

# Deep Learning Stack Ensemble To Detect Sarcasm In News Headline Dataset

## Stack Ensemble Deep Learning Untuk Mendeteksi Sarkasme Pada Dataset News Headline

Siti Nur Haliza<sup>1\*)</sup>, Mochammad Alfian Rosid<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

Email Penulis Korespondensi : [snhalisa@umsida.ac.id](mailto:snhalisa@umsida.ac.id)

**Abstract.** Sarcasm is still often used by society to make someone give up and hurt. Detecting sarcasm is a tricky problem that remains in recent sentiment analysis studies. The fact that it's widely used in many conversations also makes it a common thing to have to deal with in datasets filled with people's conversations. Lack of methods for distinguishing between these types of statements can affect the performance of sentiment analysis, especially in identifying negative, positive, or neutral sentiments. Failure to detect sarcasm can affect the classification of sentiment analysis results. Therefore, sentiment analysis was carried out to classify sarcasm sentences that contain positive sentiments but have negative meanings. This study aims to compare the performance of three deep learning methods, namely Bidirectional Gated Recurrent Unit, Convolutional Neural Network, and LightGBM in detecting sarcasm in news headlines. The dataset used is from the Kaggle website and contains news headlines in English. The results showed that the LightGBM method had the best performance with an accuracy value of 91.2% and an f1 score of 90.2%, compared to the Bidirectional Gated Recurrent Unit and Convolutional Neural Network methods. Therefore, from the explanation above, it can be concluded that the LightGBM method is the best solution for detecting sarcasm.

**Keywords** - sentiment analysis; BiGRU; CNNs; LightGBM; detection sarcasm

**Abstrak.** Sarkasme masih sering digunakan oleh masyarakat untuk membuat seseorang yang mendengarnya tersinggung dan sakit hati. Mendeteksi sarkasme merupakan masalah rumit yang masih ada dalam studi analisis sentimen baru-baru ini. Fakta bahwa itu tersebar luas digunakan dalam banyak percakapan juga membuatnya menjadi umum hal yang harus dihadapi dalam kumpulan data yang diisi dengan orang-orang percakapan. Kekurangan metode untuk membedakan jenis pernyataan ini dapat mempengaruhi kinerja analisis sentimen, terutama dalam mengidentifikasi sentimen negatif, positif, atau netral. Kegagalan dalam mendeteksi sarkasme dapat mempengaruhi klasifikasi hasil analisis sentimen. Oleh karena itu, analisis sentimen dilakukan untuk mengklasifikasikan kalimat sarkasme yang mengandung sentimen positif namun memiliki makna negatif. Penelitian ini bertujuan untuk membandingkan kinerja tiga metode deep learning, yaitu Bidirectional Gated Recurrent Unit, Convolutional Neural Network, dan LightGBM dalam mendeteksi kalimat sarkasme pada judul berita. Dataset yang digunakan berasal dari situs web Kaggle dan berisi judul berita dalam bahasa Inggris. Hasil penelitian menunjukkan bahwa metode LightGBM memiliki kinerja terbaik dengan nilai akurasi 91,2% dan skor f1 90,2%, dibandingkan dengan metode Bidirectional Gated Recurrent Unit dan Convolutional Neural Network. Oleh karena itu, dari penjelasan di atas, dapat disimpulkan bahwa metode LightGBM adalah solusi terbaik untuk mendeteksi sarkasme.

**Kata Kunci** - analisis sentimen; BiGRU; CNN; LightGBM; deteksi sarkasme

## I. PENDAHULUAN

Bahasa terbentuk melalui serangkaian aturan, kaidah, atau pola yang ditetapkan untuk mengatur tata bunyi, bentuk kata, serta susunan kalimat. Pelanggaran terhadap aturan, kaidah, atau pola ini dapat mengganggu komunikasi [1]. Gaya bahasa memainkan peran penting dalam mengartikan maksud dan tujuan seseorang. Bahasa bisa berupa kata kasar yang penuh dengan sindiran dan mengandung olok-an bisa menyakiti hati seseorang atau bisa diartikan sebagai sarkasme. Sarkasme dapat didefinisikan sebagai “ekspresi atau ucapan yang tajam, pahit, atau tajam; ejekan atau ejekan yang pahit” [2]. Menganalisis kalimat sarkasme bukan suatu hal yang mudah untuk dilakukan. Kalimat sarkasme tak jarang sekali disampaikan seseorang dalam menginformasikan sesuatu melalui sosial media.

Deteksi sarkasme sering dijadikan untuk analisis sentimen dalam mengidentifikasi sentimen negatif, positif, atau netral. Sarkasme terdapat dalam ekspresi wajah, gerak tubuh, dan bahkan dalam teks [3]. Suatu kegagalan yang terjadi saat mendeteksi sarkasme dapat mempengaruhi hasil dalam klasifikasi analisis sentimen dari dataset. Tantangan sarkasme dan manfaat deteksi sarkasme untuk analisis sentimen telah menyebabkan minat penelitian deteksi sarkasme otomatis [4]. Berdasarkan hasil penelitian Alita dan Rahman [5]. berhasil meningkatkan akurasi analisis sentimen

sebesar 16,61% dengan mendeteksi sarkasme pada tweet tentang layanan publik. Dalam penelitian yang telah dilakukan oleh Saifullah dkk [6] yang melakukan deteksi sarkasme pada sosial media twitter. Sebanyak 17% pengguna sosial media Instagram menggunakan kalimat sarkasme, menunjukkan bahwa penggunaan kalimat sarkasme masih sering ditemukan pada platform tersebut.

Pada penelitian yang telah disebutkan sebelumnya, digunakan metode base classifier machine learning dan deep learning dalam proses klasifikasi. Belum ada penelitian yang menggunakan stack ensemble deep learning untuk

mendeteksi sarkasme. Berdasarkan hal tersebut, ada peluang untuk pengembangan dalam deteksi sarkasme menggunakan metode stack ensemble deep learning.

Penelitian ini bertujuan untuk membangun model stack ensemble deep learning. Dalam penelitian ini, metode deep learning dasar yang akan digunakan meliputi Bidirectional Gated Recurrent Unit (BiGRU) dan Convolutional Neural Network (CNN). Pada penelitian ini, LightGBM digunakan untuk melakukan stack ensemble metode deep learning. Setelah melalui proses uji coba, diharapkan dapat ditemukan metode deep learning yang memiliki performa terbaik, sehingga dapat dijadikan acuan dalam pembuatan model untuk mendeteksi sarkasme pada judul berita di platform media sosial.

## II. LITERATUR REVIEW

Pada penelitian ini dilakukan analisis dari kumpulan informasi dengan penelitian yang serupa. Penelitian yang serupa akan ditelusuri keutamaan dan kekurangan dalam penelitian tersebut dengan melihat dari hasil yang terdahulu. Penelitian untuk mendeteksi sarkasme mulai banyak dipelajari dalam beberapa tahun terakhir. Penelitian yang dilakukan oleh [5] melibatkan tahap ekstraksi fitur dengan 4 set fitur, hal yang serupa juga dilakukan oleh [7] yaitu fitur yang berkaitan dengan sentimen, tanda baca, leksikal dan sintaktik, dan pola kalimat, dan hasilnya diklasifikasikan dengan menggunakan metode Random Forest Classifier. Hasil pengujian tersebut masih memiliki akurasi yang rendah. Penelitian selanjutnya yang dilakukan oleh [8] untuk mengidentifikasi sarkasme dalam bahasa indonesia. Penelitian ini menghasilkan skor F1 90% pada 1138 komentar pada data yang diuji.

Penelitian selanjutnya yang dilakukan oleh [9] proses analisis sentimen melibatkan beberapa tahapan, antara lain preprocessing data, seleksi fitur dengan menggunakan metode Query Expansion Ranking, dan klasifikasi data menggunakan algoritma Naive Bayes. Hasil seleksi fitur 75% memiliki akurasi terbaik sebesar 86.6%. Penelitian selanjutnya yang dilakukan oleh [10] menggabungkan dua metode untuk deteksi sarkasme, sehingga dapat meningkatkan akurasi deteksi sarkasme situasional.

Penelitian selanjutnya yang dilakukan oleh [11] Penelitian ini dapat diperluas dengan menambahkan lebih banyak data dan dataset pelatihan yang lebih besar. Selain itu, fitur kelas juga dapat ditambahkan agar tidak hanya membedakan antara cyberbullying dan bukan cyberbullying. Hal ini diharapkan dapat meningkatkan kinerja model dan memperluas cakupan aplikasi deteksi sentimen dalam berbagai konteks dan domain. Hasil akurasi dari metode LSTM 93.77% dan BiLSTM 95.24%. Penelitian yang dilakukan oleh [12] merepresentasikan vektor dari kata-kata, tiga model penyisipan kata pra-pelatihan yaitu GloVe, fastText, dan BERT. Hasil penelitian menunjukkan bahwa kombinasi fastText embeddings dan BiGRU sebagai pengklasifikasi menghasilkan performa terbaik dengan akurasi 93,85%.

Pada pembuatan penelitian ini penulis mereferensikan sebagian artikel jurnal sebagai pustaka riset terkait topik yang serupa. Penulis menyusun dan menerapkan beberapa teori yang digunakan dalam penelitian terdahulu agar bisa dijadikan referensi oleh penulis dan dapat dikembangkan kembali.

### 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 :

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

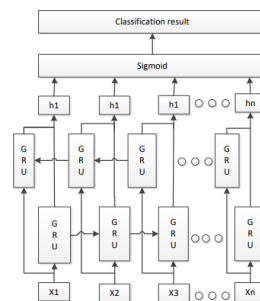
Dimana  $Z_t$  sebagai gerbang pembaruan dan  $W_z$  adalah bobot matriks.  $X_t$  sebagai input pada waktu  $t$  dan  $h_{t-1}$  adalah output pada saat sebelumnya. Hasil apakah keadaan saat ini adalah keadaan yang diperbarui atau keadaan sebelumnya dapat ditentukan oleh gerbang pembaruan. Saat gerbang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

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

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 :

$$H = \vec{h}_t \oplus \overleftarrow{h}_t \quad (4)$$

di mana  $H$  mewakili output BiGRU,  $\vec{h}_t$  dan  $\overleftarrow{h}_t$  mewakili dua GRU searah, dan merupakan elemen tambahan [13]. Struktur Jaringan BiGRU diperlihatkan oleh Gambar 1.

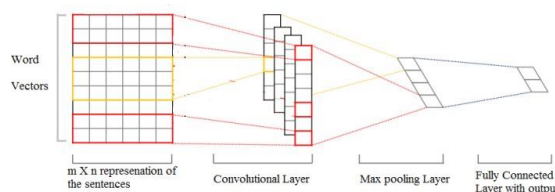


Gambar 1. Struktur Jaringan BiGRU

Sumber : [2]

## 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  $P1$  disebarkan ke lapisan yang terhubung sepenuhnya [3]. Arsitektur CNN diperlihatkan oleh Gambar 2.



Gambar 2. Arsitektur Convolutional Neural Network

Sumber : [3]

### C. LightGBM

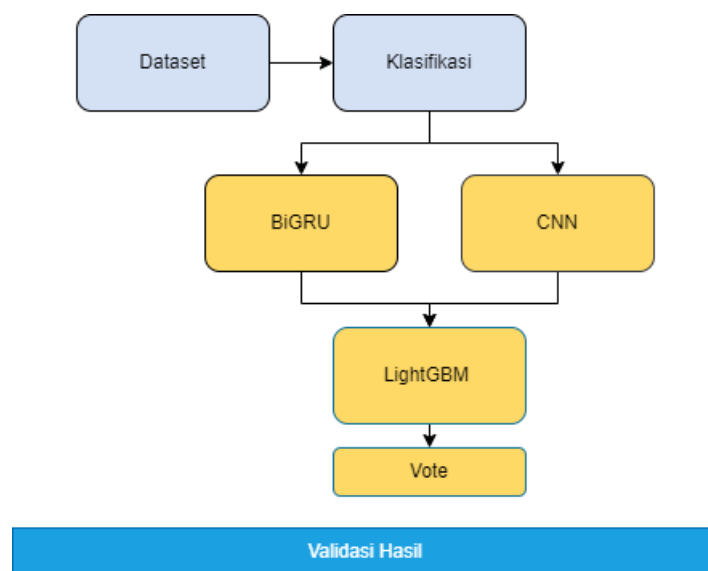
LightGBM adalah algoritma ansambel yang dikembangkan oleh Microsoft yang memberikan implementasi yang efisien dari algoritma peningkatan gradien. Keuntungan utama dari LightGBM adalah akselerasi dramatis dari proses pelatihan algoritma, yang pada banyak kasus menghasilkan model yang lebih efektif dalam waktu yang lebih cepat. LightGBM dibangun di atas algoritme pohon keputusan, menggunakan jumlah nestimator dari pohon yang ditingkatkan. Algoritma penguat pohon mengungguli yang lain untuk masalah prediksi. Algoritma pembelajaran ansambel LightGBM telah diterapkan dalam berbagai studi klasifikasi dan regresi dan mencapai hasil deteksi yang sangat baik, yang menunjukkan bahwa LightGBM adalah algoritma pengklasifikasi yang efektif. Model LightGBM yang diusulkan menyediakan sistem pendukung keputusan yang dioptimalkan untuk pengguna. Kekhasan dari pendekatan yang diusulkan adalah dalam prosedur yang digunakan untuk menghitung jumlah pohon keputusan, kedalaman maksimum pohon, dan jumlah daun pohon untuk membangun model LightGBM yang optimal [4].

## III. METODE

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

### A. Dataset

Dataset diambil dari situs web yaitu Kaggle. Data yang diperoleh tersebut adalah data berita utama. Kumpulan data berita utama memiliki kesalahan tata bahasa yang minimal, kalimat yang unik dan data yang sangat bersih. Data tersebut menggunakan Natural Language Toolkit (NLTK) untuk tokenisasi, lemmatisasi dan pra pemrosesan. Dataset headline berita tersebut terdiri dari 26.709 headline berita.



Gambar 3. Arsitektur Sistem untuk Stack Ensemble Deep Learning

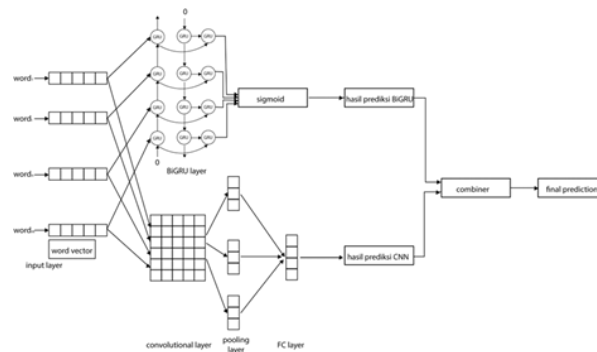
### B. Proses Klasifikasi Deep Learning

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. Penelitian ini mengusulkan metode BiGRU sebagai alat ekstraksi fitur teks yang optimal. Parameter seperti Learning Rate, Batch Size, dan Epoch diukur dalam penelitian ini. Proses ekstraksi fitur dilakukan dengan menggunakan metode BiGRU yang melakukan ekstraksi kata secara forward dan backward melalui dua arah, yaitu layer forward yang mengambil kata dari awal hingga akhir, dan backward yang mengambil kata dari akhir hingga awal. Algoritma CNN memiliki keunggulan dalam hal kecepatan proses, dimana CNN terdapat layer konvolusi dan layer pooling yang dapat mereduksi dimensi data.

### C. Prediksi Akhir

Prediksi akhir menggunakan algoritma LightGBM. LightGBM merupakan salah satu algoritma ansambel. LightGBM mengimplementasikan algoritma Gradient Boosting Decision Tree (GBDT) konvensional dengan penambahan dua teknik baru yakni Pengambilan Sampel Satu Sisi Berbasis Gradien (GOSS) dan Bundling Fitur

Eksklusif (EFB). Teknik-teknik ini dirancang untuk secara signifikan meningkatkan efisiensi dan skalabilitas GBDT. Model LightGBM yang diusulkan menyediakan sistem pendukung keputusan yang dioptimalkan untuk pengguna. Kekhasan dari pendekatan yang diusulkan adalah dalam prosedur yang digunakan untuk menghitung jumlah pohon keputusan, kedalaman maksimum pohon, dan jumlah daun pohon untuk membangun model LightGBM yang optimal. Prediksi akhir ini dilakukan setelah hasil akurasi penilaian dan F1 score yang dihitung menggunakan instrument yang disebut confusion matrix dari klasifikasi dengan metode BiGRU dan CNN ditemukan, karena data prediksi akhir dengan algoritma LightGBM diambil dari hasil metode BiGRU dan CNN.



Gambar 4. Arsitektur CNN dan BiGRU

#### 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.

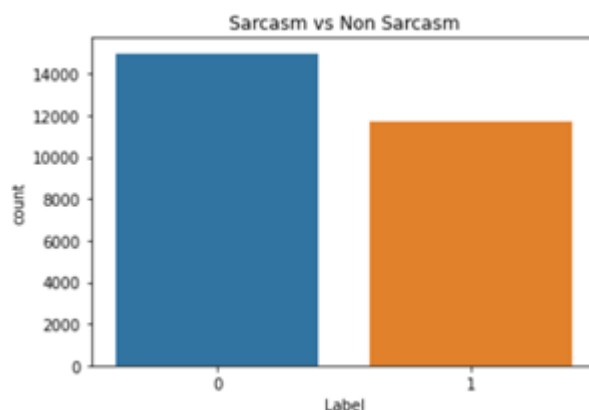
### V. HASIL DAN PEMBAHASAN

Pada bagian ini, dijelaskan tentang pengujian sistem untuk mengukur akurasi sentimen pada headline berita. Hasil pengujian akan digunakan untuk menentukan parameter yang memberikan akurasi terbaik. Alat ukur yang digunakan adalah akurasi dan f1-score.

Penelitian ini menggunakan dataset sebanyak 26.709 headline berita yang terdiri dari 11.724 yang bermakna sarkasme dan 14985 yang tidak bermakna sarkasme. Sebanyak 80% data digunakan untuk training data dan 20% data digunakan untuk testing data. Total data yang digunakan adalah 26.709 headline berita, dengan 20.431 data untuk training dan 4.007 data untuk testing. Percobaan data dimulai dari 500, 1000 dan 2000. Jumlah epoch yang dicoba adalah 50 dan 100. Data dibagi menjadi 2 label sentiment, yaitu positif (1) dan negatif (0). Percobaan dengan stack ensemble deep learning dilakukan dengan membandingkan Bidirectional Gated Recurrent Unit, Convolutional Neural Network dan LightGBM. Statistik untuk perbandingan data terhadap label sentiment terdapat pada Tabel 1. Diagram penyebaran dataset dapat dilihat pada Gambar 5. Hasil percobaan dengan Metode BiGRU, CNN, dan LightGBM untuk epoch 50 dengan 2 Label Sentimen dapat dilihat pada Tabel 2, Tabel 3 dan Tabel 4. Hasil percobaan dengan Metode BiGRU, CNN, dan LightGBM untuk epoch 100 dengan 2 Label Sentimen dapat dilihat pada Tabel 5, Tabel 6, dan Tabel 7.

Tabel 1. Perbandingan Data Terhadap Label Sentimen

Label	Jumlah Data
0 (sarkasme)	11724
1 (tidak sarkasme)	14985



Gambar 5. Diagram Penyebaran Dataset

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

Epoch 50 - BiGRU				
No.	Epoch	Dimensi	Akurasi	F1-Score
1	50	500	84,7%	83,6%
2	50	1000	85,1%	84,6%
3	50	2000	84,7%	84,1%
Rata-rata			84,8%	84,1%

Hasil terbaik percobaan pada epoch 50 dengan metode BiGRU di tabel 1 diperoleh dimensi 1000 dengan nilai akurasi 85,1% dan f1-score 84,6%, diikuti dimensi 2000 dengan nilai akurasi 84,7% dan f1-score 84,1% dan diurutan terakhir dengan dimensi 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%.

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

Epoch 50 - CNN				
No.	Epoch	Dimensi	Akurasi	F1-Score
1	50	500	86,7%	85,2%
2	50	1000	86,7%	85,4%
3	50	2000	86,3%	84,7%
Rata-rata			86,6%	85,1%

Percobaan dengan epoch 50 dengan metode CNN di Tabel 2 diperoleh hasil terbaik pada dimensi 1000 dengan nilai akurasi 86,7% dan f1-score 85,4%, diikuti dimensi 500 dengan nilai akurasi 86,7% dan f1-score 85,2% dan diurutan terakhir dengan dimensi 2000 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%.

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

Epoch 50 - LightGBM				
No.	Epoch	Dimensi	Akurasi	F1-Score
1	50	500	91,5%	90,4%
2	50	1000	91,9%	90,8%
3	50	2000	91,8%	90,8%
Rata-rata			91,7%	90,7%

Percobaan dengan epoch 50 dengan metode LightGBM di Tabel 3 diperoleh hasil terbaik pada dimensi 1000 dengan nilai akurasi 91,9% dan f1-score 90,8%, diikuti dimensi 2000 dengan nilai akurasi 91,8% dan f1-score 90,8% dan

diurutan terakhir dengan dimensi 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%.

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	1000	85,3%	84,7%
3	100	2000	84,3%	83,9%
<b>Rata-rata</b>			84,6%	84,1%

Hasil terbaik percobaan pada epoch 100 dengan metode BiGRU di tabel 5 diperoleh dimensi 1000 dengan nilai akurasi 85,3% dan f1-score 84,7%, diikuti dimensi 2000 dengan nilai akurasi 84,3 % dan f1-score 83,9% dan diurutan terakhir dengan dimensi 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%.

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

No.	Epoch	Dimensi	Akurasi	F1-Score
1	100	500	86,7%	84,9%
2	100	1000	86,5%	85,0%
3	100	2000	86,0%	84,5%
<b>Rata-rata</b>			86,4%	84,8%

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 1000 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%.

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

No.	Epoch	Dimensi	Akurasi	F1-Score
1	100	500	91,2%	90,2%
2	100	1000	91,2%	90,2%
3	100	2000	90,8%	89,7%
<b>Rata-rata</b>			91,1%	90,0%

Percobaan dengan epoch 100 dengan metode LightGBM di tabel 7 diperoleh hasil terbaik pada dimensi 500 dan 1000 dengan nilai akurasi yang sama yaitu 91,2% dan f1-score yang sama juga yakni 90,2%, diikuti dimensi 2000 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	Epoch	Dimensi	Akurasi	F1-Score
<b>LightGBM</b>	50	1000	91,2%	90,2%
<b>CNN</b>	50	1000	86,7%	85,4%
<b>BiGRU</b>	50	1000	85,1%	84,6%

Nilai rata-rata dari percobaan dimensi BiGRU, CNN, dan LightGBM terbaik diperoleh pada epoch 50. Percobaan metode terhadap dataset dilakukan untuk melihat nilai akurasi dan f1-score dari metode stack ensemble deep learning yang diusulkan. Metode yang diuji adalah BiGRU, CNN, dan LightGBM. Hasil akurasi dan f1-score untuk perbandingan dari metode dapat dilihat pada Tabel 8. Hasil terbaik diperoleh

pada dimensi 1000 dengan metode LightGBM mendapatkan hasil akurasi dan f1-score terbaik, dengan urutan kedua yaitu metode CNN, dan terakhir yaitu metode BiGRU.

## V. SIMPULAN

Metode terbaik untuk mendeteksi sarkasme adalah menggunakan metode LightGBM dengan dimensi 1000, karenamampu memberikan hasil klasifikasi yang paling akurat dibandingkan metode lain, dengan nilai akurasi 91,2% dan f1-score 90,2%. Keakuratan tersebut disebabkan oleh dimensi 1000 yang dapat merepresentasikan sentimen dari data secara tepat. Pemilihan jumlah epoch yang tepat juga berpengaruh pada hasil LightGBM, yang pada penelitian ini menggunakan epoch 50. Namun, terlalu banyak epoch dapat menghasilkan fitur-fitur yang kurang penting dan menjadinoise. Pada penelitian selanjutnya, akan dilakukan pengujian metode stack ensemble deep learning pada dataset dengan modalitas yang berbeda.

## 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. Vij, "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.
- [15] D. D. Rufo, T. G. Debelee, A. Ibenthal, and W. G. Negera, "Diagnosis of diabetes mellitus using gradient boosting machine (Lightgbm)," Diagnostics, vol. 11, no. 9, pp. 1–14, 2021

### **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.*