebesar 98,01%. Penelitian ini memberikan wawasan yang lebih mendalam tentang persepsi masyarakat terhadap resesi ekonomi, yang dapat membantu pembuat kebijakan dan industri dalam merancang strategi yang lebih efektif untuk menghadapi dampak dari krisis ekonomi global.

Kata Kunci: Resesi Ekonomi Global, Analisis Sentimen, Hybrid, Linear Regression, Naive Bayes.

PENDAHULUAN

Analisis sentimen adalah proses pembelajaran komputasi untuk memahami opini, perasaan, dan emosi yang diekspresikan dalam teks. Teknik ini sudah banyak digunakan sebagai sarana untuk melakukan analisis media sosial terhadap aspek pemasaran, sosial, dan politik (Meisya et al., n.d.). Dalam menganalisis sentimen, fase pertama adalah memahami, mengambil manfaat, dan mengolah teks yang berisi data. Tujuan dari tindakan ini adalah tentang membentuk sekumpulan informasi yang bermakna dan bermakna (Sari et al., 2020). Pengetahuan bahasa itu

sendiri diperhatikan dalam sistem bahasa alami, yang mencakup pengetahuan tentang kata-kata yang digunakan, gabungan kata yang membentuk kalimat, makna kata, peran kata dalam kalimat, dan sebagainya. Namun, fungsionalitas bahasa juga sangat penting, yaitu kemampuan manusia untuk memahami, yang dibangun atas pengetahuan yang diperoleh sepanjang hidup (Noor et al., n.d.).

Resesi berarti penurunan aktivitas ekonomi yang signifikan yang menyebar di seluruh ekonomi selama periode terpanjang lebih dari beberapa bulan. Tidak hanya diperlukan analisis untuk

menentukan komponen apa yang dapat menyebabkan resesi, tetapi juga untuk menanggapi efek negatif yang menjadi kekhawatiran masyarakat umum. Akhir-akhir ini, banyak pendapat publik tentang resesi yang akan datang pada tahun 2023, terutama di Indonesia (Halim Lubis et al., 2023). Pendapat masyarakat dapat ditemukan pada banyak platform sosial, diantaranya Kaggle yang mengandung teks Twitter.

Analisis sentimen, juga dikenal sebagai analisis pendapat, adalah salah satu ilmu di mana opini, perasaan, penilaian, sikap, dan emosi individu terhadap objek. Entitas ini bisa berbentuk barang, jasa, perusahaan, orang, acara, masalah, atau subjek yang dibahas pada teks yang di tulis. Analisis sentimen sudah membawa masalah - masalah pengkajian yang sulit yang belum pernah dipelajari ke dunia penelitian NLP. Salah satu masalah dalam analisis semantik adalah analisis sentimen. Namun, sistem analisis sentimen Cukup memahami beberapa aspek, seperti pendapat sisi positif dan negatif serta tujuan dari kalimat atau dokumen, sehingga sangat terbatas dan terfokus (Hasibuan & Heriyanto, n.d.).

Berdasarkan penjelasan sebelumnya, tujuan dari riset ini adalah untuk menganalisis sentimen masyarakat terkait dalam hal resesi ekonomi global. Penelitian ini akan menggunakan metode klasifikasi Hybrid Linear Regression dan Algoritma Naive Bayes. Algoritma Naive Bayes Classifier dimanfaatkan untuk mengelompokkan komentar dari komentar, dan Penelitian ini dimaksudkan untuk membedakan apakah opini publik positif, negatif, atau netral (Sutresno, 2023).

Fokus penelitian analisis sentimen ini adalah untuk mempelajari dan mengidentifikasi pendapat masyarakat mengenai isu berita resesi di media sosial, mengidentifikasi apakah persepsi tersebut dikelompokkan dalam kategori sentimen positif, negatif maupun netral, dan juga untuk mengetahui hasil akurasi persentase untuk masing-masing kategori. Pengujian

dilakukan menggunakan hybrid algoritma linear regression naïve bayes. Analisis regresi linier adalah teknik statistika yang menganalisis relasi antara variabel Y dan sekumpulan variabel X_1, X_p . Pendekatan ini bertujuan untuk memprediksi nilai Y dari nilai X yang telah diberikan. Model ini hanya melibatkan satu variabel independen, yaitu variabel X. Evaluasi ini melakukan sejumlah tugas, salah satunya yaitu membuat prediksi untuk variabel terikat, variabel Y (Kriswantara et al., 2021).

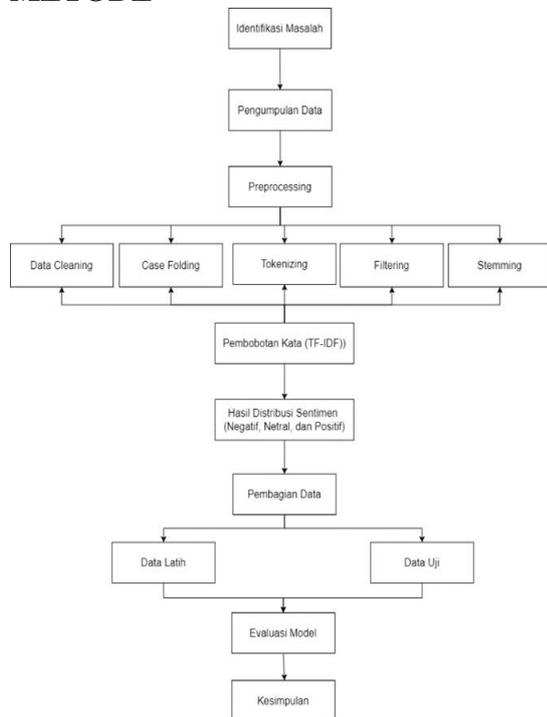
Algoritma yang diterapkan dalam penelitian ini adalah multiple regresi linier Karena melibatkan 8 atribut independen dan 1 atribut dependen. Beberapa riset terdahulu yang mencakup klasifikasi pelanggan menggunakan metode Naïve Bayes, yang sukses memprediksi pelanggan dengan precision mencapai 100% dan recall sebesar 91%, dan Tingkat akurasi mencapai 92% (Putro et al., 2020).

Penelitian lain oleh (Himawan & Eliyani, 2021) untuk Menganalisis sentimen tweet masyarakat terhadap akun Twitter resmi Pemerintah Provinsi DKI Jakarta selama pandemi COVID19 dengan TF-IDF Vectorizer untuk pembobotan kata, klasifikasi dilakukan menggunakan beberapa metode: Random Forest Classifier dengan akurasi 75,81%, Naive Bayes dengan akurasi 75,22%, dan Support Vector Machine dengan akurasi 77,58%. Analisis sentimen dilakukan pada twitter dengan persentase hasil dalam kategori negatif, netral, dan positif masing-masing yang terdiri dari 8,8%, 83,6%, dan 7,6%.

Tujuan dari penelitian ini adalah untuk memberikan wawasan mengkaji secara mendalam sentimen publik untuk resesi ekonomi global di Indonesia dengan menggunakan pendekatan hybrid yang menggabungkan algoritma Linear Regression dan Naïve Bayes. Penelitian sebelumnya cenderung menggunakan metode tunggal seperti Linear Regression atau Naïve Bayes saja, sehingga penelitian ini diharapkan dapat memberikan output

lebih valid selama mengelompokkan opini publik menjadi positif, negatif, atau netral. Dengan pemahaman yang lebih baik tentang bagaimana sentimen masyarakat berkembang terkait resesi ekonomi, pembuat kebijakan dan industri dapat merancang strategi yang lebih efektif dalam menghadapi dampak dari krisis ekonomi global

METODE



Gambar 1. Tahapan Penelitian

Identifikasi Masalah

Penelitian ini berfokus pada analisis sentimen terhadap resesi ekonomi global di Indonesia. Resesi ekonomi global merupakan fenomena yang berdampak signifikan terhadap kondisi ekonomi nasional dan kesejahteraan masyarakat. Untuk memahami bagaimana masyarakat Indonesia merespons situasi ini, kami menggunakan data dari dataset Kaggle yang berisi berbagai opini, komentar, dan diskusi mengenai resesi ekonomi. Identifikasi masalah ini melibatkan analisis sentimen untuk mengkategorikan pandangan masyarakat menjadi tiga kategori utama: negatif, netral, dan positif. Dengan mengidentifikasi sentimen ini, kami dapat memberikan wawasan yang

lebih baik tentang persepsi publik terhadap kondisi ekonomi saat ini dan bagaimana hal tersebut dapat mempengaruhi kebijakan ekonomi di masa depan.

Pengumpulan Dataset

Data yang dikumpulkan dari tweet yang diambil dari platform Kaggle digunakan dalam penelitian ini sebanyak 2.138 tweet. Dataset ini mencakup berbagai sumber data dari media sosial, forum diskusi, dan platform online lainnya yang berisi opini dan komentar mengenai resesi ekonomi global. Metode yang digunakan adalah linear regression, naive bayes, dan hybrid linear regression-naive baye untuk menganalisis sentimen. Beberapa pustaka yang digunakan dalam penelitian menggunakan bahasa pemrograman Python melalui Google Collab.

Preprocessing

Menurut Oueslati et al (Oueslati et al., n.d.), Sebelum tahap awal penelitian dimulai, proses pemrosesan teks sangat penting karena penelitian dianggap berhasil dan lancar jika tidak ada kesalahan dalam proses pemrosesan teks. Preprocessing teks melibatkan pemrosesan data awal melalui berbagai tahapan hingga data benar-benar siap untuk digunakan (Albab et al., 2023). Adapun Tahapan Preprocessing yang digunakan yaitu :

1) Tahapan Data Cleaning

Terlebih dahulu, pembersihan data harus dilakukan agar wordcloud dan analisis sentimen dapat menunjukkan hasil yang sebenarnya. Pembersihan data mencakup menghilangkan huruf besar, mention, hashtag, stopwords, emoticon, tanda baca, angka, dan url. (Albab et al., 2023).

2) Tahapan Case Folding

Case Folding yaitu suatu proses yang merubah data tweet menjadi lowercase (Pasek et al., 2022), supaya teks yang diproses dalam bentuk yang sama, case folding digunakan untuk

mengembalikan setiap kata ke dalam bentuk huruf kecil (Albab et al., 2023).

3) Tahapan Tokenizing

Tujuan tokenizing adalah untuk memberikan bobot nilai kepada setiap dokumen setelah memecah sejumlah karakter yang terdapat dalam teks menjadi satuan kata (Oryza Habibie Rahman et al., 2021). Proses memisahkan teks menjadi bagian yang disebut token, yang dapat berupa bagian huruf, kata, atau kalimat, sebelum diproses lebih lanjut. Token termasuk entitas seperti kata, angka, simbol, dan tanda baca, antara lain (Akbar & Ihsan, 2023).

4) Tahapan Filtering

Tahap di mana kata-kata penting diambil dari hasil token. Bisa menggunakan algoritma stoplist (membuang kata yang kurang penting) atau word list (menyimpan kata yang penting). Proses mengumpulkan kata-kata penting dari hasil token. Anda dapat menggunakan algoritma untuk menyimpan kata-kata yang penting atau stoplist (membuang kata-kata yang tidak penting) (Djamaludin et al., 2022). Karena jumlah kata yang diproses akan lebih sedikit, proses klasifikasi akan lebih cepat dan efisien (Saron Tandiapa & Caren Rorimpandey, n.d.).

5) Steaming

Stemming adalah proses memetik dan menggabungkan banyak kata menjadi bentuk dasar (Albab et al., 2023). Pada tahap ini, kata berimbuhan yang telah difilter akan diubah menjadi kata dasar (Saron Tandiapa & Caren Rorimpandey, n.d.).

Pembobotan Kata

Metode Inverse Document Frequency of Terms (TF-IDF) menentukan seberapa banyak kata yang mewakili isi

dokumen dengan memberikan nilai kepada setiap kata yang ada di dalamnya.

Adapun rumus dari pembobotan kata TD-IDF adalah:

$$Tf_{t,d} = \frac{N_{t,d}}{N_d}$$

$$Idf_t = \log \frac{n}{n_k}$$

Nilai TF-IDF dihasilkan dari perkalian TF dan IDF. Nilai TF dan IDF didapat dari rumus perhitungan tersebut. Dengan $Tf_{t,d}$ merupakan nilai term frequency t di dokumen d, $N_{t,d}$ merupakan jumlah munculnya term t di dokumen d, N_d merupakan total term yang terdapat pada dokumen d, Idf_t merupakan nilai IDF dari term t, n adalah jumlah dokumen, dan n_k merupakan banyaknya dokumen yang memuat term t (Farah Zhafira et al., 2021).

Hasil Distribusi Sentimen

Data yang telah melewati tahap preprocessing dan perhitungan TF-IDF digunakan sebagai data latih untuk proses pengujian. Selanjutnya, data uji diklasifikasikan menggunakan metode Hybrid Algoritma Linear Regression dan Naïve Bayes. Hasil distribusi sentimen yang dihasilkan oleh ketiga metode akan di bersifat Negatif, Netral dan Positif.

Pembagian Data

Setelah data melewati tahapan preprocessing dan perhitungan TF-IDF merupakan data latih yang akan menjadi input pada proses pengujian. Selanjutnya data uji akan melalui tahap pengklasifikasian menggunakan metode Hybrid Algoritma linear Regression dan Naïve Bayes. Data uji dan data latih yang digunakan adalah 50 :50 , 60 : 40, dan 80 :20.

Evaluasi Model

Untuk dapat mengetahui suatu performa dari metode Hybrid Algoritma linear Regression dan Naïve Bayes, maka akan dilakukan suatu pengujian terhadap model. Hasil yang akan ditampilkan dalam

bentuk confusion matrix. Dalam penelitian ini, kita mengevaluasi menggunakan nilai akurasi. Matrik evaluasi yang disebut nilai akurasi digunakan untuk mengevaluasi seberapa baik model klasifikasi atau prediksi dapat menghasilkan hasil yang tepat atau sesuai dengan data saat ini.

HASIL DAN PEMBAHASAN

Sebanyak 2138 data mentah diklasifikasikan menjadi positif, negatif, dan netral. Data mentah tersebut melalui tahapan preprocessing text dahulu agar data yang dipakai lebih akurat.

Tabel 1. Data Cleaning

Tweet	Data Cleaning
RT @xskybls: rt walaupun @Askrlfess indonesia menyusul Walaupun kayanya juga ga... Indonesia men...	
RT @jokowi: rt jumlah penduduk Jumlah penduduk, yang berkualitas yang berkualitas sekarang	

Tabel 2. Case Folding

Data Cleaning	Case Folding
rt walaupun indonesia menyusul kayanya juga ga...	rt walaupun indonesia menyusul kayanya juga ga...
rt jumlah penduduk yang berkualitas sekarang	rt jumlah penduduk yang berkualitas sekarang

Tabel 3. Tokenizing

Case Folding	Tokenizing
rt walaupun indonesia menyusul kayanya juga ga...	[rt, walaupun, indonesia, menyusul, kayanya, juga, ga...]
rt jumlah penduduk yang berkualitas sekarang	[rt, jumlah, penduduk, yang, berkualitas, sekarang]

Tabel 4. Filtering

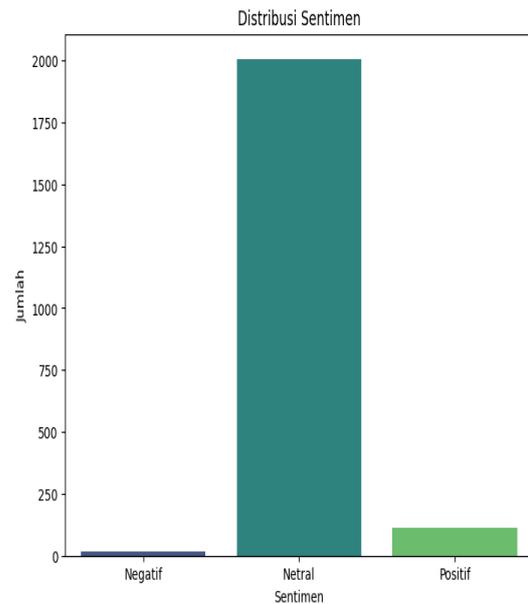
Tokenizing	Filtering
[rt, walaupun, indonesia, menyusul, kayanya, j...]	[rt, indonesia, menyusul, kayanya, gak, resesi...]

[rt, jumlah, penduduk, yang, berkualitas,	[rt, penduduk, berkualitas, kekuatan,
---	---------------------------------------

Tabel 5. Stemming

Filtering	Stemming
[rt, indonesia, menyusul, kayanya, gak, resesi...	rt indonesia menyusul kayanya gak resesi kaya ..
[rt, penduduk, berkualitas, kekuatan,	rt penduduk berkualita kekuatan ekonom

Setelah dilakukannya preprocessing text, akan menghitung label sentimennya untuk menghitung jumlah sentimen publik terhadap resesi ekonomi global di Indonesia.



Gambar 2. Hasil Distribusi Sentimen

Gambar di atas menunjukkan bahwa mayoritas pengguna Twitter bersikap netral terhadap kenaikan resesi ekonomi. Proses berikutnya adalah mengevaluasi data ke dalam algoritma Linear Regression, Naive Bayes dan Hybrid Linear Regression – Naive Bayes dengan 3 splitting data 50:50, 60:40, dan 80:20. Penelitian ini akan dilakukan menggunakan metode pembobotan kata yaitu TF-IDF.

Tabel 6. Hasil Perbandingan Akurasi

Splitting Data	TF - IDF		
	Linear Regression	Naive Bayes	Hybrid LR-NB

50:50	97,66%	97,01%	97,01%
60:40	98,01%	97,66%	97,66%
80:20	97,90%	97,20%	97,20%

Tabel 6 diatas menunjukkan bahwa algoritma Linear Regression menggunakan metode pembobotan kata TF-IDF dengan splitting data 60:40 memiliki tingkat akurasi yang paling tinggi yaitu 98,01.

PENUTUP

Kesimpulan

Dari hasil penelitian yang telah dilakukan, dapat ditarik kesimpulan bahwa pengguna Twitter dominan memberikan reaksi netral terhadap resesi ekonomi global. Penelitian ini juga membuktikan bahwa penggunaan pembobotan kata TF - IDF memiliki hasil akurasi terbaik. Menggunakan pembobotan kata TF-IDF pada splitting data 60:40 berhasil menaikkan hasil kinerja algoritma dengan akurasi 98,01%. Pada hybrid linear regression – naive bayes, splitting data 60:40 memiliki akurasi tertinggi yaitu 97,66%. Penelitian ini memberikan wawasan yang lebih mendalam tentang persepsi masyarakat terhadap resesi ekonomi, yang dapat membantu pembuat kebijakan dan industri dalam merancang strategi yang lebih efektif untuk menghadapi dampak dari krisis ekonomi global.

DAFTAR PUSTAKA

Albab, M. U., Karuniawati P, Y., & Fawaiq, M. N. (2023). Optimization of the Stemming Technique on Text preprocessing President 3 Periods Topic. 20(2), 1–10. <https://doi.org/10.26623/transformatika.v20i2.5374>

Akbar, Y., & Ihsan, A. N. (2023). Analisis Sentimen Twitter Terhadap Opini Masyarakat Pada Sea Games Kamboja 2023 Menggunakan Algoritma Support Vector Machine. *INTECOMS: Journal of Information Technology and Computer Science*, 6(2), 814–821. <https://doi.org/10.31539/intecomsv6>

i2.7670

Himawan, R. D., & Eliyani, E. (2021). Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 7(1), 58. <https://doi.org/10.26418/jp.v7i1.41728>

Djamaludin, M. A., Triayudi, A., & Mardiani, E. (2022). Analisis Sentimen Tweet KRI Nanggala 402 di Twitter menggunakan Metode Naïve Bayes Classifier. *Jurnal Teknologi Informasi Dan Komunikasi*, 6(2), 2022. <https://doi.org/10.35870/jti>

Farah Zhafira, D., Rahayudi, B., & Korespondensi, P. (2021). ANALISIS SENTIMEN KEBIJAKAN KAMPUS MERDEKA MENGGUNAKAN NAIVE BAYES DAN PEMBOBOTAN TF-IDF BERDASARKAN KOMENTAR PADA YOUTUBE (Vol. 2, Issue 1).

Halim Lubis, A., Fadillah Harahap, Y., & Studi Ilmu Komputer, P. (2023). Analisis Sentimen Masyarakat Terhadap Resesi Ekonomi Global 2023 Menggunakan Algoritma Naïve Bayes Classifier. 16(2), 442–450. <https://doi.org/10.51903/elkom.v16i2.1673>

Hasibuan, E., & Heriyanto, E. A. (n.d.). ANALISIS SENTIMEN PADA ULASAN APLIKASI AMAZON SHOPPING DI GOOGLE PLAY STORE MENGGUNAKAN NAIVE BAYES CLASSIFIER. *JTS*, 1(3).

Kriswantara, B., Kurniawati, K., & Pardede, H. F. (2021). Prediksi Harga Mobil Bekas dengan Machine Learning. *Syntax Literate; Jurnal Ilmiah Indonesia*, 6(5), 2100. <https://doi.org/10.36418/syntax-literate.v6i5.2716>

- Meisya, T., Aulia, P., Arifin, N., & Mayasari, R. (n.d.). PERBANDINGAN KERNEL SUPPORT VECTOR MACHINE (SVM) DALAM PENERAPAN ANALISIS SENTIMEN VAKSINISASI COVID-19. <https://doi.org/10.31598>
- Noor, F., #1, H., & Dwijayanti, M. (n.d.). Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naïve Bayes Classifier. In *JLK* (Vol. 4).
- Oryza Habibie Rahman, Gunawan Abdillah, & Agus Komarudin. (2021). Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 17–23. <https://doi.org/10.29207/resti.v5i1.2700>
- Oueslati, O., Cambria, E., Hajhmida, B., & Ounelli, H. (n.d.). A Review of Sentiment Analysis Research in Arabic Language. www.internetworldstats.com/stats7.htm
- Pasek, P., Mahawardana, O., Sasmita, G. A., Agus, P., & Pratama, E. (2022). Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap “Figure Pemimpin” Menggunakan Python. In *JITTER-Jurnal Ilmiah Teknologi dan Komputer* (Vol. 3, Issue 1).
- Putro, H. F., Vuldari, R. T., & Saptomo, W. L. Y. (2020). Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan. *Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN)*, 8(2). <https://doi.org/10.30646/tikomsin.v8i2.500>
- Sari, D. I., Wati, Y. F., & Widiastuti. (2020). ANALISIS SENTIMEN DAN KLASIFIKASI TWEETS BERBAHASA INDONESIA TERHADAP TRANSPORTASI UMUM MRT JAKARTA MENGGUNAKAN NAÏVE BAYES CLASSIFIER. *Jurnal Ilmiah Informatika Komputer*, 25(1), 64–75. <https://doi.org/10.35760/ik.2020.v25i1.2427>
- Saron Tandiapa, S., & Caren Rorimpandey, G. (n.d.). ANALISIS SENTIMEN ULASAN PENGGUNA PADA APLIKASI THREADS DENGAN METODE LEXICON BASED DAN NAIVE BAYES CLASSIFIER.
- Sutresno, S. A. (2023). Analisis Sentimen Masyarakat Indonesia Terhadap Dampak Penurunan Global Sebagai Akibat Resesi di Twitter. *Building of Informatics, Technology and Science (BITS)*, 4(4). <https://doi.org/10.47065/bits.v4i4.3149>