

Klasifikasi Sentimen Publik Terkait *Stunting* Di Indonesia Menggunakan BERT Dan SVM

Classification of Public Sentiment Related to Stunting in Indonesia Using BERT and SVM

Yampi R Kaesmetan^{1)*}, Wisard W Kalengkongan²⁾

¹⁾Teknik Informatika S1, STIKOM Uyelindo Kupang, NTT

²⁾Sistem Informasi, FMIPA, Universitas Sam Ratulangi, Manado, Sulawesi Selatan

Diajukan 20 Agustus 2025 / Disetujui 28 Agustus 2025

Abstrak

Stunting merupakan masalah kesehatan yang serius di Indonesia, termasuk di Provinsi Nusa Tenggara Timur (NTT), yang memerlukan analisis persepsi publik untuk mendukung kebijakan berbasis data. Penelitian ini mengusulkan analisis sentimen (sentiment analysis) menggunakan pendekatan *deep learning* dan *machine learning* untuk mengklasifikasikan opini publik terkait *Stunting* dari media sosial/*Platform online*. Dengan tujuan untuk mengevaluasi performa model BERT (*Bidirectional Encoder Representations from Transformers*) dan SVM (*Support Vector Machine*) dalam mengidentifikasi sentimen (positif, negatif, netral), dan membandingkan keunggulan BERT (berbasis transformer) dan SVM (*traditional machine learning*) untuk tugas klasifikasi sentimen, serta menganalisis faktor linguistik dan kontekstual yang memengaruhi polaritas sentimen melalui ekstraksi fitur teks. Metode penelitian meliputi pengumpulan data teks dari *Platform digital*, *preprocessing* teks, dan pelatihan model dengan *embedding* BERT sebagai fitur input SVM. Hasilnya dibandingkan dengan *baseline* tradisional (TF-IDF dan word2vec) untuk mengukur peningkatan akurasi. Hasil evaluasi menunjukkan bahwa Sentimen Negatif (86 *tweet*) *Precision*: 57,9%, *Recall*: 58%, *F1-Score*: 58%, *Accuracy*: 87%. Sentimen Netral (814 *tweet*) *Precision*: 30%, *Recall*: 20%, *F1-Score*: 25%, *Accuracy*: 100%. Sentimen Positif (100 *tweet*) *Precision*: 60%, *Recall*: 75%, *F1-Score*: 68%, *Accuracy*: 87%. Sementara itu, SVM dengan berbagai jenis kernel menunjukkan perbedaan performa dalam klasifikasi sentimen.

Kata Kunci: Analisis Sentimen, *Stunting*, BERT, SVM, Klasifikasi Teks, Media Sosial,

Abstract

Stunting is a serious health issue in Indonesia, including in East Nusa Tenggara Province (NTT), which requires an analysis of public perception to support data-driven policies. This study proposes sentiment analysis using deep learning and machine learning approaches to classify public opinions regarding Stunting from social media/online Platforms. It aims to evaluate the performance of the BERT (Bidirectional Encoder Representations from Transformers) and SVM (Support Vector Machine) models in identifying sentiment (positive, negative, neutral), compare the advantages of BERT (transformer-based) and SVM (traditional machine learning) for sentiment classification tasks, and analyze the linguistic and contextual factors influencing sentiment polarity through text feature extraction. The research methods include collecting text data from digital Platforms, text preprocessing, and model training with BERT embeddings as input features for SVM. The results are compared with traditional baselines (TF-IDF and word2vec) to measure Accuracy improvement. The evaluation results show that for Negative Sentiment (86 tweets) Precision: 58%, Recall: 58%, F1-Score: 58%, Accuracy: 87%. Neutral Sentiment (814 tweets) Precision: 30%, Recall: 20%, F1-Score: 25%, Accuracy: 100%. Positive Sentiment (100 tweets) Precision: 60%, Recall: 75%, F1-Score: 68%, Accuracy: 87%. Meanwhile, SVM with various kernel types showed performance differences in sentiment classification.

Keywords: Sentiment Analysis, *Stunting*, BERT, SVM, Text Classification, Social Media,

*Korespondensi Penulis:

E-mail: kaesmetanyampi@gmail.com

Pendahuluan

Stunting merupakan salah satu masalah kesehatan masyarakat yang paling krusial di Indonesia, dengan prevalensi yang masih cukup tinggi meskipun berbagai upaya penanganan telah dilakukan. Berdasarkan data dari Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) pada tahun 2022, sekitar 24,4% balita di Indonesia mengalami *stunting*, artinya hampir satu dari empat anak balita memiliki tinggi badan di bawah standar untuk usianya akibat kekurangan gizi kronis. Angka ini masih jauh di atas ambang batas yang ditetapkan oleh Organisasi Kesehatan Dunia (WHO), yaitu di bawah 20%. *Stunting* tidak hanya berdampak pada pertumbuhan fisik anak, tetapi juga memengaruhi perkembangan kognitif, produktivitas, dan potensi ekonomi di masa depan. Oleh karena itu, isu *Stunting* telah menjadi salah satu fokus utama pemerintah Indonesia dalam beberapa tahun terakhir, dengan berbagai program intervensi gizi yang ditujukan untuk menurunkan angka *Stunting* secara signifikan. Sentimen publik terhadap *Stunting* dapat memengaruhi keberhasilan program-program pemerintah, karena partisipasi aktif masyarakat sangat dibutuhkan dalam pencegahan dan penanganan *stunting*. Salah satunya dengan media sosial dan *Platform* online telah menjadi saluran utama bagi masyarakat untuk menyampaikan pendapat, keluhan, maupun apresiasi terkait berbagai isu, termasuk *stunting*. *Platform* seperti Twitter, Facebook, Instagram, dan forum-forum diskusi kesehatan dipenuhi dengan percakapan tentang pengalaman pribadi, kritik terhadap kebijakan pemerintah, atau sekadar berbagi informasi terkait *stunting*. Salah satu tantangan utama dalam menganalisis sentimen publik yaitu kompleksitas bahasa manusia, terutama dalam konteks bahasa Indonesia yang kaya akan variasi, termasuk penggunaan bahasa informal, slang, dan campuran bahasa daerah. Pendekatan *machine learning* konvensional seperti algoritma klasifikasi tradisional seperti *Support Vector Machine* (SVM) dapat memberikan solusi yang efisien. Penelitian ini bertujuan untuk mengimplementasikan klasifikasi sentimen publik terkait *Stunting* di Indonesia dengan memanfaatkan BERT dan SVM. Dengan mengumpulkan data dari media sosial dan *Platform* online, penelitian ini menganalisis bagaimana masyarakat merespons isu *stunting*, apakah cenderung positif (misalnya mendukung program pemerintah), negatif (misalnya mengkritik ketidakefektifan intervensi), atau netral (hanya berbagi informasi tanpa sikap jelas). Secara keseluruhan, penelitian tentang klasifikasi sentimen publik terkait *Stunting* menggunakan BERT dan SVM memiliki relevansi yang tinggi, baik dari sisi kesehatan masyarakat maupun pengembangan teknologi. Dengan memahami sentimen publik secara lebih mendalam, intervensi kebijakan dapat lebih tepat sasaran, sehingga upaya penurunan *Stunting* di Indonesia dapat berjalan lebih efektif dan berkelanjutan..

Metode Penelitian

Support Vector Machine

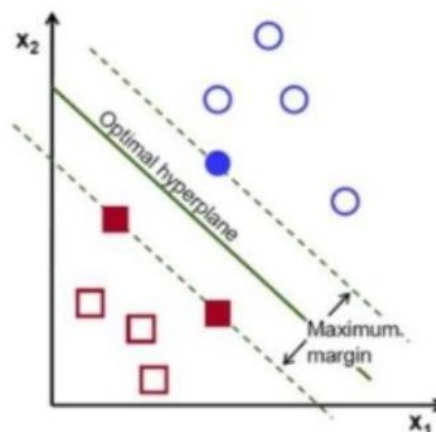
Support Vector Machine merupakan metode klasifikasi jenis (*Supervised*) karena ketika proses pelatihan memerlukan goal atau target pembelajaran tertentu. SVM merupakan algoritma yang bekerja menggunakan pemetaan *nonlinier* untuk mengubah data pelatihan menjadi dimensi yang tinggi. Pada dimensi baru akan mencari *hyperplane* untuk memisahkan secara *linier* dan dengan pemetaan *nonlinier* yang tepat ke dimensi yang lebih tinggi lagi (Aldino et al, 2022). Prinsip SVM merupakan membangun *hyperplane* yang memiliki ukuran margin yang sama dan tidak cenderung mendekati daerah dari salah satu kelas. Hal tersebut dapat dilakukan dengan pengukuran margin kemudian dilakukan pencarian titik maksimalnya. *Polynomial*, *Sigmoid*, *Linear*, dan *Radial Basis Function* (RBF) merupakan empat *kernel* dari SVM.

Dalam pembelajaran mesin, *kernel polinomial* merupakan fungsi *kernel* yang sering digunakan dengan SVM, mewakili kesamaan sampel pelatihan (vektor) dalam ruang fitur atas polinomial dari variabel asal, memungkinkan pembelajaran model non-linear. Ini cocok untuk masalah di mana semua data pelatihan dinormalisasi. Sementara itu, kernel Sigmoid diturunkan dari Jaringan Syaraf, di mana fungsi sigmoid bipolar secara rutin digunakan untuk mengaktifkan neuron buatan. Untuk kernel Linear, digunakan ketika data dapat dipisahkan. Ini paling sering digunakan

ketika ada sejumlah besar data yang dapat dipisahkan, seperti kelas dalam teks. Yang terakhir adalah kernel RBF (*Radial Basis Function*); ini adalah kernel yang sering digunakan karena biasanya memberikan hasil yang tepat. Kernel RBF adalah kernel yang secara umum dapat digunakan untuk semua jenis data (Prasetyo et al, 2020). Persamaan fungsi kernel dapat dilihat di Gambar 1.

No	Kernel Function	Equation
1	Linear	$K(x_n, x_i) = (x_n, x_i)$
2	Polynomial	$K(x_n, x_i) = (-\gamma(x_n, x_i) + r)^d$
3	Sigmoid	$K(x_n, x_i) = \tanh(\gamma(x_n, x_i) + r)$
4	RBF	$K(x_n, x_i) = \exp(-\gamma\ x_n - x_i\ ^2 + C)$

Gambar 1. Persamaan kernel *Support Vector Machine*



Gambar 2. Ilustrasi *Hyperplane*

Dalam gambar 2 terlihat bahwa pada sumbu X_1 dan X_2 menunjukkan adanya sebaran data dari masing, *Hyperplane* dengan margin terbesar dianggap paling robust. Ia seperti jalan raya yang paling lebar di antara dua desa. *Hyperplane* ini akan lebih umum (*generalize*) dan memiliki ketahanan yang lebih baik terhadap data baru atau *noise* dalam data. Model yang dihasilkan cenderung tidak *overfitting*. Kelas Merah (+1) memiliki nilai X_1 (misalnya, ukuran rumah) yang besar dan X_2 (misalnya, harga) yang tinggi. Kelas Biru (-1): Memiliki nilai X_1 yang kecil dan X_2 yang rendah.

BERT (*Representasi Encoder Bidirectional Transformers*)

Representasi Encoder Bidirectional Transformers atau disingkat BERT, merupakan model representasi bahasa terlatih yang dibuat pada tahun 2018 oleh peneliti Google AI Language. BERT dibuat menggunakan pembelajaran semi-diawasi, ELMo, ULMFiT, OpenAI Transformers, dan Transformers serta teknik *deep learning* (Devlin, et al., 2019). Bert merupakan model salah satu model kontekstual yang dirancang diatas transformer model, Oleh karena itu bert mampu menangkap dan mengklasifikasikan suatu kata dalam kontks dan level tertentu sesuai dengan struktur model yang dirancang (Lenggo, et al., 2023).

BERT memiliki 6 lapisan *Transformer* yang ditumpangkan diatas *encoder* dan *decoder* masing masing, yang menyumbang file proses pelatihan yang sangat kompleks, konfigurasi tinggi, waktu pelatihan yang banyak dan biaya yang sangat mahal. Namun, Google membuka sumber dari *pretrained* model dari BERT, yang mana dapat menjadi *opensource* tanpa harus membuat model terlebih dahulu. BERT *processing* dimulai dari sebuah kata dengan representasi *embedding* dari lapisan *embedding*. Setiap lapisan melakukan beberapa penghitungan *multi-headed attention* pada representasi kata dari lapisan sebelumnya untuk membuat representasi perantara baru. Semua representasi perantara ini memiliki ukuran yang sama. Dalam model BERT 12 layer, sebuah token

akan memiliki 12 representasi menengah (Kusuma Atmaja & Yustanti, 2021).

Media Sosial X

Media sosial adalah sebuah media online. Para penggunanya bisa dengan mudah berpartisipasi, berbagi, dan menciptakan isi meliputi *blog*, jejaring sosial, wiki, forum, dan dunia virtual (Harahap & Rakhmat., 2024). Aplikasi X atau yang dulu dikenal dengan Twitter merupakan *Platform* media sosial di mana orang dapat berbagi pesan singkat, gambar, dan video. *Platform* ini telah berkembang menjadi tempat di mana orang dari seluruh dunia berkumpul untuk membahas ide-ide baru dan tren (Maulana & Yuliana., 2024). Twitter atau X menjadi urutan ke 5 sebagai media sosial terfavorit di Indonesia pada tahun 2022 dengan jumlah penggunanya sejumlah 14,85 juta (Wijayanti & Wijayani., 2024). Menurut Rossianah, (2024) aplikasi X memiliki beberapa kekurangan dan kelebihan dibandingkan media sosial lain, berikut merupakan kelebihan media sosial Twitter / X Lebih menghemat biaya X sepenuhnya gratis, seperti kebanyakan media sosial. Selain itu, tidak ada biaya untuk memposting *tweet*, jadi kamu bebas mengirimkan postingan tentang apa pun yang disukai. Selain mudah digunakan Twitter atau X memiliki antarmuka yang ramah pengguna. Semua grafisnya dibuat dengan menarik sehingga orang awam dapat cepat memahami fiturnya. Twitter menggabungkan blogging, pesan, dan komentar. Semua informasi dan data dapat dibagikan kepada khalayak luas, bukan hanya sahabat dan keluarga. Ini adalah alasan mengapa pengguna dapat menemukan seseorang yang membagikan kronologi lengkap dari suatu kejadian atau banyak influencer yang memberikan review melalui thread Twitter. Dibalik kelebihan yang dimiliki aplikasi X juga memiliki beberapa kekurangan, berikut merupakan kekurangan media sosial twitter/ X:

Misinformasi dan Ujaran Kebencian mulai dari Penyebaran misinformasi dan ujaran kebencian tidak dapat dihindari di Twitter atau X. Penyebaran konten yang tidak dapat dipastikan kebenarannya di *Platform* terbuka seringkali mengakibatkan penyebaran informasi yang tidak benar dan retorika yang merugikan. Langganan X *Blue* dengan Kemunculan X *Blue* mendapat banyak kritik dari pengguna. Metode ini memungkinkan verifikasi pengguna, memungkinkan akses eksklusif untuk pengguna yang ingin membayar serta Adanya Batasan karakter dengan batasan 280 karakter per-*tweet*, Twitter membuat mengirim postingan menjadi sulit. Oleh karena itu, pengguna harus tahu cara memposting *tweet* dengan efektif agar apa yang disampaikan dapat dimuat semua.

TF-IDF

TF-IDF (*Term Frequency – Inverse Document Frequency*) digunakan sebagai ekstraksi dan seleksi fitur. TF (*Term Frequency*) mengukur seberapa sering istilah muncul dalam sebuah dokumen. Sementara itu, IDF (*Inverse Document Frequency*) menghitung seberapa penting suatu istilah. Ini tidak mempertimbangkan kata-kata yang muncul berulang kali, seperti ini (ini), (ke), dan (dan), serta (sama), dan seterusnya (Surya et al., 2019). TF-IDF adalah metode statistik yang menggabungkan teknik TF dan IDF. Ini mengevaluasi seberapa relevan suatu kata terhadap dokumen dalam kumpulan dokumen. Selain itu, ia dapat meningkatkan nilai *Recall* dan *Precision*. Persamaan TF dapat dilihat pada persamaan 1

$$TF(d, t) = \frac{f_{t,d}}{\sum_t f_{t,d}} \dots \dots \dots (1)$$

Dimana $\sum_t f_{t,d}$ merupakan total semua kata yang terkandung dalam dokumen (d) dan $f_{t,d}$ merupakan frekuensi setiap kata (t) yang muncul dalam dokumen (d). Untuk nilai IDF, dapat dihitung menggunakan persamaan 2 sebagai berikut :

$$IDF(t) = \log \frac{N}{df(t)} \dots \dots \dots (2)$$

Dimana f_t merupakan jumlah dokumen dimana istilah (t) muncul dalam dokumen (d), sedangkan $|D|$ merupakan total jumlah dokumen dalam koleksi/folder. Dengan informasi ini, nilai TF-IDF (W) dapat dihitung menggunakan persamaan 3 sebagai berikut :

$$TF - IDF (t, d) = TF (t, d) \times IDF (t) \dots\dots\dots(3)$$

Confusion Matrix

Confusion matrix merupakan tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah. Contoh *confusion matrix* untuk klasifikasi biner ditunjukkan pada tabel 1 berikut ini (Prayogi dan Normawati, 2021)

Tabel 1. Confussion matrix			
Kelas Sebenarnya	Kelas Prediksi		
	1	0	
	TP	FN	
0	FP	TN	

Keterangan

- *TP (True Positive)* = jumlah dokumen dari kelas 1 yang benar diklasifikasikan sebagai kelas 1
- *TN (True Negative)* = jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0
- *FP (False Positive)* = jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1
- *FN (False Negative)* = jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0

Rumus *confusion matrix* untuk menghitung *Accuracy*, *Precision*, dan *Recall* seperti berikut:

$$Accuracy = \frac{TP+TN}{Total} \dots\dots\dots(4)$$

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots(5)$$

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots(6)$$

Metrik-metrik ini akan membantu menilai performa SVM dalam mengklasifikasikan sentimen pengguna dengan lebih detail.

K-Fold Cross Validation

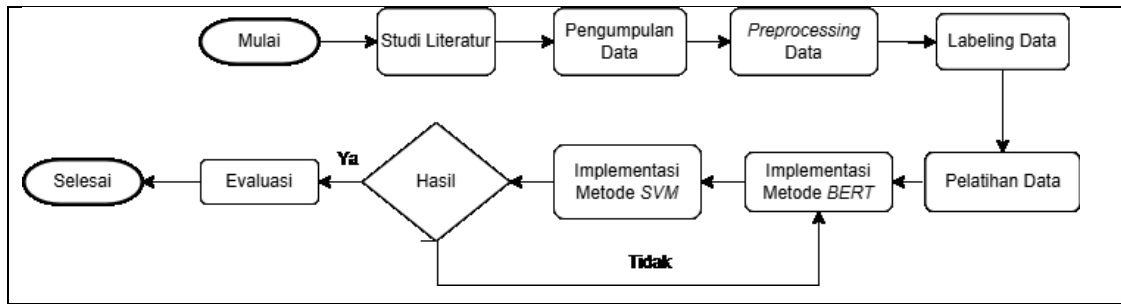
K-fold cross validation merupakan salah satu dari teknik yang difungsikan untuk memilah data menjadi *train* data serta *test* data. Teknik ini banyak diterapkan peneliti karena didapati mengurangi bias yang didapatkan didalam pengambilan sebuah sampel. *K-fold cross validation* berlaku *continue* membagi data-data menjadi *train* data dan *test* data, sehingga setiap data akan mendapat kesempatan menjadi test data (Teungku Ridwansyah, 2022).

Tahapan Penelitian

Pada gambar 3 terlihat alur penelitian yang menjadi patokan untuk dilakukan, setiap alur penelitian ini dijabarkan sebagai berikut :

Pengumpulan Data

Data dikumpulkan dari media *X* menggunakan *API* dan dengan bantuan *tools tweetharvest*. Pengumpulan data dilakukan dengan menyaring *tweet* yang mengandung kata kunci *Stunting* dalam rentang waktu Mei 2025 sampai Juni 2025. Dengan menggunakan Auth Token dari X, kemudian manegggunakan *pandas* dan *node.ds* untuk dapat membaca data.



Gambar 3. Alur Penelitian

Preprocessing Data

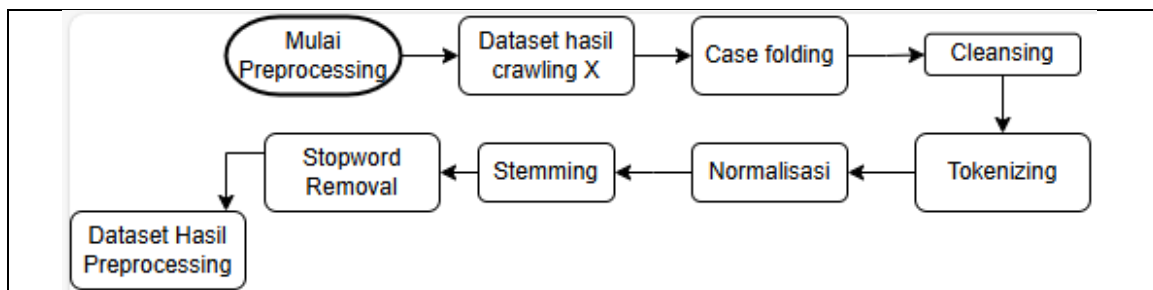
Preprocessing data dilakukan untuk membersihkan dan menyiapkan data sebelum analisis lebih lanjut. Dimana dalam data tersebut banyak mengandung tanda, karakter, hashtag dan non alfanumerik, serta mengubah teks menjadi lowercase dan menghapus spasi. Kemudian *Listing* dalam *preprocessing* ini digunakan untuk membersihkan dan memproses teks dalam *Dataset* yang berisi *tweet*. *Dataset* yang berisi *tweet* disimpan dalam file Excel. Kemudian teks dalam kolom '*tweet*' dibersihkan dari elemen-elemen yang tidak diperlukan seperti *mention*, *hashtag*, *URL*, dan karakter khusus. Setelah itu Proses *normalisasi* bertujuan untuk mengganti kata-kata tidak baku dengan kata yang lebih formal agar lebih mudah diproses.

Labeling Data

Labeling data dilakukan menggunakan *library textblob* yang menyediakan analisis sentimen berbasis *leksikon*. Setiap postingan yang diberi label sentimen sebagai positif, negatif, atau netral berdasarkan nilai popularitas dari *textblob*. *Listing code labeling* data memberikan label sentimen pada data teks hasil *preprocessing* dengan menggunakan *textBlob*. Polaritas sentimen yang digunakan berupa nilai numerik antara -1 hingga 1 yang menunjukkan apakah suatu teks memiliki sentimen negatif, netral, atau positif. Perhitungan ini dilakukan menggunakan fungsi *sentiment.polarity* dari *textBlob*. Setelah nilai polaritas dihitung, langkah berikutnya adalah memberikan label sentimen berdasarkan nilai polaritas tersebut.

Pembagian Data

Data dibagi menjadi data latih (80%) dan data uji (20%). Pembagian data ini menggunakan metode *k-fold validation* untuk memastikan model dapat belajar dari sebagian data latih dan data uji yang dibagi. Pembagian data bertujuan untuk mempersiapkan *Dataset* dalam format *BertDataset*. Membuat *DataLoader* agar bisa digunakan dalam training. Kemudian menampilkan contoh *batch* untuk memverifikasi bahwa data telah diproses dengan benar.



Gambar 4. Alur Preprocessing

Pada gambar 4 merupakan alur *preprocessing* dimana dimulai dari proses *preprocessing*, *Dataset* hasil *crawling* kemudian dilakukan *case folding* dan *cleansing*, setelah itu *tokenizing* lalu dilakukan *normalisasi* dan *stemming* dan kemudian *stopword removal* dan diperoleh *Dataset* hasil *preprocessing*.

Hasil Dan Pembahasan

Implementasi Metode BERT dan SVM

Implementasi metode *BERT* dan *SVM* ke dalam program *python* dengan menggunakan *fine-tuning BERT* untuk analisis sentimen dengan tiga kelas (positif, netral, negatif). *Dataset* diolah menjadi *tensor PyTorch*, lalu digunakan dalam *training loop*. *Optimizer AdamW* digunakan untuk memperbarui bobot model berdasarkan error. Setiap *epoch*, *loss* dan akurasi dihitung untuk melihat performa model. Model yang sudah dilatih disimpan agar bisa digunakan di masa depan tanpa perlu training ulang. Kemudian dalam Implementasi Metode *Support Vector Machine* dengan program dengan menghitung *Linear*, *RBF*, *Polynomial*, dan *Sigmoid*. Dari *listing* data yang ada maka di dapat perbedaan dari 4 *kernel Support Vector Machine*, dan terlihat pada tabel 3 setiap kelebihan dan kekurangan dari *Kernel SVM* sebagai berikut:

Tabel 3. Kelebihan dan Kekurangan *Kernel SVM*

Kernel	Data	Kelebihan	Kekurangan
<i>Linear</i>	Bisa dipisahkan dengan garis lurus	Cepat dan sederhana	Tidak cocok untuk data kompleks
<i>RBF</i>	Memiliki pola <i>non-linear</i> kompleks	Fleksibel & powerful	Butuh <i>tuning parameter gamma</i>
<i>Polynomial</i>	Pola <i>non-linear</i> tetapi masih bisa direpresentasikan sebagai <i>polynomial</i>	Bisa menangani beberapa <i>non-linearitas</i>	Lebih lambat dibanding <i>RBF</i>
<i>Sigmoid</i>	Menyerupai fungsi aktivasi <i>sigmoid</i>	Mirip dengan <i>Neural Networks</i>	Jarang digunakan, sering kurang optimal

Proses *preprocessing* dimulai dari *Dataset* hasil *crawling* yang biasanya masih mentah dan mengandung banyak *noise*. Tahap pertama adalah *case folding* dimana semua teks diubah menjadi huruf kecil untuk konsistensi, mengatasi variasi penulisan seperti "Bagus", "bagus", atau "bAGus" menjadi seragam "bagus". Ini crucial untuk memastikan model memperlakukan kata yang sama sebagai entitas yang identik. Selanjutnya *cleansing* dilakukan secara intensif untuk membersihkan berbagai jenis *noise* karakteristik data sosial media. Proses ini menghapus mention (@username) dan hashtag (#topic) yang tidak relevan untuk analisis sentimen, menghilangkan URL dan link yang biasanya tidak mengandung nilai sentiment, serta membersihkan punctuation dan karakter khusus yang dapat mengganggu proses *tokenization*. Uniknya, proses *cleansing* juga termasuk konversi emoji menjadi teks deskriptif (seperti 😊 menjadi ":smiling_face:") karena emoji sering kali membawa informasi sentiment yang kuat. Setelah teks bersih, dilakukan *tokenizing* dimana teks dipotong-potong menjadi unit kata individual. Proses ini khusus menangani kata gabung yang umum di media sosial seperti "gakpernah" menjadi ["gak", "pernah"]. Kemudian *normalization* mengubah kata-kata slang dan tidak baku menjadi bentuk formalnya menggunakan *mapping dictionary* yang komprehensif - "yg" menjadi "yang", "dg" menjadi "dengan", "gak" menjadi "tidak", dan berbagai variasi lainnya. Tahap *stemming* menggunakan dua pendekatan berbeda: Porter Stemmer untuk teks Inggris dan Sastrawi untuk teks Indonesia, yang mampu mengurangi kata ke bentuk dasarnya dengan mempertimbangkan kaidah linguistik masing-masing bahasa. Terakhir, *stopword removal* menghapus kata-kata umum yang tidak informatif seperti "dan", "atau", "pada" sambil tetap mempertahankan kata yang penting untuk analisis sentimen seperti "sangat", "buruk", atau "luar biasa". Proses ini menghasilkan teks yang terstandarisasi dan siap untuk *feature extraction* dalam model *machine learning*.

Evaluasi

Evaluasi model dilakukan menggunakan data uji. Hasil evaluasi ditampilkan menggunakan *confussion matrix*. Listing dalam evaluasi bertujuan mengevaluasi model *Support Vector Machine (SVM)* yang telah dilatih untuk klasifikasi sentimen. Evaluasi dilakukan menggunakan akurasi, *presisi*, *Recall*, dan spesifisitas berdasarkan hasil prediksi terhadap data uji.

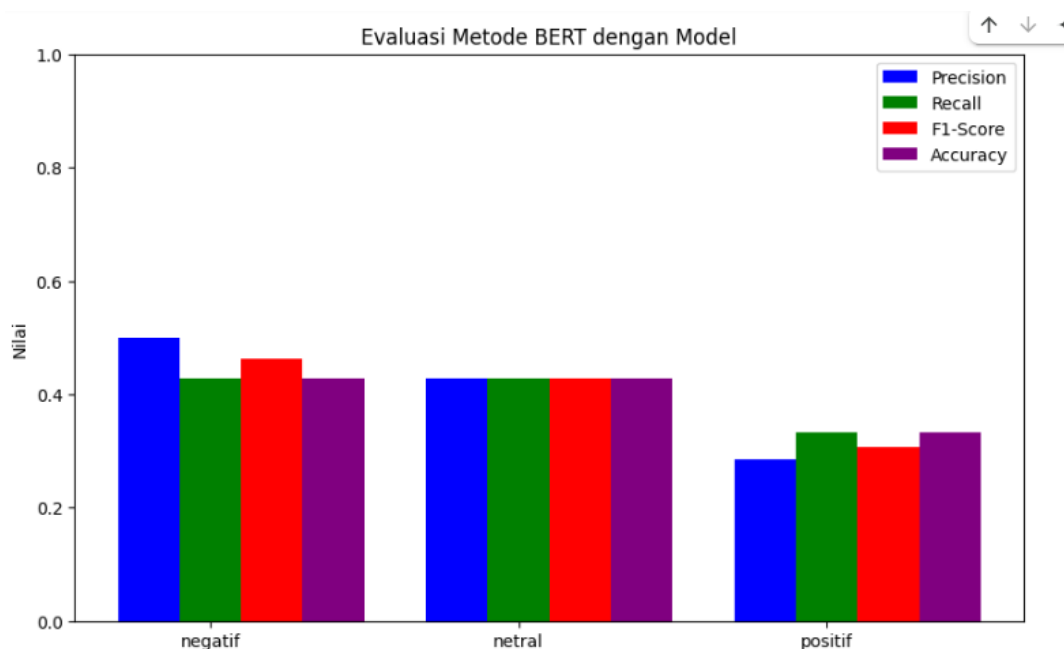
Hasil Metode BERT

Untuk memastikan eksperimen dapat direplikasi dengan hasil yang konsisten, implementasi menggunakan *Bidirectional Encoder Representations from Transformers (BERT)* dan *Support Vector Machine (SVM)*. Proses diawali dengan pengumpulan data sebanyak 1000 *tweet* yang kemudian melalui tahap *preprocessing* komprehensif meliputi *case folding*, *cleansing*, *tokenizing*, sebagai model dasar, yang secara khusus telah dilatih sebelumnya pada korpus bahasa Indonesia sehingga sangat cocok untuk analisis sentimen teks dalam bahasa Indonesia. Model ini memiliki 12 layer transformer dengan 110 juta parameter, menggunakan *vocabulary size* sebesar 30.522 token yang telah dioptimalkan untuk karakteristik linguistik bahasa Indonesia.

Parameter pelatihan diatur dengan sangat spesifik yaitu learning rate sebesar $2e-5$ (0.00002) sementara *epoch training* sebanyak 3 kali ditentukan berdasarkan observasi bahwa model sudah mencapai konvergensi tanpa menunjukkan tanda-tanda *overfitting*. Untuk *sequence length*, ditetapkan max length 128 token yang telah mencukupi untuk sebagian besar teks ulasan dalam bahasa Indonesia tanpa memotong informasi penting.

Schedule learning rate menggunakan *linear warmup* dengan *warmup steps* 0 dan *linear decay* hingga akhir training, sementara *gradient clipping* dengan max norm 1.0 diterapkan untuk mencegah *exploding gradient*. Split *Dataset* menggunakan rasio 80:20 untuk training dan validation dengan *random state* 42 agar dapat direproduksi secara konsisten. Semua parameter ini telah melalui proses tuning yang hati-hati untuk memastikan model dapat mempelajari pola sentimen dengan efektif *while maintaining generalization ability* pada data *unseen*.

Hasil analisis sentimen dan akurasi ini digambarkan pada visual grafik tabung, sebagai berikut:



Gambar 5 . Grafik Hasil Metode BERT

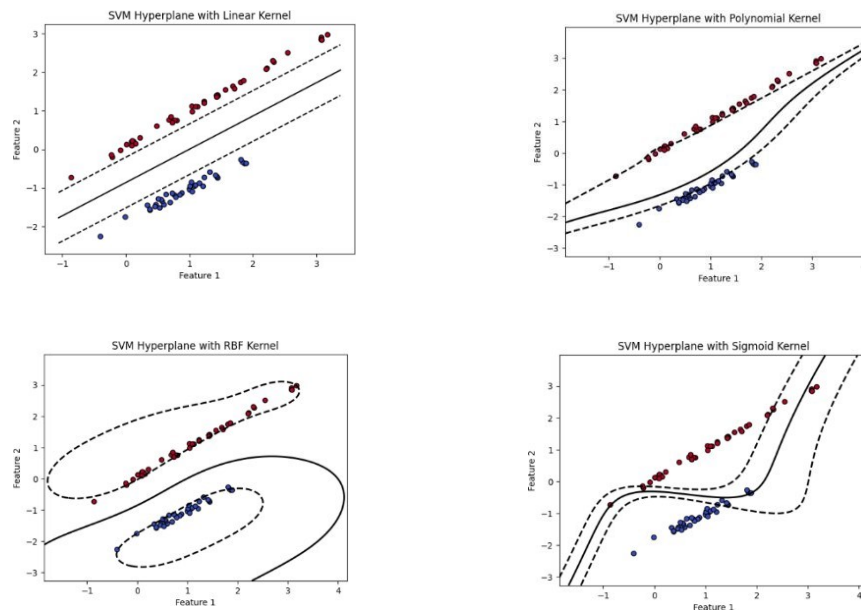
Dari gambar 5 menunjukan jika hasil untuk sentimen negatif (86 *tweet*) memiliki *presicion* 57,9%, *Recall* 58%, *fi-Score* 58%, dan *Accuracy* 100%. Untuk sentimen netral (814 *tweet*)

menunjukkan jika *Precision* 30%, *Recall* 20%, dan *fi-Score* 25%, dan *Accuracy* 100%. Sedangkan untuk sentimen positif (100 tweet) memiliki nilai *Precision* 60%, *Recall* 75%, dan *fi-Score* 68%, dan *Accuracy* 100%.

Hasil Metode *Support Vector Machine*

Berikut merupakan hasil metode *Support Vector Machine* berupa *hyperplane* dan hasil sentimen :

1. Hasil *Hyperplane Support Vector Machine*

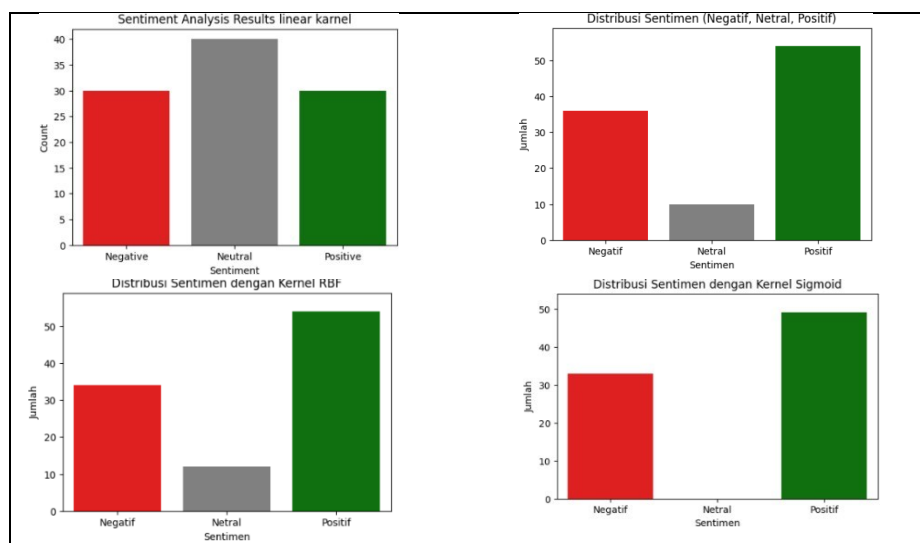


Gambar 6. Hyperplane SVM

Berdasarkan gambar 6 *hyperplane* dari model *Support Vector Machine (SVM)* dengan berbagai *kernel*, terlihat bahwa setiap jenis *kernel* memiliki karakteristik yang berbeda dalam memisahkan data ke dalam dua kelas. *Kernel* linear menunjukkan pemisahan yang sederhana dengan garis lurus, sedangkan *kernel polynomial* dan *RBF (Radial Basis Function)* mampu menangani data yang tidak terpisah secara linier dengan menghasilkan batas keputusan yang lebih kompleks. Sementara itu, *kernel sigmoid* menghasilkan pemisahan dengan bentuk kurva berbasis fungsi aktivasi yang menyerupai jaringan saraf tiruan. Karena data sentimen dari *Twitter* biasanya memiliki pola yang kompleks dan tidak linier, *kernel RBF* merupakan pilihan yang paling cocok. *Kernel RBF* memiliki kemampuan untuk menangani data dengan distribusi yang tidak beraturan dan menyesuaikan pola hubungan antar fitur, sehingga dapat meningkatkan akurasi dalam mengklasifikasikan sentimen positif, negatif, dan netral. Dengan demikian, penggunaan *kernel RBF* dalam penelitian ini dapat membantu dalam mendapatkan hasil analisis yang lebih akurat dan mendalam.

2. Hasil Sentimen *Support Vector Machine*

Berdasarkan gambar 7 menampilkan hasil dari SVM.



Gambar 7. Hasil Sentimen SVM

Simpulan

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa metode *Bidirectional Encoder Representations from Transformers* (BERT) dan *Support Vector Machine* (SVM) dapat digunakan untuk melakukan analisis sentimen terhadap sunting pada Platform X. Implementasi metode BERT menunjukkan bahwa model ini mampu menangkap konteks kata dengan baik, Hasil evaluasi menunjukkan bahwa Sentimen Negatif (86 tweet) *Precision*: 57,9%, *Recall*: 58%, *F1-Score*: 58%, *Accuracy*: 87%. Sentimen Netral (814 tweet) *Precision*: 30%, *Recall*: 20%, *F1-Score*: 25%, *Accuracy*: 100%. Sentimen Positif (100 tweet) *Precision*: 60%, *Recall*: 75%, *F1-Score*: 68%, *Accuracy*: 87%. Sementara itu, SVM dengan berbagai jenis kernel menunjukkan perbedaan performa dalam klasifikasi sentimen. Hasil analisis menunjukkan bahwa kernel *Radial Basis Function* (RBF) dan *Polynomial* memberikan pemisahan data yang lebih akurat dibandingkan kernel linear dan sigmoid, karena kemampuannya dalam menangani data dengan pola kompleks dan tidak linier. Jumlah tweet berdasarkan hasil klasifikasi SVM dengan masing-masing kernel adalah sebagai berikut, kernel Linear: Sentimen Negatif (426 tweet), sentimen Netral (146 tweet), sentimen Positif (428 tweet). Kernel RBF & *Polynomial*: Sentimen Negatif (423 tweet), Sentimen Netral (155 tweet), sentimen Positif (422 tweet). Kernel Sigmoid: Sentimen Negatif (500 tweet), sentimen Netral (0 tweet), sentimen Positif (500 tweet). Berdasarkan hasil tersebut, kernel RBF merupakan pilihan yang paling optimal dalam implementasi SVM untuk analisis sentimen dalam penelitian ini.

Daftar Pustaka

- Abdullah, M. D. & Abdulazeez, A. M., 2021. *Machine learning Applications based on SVM Classification: A Review*. *Qubahan Academic Journal*, pp. 81- 90.[Internet]. [Diakses pada 20 Juli 2025]. Tersedia pada <https://journal.qubahan.com/index.php/qaj/article/view/50>
- Alghifari, F. & Juardi, D., 2021. Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes. *Jurnal Ilmiah Informatika (JIF)*, Volume 9, pp. 76-81. [Internet]. [Diakses pada 20 Juli 2025]. Tersedia pada <https://forum.upbatam.ac.id/index.php/jif/article/view/3755>
- Ananda, D. & Suryono, R. R., 2024. Analisis Sentimen Publik Terhadap Pengungsi Rohingya di Indonesia dengan Metode *Support Vector Machine* dan Naïve Bayes. *Jurnal Media Informatika*

- Budidarma, Volume 8, pp. 748-757. [Internet]. [Diakses pada 30 Juli 2025]. Tersedia pada <https://ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/7517>
- Andriani, R., Suhwardi & Hapisah, 2022. Hubungan Tingkat Pengetahuan Dan Sikap Remaja Dengan. *Jurnal Inovasi Penelitian*, Volume Vol.2 No.10 Maret, pp. 3441-3446. [Internet]. [Diakses pada 30 Juli 2025]. Tersedia pada <https://ejournal.stpmataram.ac.id/JIP/article/view/1341>
- Atika, D., Styawati & Aldino, A. A., 2022. Term Frequency-Inverse Document Frequency Support Vector Machine Untuk Analisis Sentimen Opini Masyarakat Terhadap Tekanan Mental Pada Media Sosial Twitter. *Jurnal Teknologi dan Sistem Informasi (JTSI)*, Volume 4, pp. 86-97. [Internet]. [Diakses pada 30 Juli 2025]. Tersedia pada <https://jim.teknokrat.ac.id/index.php/sisteminformasi/article/viewFile/2054/719>
- Aziz, Z. A., Abdulqader, D. N., Sallow, A. B. & Omer, H. K., 2021. Python Parallel Processing and Multiprocessing: A Review. *Academic Journal of Nawroz University (AJNU)*, Volume 10, pp. 345-354. [Internet]. [Diakses pada 30 Juli 2025]. Tersedia pada <https://pdfs.semanticscholar.org/7337/73fdf89057322ea78489912c6f769bdfbaff.pdf>
- Darwis, D., Pratiwi, E. S. & Pasaribu, A. F. O., 2020. Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia. *Jurnal Ilmiah Edutic*, Volume 7, pp. 1-11. [Internet]. [Diakses pada 20 Juli 2024]. Tersedia pada <https://journal.trunojoyo.ac.id/edutic/article/view/8779>
- Darwis, D., Siskawati, N. & Zaenal, A., 2021. Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional. *Jurnal TEKNO KOMPAK*, Volume 15, pp. 131-145. [Internet]. [Diakses pada 30 Juli 2025]. Tersedia pada <https://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/744/0>
- Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Association for Computational Linguistics*, pp. 4171-4186. [Internet]. [Diakses pada 30 Juli 2025]. Tersedia pada <https://arxiv.org/abs/1810.04805>
- Farras, M., Mawardi, V. C. & Sutrisno, T., 2023. Aplikasi Analisis Sentimen Komentar Pengguna Genshin Impact Di Play Store. *Jurnal Ilmu Komputer dan Sistem Informasi*, pp. 1-6. [Internet]. [Diakses pada 30 Juli 2025]. Tersedia pada https://www.researchgate.net/publication/373373575_Aplikasi_Analisis_Sentimen_Komentar_Pengguna_Genshin_Impact_Di_Play_Store
- Fikri, M. I., Sabrila, T. S. & Azhar, Y., 2020. Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *SMATIKA Jurnal*, Volume 10, pp. 71-76. [Internet]. [Diakses pada 30 Juli 2025]. Tersedia pada <https://jurnal.stiki.ac.id/SMATIKA/article/view/455>
- Hartanto, R. & Siahaan, M., 2023. Multi-Method Analysis On Video Game Addiction And Academic. *Conference on Management, Business*, Volume 3, pp. 279-290. [Internet]. [Diakses pada 30 Juli 2025]. Tersedia pada <https://journal.uib.ac.id/index.php/combines/article/view/7697>
- Hatta, M. S. & Lindawati, Y. I., 2024. Hiperrealitas Pemain Game Online: Studi Pemain Genshin Impact di Kota Serang. *INNOVATIVE: Journal Of Social Science Research*, Volume 4, pp. 2938-2948. [Internet]. [Diakses pada 30 Juli 2025]. Tersedia pada <https://j-innovative.org/index.php/Innovative/article/view/12811>
- Kurniyawan, B., Aldino, A. A. & Isnain, A. R., 2022. Sentimen Analisis Terhadap Kebijakan Penyelenggara Sistem Elektronik (Pse) Menggunakan Algoritma Bidirectional Encoder Representations From Transformers (Bert). *Jurnal Teknologi dan Sistem Informasi (JTSI)*, 3(E-ISSN: 2746-369), pp. 98-106. [Internet]. [Diakses pada 30 Juli 2025]. Tersedia pada <https://jim.teknokrat.ac.id/index.php/sisteminformasi/article/view/2204>
- Kusnadi, R. et al., 2021. Analisis Sentimen Terhadap Game Genshin Impact Menggunakan Bert. *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, Volume 6, pp. 122-129. [Internet]. [Diakses pada 30 Juli 2025]. Tersedia pada https://www.researchgate.net/publication/353176928_ANALISIS_SENTIMEN_TERHADAP_GAME_GENSHIN_IMPACT_MENGUNAKAN_BERT

- Lenggo, G., Yulianti, E. & Sensuse, D. I., 2023. Sentiment Analysis of *Tweets* Before the 2024 Elections in Indonesia Using IndoBERT Language Models. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, 9(DOI: 10.26555/jiteki.v9i3.26490), pp. 746-757. [Internet]. [Diakses pada 10 April 2025]. Tersedia pada <https://journal.uad.ac.id/index.php/JITEKI/article/view/26490>
- Liu, M. & Chung, J., 2022. Research of popular success factors of game content <Genshin Impact>. *International Journal of Advanced Smart Convergence*, Volume 11, pp. 83-87. [Internet]. [Internet]. [Diakses pada 10 April 2025]. Tersedia pada <https://www.earticle.net/Article/A414550>
- Musfiroh, D., Khaira, U., Utomo, P. E. P. & Suratno, T., 2021. Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter *Dataset* Menggunakan InSet Lexicon. *MALCOM: Indonesian Journal of Machine learning and Computer Science*, Volume 1, pp. 24-33. [Internet]. [Internet]. [Diakses pada 10 April 2025]. Tersedia pada <https://www.journal.irpi.or.id/index.php/malcom/article/view/20>
- Normawati, D. & Prayogi, S. A., 2021. Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, Volume 5, pp. 697-711. [Internet]. [Internet]. [Diakses pada 10 April 2025]. Tersedia pada <https://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/view/369>
- Prastyo, P. H., Ardiyanto, I. & Hidayat, R., 2020. Indonesian Sentiment Analysis: An Experimental Study of Four Kernel Functions on SVM Algorithm with TF-IDF. *International Conference on Data Analytics for Business and Industry (ICDABI)*. [Internet]. [Internet]. [Diakses pada 10 April 2025]. Tersedia pada <https://ieeexplore.ieee.org/abstract/document/9325685>
- Putra, A. B. et al., 2024. Pengenalan Dasar-Dasar HTML dan CSS Pada Siswa/I Yayasan Al- Qaaf Tangerang Banten Dengan Menggunakan Bahasa Pemrograman Visual Studio Code Serta Mampu Membuat Website Sederhana. *APPA : Jurnal Pengabdian kepada Masyarakat*, 1(3025-0889), pp. 417-422. [Internet]. [Diakses pada 10 April 2025]. Tersedia pada <https://jurnalmahasiswa.com/index.php/appa/article/view/915>
- Putri, D. D., Nama, G. F. & Sulistino, W. E., 2022. Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (Dpr) Pada Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Informatika dan Teknik Elektro Terapan (JITET)*, Volume 10, pp. 34-40. [Internet]. [Diakses pada 10 April 2025]. Tersedia pada <https://journal.eng.unila.ac.id/index.php/jitet/article/view/2262>
- Raharjo, S., Wulandari & Dito, H. A., 2024. Pengaruh Promosi Media Sosial Dan Reputasi Perusahaan Terhadap Keputusan Pembelian Produk Virtual Genshin Impact. *Jurnal Arimbi (Applied Research In Management And Business)*, Volume 3, pp. 40-52. [Internet]. [Diakses pada 10 April 2025]. Tersedia pada <https://unkartur.ac.id/journal/index.php/arimbi/article/view/230>
- Ridwansyah, T., 2022. Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier. *KLIK: Kajian Ilmiah Informatika dan Komputer*, Volume 2, pp. 178-185. [Internet]. [Diakses pada 21 Juli 2025]. Tersedia pada <https://pdfs.semanticscholar.org/2090/abb19c9620e358e3efe0b86c9b6c974173d7.pdf>
- Ridwansyah, T., 2022. Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 2(ISSN 2723-3898 (Media Online)), pp. 178-185. [Internet]. [Diakses pada 21 April 2025]. Tersedia pada <https://pdfs.semanticscholar.org/2090/abb19c9620e358e3efe0b86c9b6c974173d7.pdf>
- Rina, Hidayat, T. & Saputri, D. U. E., 2024. Analisis Percepatan Pemulihan Ekonomi Indonesia Pasca Pandemi Dengan Big Data Dan Deep learning. *JATI (Jurnal Mahasiswa Teknik Informatika)*, Volume 8, pp. 3244-3252. [Internet]. [Diakses pada 21 April 2025]. Tersedia pada <https://ejournal.itn.ac.id/index.php/jati/article/view/9420>
- Romadhani, F., Mahbubah, N. A. & Kurniawan, M. D., 2021. Implementasi Metode Lean Six Sigma Guna Mengeliminasi Defect Proses Produksi Purified Gypsum Di PT AAA. *Jurnal Peradaban Sains Rekayasa Dan Teknologi*, Volume 9, pp. 89-103. [Internet]. [Diakses pada 21 April 2025]. Tersedia pada <https://ejournal.itn.ac.id/index.php/jati/article/view/9420>

- 2025]. Tersedia pada <https://www.researchgate.net/profile/Nina-Mahbubah/publication/357152383>
- Safitri, T., Umaidah, Y. & Iqbal, M., 2023. Analisis Sentimen Pengguna Twitter Terhadap BTS Menggunakan Algoritma *Support Vector Machine*. *Journal of Applied Informatics and Computing (JAIC)*, Volume 7, pp. 38-41. [Internet]. [Diakses pada 21 April 2025]. Tersedia pada <https://www.researchgate.net/profile/NinaMahbubah/publication/357152383>
- Sambada, F. P. & Ariatmi, S. Z., 2024. An Analysis Of Impoliteness Strategies Of Genshin Impact Player In 1st Anniversary Posted On Twitter. *Jurnal Onoma: Pendidikan, Bahasa dan Sastra*, Volume 10, pp. 322-332. [Internet]. [Diakses pada 21 April 2025]. Tersedia pada <https://e-journal.my.id/onoma/article/view/3220>
- Supriyanto, J., Alita, D. & Isnain, A., 2023. Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring. *Jurnal Informatika Dan Rekayasa Perangkat Lunak (JATIKA)*, Volume 4, pp. 74-80. [Internet]. [Diakses pada 21 April 2025] Tersedia pada <https://pdfs.semanticscholar.org/99cb/3a59bd48539d64b38494e03322d0fe9512c7.pdf>
- Vindua, R. & Zailaini, A. U., 2023. Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python. *JURIKOM (Jurnal Riset Komputer)*, Volume 10, pp. 479-487. [Internet]. [Diakses pada 21 April 2025]. Tersedia pada <https://pdfs.semanticscholar.org/8d92/ed855cae3b25d230b67f9dad9bala48b6e6c.pdf>
- Wang, Y., Guo, J., Yuan, C. & Li, B., 2022. Sentiment Analysis of Twitter Data. *MDPI*, pp. 1-14. [Internet]. [Diakses pada 21 April 2025]. Tersedia pada <https://dlwqtxts1xzle7.cloudfront.net/36584429/28>.
- Zahruputri, A. & Lusianus, K., 2021. Analisis Persepsi Pemain Terhadap Game Cross-Platform: Studi Kasus Permainan Genshin Impact. *Prosiding The 12th Industrial Research Workshop and National Seminar*, pp. 1273-1278. [Internet]. [Diakses pada 21 April 2025]. Tersedia pada <https://dlwqtxts1xzle7.cloudfront.net/36584429/28>.