



# Sarcasm Detection in News Headline Dataset with Ensemble Deep Learning Method

## Deteksi Sarkasme Pada Dataset News Headline Dengan Metode Ensemble Deep Learning

Mochammad Alfian Rosid<sup>1\*</sup>, Siti Nur Haliza<sup>2</sup>, Yulian Findawati<sup>3</sup>, Uce Indahyanti<sup>4</sup>

<sup>1)2)3)4)</sup>Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Sidoarjo, Indonesia

Email to Correspondence: <sup>1\*</sup>alfanrosid@umsida.ac.id, <sup>2</sup>snhalisa@umsida.ac.id, <sup>3</sup>yulianfindawati@umsida.ac.id, <sup>4</sup>uceindahyanti@umsida.ac.id

**Abstract.** Sarcasm, a prevalent linguistic device, is frequently used in public discourse, often causing offence and distress to the listener. The complexity inherent in detecting sarcasm is a significant and ongoing challenge in the field of sentiment analysis research. The widespread use of this phenomenon in diverse conversational contexts further complicates its identification in data sets full of human interactions. Deficiencies in methodologies for distinguishing such statements adversely affect the performance of sentiment analysis, especially in distinguishing negative, positive or neutral sentiments. Inaccuracies in sarcasm detection can affect the classification results of sentiment analysis. Therefore, sentiment analysis seeks to categorise sarcastic sentences that, despite appearing positive, actually contain negative meanings. This research aims to build a deep learning ensemble stack model. The basic deep learning methods used are Bidirectional Gated Recurrent Unit (BiGRU) and Convolutional Neural Network (CNN). LightGBM is used to perform stack ensemble of deep learning methods. The dataset used comes from the Kaggle website and consists of English headlines. The findings show that the stack ensemble method outperforms BiGRU and CNN, evidenced by an accuracy rate of 91.2% and an F1 score of 90.2%. Therefore, from the above discussion, it can be concluded that the LightGBM method emerges as the optimal solution for sarcasm detection.

**Keywords -** Sentiment analysis, BiGRU, CNN, LightGBM, Detection Sarcasm

**Abstrak.** Sarkasme, sebuah perangkat linguistik yang lazim, sering digunakan dalam wacana masyarakat, yang sering menimbulkan perasaan tersinggung dan tertekan pada pendengarnya. Kompleksitas yang melekat dalam mendeteksi sarkasme merupakan tantangan yang signifikan dan berkelanjutan dalam bidang penelitian analisis sentimen. Penggunaan fenomena ini secara luas dalam konteks percakapan yang beragam semakin memperumit identifikasinya dalam kumpulan data yang penuh dengan interaksi manusia. Kekurangan dalam metodologi untuk membedakan pernyataan semacam itu berdampak buruk pada kinerja analisis sentimen, terutama dalam membedakan sentimen negatif, positif, atau netral. Ketidakakuratan dalam deteksi sarkasme dapat mempengaruhi hasil klasifikasi analisis sentimen. Oleh karena itu, analisis sentimen berusaha untuk mengkategorikan kalimat sarkastik yang, meskipun terlihat positif, sebenarnya mengandung makna negatif. Penelitian ini bertujuan untuk membangun model stack ensemble deep learning. Metode deep learning dasar yang digunakan adalah Bidirectional Gated Recurrent Unit (BiGRU) dan Convolutional Neural Network (CNN). LightGBM digunakan untuk melakukan stack ensemble metode deep learning. Dataset yang digunakan berasal dari situs web Kaggle dan terdiri dari berita utama berbahasa Inggris. Temuan menunjukkan bahwa metode stack ensemble mengungguli BiGRU dan CNN, dibuktikan dengan tingkat akurasi sebesar 91.2% dan skor F1 sebesar 90.2%. Oleh karena itu, dari pembahasan di atas, dapat disimpulkan bahwa metode LightGBM muncul sebagai solusi optimal untuk deteksi sarkasme.

**Kata kunci -** Analisis Sentimen, BiGRU, CNN, LightGBM, Deteksi Sarkasme

## PENDAHULUAN

Bahasa dibangun melalui serangkaian peraturan, pedoman, atau pola yang dibuat untuk mengatur sistem fonologi, struktur morfologi, dan susunan sintaksis. Penyimpangan dari peraturan, pedoman, atau pola ini berpotensi mengganggu komunikasi yang efektif [1]. Gaya bahasa memainkan peran penting dalam mengartikan maksud seseorang. Bahasa dapat berbentuk kata-kata kasar yang mengandung makna tersirat dan ejekan, yang dapat menimbulkan rasa sakit secara emosional atau diartikan sebagai sarkasme. Sarkasme

dapat didefinisikan sebagai "ekspresi atau ucapan yang tajam, pedas, atau menggigit; cemoohan atau ejekan yang pedas" [2]. Memeriksa pernyataan sarkastik memerlukan tugas yang kompleks. Sarkasme sering kali disampaikan oleh individu ketika menyampaikan informasi melalui media sosial.

Deteksi sarkasme sering digunakan dalam analisis sentimen untuk membedakan sentimen negatif, positif, atau netral. Sarkasme dapat muncul dalam ekspresi wajah, gerak tubuh, dan bahkan konten tekstual [3]. Kegagalan dalam mengidentifikasi sarkasme dapat berdampak pada hasil analisis sentimen dalam mengklasifikasikan dataset. Tantangan yang terkait

dengan sarkasme dan keuntungan dari deteksi sarkasme untuk analisis sentimen telah mendorong minat penelitian dalam deteksi sarkasme otomatis [4]. Berdasarkan hasil penelitian Alita dan Rahman [5]. Peningkatan yang signifikan dalam akurasi analisis sentimen sebesar 16,61% dicapai dengan mendeteksi sarkasme pada tweet yang berkaitan dengan layanan publik. Dalam penelitian yang dilakukan oleh Saifullah dkk [6] yang melakukan deteksi sarkasme pada platform media sosial Twitter, terungkap bahwa sebanyak 17% pengguna media sosial Instagram menggunakan kalimat sarkasme, sehingga mengindikasikan bahwa penggunaan kalimat semacam itu masih lazim digunakan di platform tersebut.

Dalam penelitian-penelitian yang disebutkan di atas, proses klasifikasi menggunakan metode pembelajaran mesin dan pengklasifikasi dasar pembelajaran mendalam. Belum ada penelitian yang menggunakan pendekatan deep learning stack ensemble untuk mengidentifikasi sarkasme. Oleh karena itu, ada potensi untuk kemajuan di bidang deteksi sarkasme dengan memanfaatkan teknik stack ensemble deep learning.

Penelitian ini bertujuan untuk membangun model pembelajaran mendalam stack ensemble. Teknik deep learning fundamental yang digunakan dalam penelitian ini meliputi Bidirectional Gated Recurrent Unit (BiGRU) dan Convolutional Neural Network (CNN). Metode pembelajaran dalam stack ensemble deep learning dijalankan menggunakan LightGBM. Setelah tahap pengujian selesai, diharapkan pendekatan deep learning dengan performa terbaik akan teridentifikasi, sehingga dapat menjadi tolok ukur untuk membangun model yang dirancang untuk mendeteksi sarkasme dalam berita utama di platform media sosial.

## LITERATUR REVIEW

Penelitian ini mengkaji proses pengumpulan informasi dari penelitian-penelitian serupa. Kekuatan dan kelemahan dari penelitian serupa akan diselidiki melalui analisis temuan-temuan sebelumnya. Dalam beberapa tahun terakhir, telah banyak penelitian ilmiah yang dilakukan untuk mendeteksi sarkasme. [5] melakukan penelitian yang melibatkan pemanfaatan empat set fitur dalam tahap ekstraksi fitur. Pendekatan serupa juga dilakukan oleh [7], yang menggabungkan fitur-fitur yang berkaitan dengan sentimen, tanda baca, leksikal dan sintaksis, serta pola kalimat. Hasilnya kemudian diklasifikasikan menggunakan metode Random Forest Classifier. Terlepas dari upaya-upaya tersebut, hasil pengujian masih menunjukkan tingkat akurasi yang relatif rendah. Eksplorasi lebih lanjut mengenai identifikasi sarkasme dalam konteks bahasa Indonesia dilakukan oleh [8] Penelitian ini menghasilkan skor F1 sebesar 90% berdasarkan analisis terhadap 1138 komentar dari dataset yang diuji.

Investigasi lebih lanjut yang dilakukan oleh para peneliti [9] telah menunjukkan bahwa prosedur analisis sentimen mencakup beberapa tahap, seperti preprocessing data, pemilihan fitur menggunakan metode Query Expansion Ranking, dan klasifikasi data melalui penerapan algoritma Naive Bayes. Hasil dari proses seleksi fitur, yang memberikan hasil 75%, menunjukkan akurasi yang paling luar biasa, yaitu 86,6%. Selain itu, penelitian selanjutnya yang dilakukan oleh [10] telah menggabungkan dua metodologi yang berbeda untuk

deteksi sarkasme, sehingga meningkatkan ketepatan identifikasi sarkasme situasional.

Penelitian yang dilakukan oleh [11] dapat dikembangkan lebih lanjut dengan memasukkan dataset pelatihan yang lebih besar dan meningkatkan jumlah data. Selain itu, kemampuan klasifikasi juga dapat ditingkatkan dengan memasukkan fitur kelas yang dapat secara efektif membedakan antara kasus cyberbullying dan noncyberbullying. Penambahan ini diharapkan dapat menghasilkan peningkatan dalam kinerja model dan memperluas penerapan deteksi sentimen dalam berbagai domain dan konteks. Metode LSTM mencapai akurasi 93,77%, sedangkan metode BiLSTM mencapai akurasi 95,24%. [12] melakukan penelitian yang berfokus pada vektor kata, secara khusus mengeksplorasi tiga model penyisipan kata pra-pelatihan: GloVe, fastText, dan BERT. Hasil penelitian menunjukkan bahwa kombinasi penyisipan fastText dan BiGRU sebagai pengklasifikasi menghasilkan kinerja tertinggi, dengan akurasi 93,85%.

### A. Bidirectional Gated Recurrent Unit

Bidirectional GRU atau BiGRU memiliki struktur lebih sederhana dibandingkan dengan LSTM. Jadi, pelatihannya lebih mudah. Semua sumber daya yang mengalir melalui model jaringan ini secara bersamaan digunakan oleh dua lapisan GRU. BiGRU memiliki keunggulan ketergantungan kecil pada vektor kata, kompleksitas rendah, dan waktu respons yang cepat. Keluaran dari dua arah BiGRU yaitu menghasilkan etimasi kemungkinan sarkasme [13]. BiGRU adalah model pemrosesan urutan yang terdiri dari dua GRU yang searah dan berlawanan arah, digabungkan untuk membentuk model jaringan saraf [10].

Model GRU mencakup gerbang reset dan gerbang pembaruan. Gerbang reset menentukan seberapa banyak informasi yang digunakan. Rumus perhitungannya sebagai berikut :

$$R_t = \sigma(W_r \cdot [h_{t-1}, X_t]) \quad (1)$$

Dimana  $R_t$  sebagai gerbang reset dan  $W_r$  adalah bobot matriks.  $X_t$  sebagai input pada waktu  $t$  dan  $h_{t-1}$  adalah output pada saat sebelumnya. Pada rumus ini, dihitung untuk menentukan informasi yang perlu dibuang dan informasi yang dipertahankan.

Gerbang pembaruan menentukan kapan harus memperbarui keadaan sel. Rumus perhitungannya sebagai berikut :

$$Z_t = \sigma(W_z \cdot [h_{t-1}, X_t]) \quad (2)$$

Dimana  $Z_t$  sebagai gerbang pembaruan dan  $W_z$  adalah bobot matriks.  $X_t$  sebagai input pada waktu  $t$  dan  $h_{t-1}$  adalah output pada saat sebelumnya. Hasil apakah keadaan saat ini adalah keadaan yang diperbarui atau keadaan sebelumnya dapat ditentukan oleh gerbang pembaruan. Saat gerbang pembaruan adalah 1, status berubah, dan saat gerbang pembaruan adalah 0, status saat ini dipertahankan dan transmisi berlanjut. Rumus GRU adalah sebagai berikut :

$$\tilde{h}_t = \tanh(W^h X_t + U^h (h_{t-1} * r_t)) \quad (3)$$

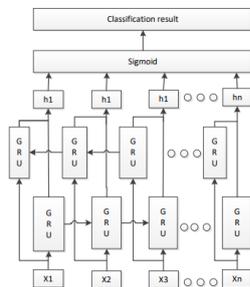
$$h_t = (1 - z_t) * \tilde{h}_t + z_t * h_{t-1}$$

Model BiGRU menggabungkan dua GRU searah. Setiap saat, ada input ke dua GRU yang berlawanan arah pada saat bersamaan, dan output ditentukan

bersama untuk membuat hasilnya lebih akurat. Output dari BiGRU dapat digambarkan sebagai berikut :

$$H = h_t \oplus \overleftarrow{h_t} \quad (4)$$

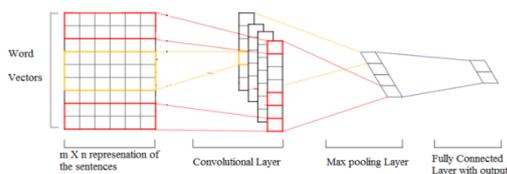
Di mana H mewakili output BiGRU, dan mewakili dua GRU searah, dan merupakan elemen tambahan [13]. Struktur Jaringan BiGRU diperlihatkan oleh Gambar 1.



**Gambar 1.** Struktur Jaringan BiGRU  
Sumber : [13]

### B. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) telah memainkan peran penting dalam perancangan model prediksi selama beberapa tahun terakhir. CNN terdiri dari dua lapisan yakni konvolusi dan pooling. Dimana lapisan konvolusi dibentuk oleh tumpukan berbagai lapisan independen yang mengubah volume input menjadi volume output melalui fungsi yang berbeda. Model CNN membantu menangkap fitur spesial dari input yang diberikan dalam bentuk matriks. Lapisan pooling mengambil setiap output setiap fitur dari lapisan konvolusi dan menyiapkan peta fitur yang diringkas. Lapisan Konvolusi dibentuk oleh tumpukan berbagai lapisan independen yang mengubah volume input menjadi volume output melalui fungsi yang berbeda. Model CNN membantu menangkap fitur spasial dari input yang diberikan dalam bentuk matriks. Bobot disusun sebagai jendela dengan bentuk  $m \times n$ . di mana "m" dan "n" adalah hyperparameter yang ditentukan pengguna. Matriks bobot  $m \times n$  ini, disebut filter, dijalankan pada matriks input untuk mendapatkan matriks lain (C1) dengan baris  $xr - m + 1$  dan kolom  $xc - n + 1$ . Matriks baru berada di bawah operasi gabungan (dijelaskan di bagian selanjutnya) yang memilih neuron aktif maksimal dari matriks bentuk yang ditentukan pengguna. Jendela baru ini bergerak satu langkah di atas matriks C1 untuk membentuk matriks lain, misalkan P1. C1 dan P1 membentuk lapisan konvolusional. Kemudian matriks P1 disebar ke lapisan yang terhubung sepenuhnya [3]. Arsitektur CNN diperlihatkan oleh Gambar 2.



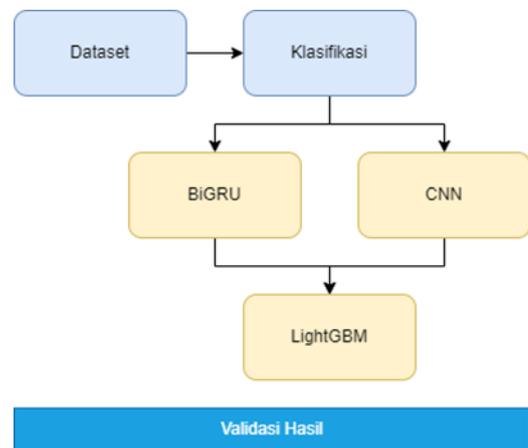
**Gambar 2.** Arsitektur Convolutional Neural Network  
Sumber : [14]

### C. LightGBM

LightGBM, algoritma pembelajaran ansambel yang dirancang oleh Microsoft, mewakili implementasi kerangka peningkatan gradien yang sangat efisien. Atribut utama algoritme ini adalah kemampuannya untuk mempercepat proses pelatihan secara signifikan, seringkali menghasilkan model yang lebih mahir dalam jangka waktu yang lebih singkat. LightGBM beroperasi berdasarkan prinsip algoritme pohon keputusan, menggabungkan jumlah nestimator yang ditingkatkan yang berasal dari struktur pohon. Ini menunjukkan kinerja unggul dalam analisis prediktif, melampaui algoritma lain dalam domain ini. Penerapan algoritme LightGBM dalam berbagai studi klasifikasi dan regresi secara konsisten menghasilkan hasil deteksi yang patut dicontoh, yang menegaskan kemajuannya sebagai algoritme klasifikasi. Model yang diusulkan menggunakan LightGBM menawarkan mekanisme pendukung keputusan yang dioptimalkan untuk pengguna akhir. Pendekatannya berbeda-beda, model ini melibatkan prosedur khusus untuk menghitung jumlah pohon keputusan yang optimal, menentukan kedalaman maksimum pohon-pohon tersebut, dan memastikan jumlah daun pohon yang diperlukan untuk membangun model LightGBM yang efektif [4].

## METODE PENELITIAN

Tahapan penelitian diawali dari pengumpulan data, kemudian dilakukan pemodelan klasifikasi menggunakan BiGRU dan CNN, kemudian dilakukan prediksi akhir dengan LightGBM. Arsitektur sistem untuk stack ensemble deep learning ditunjukkan oleh Gambar 3.



**Gambar 3.** Arsitektur Sistem untuk Ensemble Deep Learning

### A. Dataset

Dataset ini bersumber dari Kaggle, sebuah situs web terkemuka. Data yang dikumpulkan berkaitan dengan berita utama. Kumpulan data berita utama menunjukkan jumlah kesalahan tata bahasa yang sangat rendah dan terdiri dari kalimat-kalimat yang berbeda. Selain itu, dataset ini memiliki kualitas yang sangat baik. Untuk memfasilitasi pemrosesan data, Natural Language Toolkit (NLTK) digunakan untuk tokenisasi, lemmatisasi, dan pra-pemrosesan. Kumpulan berita utama mencakup total 26.709 contoh.

## B. Proses Klasifikasi Deep Learning

Proses klasifikasi dimulai dengan memastikan hasil yang diperoleh dari Metode Unit Berulang Dua Arah dan Jaringan Syaraf Tiruan. Hasil klasifikasi dari kedua metodologi tersebut kemudian digabungkan dengan menggunakan algoritma LightGBM. Penelitian ini mendalilkan metode BiGRU sebagai alat yang sangat baik untuk mengekstraksi fitur-fitur tekstual. Parameter seperti Learning Rate, Batch Size, dan Epoch dinilai dalam kerangka kerja investigasi ini. Proses ekstraksi fitur dijalankan dengan metode BiGRU, yang melakukan ekstraksi kata dengan cara dua arah, yang mencakup arah maju dan mundur. Lapisan maju mengambil kata-kata dari awal hingga akhir, sedangkan lapisan mundur mengekstrak kata-kata dari akhir hingga awal. Algoritma CNN menawarkan keunggulan dalam hal kecepatan pemrosesan, karena penggabungan lapisan konvolusi dan lapisan penggabungan yang secara efektif mengurangi dimensi data. Proses klasifikasi diawali dengan mencari hasil dari Metode Bidirectional Gated Recurrent Unit dan Convolutional Neural Network. Hasil klasifikasi dari kedua metode kemudian digabungkan dengan algoritma LightGBM.

## C. Prediksi Akhir

Prediksi akhir dalam penelitian ini diperoleh dengan menggunakan algoritme Light Gradient Boosting Machine (LightGBM), sebuah anggota penting dari rangkaian algoritme ensemble. LightGBM menyempurnakan kerangka kerja Gradient Boosting Decision Tree (GBDT) standar dengan menggabungkan dua teknik inovatif: Gradient-Based One-Sided Sampling (GOSS) dan Exclusive Feature Bundling (EFB). Metodologi ini dikembangkan secara strategis untuk meningkatkan efisiensi dan skalabilitas algoritme GBDT secara substansial. Model LightGBM yang disajikan dalam penelitian ini berfungsi sebagai mekanisme pendukung keputusan yang dioptimalkan untuk pengguna. Keunikannya terletak pada metodologi yang digunakan untuk menentukan jumlah pohon keputusan yang optimal, kedalaman pohon maksimal, dan jumlah daun, yang merupakan parameter penting dalam membangun model LightGBM yang dioptimalkan. Prediksi akhir diformulasikan setelah evaluasi akurasi penilaian dan skor F1. Metrik-metrik ini dipastikan dengan menggunakan alat yang dikenal sebagai matriks kebingungan, yang diterapkan dalam proses klasifikasi yang melibatkan metode Bidirectional Gated Recurrent Unit (BiGRU) dan Convolutional Neural Network (CNN). Khususnya, data untuk prediksi akhir dengan algoritme LightGBM diekstrapolasi dari hasil yang diperoleh dari metodologi BiGRU dan CNN.

## D. Validasi Hasil

Validasi hasil dari kinerja metode BiGRU, CNN, ansambel dengan algoritma LightGBM dilakukan setelah selesainya pembagian terstruktur mengenai akurasi penilaian dan F1 score yang dihitung menggunakan instrument yang disebut confusion matrix. Nilai akurasi lebih tinggi menunjukkan bahwa algoritma tersebut adalah pengklasifikasi teks yang baik untuk pengujian sentimen sarkasme.

## HASIL DAN PEMBAHASAN

Pada bagian ini, menguraikan metodologi yang digunakan untuk pengujian sistem guna mengevaluasi akurasi sentimen dalam konteks judul berita. Tujuan dari pengujian ini adalah untuk memastikan parameter optimal yang menghasilkan akurasi tertinggi. Metrik yang digunakan untuk evaluasi ini adalah akurasi dan skor F1.

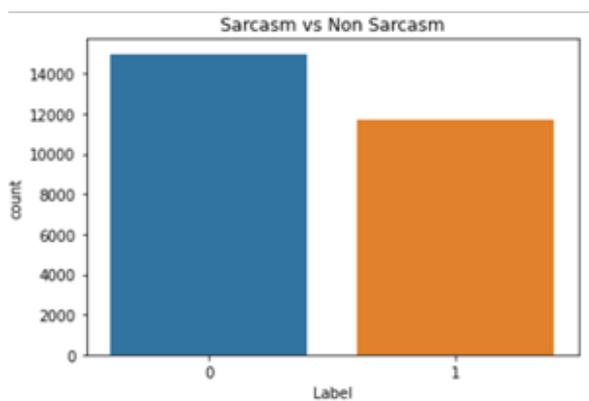
Penelitian ini menggunakan dataset yang terdiri dari 26.709 judul berita, yang terbagi menjadi 11.724 judul berita yang menunjukkan sarkasme dan 14.985 judul berita yang tidak mengandung sarkasme. Pembagian dataset mengalokasikan 80% (20.431 judul berita) untuk tujuan pelatihan dan 20% (4.007 judul berita) untuk pengujian. Eksperimen ini bervariasi ukuran dataset, dimulai dari 500, diperluas menjadi 1.000, dan berakhir pada 2.000 judul berita. Selanjutnya, penelitian ini mengeksplorasi dampak dari berbagai jumlah epoch, khususnya 50 dan 100 epoch. Judul berita dikategorikan ke dalam dua label sentimen: positif (1) dan negatif (0).

Penelitian ini melakukan eksperimen menggunakan pendekatan pembelajaran mendalam ansambel bertumpuk, yang mencakup analisis komparatif dari Bidirectional Gated Recurrent Unit (BiGRU), Convolutional Neural Network (CNN), dan LightGBM. Ikhtisar statistik untuk distribusi data di seluruh label sentimen disajikan dalam Tabel 1.

Tabel 1. Perbandingan Data Terhadap Label Sentimen

Label	Jumlah Data
0 (sarkasme)	11.724
1 (tidak sarkasme)	14.985

Penyebaran dataset label diperlihatkan pada Gambar 4.



Gambar 4. Diagram Penyebaran Dataset

Hasil dari eksperimen yang menggunakan BiGRU, CNN, dan LightGBM untuk 50 epoch di dua label sentimen ditampilkan dalam Tabel 2, 3, dan 4, masing-masing. Demikian pula, hasil untuk uji coba 100 epoch dengan metode yang sama dan label sentimen didokumentasikan dalam Tabel 5, 6, dan 7.

**Tabel 2.** Hasil Percobaan dengan Metode BiGRU untuk *epoch* 50 dengan 2 Label Sentimen

Epoch 50 - BiGRU				
No.	Epoch	Data	Akurasi	F1-Score
1	50	500	84.7%	83.6%
2	50	1.000	85.1%	84.6%
3	50	2.000	84.7%	84.1%
Rata-rata			84.8%	84.1%

**Tabel 3.** Hasil Percobaan dengan Metode CNN untuk *epoch* 50 dengan 2 Label Sentimen

Epoch 50 - CNN				
No.	Epoch	Data	Akurasi	F1-Score
1	50	500	86.7%	85.2%
2	50	1.000	86.7%	85.4%
3	50	2.000	86.3%	84.7%
Rata-rata			86.6%	85.1%

**Tabel 4.** Hasil Percobaan dengan Metode LightGBM untuk *epoch* 50 dengan 2 Label Sentimen

Epoch 50 - LightGBM				
No.	Epoch	Data	Akurasi	F1-Score
1	50	500	91.5%	90.4%
2	50	1.000	91.9%	90.8%
3	50	2.000	91.8%	90.8%
Rata-rata			91.7%	90.7%

**Tabel 5.** Hasil Percobaan dengan Metode BiGRU untuk *epoch* 100 dengan 2 Label Sentimen

Epoch 100 - BiGRU				
No.	Epoch	Dimensi	Akurasi	F1-Score
1	100	500	84.2%	83.7%
2	100	1.000	85.3%	84.7%
3	100	2.000	84.3%	83.9%
Rata-rata			84.6%	84.1%

**Tabel 6.** Hasil Percobaan dengan Metode BiGRU untuk *epoch* 100 dengan 2 Label Sentimen

Epoch 100 - CNN				
No.	Epoch	Data	Akurasi	F1-Score
1	100	500	86.7%	84.9%
2	100	1.000	86.5%	85.0%
3	100	2.000	86.0%	84.5%
Rata-rata			86.4%	84.8%

**Tabel 7.** Hasil Percobaan dengan Metode LightGBM untuk *epoch* 100 dengan 2 Label

No.	Epoch	Data	Akurasi	F1-Score
1	100	500	91.2%	90.2%
2	100	1.000	91.2%	90.2%
3	100	2.000	90.8%	89.7%
Rata-rata			91.1%	90%

Hasil terbaik percobaan pada *epoch* 50 dengan metode BiGRU di tabel 1 diperoleh jumlah data 1.000 dengan nilai akurasi 85.1% dan f1-score 84.6%, diikuti jumlah data 2.000 dengan nilai akurasi 84,7% dan f1-score 84.1% dan diurutan terakhir dengan jumlah data 500 dengan nilai akurasi 84.7% dan f1-score 83.6%. Rata-rata akurasi dari hasil percobaan yang dilakukan dengan *epoch* 50 adalah 84.8% dan rata-rata f1-score 84.1%.

Percobaan dengan *epoch* 50 dengan metode CNN di Tabel 2 diperoleh hasil terbaik pada jumlah data 1.000 dengan nilai akurasi 86.7% dan f1-score 85.4%, diikuti jumlah data 500 dengan nilai akurasi 86.7% dan f1-score 85.2% dan diurutan terakhir dengan jumlah data 2.000 dengan nilai akurasi 86.3% dan f1-score 84.7%. Rata-rata akurasi dari hasil percobaan yang dilakukan dengan *epoch* 50 adalah 86.6% dan rata-rata f1-score 85.1%.

Percobaan dengan *epoch* 50 dengan metode LightGBM di Tabel 3 diperoleh hasil terbaik pada jumlah data 1.000 dengan nilai akurasi 91.9% dan f1-score 90.8%, diikuti jumlah data 2.000 dengan nilai akurasi 91.8% dan f1-score 90.8% dan diurutan terakhir dengan jumlah data 500 dengan nilai akurasi 91.5% dan f1-score 90.4%. Rata-rata akurasi dari hasil percobaan yang dilakukan dengan *epoch* 50 adalah 91.7% dan rata-rata f1-score 90.7%.

Hasil terbaik percobaan pada *epoch* 100 dengan metode BiGRU di tabel 5 diperoleh jumlah data 1.000 dengan nilai akurasi 85.3% dan f1-score 84.7%, diikuti jumlah data 2.000 dengan nilai akurasi 84.3 % dan f1-score 83.9% dan diurutan terakhir dengan jumlah data 500 dengan nilai akurasi 84.2% dan f1-score 83.7%. Rata-rata akurasi dari hasil percobaan yang dilakukan dengan *epoch* 100 adalah 84.6% dan rata-rata f1-score 84.1%.

Hasil terbaik percobaan pada *epoch* 100 dengan metode CNN di tabel 6 diperoleh dimensi 500 dengan nilai akurasi 86.7% dan f1-score 84.9%, diikuti dimensi 1.000 dengan nilai akurasi 85.5 % dan f1-score 85.0% dan diurutan terakhir dengan dimensi 2000 dengan nilai akurasi 86.0% dan f1-score 84.5%. Rata-rata akurasi dari hasil percobaan yang dilakukan dengan *epoch* 100 adalah 86.4% dan rata-rata f1-score 84.8%.

Percobaan dengan *epoch* 100 dengan metode LightGBM di tabel 7 diperoleh hasil terbaik pada jumlah data 500 dan 1.000 dengan nilai akurasi yang sama yaitu 91.2% dan f1-score yang sama juga yakni 90.2%, diikuti jumlah data 2.000 dengan nilai akurasi 90.8% dan f1-score 89.7%. Rata-rata akurasi dari hasil percobaan yang dilakukan dengan *epoch* 100 adalah 91.1% dan rata-rata f1-score 90.0%.

**Tabel 8.** Perbandingan dari metode BiGRU, CNN, dan LightGBM

Metode	Akurasi	F1-Score
LightGBM	91.2%	90.2%
CNN	86.7%	85.4%
BiGRU	85.1%	84.6%

Penilaian empiris terhadap konfigurasi dimensi BiGRU, CNN, dan LightGBM yang optimal dilakukan pada *epoch* ke-50. Evaluasi ini merupakan bagian dari kerangka kerja eksperimental yang lebih luas yang dirancang untuk memastikan ketepatan dan metrik F1-score dari pendekatan pembelajaran mendalam

ensemble bertumpuk yang diusulkan. Metodologi yang menjadi sasaran pengujian meliputi BiGRU, CNN, dan LightGBM. Analisis komparatif dari metode-metode ini, khususnya yang berfokus pada metrik akurasi dan F1-score, secara sistematis disajikan pada Tabel 8. Secara khusus, hasil yang paling baik diamati dengan ukuran dataset 1.000 entri. Dalam konteks ini, pendekatan LightGBM menunjukkan kinerja yang lebih unggul dalam hal akurasi dan F1-score, diikuti oleh CNN, dan kemudian BiGRU, dalam urutan efektivitas yang menurun.

## KESIMPULAN

Dalam bidang deteksi sarkasme, penggunaan metode LightGBM, khususnya dengan set data yang terdiri dari 1.000 entri, telah diidentifikasi sebagai pendekatan yang paling efektif. Metode ini mencapai akurasi klasifikasi yang unggul dibandingkan dengan metodologi alternatif, dibuktikan dengan tingkat akurasi 91,2% dan skor F1 90,2%. Akurasi yang tinggi disebabkan oleh ukuran dataset, di mana 1.000 titik data cukup untuk merangkum spektrum sentimen korpus. Selain itu, pemilihan jumlah epoch yang optimal, yang ditetapkan pada 50 dalam analisis ini, secara signifikan mempengaruhi kinerja LightGBM. Meskipun demikian, jumlah epoch yang berlebihan dapat menyebabkan penggabungan fitur-fitur yang kurang relevan, sehingga memperkenalkan noise ke dalam model. Ke depannya, penelitian di masa depan diantisipasi untuk mengeksplorasi kemampuan pendekatan pembelajaran mendalam stack ensemble dalam set data yang ditandai dengan modalitas yang beragam.

## REFERENSI

- [1] A. Heru, "Gaya Bahasa Sindiran Ironi, Sinisme Dan Sarkasme Dalam Berita Utama Harian Kompas," *J. Pembahsi (Pembelajaran Bhs. Dan Sastra Indones.*, vol. 8, no. 2, p. 43, 2018.
- [2] S. Poria, E. Cambria, D. Hazarika, and P. Viji, "A deeper look into sarcastic tweets using deep convolutional neural networks," *COLING 2016 - 26th Int. Conf. Comput. Linguist. Proc. COLING 2016 Tech. Pap.*, pp. 1601–1612, 2016.
- [3] V. Govindan and V. Balakrishnan, "A machine learning approach in analysing the effect of hyperboles using negative sentiment tweets for sarcasm detection," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 5110–5120, 2022.
- [4] S. M. Sarsam, H. Al-Samarraie, A. I. Alzahrani, and B. Wright, "Sarcasm detection using machine learning algorithms in Twitter: A systematic review," *Int. J. Mark. Res.*, vol. 62, no. 5, pp. 578–598, 2020.
- [5] D. Alita and A. R. Isnain, "Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier," *J. Komputasi*, vol. 8, no. 2, pp. 50–58, 2020.
- [6] M. S. Razali, A. A. Halin, L. Ye, S. Doraisamy, and N. M. Norowi, "Sarcasm Detection Using Deep Learning with Contextual Features," *IEEE Access*, vol. 9, pp. 68609–68618, 2021.
- [7] Y. Yunitasari, A. Musdholifah, and A. K. Sari, "Sarcasm Detection For Sentiment Analysis in Indonesian Tweets," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 1, p. 53, 2019.
- [8] K. S. Ranti and A. S. Girsang, "Indonesian sarcasm detection using convolutional neural network," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 9, pp. 4952–4955, 2020.
- [9] S. Fanissa, M. A. Fauzi, and S. Adinugroho, "Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking," *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, pp. 2766–2770, 2018.
- [10] C. Jia and H. Zan, "Context-Based Sarcasm Detection Model in Chinese Social Media Using BERT and Bi-GRU Models," *CEUR Workshop Proc.*, vol. 3150, pp. 42–50, 2022.
- [11] H. F. Fadli and A. F. Hidayatullah, "Identifikasi Cyberbullying Pada Media Sosial Twitter Menggunakan Metode Klasifikasi Random Forest," *Automata*, 2019.
- [12] M. A. Rosid, D. Siahaan, and A. Saikhu, "Pre-Trained Word Embeddings for Sarcasm Detection in Indonesian Tweets: A Comparative Study," in *2022 9th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, 2022, pp. 281–286.
- [13] C. N. N. Bigru, "applied sciences Short Text Aspect-Based Sentiment Analysis Based on," 2022.
- [14] P. Mehndiratta, S. Sachdeva, and D. Soni, "Detection of sarcasm in text data using deep convolutional neural networks," *Scalable Comput.*, vol. 18, no. 3, pp. 219–228, 2017.

### Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

### Article History:

Received: 30 July 2023 | Accepted: 25 September 2023 | Published: 30 November 2023