

Klasifikasi Tweet Sarkasme Pada Platform X Menggunakan Bidirectional Encoder Representations from Transformers

Rizal Akbar Fitrianto^a, Arda Surya Editya^b

^aTeknik informatika, fakultas ilmu komputer, universitas nahdlatul ulama sidoarjo, rizal.akbar.fitrianto@gmail.com

^bTeknik informatika, fakultas ilmu komputer, universitas nahdlatul ulama sidoarjo, ardasurya.tif@unusida.ac.id

Submitted: 21-03-2024, Reviewed: 22-04-2024, Accepted 13-05-2024

<https://doi.org/10.47233/jteksis.v6i3.1344>

Abstract

X is a digital platform that facilitates the sharing of thoughts and criticisms through written content. Several individuals and organizations or companies utilize social media platforms like X to gauge public sentiment when making policies. However, due to the limitations of writing, which cannot convey nonverbal signals such as body movements, facial expressions, and intonations, it is often difficult to identify implicit signs like sarcasm. Therefore, it is important to detect sarcasm, as it can be used as evidence in legal matters and in the implementation of NLP to enhance sentiment analysis. This study examines the performance of deep learning methods such as IndoBERT, RoBERTa, BERT, and BERT Multilingual in classifying Indonesian sarcasm on platform X. The experimental results show that the IndoBERT model, which is specially designed for this task, achieved an impressive F1 score of 95%.

Keywords: sarcasm, deep learning, tweet.

Abstrak

X adalah platform digital yang memfasilitasi berbagi pemikiran dan kritik melalui konten tertulis. Sejumlah individu dan organisasi atau perusahaan memanfaatkan media sosial seperti X untuk mengetahui sentimen masyarakat umum saat akan mengambil kebijakan. Namun, karena keterbatasan menulis, yang tidak mampu mentransmisikan isyarat nonverbal seperti gerakan tubuh, ekspresi wajah, dan intonasi, seringkali sulit untuk mengidentifikasi tanda-tanda implisit seperti sarkas. Oleh karena itu, penting untuk melakukan deteksi sarkas, karena hal itu dapat digunakan sebagai alat bukti hukum dan Pada penerapan NLP dapat meningkatkan nilai sentimen. Studi ini melihat performa metode deep learning seperti, IndoBERT, RoBERTa, BERT, dan BERT Multilingual, dalam mengklasifikasi sarkas berbahasa Indonesia di platform X. Hasil eksperimen menunjukkan bahwa model IndoBERT, yang telah dirancang khusus untuk tugas tersebut, mendapatkan skor F1 yang mengesankan sebesar 95%.

Keywords: sarkasme, deep learning, tweet.

This work is licensed under Creative Commons Attribution License 4.0 CC-BY International license



PENDAHULUAN

Bersamaan dengan kemajuan teknologi informasi dan komunikasi, akses internet telah menjadi lebih mudah. Pada tahun 2017, jumlah pengguna internet di Indonesia mencapai sepertiga dari total populasi, sekitar 85,25 juta dari total 261,89 juta orang [1] [2]. Sementara itu, berbagai platform media sosial semakin berkembang, seperti Facebook, Instagram, Snapchat, dan X. Akun-akun di platform ini dapat mewakili individu, organisasi, atau produk.

Di antara berbagai platform media sosial tersebut, X, sebelumnya dikenal sebagai Twitter, adalah media sosial berbasis microblogging di mana pengguna dapat membaca dan menulis teks dengan batas maksimum 280 karakter, yang disebut dengan tweet. Pada kuartal ketiga tahun 2021, terdapat 288 juta pengguna aktif harian [3]. Pengguna juga dapat membuat utas untuk membagi pikiran atau pendapat menjadi bagian-bagian. Namun, komunikasi berbasis teks tidak memiliki isyarat non-verbal untuk disampaikan [4]. Hal-hal seperti gerakan

tubuh, ekspresi wajah, dan intonasi tidak tersedia dalam model komunikasi ini. Selain itu, karena penggunaan bahasa informal di X dan batasan jumlah karakter (280 karakter per tweet), hal-hal halus seperti sarkas menjadi sulit dideteksi.

Sarkas adalah bentuk bahasa yang sering dimaksudkan untuk menyampaikan niat yang bertentangan dengan apa yang sedang diungkapkan [5]. Perbedaan ini digunakan untuk menunjukkan ketidakpuasan terhadap proposisi sebelumnya, seringkali dalam bentuk penghinaan. Sarkas dapat ditujukan kepada individu atau kelompok dan memiliki berbagai tujuan, seperti mendukung analisis sentimen [6], yang dibagi menjadi tiga kategori: kecerdasan, keluhan, dan penghindaran [7].

Banyak penelitian tentang klasifikasi sindiran telah dilakukan menggunakan berbagai metode. Sebuah model yang berfokus pada konteks untuk mengidentifikasi sarkas dalam tweet telah diusulkan menggunakan algoritma Support Vector Machine (SVM) sebagai pendekatan klasifikasi berurutan.

Dengan memperhatikan bagaimana Twitter berfungsi, informasi kontekstual dibagi menjadi tiga kategori: konteks yang terkait dengan percakapan, sejarah, dan topik. Berdasarkan eksperimen menggunakan dataset yang dihasilkan secara otomatis, hasil menunjukkan bahwa konteks berbasis sejarah memberikan kinerja terbaik dibandingkan dengan konteks lainnya, bahkan jika tidak ada konteks yang digunakan sama sekali [10].

Mengidentifikasi sarkas berdasarkan hanya kata-kata dalam teks dapat membingungkan. Mukherjee & Bala [11] menyatakan bahwa ini memerlukan pemahaman tentang gaya bahasa yang digunakan oleh penulis. Oleh karena itu, mereka menganalisis berbagai set fitur yang relatif tidak berhubungan dengan teks itu sendiri: seperti kata fungsi dan n-gram dari kalimat, dalam penggunaan pengelompokan Fuzzy C-means dan Naïve Bayes sebagai alat klasifikasi untuk mendeteksi sarkas. Dari hasil tersebut, dapat disimpulkan bahwa fitur-fitur yang independen dari teks dapat meningkatkan akurasi dalam mendeteksi sarkas secara umum, dan kata-kata fungsi adalah salah satu fitur yang dapat meningkatkan performa. Namun, Naïve Bayes lebih unggul daripada pengelompokan C-means ketika digunakan pada dataset berisi 2000 tweet. Namun, ketika jumlah data lebih besar dari itu, Pengelompokan Fuzzy C-means umumnya memberikan hasil yang lebih baik.

Deteksi sarkasme juga diterapkan sebagai dukungan dalam analisis sentimen pada cuitan berbahasa Indonesia dengan menggunakan ekstraksi fitur unigram bersama dengan empat set fitur yang diambil dari Bouazizi [7], termasuk fitur terkait tanda baca, fitur sentimen, fitur sintaksis, dan fitur kunci. Pendekatan ini menghasilkan akurasi 80,4%, presisi 83,2%, dan recall 91,3%.

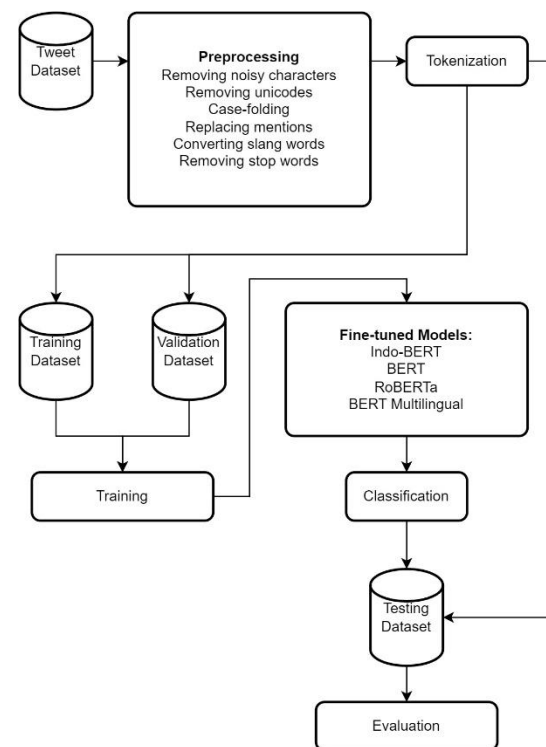
Mendeteksi sarkas adalah komponen penting dari klasifikasi teks dengan implikasi luas di berbagai sektor seperti keamanan, kesehatan, dan penjualan. Misalnya, dalam analisis umpan balik pelanggan, perusahaan dapat menilai sentimen pelanggan terhadap produk mereka. Wawasan ini sangat berharga untuk meningkatkan kualitas produk. Dalam analisis sentimen, klasifikasi sindiran sangat penting untuk menginterpretasikan pesan tersirat dalam tweet. Selain itu, struktur tweet itu sendiri kadang-kadang dapat menunjukkan sindiran, seperti ketika polaritas pernyataan dibalik (misalnya, mengubah pernyataan positif menjadi bentuknya yang berlawanan) [8].

Dalam penelitian ini, kami akan berfokus pada penggunaan empat model turunan Bidirectional Encoder Representations from Transformers (BERT) yaitu IndoBERT, BERT, RoBERTa, dan BERT Multilingual, dalam mengklasifikasikan sarkas dengan menggunakan dataset cuitan berbahasa Indonesia. Model-model ini disesuaikan

menggunakan metode pembelajaran terawasi, di mana mereka diberikan contoh-contoh yang telah dilabeli sebagai sarkasme dan bukan sarkasme. Selama proses pelatihan, model-model tersebut belajar untuk mengenali pola dan fitur-fitur yang membedakan antara dua jenis tweet tersebut.

Keterbaruan dari penelitian ini adalah pemanfaatan metode deep learning untuk deteksi sarkasme, dimana kalimat sarkasme itu memiliki bentuk kata positif namun memiliki arti negatif.

METODE PENELITIAN



Gambar 1. Alur penelitian

Seperti yang diilustrasikan dalam Gambar 1, fokus kami adalah untuk menentukan apakah sekelompok tweet tertentu mengandung sarkas. Untuk mencapai tujuan ini, kami menggunakan IndoBERT, BERT, RoBERTa, dan BERT Multilingual, dan memproses pra-pemrosesan tweet untuk mendapatkan dataset yang terstruktur dengan baik. Setelah proses tokenisasi data, kami melatih model yang telah dipretrain, melakukan fine-tuning pada model. Model-model yang sudah di-tuning kemudian dievaluasi menggunakan dataset uji dan dibandingkan dengan kinerja model pretrain asli.

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari Khotijah dkk. [19] dan terdiri dari 8700 tweet berbahasa Indonesia yang dikumpulkan dari Maret 2013 hingga Februari 2020. Dataset ini dibagi menjadi dua klasifikasi, yaitu tweet yang mengandung sindiran dan tweet yang tidak

mengandung sindiran. Penandaan tweet ini dilakukan secara manual oleh lima individu yang tidak memiliki informasi sebelumnya tentang asal atau penulis tweet. Ketika sebuah tweet tidak jelas, seorang ahli bahasa dihubungi untuk membantu dalam proses penandaan.

2.2. Pre-processing

Kami melakukan beberapa langkah untuk membersihkan dataset:

1. Menghapus karakter bising yang berasal dari emoji yang diurai menjadi encode UTF-8.
2. Menghilangkan karakter unicode, karena beberapa di antaranya mungkin ditafsirkan sebagai spasi putih, yang bisa mengakibatkan tokenisasi yang tidak tepat.
3. Case-folding, di mana seluruh teks diubah menjadi huruf kecil guna memastikan keseragaman dan mengurangi jumlah kata dan frasa yang unik.
4. Mengganti mention (contohnya, "@tvOneNews") dengan "@USER" dan URL (misalnya, "https://t.co/2G2azHJx3g") dengan "HTTPURL".
5. Mengonversi kata-kata slang dari bahasa informal yang umumnya digunakan dalam percakapan santai ke bahasa yang lebih formal atau standar.
6. Menghapus stop words, yaitu kata-kata umum yang memiliki makna minimal untuk menyederhanakan data dan mengurangi dimensinya. Proses ini dilakukan menggunakan pustaka NLTK dan Sastrawi.

2.3. Tokenisasi

Tokenisasi adalah proses mengubah teks menjadi potongan-potongan kecil yang disebut "token". Token bisa berupa kata, frasa, atau karakter, tergantung pada jenis tokenisasi yang digunakan. Tujuan utama tokenisasi adalah untuk memecah teks menjadi bagian-bagian yang lebih kecil sehingga lebih mudah untuk diproses oleh model bahasa atau algoritma lainnya. Prosedur tokenisasi pada BERT, RoBERTa, BERT Multilingual, dan IndoBERT memanfaatkan menggunakan WordsPiece dengan metode enkripsi byte-pair (BPE) yang diperkenalkan oleh [22].

Sebagai contoh perbedaan dalam proses tokenisasi, kita bisa menggunakan kalimat: "proses digitalisasi dapat dinormalisasi agar proses administrasi berjalan dengan optimum". Rincian perbedaan hasil tokenisasi dapat dilihat dalam Tabel 1.

Tabel 1. Perbandingan tokenisasi

Model	Hasil Tokenisasi
IndoBERT	['proses', 'digital', 'isasi', 'dapat', 'di', 'normal', 'isa', 'si', 'agar', 'proses', 'administrasi', 'berjalan', 'dengan', 'optimum']
BERT	['proses', 'digital', 'isasi', 'dapat', 'di', 'normal', 'isasi', 'agar', 'proses', 'administrasi', 'berjalan', 'dengan', 'optimum']
RoBERTa	['proses', 'digital', 'isasi', 'dapat', 'di', 'normal', 'isasi', 'agar', 'proses', 'administrasi', 'berjalan', 'dengan', 'optimum']
BERT Multilingual	['proses', 'digitalisasi', 'dapat', 'di', 'normal', 'isasi', 'agar', 'proses', 'administrasi', 'berjalan', 'dengan', 'optimum']

2.4. Melatih Model

Untuk menemukan hyperparameter optimum pada setiap model, kami menyetel kembali IndoBERT, BERT base, RoBERTa, dan BERT Multilingual sebagai berikut:

1. Menghapus karakter bising yang berasal dari emoji yang diurai menjadi encode UTF-8.
2. Loss (CategoricalCrossentropy): Digunakan untuk menyelesaikan permasalahan klasifikasi biner dengan nilai target yang berada dalam set (0, 1).
3. Validation_split: Digunakan untuk membagi data uji dan data latih dengan proporsi 30% untuk data uji dan 70% untuk data latih. Kinerja model diukur dengan menghitung akurasi. Validasi split memiliki dua konsep, yaitu kesalahan pelatihan (training error) yang dihitung dengan kesalahan klasifikasi model pada data pelatihan, dan kesalahan uji (test error) yang menggunakan dua set data yang benar-benar terpisah: satu untuk pelatihan dan yang lainnya untuk pengujian.
4. Ukuran Batch (Batch_size = 32): Parameter yang menentukan jumlah sampel data yang akan diproses dalam setiap iterasi. Pada kasus ini, 32 sampel data akan diproses dalam satu batch sebelum model diperbarui.
5. Epochs = 5: Penggunaan 5 epochs dipilih karena mampu menghasilkan tingkat akurasi yang baik.

2.5. Evaluasi

Dalam studi ini, kami menilai efektivitas beberapa model deep learning untuk mengklasifikasikan sarkasme dalam tweet berbahasa Indonesia. Spesifiknya, kami membandingkan bagaimana empat tipe model deep learning (IndoBERT, BERT, RoBERTa, dan BERT Multilingual) berkinerja. Dataset yang digunakan untuk pelatihan dan validasi dibagi dalam rasio 7:3 dan terdiri dari total 8700 tweet berbahasa

Indonesia. Dataset pelatihan terdiri dari 6088 tweet, sedangkan dataset validasi berisi 1748 tweet. Sedangkan untuk validasi berisi 890 tweet. Dalam pengujian, kami menggunakan dataset terpisah yang berjumlah 120 tweet. Dataset pengujian terdiri dari 64 tweet yang bukan sarkas dan 56 cuitan yang bersifat sarkas.

HASIL DAN PEMBAHASAN

Pelatihan tidak dilakukan dari awal, tetapi dengan menggunakan model yang telah dilatih sebelumnya dan hanya mempelajari sedikit lagi untuk mencapai titik optimal pada tugas baru. Teknik pelatihan ini disebut fine-tuning. Kami melakukan proses fine-tuning dengan menyesuaikan hiperparameter yang digunakan dalam model dalam proyek ini seperti yang diilustrasikan dalam Tabel 2 di bawah ini, yaitu ukuran batch 32, laju pembelajaran (Adam) 3×10^{-6} , dan epoch 5.

Tabel 2. Hyperparameter setup

Model	Learning Rate	Training Batch Size	Epochs
IndoBERT	3×10^{-6}	32	32
BERT	3×10^{-6}	32	32
RoBERTa	3×10^{-6}	32	32
BERT Multilingual	3×10^{-6}	32	32

Perbedaan performa antara model-model yang berbeda diukur menggunakan Matriks Konfusi. Matriks ini memberikan gambaran hasil prediksi dalam masalah klasifikasi. Fungsinya adalah untuk menilai keandalan pendekatan yang diusulkan dan efektivitas metode dalam mengenali sarkas. Rumus-rumus untuk matriks konfusi dijelaskan dalam Persamaan 1, Persamaan 2, Persamaan 3, dan Persamaan 4, di mana TP adalah true positive, TN adalah true negative, FP adalah false positive, dan FN adalah false negative.

$$Accuracy = \frac{TP}{TN+FP+FN+TP} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\ Score = \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Kinerja model NLP pada dataset validasi dan pengujian menunjukkan dampak berbagai faktor terhadap kinerja model, termasuk algoritma tokenisasi, strategi pretraining, dan adaptasi domain

husus. Hasilnya ditampilkan pada Tabel 3 dan Tabel 4.

Tabel 3. Performa model terhadap validasi dataset

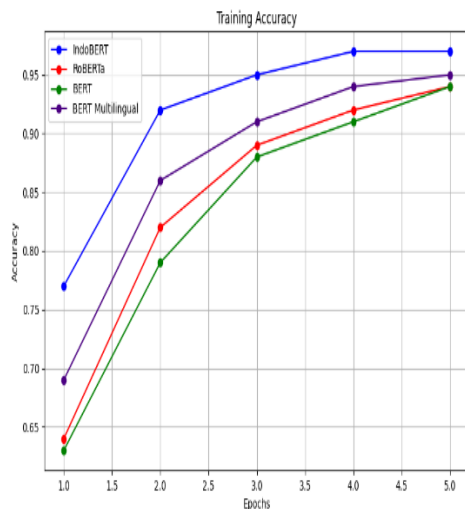
Model	Accuracy	Precision	Recall	F1 Score
IndoBERT	95%	95%	95%	95%
BERT	90%	92%	92%	92%
RoBERTa	93%	93%	93%	93%
BERT Multilingual	94%	94%	94%	94%

Tabel 4. Performa model terhadap testing dataset

Model	Accuracy	Precision	Recall	F1 Score
IndoBERT	95%	95%	95%	95%
BERT	91%	92%	92%	92%
RoBERTa	94%	94%	94%	94%
BERT Multilingual	92%	92%	93%	92%

Model IndoBERT menunjukkan kinerja yang kuat secara konsisten di kedua dataset, dengan mempertahankan akurasi dan skor F1 sebesar 95%. Hal ini dapat dikaitkan dengan model yang telah di pre-train menggunakan dataset yang lebih besar dan menggunakan tiga bahasa yaitu Indonesia, Sunda, dan Jawa yang akhirnya turut membantu model menangkap fitur linguistik penting saat disesuaikan kembali pada dataset. Penggunaan SentencePiece dengan algoritma tokenisasi BPE juga mungkin berkontribusi pada kinerjanya. Skor F1 model pada dataset pengujian juga menunjukkan kinerja yang kuat, menunjukkan kemampuannya untuk menggeneralisasi ke dataset yang berbeda dan keahliannya dalam mengidentifikasi tweet sindiran. Sebaliknya, baik RoBERTa maupun BERT Multilingual menunjukkan kinerja yang mirip dalam akurasi dan skor F1 di kedua dataset validasi dan pengujian.

Kinerja model NLP pada dataset validasi dan pengujian menyoroti pengaruh berbagai faktor seperti algoritma tokenisasi, strategi pemutasi pra, dan adaptasi khusus domain. Model BERT secara konsisten memberikan kinerja yang baik pada kedua dataset, mencapai akurasi di atas 90% dan skor F1 sebesar 92%. Namun, model ini kalah dari model deep learning berbasis transformer yang lebih canggih seperti IndoBERT, RoBERTa, dan BERT Multilingual.



Gambar 2. Perbandingan akurasi kurva belajar

Dari perbandingan kurva pembelajaran akurasi pada dataset pelatihan yang ditunjukkan dalam Gambar 2, dapat dilihat bahwa IndoBERT memiliki kurva pembelajaran yang lebih unggul. Hal ini mungkin disebabkan oleh fakta bahwa model dasar IndoBERT telah dilatih menggunakan basis data bahasa Indonesia yang sangat besar, yang menyederhanakan proses pelatihan dan memungkinkannya mencapai tingkat akurasi yang tinggi.

SIMPULAN

Hasil kami menunjukkan bahwa model deep learning yang telah disesuaikan dalam arsitektur memberikan kinerja yang lebih baik daripada model BERT dasar. Model IndoBERT yang disesuaikan kembali mencapai kinerja tertinggi dan sangat konsisten dengan akurasi 95%. Oleh karena itu, temuan dari penelitian ini menunjukkan bahwa model deep learning adalah pendekatan yang menjanjikan untuk mengklasifikasikan sindiran dalam data X berbahasa Indonesia.

Selain itu, penelitian lebih lanjut dapat dilakukan untuk mengeksplorasi potensi mereka dalam dataset lain, dan model yang dikembangkan dapat diuji pada tugas klasifikasi lainnya selain pengenalan sindiran. Penggunaan model dalam tugas-tugas seperti klasifikasi topik, deteksi spam, atau tugas NLP lainnya dapat memberikan wawasan tambahan tentang kegunaan dan fleksibilitas model tersebut.

UCAPAN TERIMA KASIH

Terima kasih yang sebesar-besarnya saya ucapkan kepada kedua orang tua yang selalu mendukung saya selama perkuliahan. Terima kasih juga saya ucapkan kepada bapak Arda Surya

Editya selaku dosen pembimbing yang selalu membimbing saya dalam mengerjakan penelitian ini. Terakhir, terima kasih saya ucapkan kepada rekan-rekan sesama mahasiswa yang juga membantu menyemangati saya dalam mengerjakan penelitian.

DAFTAR PUSTAKA

Penulisan naskah dan sitasi yang diacu dalam naskah ini disarankan menggunakan aplikasi referensi (*reference manager*) seperti Mendeley, Zotero, Reffwork, Endnote dan lain-lain. Artikel wajib merujuk 10-20 jurnal dan tiga artikel pada Jurnal JTEKSIS [Times New Roman, 8, normal], spasi 1. (rujuk pada pedoman penulisan).

- [1] Ritchie, H., Mathieu, E., Roser, M., & Ortiz-Ospina, E. (n.d.). Internet. Our World in Data.
- [2] Hakiki, G., & Ulfa, A. 2018. Perempuan dan laki-laki di Indonesia 2017.
- [3] X. 2022. Q4 and fiscal year 2021: Letter to shareholders.
- [4] Hancock, J. T., Landrigan, C., & Silver, C. 2007. Expressing emotion in text-based communication. *ACM*, 929–932. doi:10.1145/1240624.1240764
- [5] Filik, R., Turcan, A. T., Thompson, D., Harvey, N., Davies, H., & Turner, A. 2016. Sarcasm and emoticons: Comprehension and emotional impact. *Quarterly Journal of Experimental Psychology*, 69, 2130–2146. doi:10.1080/17470218.2015.1106566
- [6] Yunitasari, Y., Musdholifah, A., & Sari, A. K. 2019. Sarcasm detection for sentiment analysis in Indonesian tweets. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13, 53. doi:10.22146/ijccs.41136
- [7] Bouazizi, M., & Ohtsuki, T. 2017. A pattern-based approach for multi-class sentiment analysis in Twitter. *IEEE Access*, 5, 20617–20639. doi:10.1109/ACCESS.2017.2740982
- [8] Sarsam, S. M., Al-Samarraie, H., Alzahrani, A. I., & Wright, B. 2020. Sarcasm detection using machine learning algorithms in Twitter: A systematic review. *International Journal of Market Research*, 62, 578–598. doi:10.1177/1470785320921779
- [9] Parmar, K., Limbasiya, N., & Dhamecha, M. 2018. Feature based composite approach for sarcasm detection using mapreduce. *IEEE*, 587–591. doi:10.1109/ICCMC.2018.8488096
- [10] Wang, Z., Wu, Z., & Wang, R. 2015. Twitter Sarcasm Detection Exploiting a Context-Based Model, 77–91. doi:10.1007/978-3-319-26190-46
- [11] Mukherjee, S., & Bala, P. K. 2017. Sarcasm detection in microblogs using naïve bayes and fuzzy clustering. *Technology in Society*, 48, 19–27. doi:10.1016/j.techsoc.2016.10.003
- [12] Ren, Y., Ji, D., & Ren, H. 2018. Context-augmented convolutional neural networks for Twitter sarcasm detection. *Neurocomputing*, 308, 1–7. doi:10.1016/j.neucom.2018.03.047
- [13] Barbieri, F., & Saggion, H. 2014. Modelling irony in Twitter. *Association for Computational Linguistics*, 56–64. doi:10.3115/v1/E14-3007
- [14] Rahayu, D. A. P., Kuntur, S., & Hayatin, N. 2018. Sarcasm detection on Indonesian Twitter feeds. *IEEE*, 137–141. doi:10.1109/EECSL.2018.8752913
- [15] Son, L. H., Kumar, A., & Sangwan, S. R. 2019. Sarcasm detection using soft attention-based bidirectional long short-term memory model with convolution network. *IEEE Access*, 7, 23319–23328. doi:10.1109/ACCESS.2019.2899260
- [16] Ghosh, A., Li, G., Veale, T., Rosso, P., Shutova, E., Barnden, J., & Reyes, A. 2015. Semeval-2015 task 11: Sentiment

- analysis of figurative language in twitter. Association for Computational Linguistics, 470–478. doi:10.18653/v1/S15-2080
- [17] Majumder, N., Poria, S., Peng, H., Chhaya, N., Cambria, E., & Gelbukh, A. 2019. Sentiment and sarcasm classification with multitask learning. IEEE Intelligent Systems, 34, 38–43. doi:10.1109/MIS.2019.2904691
 - [18] Mishra, A., & Bhattacharyya, P. 2018. Predicting Readers' Sarcasm Understandability by Modeling Gaze Behavior, 99–115. doi:10.1007/978-981-13-1516-95
 - [19] Wijaya, W., Murwantara, I. M., & Mitra, A. R. 2020. A simplified method to identify the sarcastic elements of bahasa indonesia in youtube comments. IEEE, 1–6. doi:10.1109/ICoICT49345.
 - [20] Khotijah, S., Tirtawangsa, J., & Suryani, A. A. 2020. Using LSTM for context-based approach of sarcasm detection in Twitter. ACM, 1–7. doi:10.1145/3406601.3406624.
 - [21] Handoyo, A. T., Rahman, H., Setiadi, C. J., & Suhartono, D. 2022. Sarcasm detection in Twitter - performance impact while using data augmentation: Word embeddings. International Journal of Fuzzy Logic and Intelligent Systems, 22(4), 401–413. doi:10.5391/IJFIS.2022.22.4.401.
 - [22] Wu, et. all. Google's Neural Machine Translation System: Bridging The Gap Between Human and Machine Translation.
 - [23] Sudirjo, F., Diantoro, K., Al-Gasawneh, J. A., Azzaakiyyah, H. K., & Ausat, A. M. A. 2023. Application of ChatGPT in improving customer sentiment analysis for businesses. Jurnal Teknologi Dan Sistem Informasi Bisnis, 5(3), 283-288.
 - [24] Guntara, R. G. 2023. Pemanfaatan google colab untuk aplikasi pendeteksian masker wajah menggunakan Algoritma Deep Learning YOLOv7. Jurnal Teknologi Dan Sistem Informasi Bisnis, 5(1), 55-60.
 - [25] Pamungkas, M. R. S. P., Asyivadibrata, A., Susilawati, T., & Huda, M. N. 2023. Unleashing the Potentials of Artificial Intelligence for Micro, Small, and Medium Enterprises: A Systematic Literature Review. Jurnal Teknologi Dan Sistem Informasi Bisnis, 5(3), 303-310.