

ANALISIS KOMPARATIF KLASIFIKASI SENTIMEN PENGGUNA APLIKASI INVESTASI MENGGUNAKAN ALGORITMA HYBRID CNN-LSTM, CNN-GRU DENGAN IMPLEMENTASI SMOTE

(Comparative Analysis of Investment Application User Sentiment Classification Using Hybrid CNN-LSTM, CNN-GRU Algorithms with SMOTE Implementation)

I Putu Agus Ari Mahendra¹⁾, Kusrini Kusrini²⁾

^{1, 2)} Magister Teknik Informatika Universitas Amikom Yogyakarta

Jl. Ring Road Utara, Ngringin, Condongcatur, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta

e-mail: arimahendra@students.amikom.ac.id¹⁾, kusrini@amikom.ac.id²⁾

ABSTRAK

Analisis sentimen, yang merupakan bagian dari Natural Language Processing (NLP), berperan penting dalam mengidentifikasi dan mengklasifikasikan opini yang terkandung dalam teks. Penelitian ini bertujuan membandingkan performa model hybrid CNN-LSTM dan CNN-GRU pada analisis sentimen ulasan pengguna aplikasi investasi di Google Play Store, dengan menggunakan Synthetic Minority Over-sampling Technique (SMOTE) untuk mengatasi ketidakseimbangan data. Sebanyak 15.000 data ulasan dikumpulkan melalui web scraping, kemudian diproses menggunakan metode TF-IDF dan teknik pembersihan teks lainnya. Model CNN-LSTM dan CNN-GRU diuji dengan pembagian data 80% untuk pelatihan dan 20% untuk pengujian. Hasil evaluasi menunjukkan bahwa CNN-GRU memiliki keunggulan dalam presisi (91,62%), skor F1 (90,45%), dan akurasi keseluruhan (87,60%), sementara CNN-LSTM unggul dalam recall (91,08%) untuk mendeteksi ulasan positif. CNN-GRU dinilai lebih seimbang dalam mendeteksi sentimen positif dan negatif, menjadikannya pilihan yang lebih andal untuk analisis sentimen yang membutuhkan performa seragam.

Kata Kunci: Analisis Sentimen, Algoritma Hybrid, CNN-GRU, CNN-LSTM, Ketidakseimbangan Kelas.

ABSTRACT

Sentiment analysis, a branch of Natural Language Processing (NLP), plays a crucial role in identifying and classifying opinions embedded in text. This study aims to compare the performance of hybrid CNN-LSTM and CNN-GRU models in sentiment analysis of user reviews for investment applications on Google Play Store, utilizing the Synthetic Minority Over-sampling Technique (SMOTE) to address data imbalance. A total of 15,000 user reviews were collected through web scraping, preprocessed using the TF-IDF method and various text cleaning techniques. The CNN-LSTM and CNN-GRU models were evaluated using an 80%-20% train-test split. The evaluation results showed that CNN-GRU outperformed in terms of precision (91.62%), F1 score (90.45%), and overall accuracy (87.60%), while CNN-LSTM excelled in recall (91.08%) for detecting positive reviews. CNN-GRU was deemed more balanced in detecting both positive and negative sentiments, making it a more reliable choice for sentiment analysis requiring uniform performance.

Keywords: CNN-GRU, CNN-LSTM, Class Imbalance, Hybrid Algorithm, Sentiment Analysis

I. PENDAHULUAN

Analisis sentimen, atau yang sering disebut sebagai penambangan opini, merupakan salah satu cabang dalam pemrosesan bahasa alami (Natural Language Processing/NLP) bertujuan untuk secara otomatis mengidentifikasi dan mengklasifikasikan sentimen yang terkandung dalam teks. Proses tersebut mencakup berbagai aspek, dari pengklasifikasian sudut pandang penulis, memprediksi orientasi semantik dari kata-kata tertentu, hingga menilai subjektivitas dan elemen-elemen lain yang lebih bernuansa dalam teks[1].

Teknologi analisis sentimen tidak hanya terbatas pada identifikasi positif atau negatif, tetapi juga mampu mendeteksi emosi yang lebih kompleks, seperti kebahagiaan, kemarahan, atau kesedihan. Metode ini memungkinkan analisis terhadap opini secara mendalam pada level kalimat, paragraf, atau dokumen. Hal ini membuatnya sangat berguna dalam berbagai hal, seperti analisis ulasan aplikasi, analisis media sosial, dan penilaian persepsi publik terhadap isu tertentu[2].

Perkembangan analisis sentimen terlihat dari sejumlah penelitian dan aplikasi nyata yang membuktikan keefektifannya dalam memahami sentimen yang bermunculan dari teks, dalam

konteks formal maupun informal, seperti halnya dalam ulasan pengguna aplikasi, media sosial, berita, serta forum diskusi daring[3].

Investasi merupakan elemen penting dalam dunia bisnis dan ekonomi, pada dasarnya individu atau institusi menempatkan dana atau aset dengan tujuan memperoleh keuntungan di masa depan[4]. Di Indonesia, kesadaran masyarakat terhadap investasi berkembang begitu pesat dan jumlah investor terus meningkat, terutama di era digital yang menawarkan kemudahan akses melalui aplikasi investasi. Jenis investasi yang populer meliputi saham, obligasi, properti, hingga mata uang kripto, baik di pasar domestik maupun internasional. Tren ini didukung oleh upaya pemerintah sejak lama dalam mendorong investasi melalui regulasi[5].

Google Play Store adalah *platform* resmi *Google* untuk mendistribusikan aplikasi *Android*, termasuk aplikasi investasi[6]. Ulasan pengguna di *platform* ini menjadi sumber data penting untuk memahami opini pengguna aplikasi. Data tersebut dikumpulkan menggunakan web scraping dengan *API Google Play-Scraper*, yang memungkinkan pengambilan informasi aplikasi dan ulasan secara efisien tanpa perlu layanan pihak ketiga. Adapun berbagai gambaran tentang penerapan analisis sentimen pada aplikasi dan layanan berbasis teknologi. [7] Dalam penelitiannya menguji metode *Random Forest* untuk analisis sentimen ulasan pengguna. Penelitian ini melaporkan akurasi sebesar 84%, namun mengalami keterbatasan dalam jumlah data yang digunakan, yang dapat memengaruhi keakuratan model.

Pada penelitian [8] membuktikan bahwa teknik penyeimbangan data seperti *SMOTE* dapat meningkatkan performa model yang digunakan dalam klasifikasi, dengan mengatasi ketidakseimbangan kelas. Meski penelitian ini menunjukkan efektivitas *SMOTE*, kelemahan terletak pada aplikasi hanya disatu domain, yaitu spam email, sehingga diperlukan pengujian lebih lanjut di konteks analisis sentimen pada ulasan pengguna aplikasi investasi. Pada penelitian yang dilakukan oleh[9] juga menunjukkan bahwa pengaplikasian *SMOTE* bersama dengan algoritma *Decision Tree (DT)* dan *SVM* dapat meningkatkan performa model, terutama dalam menangani data yang tidak seimbang.

Penelitian [10] mengenai analisis sentimen terhadap ulasan restaurant dalam Bahasa Bengali

menunjukkan hasil bahwa model gabungan *CNN-LSTM* mencapai akurasi 94.22% dalam analisis sentimen ulasan restoran, menunjukkan efektivitas metode ini dibandingkan dengan pendekatan sebelumnya yang menjadi rujukan penelitian tersebut. Namun, lain dengan penelitian [11] menunjukkan bahwa model *CNN-GRU* mencapai akurasi tertinggi sebesar 97.77%, menandakan bahwa kombinasi metode ini lebih efektif dalam analisis sentimen pada Twitter dibandingkan algoritma *CNN*, *GRU* dan *GRU-CNN*.

Berbagai penelitian sebelumnya telah menunjukkan bahwa metode analisis sentimen mampu memberikan wawasan berharga bagi penyedia layanan. Namun, beberapa penelitian tersebut masih memiliki keterbatasan, seperti penanganan data tidak seimbang, keterbatasan *dataset*, kurangnya eksplorasi kombinasi algoritma, dan perbedaan pendapat tentang efektivitas algoritma *hybrid Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM)* dan *Convolutional Neural Network-Gated Recurrent Unit (CNN-GRU)*. Berkaitan dengan penelitian yang telah dilakukan tersebut, pada penelitian ini menggunakan algoritma *hybrid Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM)* dan *Convolutional Neural Network-Gated Recurrent Unit (CNN-GRU)* Oleh karena itu, penelitian ini bertujuan menganalisis kinerja arsitektur *hybrid Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM)* dan *Convolutional Neural Network-Gated Recurrent Unit (CNN-GRU)* pada ulasan platform aplikasi investasi yang lebih efektif serta menerapkan teknik penyeimbangan data *SMOTE*. Dengan mengintegrasikan algoritma *hybrid CNN-LSTM* dan *CNN-GRU*, penelitian ini bertujuan untuk membandingkan akurasi klasifikasi sentimen pada *dataset* ulasan pengguna aplikasi investasi. Selain itu, penggunaan *SMOTE* sebagai teknik penanganan ketidakseimbangan data akan memastikan distribusi kelas yang lebih seimbang, sehingga hasil analisis menjadi lebih representatif dan dapat diandalkan.

II. STUDI PUSTAKA

Penelitian tentang analisis sentimen telah dilakukan dengan berbagai pendekatan, algoritma, dan konteks untuk memahami opini masyarakat terhadap produk, layanan, atau fenomena tertentu

[8]Meneliti pengaruh kombinasi *TF-IDF* dan *SMOTE* pada performa model *Deep Learning* untuk mendeteksi spam email. Menunjukkan peningkatan akurasi hingga 44,14% dan *f1-score* hingga 73,75%, namun penelitian tersebut menghadapi risiko *overfitting* dan kurangnya pengujian pada algoritma lain untuk validasi hasil yang lebih luas.

[9]Menggunakan algoritma *Decision Tree (DT)* dan *SVM* dengan *SMOTE* untuk analisis sentimen ulasan wisatawan tentang *Garden by the Bay*. Studi ini mencatat peningkatan akurasi dan *F1-score* secara signifikan, tetapi terbatas pada generalisasi karena fokus hanya pada satu domain tertentu.

[11]Membandingkan berbagai kombinasi algoritma *CNN*, *GRU*, *CNN-GRU*, dan *GRU-CNN* untuk analisis sentimen *tweet* dalam bahasa Indonesia dengan *SMOTE* untuk penyeimbangan data. Model *CNN-GRU* menunjukkan akurasi tertinggi sebesar 97,77%, namun penelitian ini masih menghadapi tantangan terkait pengumpulan dan pelabelan data.

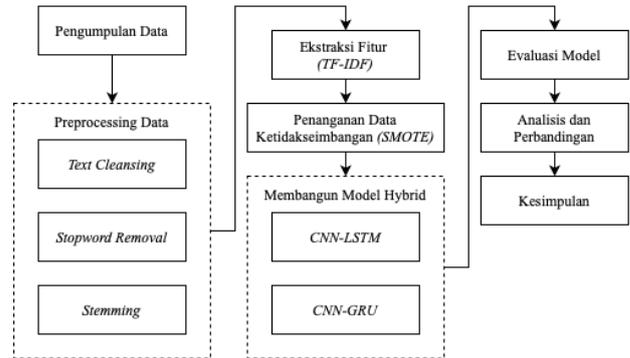
[10]Mengembangkan model *CNN-LSTM* untuk analisis sentimen ulasan restoran dalam bahasa *Bengali*, dengan akurasi mencapai 94,22%. Meski hasilnya menjanjikan, penelitian ini terbatas pada dataset kecil dan belum menggunakan teknik penyeimbangan data seperti *SMOTE* untuk mengatasi bias distribusi kelas.

[12] Mengembangkan model prediksi pergerakan harga saham menggunakan *CNN-LSTM* yang memadukan indikator teknis dan analisis sentimen *tweet*. Dengan akurasi mencapai 97%, model ini unggul dibandingkan *LSTM* dan *GRU*. Namun, penelitian ini terbatas pada *dataset* kecil yang hanya berfokus pada sentimen saham Tesla. Penelitian ini bertujuan mengatasi keterbatasan penelitian sebelumnya, seperti data tidak seimbang, keterbatasan *dataset*, dan kurangnya eksplorasi kombinasi algoritma. Dengan menggunakan algoritma *hybrid CNN-LSTM* dan *CNN-GRU* serta teknik penyeimbangan data *SMOTE*, penelitian ini menganalisis dan membandingkan kinerja kedua arsitektur dalam klasifikasi sentimen pada ulasan pengguna aplikasi investasi.

III. METODE PENELITIAN

Penelitian ini menggunakan metode eksperimen dan menggunakan metode kuantitatif, sebagai bagian dari pendekatan ini, penelitian berfokus pada analisis pola hasil eksperimen yang

dikombinasikan dengan interpretasi mendalam terhadap pengaruh setiap metode dan algoritma yang diuji [13]. Gambar 1 merupakan diagram alur dari langkah penelitian ini.



Gambar 1. Diagram Alur Penelitian

A. Pengumpulan Data

Data dalam penelitian ini dikumpulkan langsung dari *Google Play Store* menggunakan teknik *crawling* berbasis *Python* dengan bantuan *API Google-Play-Scraper*. Pada 14 Oktober 2024 sampai dengan 5 Desember 2024, sebanyak 15.000 sampel data, termasuk nama pengguna, isi ulasan (*content*), dan skor rating (1–5), berhasil dikumpulkan dan disimpan dalam format *CSV*. Tabel 1 menunjukkan contoh sampel data ulasan yang diperoleh.

Tabel 1. Contoh Sampel Data Ulasan.

<i>userName</i>	<i>skor</i>	<i>content</i>
Ugun	5	Membantu yg penghasilan sederhana bisa ikut investasi
Nahwan Firdaus	2	Terlalu rumit pendaftaran untuk trading saham

B. Preprocessing Data

Preprocessing data untuk analisis teks dimulai dengan pembersihan teks, termasuk *case folding*, penerjemahan teks bahasa Inggris ke bahasa Indonesia, serta penghapusan angka dan simbol yang mengganggu. Selanjutnya, *stopword removal* menggunakan *Natural Language Toolkit* untuk menghilangkan istilah umum dalam bahasa Indonesia yang tidak relevan bagi analisis sentimen. Proses ini diikuti oleh *stemming* dengan *Sastrawi*, yang mereduksi kata ke bentuk dasar untuk menyederhanakan data dan meningkatkan akurasi klasifikasi sentimen. Contoh hasil *preprocessing* data ditunjukkan pada Tabel 2.

Tabel 2. Contoh Hasil *Preprocessing* Data.

Sentimen Asli	Sentimen Telah diproses
---------------	-------------------------

Membantu yg penghasilan sederhana bisa ikut investasi Terlalu rumit pendaftaran untuk trading saham	membantu yg penghasilan sederhana bisa ikut investasi terlalu rumit pendaftaran untuk trading saham
--	--

C. Ekstraksi Fitur TF-IDF

TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah metode statistik untuk menilai pentingnya sebuah kata dalam dokumen di dalam suatu korpus, yang banyak digunakan dalam analisis sentimen, *text mining*, dan pencarian informasi[14]. Metode ini efektif untuk mengolah dan menganalisis data teks dalam jumlah besar[15].

D. Penanganan Data Ketidakseimbangan (SMOTE)

Untuk mengatasi ketidakseimbangan dalam dataset, seperti 15.000 ulasan dengan 9.903 ulasan positif dan 5.097 ulasan negatif, teknik *Synthetic Minority Over-sampling Technique* (*SMOTE*) terbukti sangat efektif [8]. *SMOTE* bekerja dengan menghasilkan sampel sintetis untuk kelas minoritas dalam hal ini, ulasan negatif dengan cara mengidentifikasi kesamaan dalam ruang fitur antara contoh kelas minoritas dan membuat interpolasi dari titik-titik terdekat[16]. Dengan menyeimbangkan distribusi kelas, *SMOTE* meningkatkan kinerja model *machine learning*, memastikan model tidak bias terhadap kelas mayoritas dan mampu menggeneralisasi dengan lebih baik pada data baru. Pendekatan ini sangat relevan dalam analisis sentimen, di mana akurasi dalam memprediksi sentimen minoritas sama pentingnya dengan mengidentifikasi sentimen mayoritas.

E. Membangun Model Hybrid

Pada proses klasifikasi, *dataset* dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Penelitian ini berfokus pada penerapan dan evaluasi kinerja dua model *deep learning*, yaitu *Convolutional Neural Network-Long Short-Term Memory* (*CNN-LSTM*) dan *Convolutional Neural Network-Gated Recurrent Unit* (*CNN-GRU*), dalam analisis sentimen.

Model *CNN-LSTM* mengombinasikan kemampuan *CNN* untuk mengekstraksi fitur penting dari teks, seperti pola kata dan frasa, dengan kemampuan *LSTM* untuk memahami urutan kata atau hubungan temporal dalam data. Pendekatan ini memungkinkan model mengenali konteks lokal sekaligus global secara lebih mendalam[10]. Sementara itu, *CNN-GRU*

memiliki kemampuan serupa, tetapi dengan efisiensi komputasi yang lebih tinggi, karena *GRU* merupakan versi *LSTM* yang lebih sederhana dan cepat tanpa mengorbankan akurasi[17].

F. Evaluasi Model

Evaluasi model dimana performa tiap model diuji menggunakan metrik akurasi, presisi, *recall*, dan *f1-score* untuk mengevaluasi efektivitas prediksi sentimen. Analisis komparatif membantu menentukan algoritme terbaik berdasarkan ketidakseimbangan kelas, kompleksitas fitur, dan karakteristik data tekstual. Pendekatan ini memberikan pemahaman menyeluruh tentang keunggulan dan keterbatasan tiap model dalam klasifikasi sentimen.

G. Analisis dan Perbandingan

Bagian ini menganalisis performa algoritma *hybrid Convolutional Neural Network-Long Short-Term Memory* (*CNN-LSTM*) dan *Convolutional Neural Network-Gated Recurrent Unit* (*CNN-GRU*) untuk mengevaluasi kemampuan masing-masing model dalam analisis sentimen. Fokus analisis adalah mengidentifikasi kekuatan dan kelemahan komparatif kedua model, termasuk efektivitas dalam menangkap pola data sekuensial, kemampuan generalisasi terhadap berbagai jenis data *teks*, serta dampaknya pada akurasi, *presisi*, *recall*, dan *f1-score*. Pendekatan ini memberikan wawasan mendalam mengenai keunggulan dan batasan masing-masing algoritma dalam mengolah data tekstual secara optimal.

H. Kesimpulan

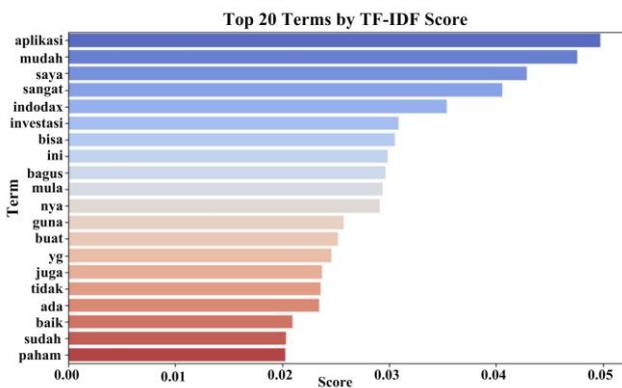
Menarik kesimpulan dari penelitian berdasarkan hasil evaluasi dan memberikan saran untuk penelitian lebih lanjut.

IV. HASIL DAN PEMBAHASAN

A. Ekstraksi Fitur TF-IDF

Hasil analisis *TF-IDF* menunjukkan bahwa istilah-istilah dengan skor tertinggi memberikan informasi penting mengenai kata-kata yang paling relevan dalam analisis sentimen. Kata-kata seperti "aplikasi", "mudah", "saya", dan "sangat" memiliki skor *TF-IDF* tertinggi, mengindikasikan kemunculan yang sering dan relevansi yang kuat dalam berbagai ulasan, khususnya terkait pengalaman pengguna. Selain itu, istilah khusus seperti "indodax" dan "investasi" menyoroti topik utama yang menjadi fokus pembahasan, yaitu aplikasi dan layanan investasi. Beberapa kata,

seperti "mudah", "bagus", "baik", dan "guna", sangat terkait dengan sentimen positif, menunjukkan pengaruh yang besar dalam mengklasifikasikan sentimen positif. Sementara itu, kata-kata seperti "tidak", "buat", dan "yg" memberikan konteks tambahan yang bisa bersifat netral atau negatif. Analisis ini menegaskan bahwa *TF-IDF* efektif dalam mengidentifikasi dan memprioritaskan kata-kata yang relevan untuk klasifikasi sentimen, sehingga membantu meningkatkan akurasi model. Dengan demikian, *TF-IDF* menjadi alat penting untuk menangkap istilah-istilah kunci yang mendukung pemahaman sentimen pengguna dalam *dataset* ulasan. Gambar 2 merupakan 20 istilah dengan skor tertinggi *TF-IDF*.

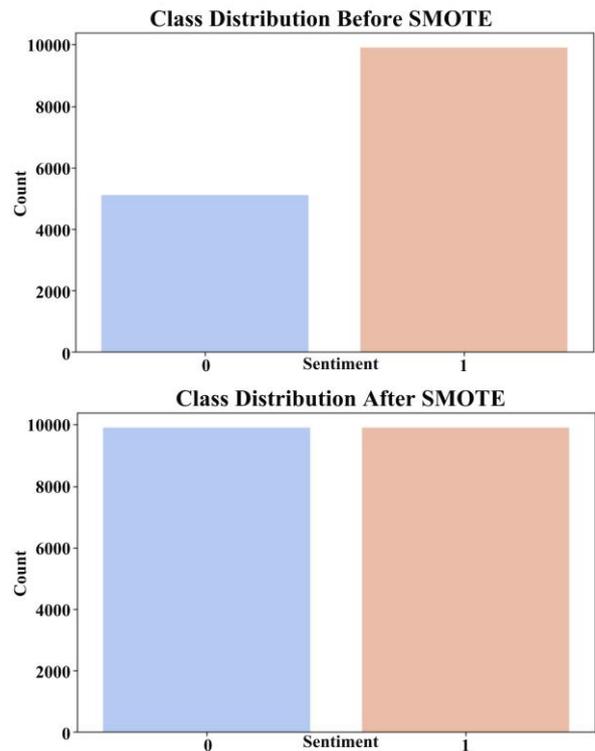


Gambar 2. 20 Istilah Dengan Skor Tertinggi

B. Penanganan Data Ketidakseimbangan (*SMOTE*)

Dalam mengatasi ketidakseimbangan kelas awal, dimana *dataset* terdiri dari 5.097 contoh kelas negatif dan 9.903 contoh kelas positif, penerapan *SMOTE* berhasil menyamakan distribusi kelas. Dengan menciptakan sampel sintetis untuk kelas negatif yang lebih sedikit, *SMOTE* meningkatkan jumlahnya hingga setara dengan kelas positif, yaitu 9.903 contoh, sebagaimana ditampilkan pada Gambar 3. Pendekatan ini mengurangi risiko bias dalam prediksi model yang cenderung menguntungkan kelas mayoritas, sehingga memastikan proses pelatihan yang lebih andal. Dengan memperbaiki keseimbangan dalam *dataset*, *SMOTE* memungkinkan model untuk belajar secara efektif dari kedua kelas dan menghasilkan klasifikasi yang akurat, yang pada akhirnya meningkatkan performa dan keandalan prediksi. Hasil ini membuktikan bahwa *SMOTE* telah berhasil mengatasi ketidakseimbangan kelas, menghasilkan metrik performa model yang lebih seimbang, yang sangat penting dalam aplikasi yang

memerlukan sensitivitas dan spesifisitas yang setara. Gambar 3 merupakan distribusi kelas sebelum dan sesudah *SMOTE*.



Gambar 3. Grafik Distribusi Kelas Sebelum dan Seudah *SMOTE*

C. Membangun Model Hybrid

Pada dua model *deep learning* yang digunakan, *Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM)* dan *Convolutional Neural Network-Gated Recurrent Unit (CNN-GRU)*, digunakan untuk klasifikasi data biner. *CNN-LSTM* menggabungkan *CNN* untuk ekstraksi fitur dan *LSTM* untuk menangkap hubungan temporal, sementara *CNN-GRU* menggantikan *LSTM* dengan *GRU* untuk efisiensi komputasi. Kedua model menggunakan embedding dengan input dimensi 5000, fungsi aktivasi *sigmoid* pada output, *optimizer Adam*, dan diukur berdasarkan akurasi. Setelah dilatih selama lima *epoch*, *CNN-LSTM* mencapai akurasi pelatihan 97,45% dengan *val_accuracy* 88,87%, sedangkan *CNN-GRU* mencapai akurasi pelatihan 97,08% dengan *val_accuracy* lebih tinggi, yaitu 89,37%. *CNN-LSTM* unggul dalam akurasi pelatihan, tetapi *CNN-GRU* menunjukkan kestabilan lebih baik pada data validasi, menjadikannya lebih andal dalam generalisasi. Gambar 4 merupakan hasil pelatihan model dalam lima *epoch*

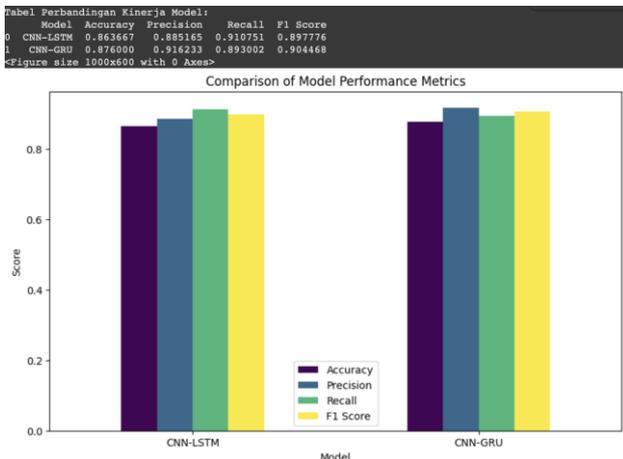
```

/usr/local/lib/python3.11/dist-packages/keras/src/layers/core/embedding.py:98: UserWarning: Argument 'input_length' is
warnings.warn(
Epoch 1/5 42s 182ms/step - accuracy: 0.7952 - loss: 0.4862 - val_accuracy: 0.8923 - val_loss: 0.2567
375/375
Epoch 2/5 37s 99ms/step - accuracy: 0.9182 - loss: 0.2838 - val_accuracy: 0.8887 - val_loss: 0.2598
375/375
Epoch 3/5 37s 99ms/step - accuracy: 0.9423 - loss: 0.1589 - val_accuracy: 0.8887 - val_loss: 0.2883
375/375
Epoch 4/5 39s 95ms/step - accuracy: 0.9618 - loss: 0.1119 - val_accuracy: 0.8797 - val_loss: 0.3263
375/375
Epoch 5/5 41s 96ms/step - accuracy: 0.9745 - loss: 0.0785 - val_accuracy: 0.8637 - val_loss: 0.4855
375/375
Epoch 1/5 37s 88ms/step - accuracy: 0.7924 - loss: 0.4153 - val_accuracy: 0.8937 - val_loss: 0.2575
375/375
Epoch 2/5 33s 88ms/step - accuracy: 0.9168 - loss: 0.2897 - val_accuracy: 0.8923 - val_loss: 0.2656
375/375
Epoch 3/5 48s 86ms/step - accuracy: 0.9457 - loss: 0.1465 - val_accuracy: 0.8817 - val_loss: 0.2897
375/375
Epoch 4/5 41s 87ms/step - accuracy: 0.9708 - loss: 0.0853 - val_accuracy: 0.8880 - val_loss: 0.3827
375/375
Epoch 5/5 34s 89ms/step - accuracy: 0.9871 - loss: 0.0418 - val_accuracy: 0.8768 - val_loss: 0.4363
375/375
    
```

Gambar 4. Hasil Pelatihan Model Epoch

D. Evaluasi Model

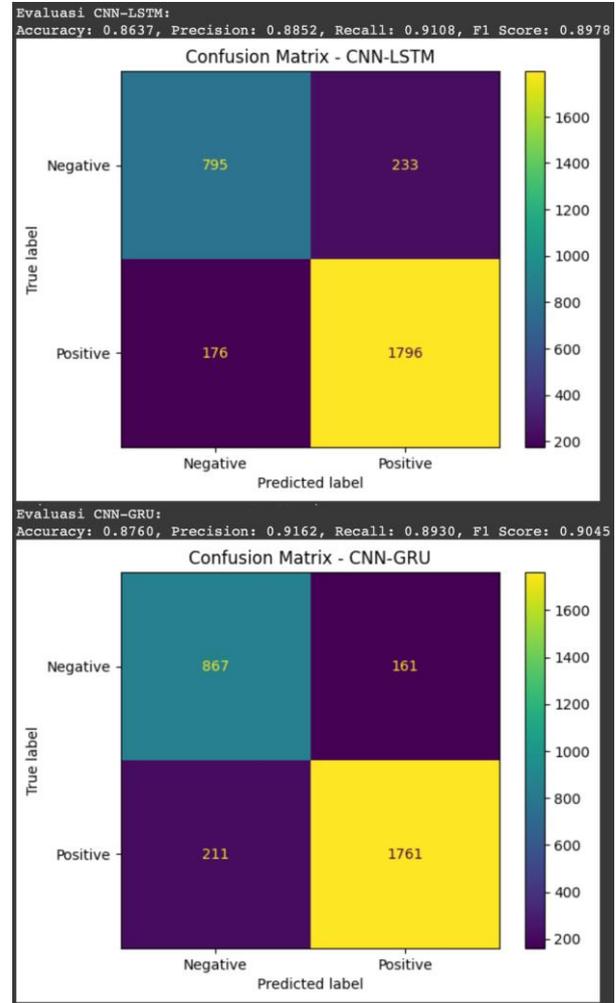
Evaluasi terhadap dua model, *Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM)* dan *Convolutional Neural Network-Gated Recurrent Unit (CNN-GRU)*, menunjukkan performa berdasarkan empat metrik utama: akurasi, presisi, *recall*, dan *f1-score*. Model *CNN-LSTM* mencapai akurasi sebesar 86,37%, presisi 88,52%, *recall* 91,08%, dan skor *f1-score* 89,78%. Sementara itu, *CNN-GRU* menunjukkan kinerja yang lebih baik secara keseluruhan dengan akurasi 87,60%, presisi 91,62%, *recall* 89,30%, dan skor *f1-score* 90,45%. Hasil ini menunjukkan bahwa *CNN-GRU* memiliki keunggulan terutama dalam presisi dan *f1-score*, meskipun *CNN-LSTM* sedikit lebih unggul dalam *recall* untuk ulasan positif. Gambar 5 merupakan perbandingan metrik kinerja model.



Gambar 5. Perbandingan Metrik Kinerja Model

Berdasarkan analisis *confusion matrix*, *CNN-GRU* lebih efisien dalam mendeteksi ulasan negatif, ditunjukkan dengan nilai *true negative* sebesar 867 dan *false positive* yang lebih rendah, yaitu 161. Di sisi lain, *CNN-LSTM* unggul dalam mendeteksi ulasan positif dengan *true positive* sebesar 1796 dan *false negative* yang lebih kecil, yaitu 176. Hal ini mencerminkan bahwa *CNN-LSTM* lebih akurat dalam menangkap sentimen

positif, sementara *CNN-GRU* menunjukkan kinerja yang lebih seimbang antara mendeteksi sentimen negatif dan positif. Gambar 6 merupakan perbandingan analisis *confusion matrix*.



Gambar 6. Perbandingan Analisis Confusion Matrix

E. Analisis dan Perbandingan

Secara keseluruhan, *CNN-GRU* tampil sebagai model yang lebih unggul untuk sebagian besar metrik evaluasi, memberikan keseimbangan yang baik antara presisi, *recall*, dan *f1-score*. Namun, pemilihan model tetap harus disesuaikan dengan tujuan spesifik aplikasi. Jika fokus utama adalah mengurangi kesalahan pada ulasan negatif, *CNN-GRU* menjadi pilihan yang lebih baik. Namun, jika prioritasnya adalah memastikan semua ulasan positif terdeteksi dengan akurasi tinggi, *CNN-LSTM* mungkin menjadi model yang lebih cocok.

V. KESIMPULAN [7]

Hasil evaluasi menunjukkan bahwa model *Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM)* dan *Convolutional Neural Network-Gated Recurrent Unit (CNN-GRU)* sama-sama memiliki performa yang baik dalam analisis sentimen, dengan keunggulan masing-masing. *CNN-LSTM* unggul dalam *recall*, terutama untuk mendeteksi ulasan positif, dengan akurasi sebesar 86,37% dan *f1-score* sebesar 89,78%. Sementara itu, *CNN-GRU* memiliki presisi dan *f1-score* yang lebih tinggi, yakni 91,62% dan 90,45%, serta akurasi keseluruhan sebesar 87,60%, menjadikannya lebih seimbang dalam mendeteksi ulasan negatif dan positif. Berdasarkan hasil ini, *CNN-GRU* lebih direkomendasikan untuk aplikasi yang membutuhkan keseimbangan performa di kedua kategori sentimen. Penelitian selanjutnya, disarankan untuk melakukan penyetelan parameter (*hyperparameter tuning*) dan menguji kedua model pada *dataset* yang lebih besar dan beragam untuk mengukur konsistensi kinerja. Selain itu, eksplorasi model berbasis *transformer*, seperti *BERT*, dapat dilakukan untuk membandingkan efektivitas dan meningkatkan akurasi analisis sentimen. [13]

DAFTAR PUSTAKA [14]

- [1] R. Wijayanti and A. Arisal, "Automatic Indonesian sentiment lexicon curation with sentiment valence tuning for social media sentiment analysis," *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, vol. 20, no. 1, pp. 1–16, 2021. [18]
- [2] D. Fimoza, A. Amalia, and T. H. F. Harumy, "Sentiment analysis for movie review in Bahasa Indonesia using BERT," in *2021 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA)*, IEEE, 2021, pp. 27–34. [16]
- [3] Y. Fauziah, B. Yuwono, and A. S. Aribowo, "Lexicon Based Sentiment Analysis in Indonesia Languages: A Systematic Literature Review," in *RSF Conference Series Engineering and Technology*, 2021, pp. 363–367. [17]
- [4] M. D. Mahardhika and N. Asandimitra, "Pengaruh overconfidence, risk tolerance, return, financial literacy, financial technology terhadap keputusan investasi yang dilakukan mahasiswa Surabaya," *Jurnal Ilmu Manajemen*, pp. 602–613, 2023.
- [5] H. Hasanudin, A. Nurwulandari, and R. K. Safitri, "Pengaruh pengetahuan investasi, motivasi dan pelatihan pasar modal terhadap keputusan investasi yang dimediasi oleh minat investasi," *Jurnal Ilmiah Manajemen, Ekonomi, & Akuntansi (MEA)*, vol. 5, no. 3, pp. 494–512, 2021.
- [6] E. Eskiyaturrofikoh and R. R. Suryono, "Analisis Sentimen Aplikasi X Pada Google Play Store Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine (SVM)," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 3, pp. 1408–1419, 2024.
- F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, vol. 6, no. 9, pp. 4305–4313, 2022.
- [8] M. Fadli, V. Wijaya, M. R. Pribadi, and W. Widhiarso, "Effect of TF-IDF Extraction and Application of SMOTE on Model Performance in Detecting Spam Email," in *2023 10th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, IEEE, 2023, pp. 637–641.
- [9] Y. A. Singgalen, "Enhancing Sentiment Analysis of Garden by the Bay Reviews on TripAdvisor Platform Using CRISP-DM through DT and SVM with SMOTE," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 2, p. 1188, Apr. 2024, doi: 10.30865/mib.v8i2.7485.
- [10] N. Hossain, M. R. Bhuiyan, Z. N. Tumpa, and S. A. Hossain, "Sentiment analysis of restaurant reviews using combined CNN-LSTM," in *2020 11th International conference on computing, communication and networking technologies (ICCCNT)*, IEEE, 2020, pp. 1–5.
- [11] A. Z. R. Adam and E. B. Setiawan, "Social media sentiment analysis using convolutional neural network (CNN) dan gated recurrent unit (GRU)," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 1, pp. 119–131, 2023.
- [11] F. Moodi, A. J. Rafsanjani, S. Zarifzadeh, and M. A. Z. Chahooki, "Fusion of technical indicators and sentiment analysis in a hybrid framework of deep learning models for stock price movement prediction," *IEEE Access*, 2024.
- E. Werdiningsih, "B, AH (2022). Lima Pendekatan dalam Penelitian Kualitatif," *Likhitaprajna Jurnal Ilmiah*, vol. 24, no. 1, pp. 39–50.
- [12] S. Singh, K. Kumar, and B. Kumar, "Sentiment analysis of Twitter data using TF-IDF and machine learning techniques," in *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, IEEE, 2022, pp. 252–255.
- [13] W. A. Prabowo and F. Azizah, "Sentiment analysis for detecting cyberbullying using tf-idf and svm," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 6, pp. 1142–1148, 2020.
- [14] D. P. Chatterjee, S. Mukhopadhyay, S. Goswami, and P. K. Panigrahi, "Efficacy of oversampling over machine learning algorithms in case of sentiment analysis," in *Data Management, Analytics and Innovation: Proceedings of ICDMAI 2020, Volume 2*, Springer, 2021, pp. 247–260.
- [15] M. F. Avidiansyah and E. B. Setiawan, "The Influence of Sentiment on the Movement of Bank BCA (BBCA) Shares with the CNN-GRU Classification Model and Feature Expansion with GloVe," in *2023 International Conference on Data Science and Its Applications (ICoDSA)*, IEEE, 2023, pp. 105–110.