

Optimalisasi Model BioBERT untuk Pengenalan Entitas pada Teks Medis dengan Conditional Random Fields (CRF)

Cynthia Dwi Nafanda, Abu Salam*

Fakultas Ilmu Komputer, Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia

Email: ¹111202113699@mhs.dinus.ac.id, ^{2,*}abu.salam@dsn.dinus.ac.id

Email Penulis Korespondensi: abu.salam@dsn.dinus.ac.id

Submitted: 21/02/2025; Accepted: 05/03/2025; Published: 07/03/2025

Abstrak—Penelitian ini mengevaluasi kinerja berbagai model dalam tugas Named Entity Recognition (NER) untuk entitas medis, dengan fokus pada dataset yang tidak seimbang. Enam konfigurasi model BioBERT diuji, yang melibatkan teknik optimasi seperti Class Weight, Conditional Random Fields (CRF), dan Hyperparameter Tuning. Evaluasi dilakukan menggunakan metrik Precision, Recall, dan F1-Score, yang relevan dalam konteks NER, terutama untuk mengatasi ketidakseimbangan kelas dalam data. Dataset yang digunakan adalah BC5CDR, yang berfokus pada entitas chemical (zat kimia) dan disease (penyakit) dalam teks medis tidak terstruktur dari PubMed. Data telah dibagi menjadi tiga bagian: train dataset untuk pelatihan, valid dataset untuk validasi, dan test dataset untuk pengujian. Pembagian dataset dilakukan secara seimbang untuk memastikan model diuji tanpa bias, sehingga hasil yang diperoleh lebih akurat dan dapat digunakan sebagai acuan dalam pengembangan sistem NER medis yang lebih efisien. Berdasarkan hasil evaluasi, menunjukkan bahwa BioBERT + CRF merupakan model dengan F1-Score yang mencerminkan keseimbangan optimal antara Precision (peringkat 3, 0.6067 pada B-Chemical, 0.5594 pada B-Disease, 0.4600 pada I-Disease, dan 0.5083 pada I-Chemical) dan Recall (peringkat 3, 0.5580 pada B-Chemical, 0.4491 pada B-Disease, 0.5718 pada I-Disease, dan 0.3840 pada I-Chemical) dibandingkan dengan model lainnya. Model ini terbukti lebih akurat dalam mendeteksi entitas medis tanpa mengorbankan ketepatan prediksi. Stabilitas model juga lebih baik dengan selisih Precision dan Recall yang lebih kecil, menjadikannya pilihan terbaik untuk NER dalam teks medis. Penerapan teknik early stopping efektif mencegah overfitting, memastikan model belajar optimal tanpa kehilangan generalisasi. Dengan keseimbangan yang lebih baik dalam mengenali entitas medis dari teks tidak terstruktur, model ini menjadi pendekatan yang paling efektif untuk sistem NER di domain medis.

Kata Kunci: Named Entity Recognition (NER); BioBERT; Conditional Random Fields (CRF); Class Weight; Hyperparameter Tuning

Abstract—This research evaluates the performance of various models in the Named Entity Recognition (NER) task for medical entities, focusing on imbalanced datasets. Six BioBERT model configurations were tested, incorporating optimization techniques such as Class Weight, Conditional Random Fields (CRF), and Hyperparameter Tuning. The evaluation was conducted using Precision, Recall, and F1-Score metrics, which are particularly relevant in the context of NER, especially for addressing class imbalance in the data. The dataset used is BC5CDR, which targets chemical and disease entities in unstructured medical texts from PubMed. The data was divided into three parts: a training dataset for model training, a validation dataset for model tuning, and a test dataset for performance evaluation. The dataset was split evenly to ensure unbiased model testing, leading to more accurate results that can serve as a reference for developing more efficient medical NER systems. The evaluation results indicate that BioBERT + CRF is the model with an F1-Score that reflects an optimal balance between Precision (ranked 3rd, 0.6067 for B-Chemical, 0.5594 for B-Disease, 0.4600 for I-Disease, and 0.5083 for I-Chemical) and Recall (ranked 3rd, 0.5580 for B-Chemical, 0.4491 for B-Disease, 0.5718 for I-Disease, and 0.3840 for I-Chemical) compared to other models. This model proved to be more accurate in detecting medical entities without compromising prediction precision. The model's stability is also enhanced by a smaller gap between Precision and Recall, making it the best choice for NER in medical texts. The application of early stopping techniques effectively prevented overfitting, ensuring the model learned optimally without losing generalization. With better balance in recognizing medical entities from unstructured texts, this model presents the most effective approach for NER systems in the medical domain.

Keywords: Named Entity Recognition (NER); BioBERT; Conditional Random Fields (CRF); Class Weight; Hyperparameter Tuning

1. PENDAHULUAN

Named Entity Recognition (NER) adalah teknik dalam Natural Language Processing (NLP) yang digunakan untuk mengidentifikasi dan mengklasifikasikan entitas penting dalam teks, seperti nama orang, organisasi, tempat, serta entitas medis seperti penyakit dan zat kimia [1], [2]. Dalam konteks teks medis, NER sangat penting karena dapat mengotomatisasi proses ekstraksi informasi medis dari teks yang tidak terstruktur, seperti yang ditemukan dalam Rekam Medis Elektronik [3]. EHR berfungsi sebagai database medis yang menyimpan informasi vital tentang pasien, termasuk riwayat kesehatan, diagnosis, pengobatan, dan prosedur medis [4]. Namun, sekitar 80% dari data dalam EHR adalah informasi tidak terstruktur, yang sulit dianalisis secara otomatis [5]. Oleh karena itu, penggunaan teknik NER untuk mengekstrak entitas seperti chemical (zat kimia) dan disease (penyakit) secara otomatis menjadi sangat penting untuk meningkatkan efisiensi dan akurasi dalam diagnosis dan pengelolaan data medis.

Urgensi penerapan NER pada EHR semakin tinggi, karena pengenalan entitas medis dalam data yang tidak terstruktur dapat mempercepat pengambilan keputusan klinis yang lebih baik [5], [6]. Salah satu tantangan utama adalah pengenalan entitas medis seperti chemical (zat kimia) dan disease (penyakit), yang sering tercatat dengan variasi istilah, singkatan, atau kesalahan penulisan [1]. Oleh karena itu, sangat penting untuk menggunakan sistem

otomatis yang mampu mengenali dan mengekstrak entitas-entitas ini secara akurat dalam teks medis, yang berperan penting dalam pengelolaan pasien dan diagnosis berbasis bukti.

Pada penelitian ini, untuk menghadapi tantangan ini, BERT (Bidirectional Encoder Representations from Transformers) merupakan model transformer-based yang sangat efektif dalam NLP, terutama untuk tugas-tugas seperti NER [4]. BERT dilatih menggunakan pendekatan bidirectional, yang memungkinkan model untuk memahami konteks kata dalam kalimat dengan mempertimbangkan informasi dari kata sebelumnya dan sesudahnya [7]. Ini memberikan keuntungan besar dibandingkan model sebelumnya yang hanya memproses teks secara unidirectional. Sebagai model yang fleksibel, BERT digunakan dalam berbagai tugas NLP, seperti klasifikasi teks, analisis sentimen, dan tentunya pengenalan entitas, tanpa perlu penyesuaian besar pada arsitektur model [7]. Dalam konteks data medis, BERT dapat mengotomatisasi proses ekstraksi entitas seperti chemical (zat kimia) dan disease (penyakit) dari teks medis, yang sangat berguna dalam pengelolaan Rekam Medis Elektronik yang tidak terstruktur [2]. Namun, meskipun BERT telah terbukti efektif, ia masih menghadapi beberapa tantangan dalam mengelola konteks medis yang lebih kompleks.

Sebagai pengembangan dari BERT, BioBERT dirancang khusus untuk teks biomedis dengan menggunakan data pelatihan dari PubMed dan PMC. BioBERT dirancang untuk menangani teks biomedis yang lebih kompleks, dan telah terbukti memberikan performa yang lebih baik dibandingkan dengan BERT dalam mengenali entitas medis, termasuk chemical (zat kimia) dan disease (penyakit) [6]. Dengan pemahaman yang lebih mendalam mengenai konteks medis, BioBERT mampu meningkatkan akurasi dalam mengekstrak informasi medis, yang sangat penting dalam proses diagnosis berbasis bukti [7].

Namun, meskipun BioBERT memberikan hasil yang lebih baik dibandingkan dengan BERT, ia masih memiliki kekurangan, terutama dalam mengenali entitas medis yang lebih kompleks atau entitas *nested* (yang saling terkandung) serta menghadapi kesulitan dalam menangani class imbalance dalam dataset medis [1]. Ketidakseimbangan kelas ini menyebabkan model kesulitan untuk mengenali entitas yang jarang ditemukan dalam data, seperti beberapa penyakit langka atau zat kimia tertentu [6]. Penelitian sebelumnya menunjukkan bahwa meskipun berbagai model berbasis pembelajaran mesin dapat digunakan untuk memproses teks medis yang tidak terstruktur, tantangan besar tetap ada dalam menangani kompleksitas dan ketidakpastian dalam data tersebut [5]. Oleh karena itu, pengolahan informasi yang lebih canggih diperlukan untuk membantu pengambilan keputusan berbasis bukti dalam dunia medis. Berbagai pendekatan berbasis NER telah banyak dikembangkan untuk mengidentifikasi dan mengklasifikasikan entitas medis penting dalam teks medis.

Berdasarkan berbagai tantangan tersebut, penelitian ini bertujuan untuk mengembangkan model NER berbasis BioBERT yang dioptimalkan dengan teknik tambahan seperti Conditional Random Fields (CRF), class weight, dan hyperparameter tuning. Penelitian ini bertujuan untuk meningkatkan kemampuan model dalam mengenali entitas medis dengan lebih akurat, khususnya dalam dataset yang tidak seimbang dan pada entitas yang kompleks atau jarang muncul. CRF berfungsi untuk menangani ketergantungan antar token dalam urutan teks, yang sangat relevan dalam tugas NER [8]. Dalam teks medis, entitas medis sering saling bergantung satu sama lain, sehingga penambahan CRF dapat meningkatkan konsistensi dalam mengenali entitas-entitas tersebut [9].

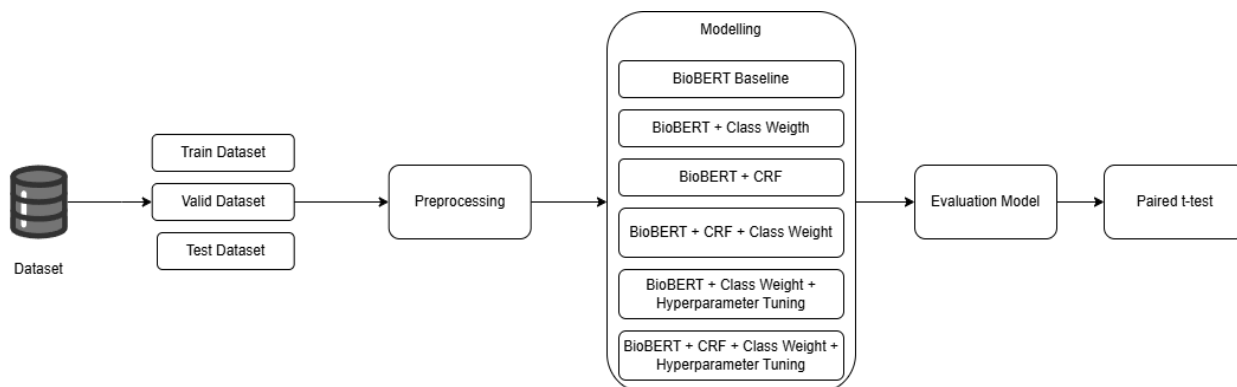
Selain itu, penerapan class weight juga sangat penting untuk mengatasi masalah ketidakseimbangan kelas [10], [11]. Dengan memberikan bobot lebih pada kelas minoritas, model dapat lebih sensitif dalam mengenali entitas yang jarang ditemukan dalam data, seperti penyakit atau zat kimia tertentu. Hyperparameter tuning juga digunakan untuk menyesuaikan parameter model agar lebih optimal dalam menghadapi kompleksitas data medis yang lebih beragam, meningkatkan akurasi model dalam mengenali entitas medis yang lebih kompleks dan beragam [12]. Dengan penyesuaian yang tepat pada hyperparameter, model dapat lebih responsif terhadap variasi dalam data, seperti entitas medis yang berlapis atau tidak terstruktur, sehingga meningkatkan kemampuan model untuk mengidentifikasi entitas yang relevan dengan lebih akurat dalam konteks medis yang dinamis [13].

Dengan mengombinasikan BioBERT, CRF, class weight, dan hyperparameter tuning, model ini dapat dioptimalkan untuk memberikan akurasi yang lebih tinggi dalam mengekstrak entitas medis dari teks medis yang tidak terstruktur. Kontribusi utama dari penelitian ini adalah mengusulkan metode optimasi pada model BioBERT untuk tugas NER pada teks medis yang tidak terstruktur, serta memberikan solusi konkret terhadap masalah class imbalance dan kompleksitas entitas medis. Selain itu, penelitian ini diharapkan memberikan kontribusi nyata dalam meningkatkan kualitas pengelolaan data medis, mendukung proses diagnosis yang lebih akurat, dan mempercepat pengambilan keputusan klinis berbasis bukti. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi dalam pengembangan model NER yang lebih baik, tetapi juga mendukung peningkatan layanan kesehatan melalui pemanfaatan teknologi NLP dalam bidang medis.

2. METODOLOGI PENELITIAN

Metodologi penelitian menggambarkan serangkaian langkah yang diambil oleh peneliti dalam melaksanakan suatu penelitian. Hal ini menjelaskan urutan proses yang dilakukan dengan tahapan-tahapan yang digunakan dalam penelitian ini yang digambarkan pada Gambar 1 mencakup serangkaian tahapan mulai dari pengolahan dataset hingga evaluasi kinerja model. Pada Gambar 1, proses dimulai dengan dataset BCDCDR, yang dibagi menjadi tiga bagian: Train Dataset, Valid Dataset, dan Test Dataset. Selanjutnya, data melalui tahap Preprocessing untuk mempersiapkan

data sebelum dimasukkan ke dalam model. Beberapa model diterapkan, dimulai dengan BioBERT Baseline, kemudian menggabungkan beberapa teknik, seperti Class Weight untuk menangani ketidakseimbangan kelas, dan CRF untuk menangani ketergantungan antar token dalam teks. Selanjutnya, dilakukan penyesuaian Hyperparameter Tuning untuk meningkatkan performa model. Setelah itu, semua model yang diterapkan dievaluasi menggunakan Uji paired t-test untuk membandingkan hasil model dan memilih yang terbaik.



Gambar 1. Alur Penelitian

2.1 Dataset

Dataset yang digunakan adalah BC5CDR, yang berfokus pada entitas chemical (zat kimia) dan disease (penyakit) dalam teks medis tidak terstruktur yang terdiri dari abstrak-artikel ilmiah dari PubMed, yang telah dianotasi dengan format BIO tagging. Seperti yang dijelaskan pada Tabel 1, dataset ini terdiri dari lima label utama untuk analisis NER, yaitu label "O" untuk token yang bukan entitas, "B-Chemical" dan "B-Disease" untuk menandai awal entitas bahan kimia dan penyakit, serta "I-Disease" dan "I-Chemical" yang menunjukkan bagian dalam entitas penyakit dan bahan kimia. Jumlah kemunculan setiap label dapat dilihat pada Tabel 1.

Dataset ini telah dibagi menjadi tiga bagian utama, yaitu data pelatihan (train dataset), data validasi (valid dataset), dan data pengujian (test dataset), yang masing-masing digunakan untuk melatih model, menyesuaikan parameter, serta mengevaluasi kinerja model setelah pelatihan. Rincian jumlah sampel dalam setiap bagian dapat dilihat pada Tabel 2, yang menggambarkan pembagian dataset secara proporsional agar model dapat diuji secara adil tanpa bias. Proses pembagian dataset dilakukan menggunakan teknik random split untuk memastikan distribusi data yang merata dan mencegah bias dalam pelatihan model. Teknik random split ini membagi dataset secara acak ke dalam train, valid, dan test dataset sesuai proporsi yang telah ditentukan. Selain itu, pengujian model juga dilakukan dengan menggunakan data validasi yang telah disediakan oleh public dataset BCD5CDR. Hal ini bertujuan untuk mengevaluasi performa model dalam skenario dunia nyata dan memastikan model tidak hanya bekerja baik pada data pelatihan tetapi juga pada data eksternal yang belum pernah dilihat sebelumnya. Penggunaan data validasi publik ini memberikan standar pengujian yang lebih objektif dan memastikan generalisasi model yang lebih baik.

Tabel 1. Deskripsi Label pada Dataset

| Label | Deskripsi | Kode Label | Train | Valid | Test |
|------------|-------------------------------|------------|--------|--------|---------|
| O | Token yang bukan entitas | 0 | 96.796 | 96.413 | 103.684 |
| B-Chemical | Awal dari entitas kimia | 1 | 5.203 | 5.347 | 5.385 |
| B-Disease | Awal dari entitas penyakit | 2 | 4.182 | 4.244 | 4.424 |
| I-Disease | Bagian dalam entitas penyakit | 3 | 2.570 | 2.416 | 2.424 |
| I-Chemical | Bagian dalam entitas kimia | 4 | 571 | 538 | 401 |

Tabel 2. Pembagian Dataset

| Dataset | Jumlah Data |
|---------------|-------------|
| Train Dataset | 5228 |
| Valid Dataset | 5330 |
| Test Dataset | 5865 |

2.2 Preprocessing

Pada tahap preprocessing, teks dalam dataset di-tokenisasi menggunakan BERTTokenizer, yang mengimplementasikan teknik WordPiece Tokenization. Teknik ini memecah kata-kata yang tidak ada dalam kosakata standar menjadi sub-kata. Hal ini memungkinkan model untuk memecah kata-kata yang tidak ada dalam kosakata standar menjadi sub-kata [14]. Berbeda dengan analisis teks umum, dalam teks medis, stopwords dipertahankan karena sering mengandung informasi penting, seperti kata penghubung yang menjelaskan hubungan antar gejala, diagnosis, atau prosedur medis. Teknik ini diterapkan dalam penelitian ini untuk memastikan bahwa model BioBERT dapat memproses teks medis dengan konteks yang lebih tepat. Penghapusan stopwords dalam teks medis dapat

menghilangkan informasi yang berharga, sebagaimana ditemukan pada penelitian sebelumnya yang menunjukkan bahwa dalam teks biomedis, stopwords sering mengandung konteks yang relevan dan tidak seharusnya dihapus [15].

2.3 Model BioBERT

Setelah melalui tahap preprocessing, model-model pretrained seperti BioBERT dan variasinya diterapkan untuk tugas Named Entity Recognition (NER) pada data medis. Model-model ini dipilih karena kemampuannya dalam mengenali entitas medis yang kompleks seperti chemical (zat kimia) dan disease (penyakit). BioBERT, yang merupakan pengembangan dari BERT, telah terbukti efektif dalam memahami konteks medis berkat pelatihan menggunakan data biomedis, seperti yang ditemukan di PubMed [7]. Meskipun BioBERT memberikan kinerja yang baik pada entitas medis yang sering muncul, model ini menunjukkan kekurangan, terutama pada kelas B-Chemical, B-Disease, dan I-Disease, dengan potensi overfitting pada kelas 0 atau non entity. Hal ini menunjukkan bahwa BioBERT kesulitan dalam mengenali entitas yang jarang muncul atau lebih kompleks, yang dapat mempengaruhi akurasi prediksi model secara keseluruhan.

2.4 Penerapan Class Weight

Pada penelitian ini, untuk mengatasi masalah ketidakseimbangan kelas, yang menyebabkan kinerja kurang optimal pada beberapa kelas, seperti entitas medis tertentu (misalnya B-Chemical, B-Disease, dan I-Chemical) penerapan class weight dilakukan. Menurut penelitian sebelumnya, penggunaan class weight terbukti efektif dalam menangani ketidakseimbangan kelas dengan memberikan bobot lebih pada kelas yang jarang muncul [16].

Pembobotan kelas dihitung menggunakan fungsi `compute_class_weight` dari scikit-learn, yang memberikan bobot lebih pada kelas yang lebih jarang muncul, sehingga model lebih fokus pada kelas tersebut. Bobot kelas ini kemudian dikonversi menjadi tensor dengan tipe float, yang memungkinkan penggunaannya dalam fungsi loss saat pelatihan. Dalam implementasi ini, fungsi `CrossEntropyLoss` menerima parameter `weight`, yang mengatur seberapa besar pengaruh setiap kelas dalam perhitungan loss, dengan kelas yang lebih jarang mendapatkan bobot lebih tinggi. Penerapan class weight ini terbukti efektif dalam meningkatkan Recall pada kelas minoritas, karena model diberi insentif untuk lebih fokus mengenali entitas medis yang kurang terwakili dalam dataset pelatihan. Namun, meskipun ada peningkatan Recall, ada sedikit penurunan pada Precision, yang menunjukkan bahwa meskipun model lebih sensitif dalam mendeteksi entitas yang relevan, hal ini dapat menyebabkan lebih banyak kesalahan prediksi atau false positives. Secara keseluruhan, penggunaan class weight ini berhasil meningkatkan kinerja model pada kelas-kelas yang kurang terwakili, meskipun dengan trade-off pada Precision.

2.5 Penggunaan CRF

Penggunaan Conditional Random Fields (CRF) untuk menangani ketergantungan antar label dalam urutan teks yang memungkinkan model untuk memprediksi hubungan antar entitas yang lebih kompleks, yang sangat penting dalam NER, di mana entitas medis sering kali saling berhubungan dalam sebuah kalimat [9]. Dengan menambahkan CRF setelah output BioBERT, model dapat memanfaatkan konteks antar token dan meningkatkan konsistensi serta akurasi dalam prediksi entitas. Penerapan kombinasi BioBERT + CRF terbukti memberikan peningkatan signifikan pada Precision dan Recall, terutama untuk kelas B-Chemical dan B-Disease, yang menunjukkan bahwa CRF berperan penting dalam mengoptimalkan kinerja model pada entitas yang lebih sulit dikenali. Hal ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa penggunaan CRF dalam NER untuk teks medis dapat mengatasi ketergantungan antar entitas medis dan meningkatkan akurasi pengenalan entitas [8], [9].

2.6 Fine-tuning dan Optimisasi

Dalam penelitian ini, dilakukan fine-tuning pada model dengan kombinasi BioBERT + Class Weighting + Hyperparameter Tuning dan BioBERT + CRF + Class Weighting + Hyperparameter Tuning, untuk meningkatkan performa model lebih lanjut. Fine-tuning ini dilakukan untuk menyesuaikan parameter model agar lebih spesifik terhadap dataset medis yang digunakan, memungkinkan model untuk lebih baik dalam mengenali entitas medis yang lebih kompleks dan beragam. Proses fine-tuning ini diharapkan dapat memperbaiki F1-Score pada prediksi entitas medis. Penelitian sebelumnya menunjukkan bahwa fine-tuning pada model BERT dapat meningkatkan performa dalam pengenalan entitas berbasis materi [13].

Dalam fine-tuning model, beberapa hyperparameter dioptimalkan untuk mencapai hasil terbaik. Learning rate (lr) disesuaikan untuk memastikan model dapat belajar dengan efisien. Nilai lr dicari dalam rentang $1e-5$ hingga $5e-5$ melalui pencarian hyperparameter menggunakan Optuna, yang memungkinkan penyesuaian dinamis terhadap kecepatan pembelajaran selama pelatihan. Batch size yang digunakan dipilih antara 8, 16, dan 32, untuk memastikan keseimbangan antara kecepatan pelatihan dan penggunaan memori, sementara dropout rate dicari dengan nilai 0.1, 0.3, dan 0.5 untuk mengurangi overfitting dengan menonaktifkan sebagian neuron selama pelatihan, yang membantu meningkatkan kemampuan generalisasi model.

Selain itu, untuk mencegah overfitting selama proses pelatihan, diterapkan teknik early stopping. Teknik ini menghentikan pelatihan apabila tidak ada peningkatan signifikan pada accuracy atau loss dalam beberapa epoch berturut-turut, sehingga model tidak terlalu terfokus pada data pelatihan dan tetap dapat melakukan generalisasi yang

baik [17]. Hal ini sejalan dengan penelitian sebelumnya yang mengungkapkan bahwa early stopping efektif untuk menghindari overfitting pada model pembelajaran mesin [12], [18].

Optimisasi lebih lanjut dilakukan dengan menggunakan AdamW sebagai optimizer. AdamW memberikan stabilitas lebih baik dalam pelatihan dibandingkan dengan Adam standar, terutama ketika menggunakan weight decay. AdamW bekerja dengan memisahkan pembaruan weight decay dari aturan pembaruan gradien asli optimizer Adam, yang terbukti lebih efektif dalam mencegah overfitting dan meningkatkan generalisasi model, khususnya pada deep neural networks [19].

Untuk menangani ketidakseimbangan kelas yang lebih ekstrem, diterapkan focal loss yang memberi bobot lebih pada contoh yang lebih sulit dikenali [15]. Teknik ini meningkatkan akurasi prediksi untuk entitas medis yang jarang muncul. Selain itu, learning rate scheduler berbasis linear decay digunakan untuk menurunkan learning rate secara dinamis selama pelatihan, memastikan model belajar secara efisien tanpa kehilangan kestabilan dalam konvergennya. Kombinasi dari pengaturan parameter ini bertujuan untuk meningkatkan akurasi model, mengurangi overfitting, dan mempercepat konvergensi selama proses pelatihan, sesuai dengan studi yang disarankan pada penelitian sebelumnya [20].

2.7 Evaluasi Model

Evaluasi kinerja model dilakukan menggunakan tiga metrik utama, yaitu Precision, Recall, dan F1-Score, yang sangat relevan dalam tugas Named Entity Recognition (NER), terutama ketika menghadapi dataset yang tidak seimbang. Precision mengukur seberapa banyak entitas yang diprediksi benar-benar relevan dibandingkan dengan total entitas yang diprediksi sebagai relevan (1).

Recall mengukur seberapa banyak entitas relevan yang berhasil ditemukan oleh model dibandingkan dengan seluruh entitas relevan yang ada (2). Dan F1-Score menggabungkan kedua metrik tersebut untuk memberikan gambaran menyeluruh tentang keseimbangan antara Precision dan Recall (3). Selain itu, untuk memastikan hasil yang signifikan secara statistik, dilakukan uji paired t-test pada nilai F1-Score dari setiap konfigurasi model (4).

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (1)$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (2)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

$$t = \frac{\bar{d}}{s_d / \sqrt{n}} \quad (4)$$

Selain itu, evaluasi model untuk menghitung Rata-rata (Mean) dengan menjumlahkan semua nilai metrik (Precision, Recall, F1-Score) untuk setiap model dan label, lalu membaginya dengan jumlah total entri (jumlah model x jumlah label). Rumusnya adalah menjumlahkan semua nilai dan membaginya dengan total entri untuk mendapatkan gambaran umum kinerja model (5). Sementara itu, standar deviasi (Standard Deviation) mengukur seberapa besar variasi nilai metrik dari rata-ratanya. Ini dihitung dengan mengukur selisih antara setiap nilai dan rata-rata, mengkuadratkan selisih tersebut, lalu mengambil akar kuadrat dari rata-rata kuadrat selisih tersebut (7). Standar deviasi memberikan indikasi sejauh mana nilai-nilai tersebut tersebar atau konsisten di seluruh model dan label.

$$Rata - rata = \frac{\sum_{i=1}^n x_i}{n} \quad (5)$$

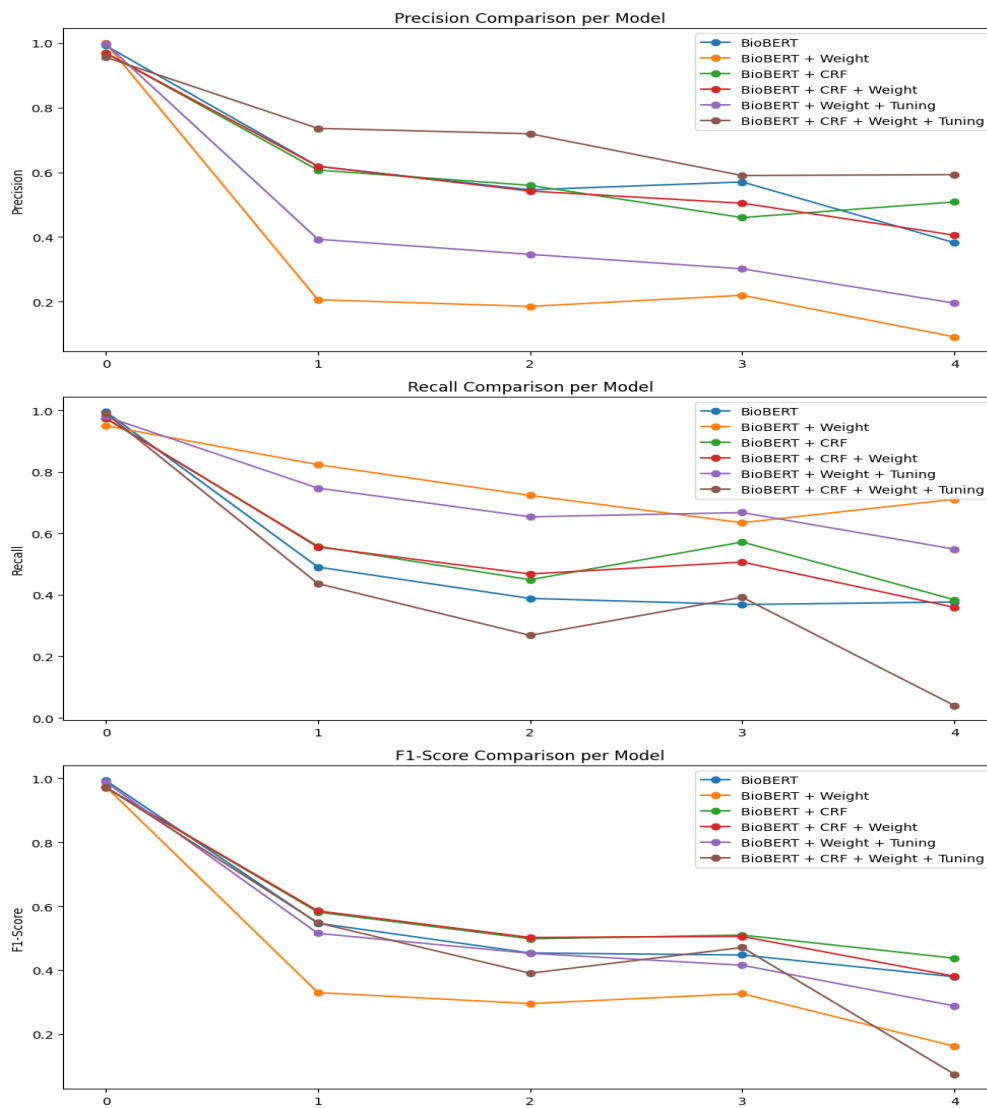
$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n \{x_i - \mu\}^2} \quad (6)$$

3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengevaluasi performa berbagai model dalam tugas Named Entity Recognition (NER) pada teks medis, dengan fokus pada identifikasi entitas chemical dan disease. Enam konfigurasi model BioBERT diuji, masing-masing menggabungkan teknik-teknik seperti Class Weight, Conditional Random Fields (CRF), dan Hyperparameter Tuning. Tiga metrik utama yang digunakan untuk evaluasi adalah Precision, Recall, dan F1-Score, yang sangat relevan dalam konteks NER, terutama untuk menghadapi dataset dengan ketidakseimbangan kelas.

3.1 Evaluasi Performa Model

Enam konfigurasi model BioBERT diuji, yang menggabungkan teknik-teknik seperti Class Weight, Conditional Random Fields (CRF), dan Hyperparameter Tuning. Masing-masing diuji dengan kombinasi teknik yang berbeda untuk meningkatkan kinerja model dalam mengenali entitas medis. Pada Gambar 2, menunjukkan grafik untuk mengukur perbandingan kinerja berbagai model yang diukur untuk setiap label yaitu label O, B-Chemical, B-Disease, I-Disease hingga I-Chemical .



Gambar 2. Grafik Comparison Metrik

Tiga metrik utama yang digunakan untuk evaluasi adalah Precision, Recall, dan F1-Score, yang sangat relevan dalam konteks NER, terutama ketika menghadapi dataset dengan ketidakseimbangan label. Selain tiga metrik utama tersebut, penelitian ini juga menggunakan akurasi untuk mengukur prediksi secara keseluruhan. Seperti yang digambarkan pada Tabel 3, model-model yang diuji menunjukkan kinerja yang bervariasi, dan berdasarkan hasil evaluasi, beberapa model menunjukkan keunggulan dalam aspek yang berbeda, yaitu Precision, Recall, F1-Score dan Akurasi.

Tabel 3. Hasil Metrik Per Label

| Model | Label | Precision | Recal | F1-Score |
|-------------------------------|------------|-----------|--------|----------|
| BioBERT Baseline | O | 0.9913 | 0.9956 | 0.9934 |
| | B-Chemical | 0.6183 | 0.4903 | 0.5469 |
| | B-Disease | 0.5457 | 0.3883 | 0.4538 |
| | I-Disease | 0.5699 | 0.3684 | 0.4475 |
| | I-Chemical | 0.3823 | 0.3766 | 0.3794 |
| BioBERT + <i>Class Weight</i> | O | 0.9991 | 0.9501 | 0.9740 |
| | B-Chemical | 0.2062 | 0.8236 | 0.3299 |
| | B-Disease | 0.1857 | 0.7233 | 0.2956 |
| | I-Disease | 0.2199 | 0.6345 | 0.3266 |
| | I-Chemical | 0.0915 | 0.7107 | 0.1622 |
| BioBERT + CRF | O | 0.9692 | 0.9740 | 0.9716 |
| | B-Chemical | 0.6067 | 0.5580 | 0.5814 |
| | B-Disease | 0.5594 | 0.4491 | 0.4982 |
| | I-Disease | 0.4600 | 0.5718 | 0.5098 |
| | I-Chemical | 0.5083 | 0.3840 | 0.4375 |

| | | | | |
|---|------------|--------|--------|--------|
| BioBERT + CRF + <i>Class Weight</i> | O | 0.9680 | 0.9749 | 0.9714 |
| | B-Chemical | 0.6184 | 0.5556 | 0.5853 |
| | B-Disease | 0.5419 | 0.4679 | 0.5022 |
| | I-Disease | 0.5043 | 0.5066 | 0.5055 |
| | I-Chemical | 0.4056 | 0.3591 | 0.3810 |
| BioBERT + <i>Class Weight</i> + <i>Hyperparameter Tuning</i> | O | 0.9968 | 0.9800 | 0.9883 |
| | B-Chemical | 0.3930 | 0.7471 | 0.5151 |
| | B-Disease | 0.3463 | 0.6537 | 0.4527 |
| | I-Disease | 0.3016 | 0.6675 | 0.4155 |
| | I-Chemical | 0.1956 | 0.5486 | 0.2883 |
| BioBERT + CRF + <i>Class Weight</i> + <i>Hyperparameter Tuning</i> | O | 0.9553 | 0.9908 | 0.9728 |
| | B-Chemical | 0.7356 | 0.4366 | 0.5480 |
| | B-Disease | 0.7188 | 0.2681 | 0.3905 |
| | I-Disease | 0.5897 | 0.3919 | 0.4709 |
| | I-Chemical | 0.5926 | 0.0399 | 0.0748 |

Tabel 4. Hasil Akurasi Keseluruhan

| Model | Akurasi |
|--|---------|
| BioBERT Baseline | 0.9861 |
| BioBERT + Class Weight | 0.9467 |
| BioBERT + CRF | 0.9416 |
| BioBERT + CRF + Class Weight | 0.9419 |
| BioBERT + Class Weight + Hyperparameter Tuning | 0.9752 |
| BioBERT + CRF + Class Weight + Hyperparameter Tuning | 0.9459 |

Precision menunjukkan seberapa tepat model dalam mendeteksi entitas yang relevan tanpa menghasilkan kesalahan positif. Dengan kata lain, precision menunjukkan seberapa banyak entitas yang terdeteksi benar oleh model dibandingkan dengan semua entitas yang diprediksi sebagai relevan. Nilai Precision yang lebih tinggi mengindikasikan bahwa model memiliki kemampuan yang lebih baik dalam menghindari kesalahan dalam memprediksi entitas medis. Berdasarkan hasil evaluasi, rata-rata Precision yang diperoleh adalah 0.559, dengan standar deviasi 0.262. Hal ini menunjukkan variasi yang signifikan antar model dalam hal ketepatan prediksi entitas medis.

Recall mengukur kemampuan model untuk mendeteksi semua entitas yang relevan. Metrik ini menunjukkan seberapa banyak entitas relevan yang berhasil dikenali oleh model dari seluruh entitas yang ada. Nilai Recall yang lebih tinggi mengindikasikan bahwa model lebih efektif dalam mendeteksi entitas yang relevan, meskipun bisa terjadi beberapa kesalahan dalam hal prediksi entitas yang tidak relevan. Rata-rata Recall yang diperoleh adalah 0.600, dengan standar deviasi 0.243. Hal ini menunjukkan bahwa model-model ini cukup efektif dalam mendeteksi entitas yang benar, meskipun masih ada beberapa entitas yang tidak dikenali, yang disebut sebagai false negatives.

F1-Score merupakan metrik gabungan yang mencerminkan keseimbangan antara Precision dan Recall. F1-Score memberikan gambaran umum mengenai kemampuan model dalam mengenali entitas dengan benar (Precision) dan mendeteksi semua entitas yang relevan (Recall), dengan fokus pada keseimbangan antara keduanya. F1-Score yang lebih tinggi menunjukkan bahwa model tersebut berhasil menjaga keseimbangan antara keduanya, menghindari ketidakseimbangan yang bisa terjadi jika model terlalu fokus pada satu metrik saja. Nilai rata-rata F1-Score yang diperoleh adalah 0.532, dengan standar deviasi 0.249. Metrik ini memberikan gambaran umum mengenai kemampuan model dalam mengenali entitas dengan benar, serta mendeteksi semua entitas yang relevan.

Hasil evaluasi yang ditunjukkan dalam Tabel 3 menunjukkan variasi yang signifikan antar model dalam hal metrik Precision, Recall, F1-Score, dan pada Tabel 4 menunjukkan Akurasi. Model BioBERT Baseline menunjukkan performa terbaik untuk label O atau non entity, dengan nilai Precision yang sangat tinggi (0.9913) dan Recall yang juga sangat baik (0.9956), menghasilkan F1-Score yang luar biasa (0.9934). Namun, untuk label lainnya (B-Chemical, B-Disease, I-Disease, dan I-Chemical), model ini mengalami penurunan yang signifikan, baik dalam Precision, Recall, maupun F1-Score, dengan label 4 yang menunjukkan nilai terendah untuk semua metrik. Akurasi model ini cukup tinggi secara keseluruhan (0.9861), yang menunjukkan bahwa meskipun ada penurunan pada label tertentu, model ini tetap cenderung melakukan prediksi yang benar secara dominan untuk label yang lebih sering muncul.

Model BioBERT + Class Weight bertujuan untuk menyeimbangkan bobot kelas, yang menghasilkan Recall yang lebih tinggi untuk label lainya (B-Chemical, B-Disease, I-Disease, dan I-Chemical), terutama label B-Chemical yang mencapai nilai 0.8236. Namun, hal ini datang dengan pengorbanan pada Precision, yang menunjukkan penurunan signifikan pada label B-Chemical, B-Disease, I-Disease, dan I-Chemical, terutama pada I-Chemical yang memiliki Precision sangat rendah (0.0915). Ini mengindikasikan bahwa meskipun model ini lebih sensitif dalam mendeteksi entitas yang relevan, hal ini menyebabkan lebih banyak kesalahan prediksi atau false positives.

Model BioBERT + CRF yang mengintegrasikan CRF untuk mempertimbangkan hubungan antar label, menunjukkan F1-Score yang sangat baik pada label O atau non entity (0.9716) dan performa yang lebih seimbang

pada label B-Chemical, B-Disease, I-Disease, dan I-Chemical. Model ini mampu menjaga keseimbangan yang lebih baik antara Precision dan Recall, dengan penurunan yang tidak terlalu signifikan di label yang lebih sulit. Namun, meskipun menunjukkan perbaikan dalam hal keseimbangan ini, nilai Akurasi model ini sedikit lebih rendah (0.9416) dibandingkan dengan BioBERT *Baseline*.

Dengan menggabungkan CRF, Class Weight, dan Hyperparameter Tuning, model BioBERT + CRF + Class Weight + Hyperparameter Tuning memperoleh Precision tertinggi untuk label O atau non entity (0.9553), yang menunjukkan bahwa model ini sangat efektif dalam mengurangi kesalahan prediksi. Namun, nilai Recall untuk label B-Chemical, B-Disease, I-Disease, dan I-Chemical menurun, dengan label B-Chemical mencapai nilai Recall tertinggi (0.7471), yang menunjukkan pengorbanan terhadap kemampuan mendeteksi entitas yang lebih relevan demi meningkatkan ketepatan deteksi. Meskipun demikian, model ini masih memperlihatkan Akurasi yang cukup tinggi (0.9752), menandakan bahwa secara keseluruhan, model ini memberikan hasil yang dapat diandalkan meskipun ada penurunan dalam Recall.

Secara keseluruhan, model BioBERT + CRF menunjukkan keseimbangan terbaik antara Precision dan Recall dibandingkan model lain, seperti yang terlihat dari F1-Score rata-rata (0.5998) yang lebih tinggi dibanding beberapa konfigurasi lainnya. Pada label O, model ini mencapai Precision 0.9692 dan Recall 0.9748 (F1-Score 0.972), menunjukkan tingkat deteksi yang sangat baik. Selain itu, pada label B-Chemical (Precision: 0.6067, Recall: 0.5580, F1-Score: 0.5814) dan I-Disease (Precision: 0.4600, Recall: 0.5718, F1-Score: 0.5098), meskipun Precision dan Recall tidak selalu sama, selisihnya tidak terlalu jauh, sehingga model tetap dapat mengenali entitas dengan baik tanpa terlalu banyak kesalahan. Dibandingkan dengan model lain yang mungkin lebih condong ke Precision tinggi tetapi Recall rendah (atau sebaliknya), BioBERT + CRF tetap konsisten dalam menyeimbangkan keduanya, menjadikannya pilihan terbaik untuk mendeteksi entitas secara akurat dan stabil di berbagai kategori.

Tabel 5. Ranking Model

| Model | Precision | Recall | F1-Score |
|--|-----------|--------|----------|
| BioBERT <i>Baseline</i> | 6 | 2 | 3 |
| BioBERT + <i>Class Weight</i> | 1 | 5 | 4 |
| BioBERT + CRF | 3 | 3 | 1 |
| BioBERT + CRF + <i>Class Weight</i> | 4 | 4 | 5 |
| BioBERT + <i>Class Weight</i> + <i>Hyperparameter Tuning</i> | 5 | 1 | 6 |
| BioBERT + CRF + <i>Class Weight</i> + <i>Hyperparameter Tuning</i> | 2 | 5 | 2 |

Pada Tabel 5, menunjukkan peringkat model berdasarkan Precision, Recall, dan F1-Score. Model BioBERT + CRF menduduki peringkat pertama untuk F1-Score, yang menunjukkan bahwa model ini memiliki keseimbangan terbaik antara Precision dan Recall. Meskipun model ini tidak unggul di kedua metrik secara terpisah, ia berhasil menjaga keseimbangan yang baik di antara keduanya. Model BioBERT *Baseline*, yang berada di peringkat pertama dalam Precision, menunjukkan bahwa model ini sangat baik dalam mendeteksi entitas relevan dengan sedikit kesalahan prediksi, namun memiliki Recall yang lebih rendah, sehingga beberapa entitas relevan mungkin terlewat.

Model BioBERT + CRF + Class Weight + Hyperparameter Tuning, yang menempati peringkat kedua untuk Precision dan Recall, menunjukkan performa yang solid di kedua metrik meskipun tidak seimbang seperti model BioBERT + CRF. Model BioBERT + CRF + Class Weight memiliki peringkat rata-rata, berada di posisi keempat dalam Precision, keempat dalam Recall, dan kelima dalam F1-Score, yang menunjukkan kinerja yang lebih konsisten namun tidak luar biasa dibandingkan model lainnya. Model BioBERT + Class Weight, yang berada di peringkat kedua dalam Recall, cenderung lebih sensitif dalam mendeteksi entitas yang relevan meskipun dengan pengorbanan pada Precision. Sementara itu, Model memiliki kinerja yang lebih rendah dengan peringkat kelima untuk Precision, keempat untuk Recall, dan keenam untuk F1-Score, menandakan kinerja yang lebih lemah di semua metrik.

3.2 Pembahasan

Hasil yang diperoleh menunjukkan bahwa F1-Score adalah metrik yang paling efektif untuk menilai kinerja model dalam tugas NER, karena memberikan gambaran seimbang antara kemampuan model dalam mengidentifikasi entitas yang benar (Precision) dan mendeteksi semua entitas yang relevan (Recall). Berdasarkan ranking performa model, BioBERT + CRF berada pada peringkat pertama dalam hal F1-Score, menunjukkan bahwa model ini sangat baik dalam mengenali entitas medis dengan akurasi tinggi dan meminimalkan kesalahan prediksi. Model ini diikuti oleh BioBERT + CRF + Class Weight di peringkat kedua, yang juga menunjukkan performa yang kuat dalam mendeteksi entitas medis secara konsisten. Peringkat ketiga ditempati oleh BioBERT + Class Weight + Hyperparameter Tuning, yang memberikan keseimbangan antara precision dan recall tetapi masih tertinggal dibandingkan model CRF.

Penelitian ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa penggunaan CRF dalam NER untuk teks medis dapat mengatasi ketergantungan antar entitas medis dan meningkatkan akurasi pengenalan entitas [12], [21], [22]. Namun, penelitian ini menunjukkan bahwa dengan penggabungan Hyperparameter Tuning dan Class

Weight, model dapat lebih dioptimalkan untuk menghadapi dataset yang tidak seimbang dan entitas yang lebih kompleks.

Implementasi NER pada entitas chemical dan disease memiliki banyak potensi dalam berbagai bidang. Dalam sektor kesehatan, model ini dapat digunakan oleh peneliti biomedis untuk mengekstraksi informasi penting dari artikel ilmiah, misalnya dalam identifikasi interaksi obat, komposisi kimia, atau diagnosis penyakit tertentu. Selain itu, penyedia layanan kesehatan dapat mengintegrasikan model ini ke dalam sistem Rekam Medis Elektronik (EHR) untuk membantu tenaga medis dalam menelusuri riwayat kesehatan pasien berdasarkan entitas medis yang terdeteksi secara otomatis. Di sektor farmasi, perusahaan obat dapat memanfaatkan NER untuk menganalisis literatur medis guna menemukan potensi bahan kimia baru atau memverifikasi keamanan penggunaan obat tertentu dalam kondisi kesehatan yang berbeda.

Keterbatasan penelitian ini terletak pada penggunaan dataset yang terbatas hanya pada entitas chemical dan disease, yang berarti model mungkin tidak sepenuhnya dapat diterapkan pada entitas medis lainnya. Penelitian selanjutnya dapat menggunakan dataset yang lebih besar dan beragam, serta memanfaatkan cloud computing untuk mengatasi keterbatasan sumber daya komputasi yang ada. Dengan pengembangan lebih lanjut, diharapkan model ini dapat diadaptasi untuk mendukung pengenalan entitas medis yang lebih luas, seperti prosedur medis, alat kesehatan, atau bahkan gejala klinis yang lebih spesifik.

4. KESIMPULAN

Berdasarkan hasil evaluasi dan peringkat performa model, BioBERT + CRF menempati peringkat pertama dalam F1-Score, menjadikannya model terbaik dengan keseimbangan optimal antara Precision (peringkat 3, 0.6067 pada B-Chemical, 0.5594 pada B-Disease, 0.4600 pada I-Disease, dan 0.5083 pada I-Chemical) dan Recall (peringkat 3, 0.5580 pada B-Chemical, 0.4491 pada B-Disease, 0.5718 pada I-Disease, dan 0.3840 pada I-Chemical). Dibandingkan model lain, seperti BioBERT + CRF + Class Weight + Hyperparameter Tuning yang berada di peringkat kedua dalam F1-Score, model ini lebih seimbang karena tidak memiliki perbedaan besar antara Precision dan Recall. Model terbaik ini menunjukkan kinerja yang lebih stabil dalam mengenali berbagai entitas medis dibandingkan model lainnya, menjadikannya pilihan optimal untuk Named Entity Recognition (NER) dalam teks medis.

REFERENCES

- [1] M. Sumarudin and M. Syafrullah, "Named Entity Recognition In Electronic Medical Records Based On Hybrid Neural Network And Transformer," *Eduvest - J. Univers. Stud.*, vol. 4, no. 6, pp. 5263–5279, Jun. 2024, doi: 10.59188/eduvest.v4i6.1473.
- [2] S. Liu, A. Wang, X. Xiu, M. Zhong, and S. Wu, "Evaluating Medical Entity Recognition in Health Care: Entity Model Quantitative Study," *JMIR Med. Inform.*, vol. 12, p. e59782, Oct. 2024, doi: 10.2196/59782.
- [3] M. D. R. Sutrisno, D. Richasdy, and A. F. Ihsan, "The Organization Entity Extraction Telkom University Affiliated using Recurrent Neural Network (RNN)," *Build. Inform. Technol. Sci. BITS*, vol. 4, no. 2, pp. 483–489, Sep. 2022, doi: 10.47065/bits.v4i2.1956.
- [4] P. Chen, M. Zhang, X. Yu, and S. Li, "Named entity recognition of Chinese electronic medical records based on a hybrid neural network and medical MC-BERT," *BMC Med. Inform. Decis. Mak.*, vol. 22, no. 1, p. 315, Dec. 2022, doi: 10.1186/s12911-022-02059-2.
- [5] V. Kocaman and D. Talby, "Accurate Clinical and Biomedical Named Entity Recognition at Scale," *Softw. Impacts*, vol. 13, p. 100373, Aug. 2022, doi: 10.1016/j.simpa.2022.100373.
- [6] A. F. Abdillah, D. Purwitasari, S. Juanita, and M. H. Purnomo, "Pengenalan Entitas Biomedis dalam Teks Konsultasi Kesehatan Online Berbahasa Indonesia Berbasis Arsitektur Transformers," *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 10, no. 1, pp. 131–140, Feb. 2023, doi: 10.25126/jtiik.20231016337.
- [7] J. Lee *et al.*, "BioBERT: a pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, Feb. 2020, doi: 10.1093/bioinformatics/btz682.
- [8] Q. Qin, S. Zhao, and C. Liu, "A BERT-BiGRU-CRF Model for Entity Recognition of Chinese Electronic Medical Records," *Complexity*, vol. 2021, no. 1, p. 6631837, Jan. 2021, doi: 10.1155/2021/6631837.
- [9] J. Jiang, M. Cheng, Q. Liu, Z. Li, and E. Chen, "Nested Named Entity Recognition from Medical Texts: An Adaptive Shared Network Architecture with Attentive CRF," Nov. 09, 2022, *arXiv: arXiv:2211.04759*. doi: 10.48550/arXiv.2211.04759.
- [10] A. Makarova *et al.*, "Overfitting in Bayesian Optimization: an empirical study and early-stopping solution," May 2021, doi: 10.3929/ETHZ-B-000521574.
- [11] T. D. Tran, M. N. Ha, L. H. B. Nguyen, and D. Dinh, "Improving Multi-Grained Named Entity Recognition with BERT and Focal Loss," *ICIC International*, 2021, doi: 10.24507/icicelb.12.03.291.
- [12] Y. Tong *et al.*, "Improving biomedical named entity recognition by dynamic caching inter-sentence information," *Bioinformatics*, vol. 38, no. 16, pp. 3976–3983, Aug. 2022, doi: 10.1093/bioinformatics/btac422.
- [13] X. Zhao, J. Greenberg, Y. An, and X. T. Hu, "Fine-Tuning BERT Model for Materials Named Entity Recognition," in *2021 IEEE International Conference on Big Data (Big Data)*, Orlando, FL, USA: IEEE, Dec. 2021, pp. 3717–3720. doi: 10.1109/BigData52589.2021.9671697.
- [14] S. Belkadi, L. Han, Y. Wu, and G. Nenadic, "Exploring the Value of Pre-trained Language Models for Clinical Named Entity Recognition," Oct. 30, 2023, *arXiv: arXiv:2210.12770*. doi: 10.48550/arXiv.2210.12770.
- [15] S. Sarica and J. Luo, "Stopwords in technical language processing," *PLOS ONE*, vol. 16, no. 8, p. e0254937, Aug. 2021, doi: 10.1371/journal.pone.0254937.



- [16] S. Nemoto, S. Kitada, and H. Iyatomi, “Majority or Minority: Data Imbalance Learning Method for Named Entity Recognition,” *IEEE Access*, vol. 13, pp. 9902–9909, 2025, doi: 10.1109/ACCESS.2024.3522972.
- [17] K. Anam, “Early Stopping on CNN-LSTM Development to Improve Classification Performance,” *J. Appl. Data Sci.*, vol. 5, no. 3, pp. 1175–1188, Sep. 2024, doi: 10.47738/jads.v5i3.312.
- [18] Y. Yin *et al.*, “Augmenting biomedical named entity recognition with general-domain resources,” *J. Biomed. Inform.*, vol. 159, p. 104731, Nov. 2024, doi: 10.1016/j.jbi.2024.104731.
- [19] L. K. Meng, H. H. Yi, and N. B. Wei, “A Machine Learning Approach for Face Mask Detection System with AdamW Optimizer,” *J Appl Technol Innov*, vol. 7, no. 3, 2023, doi: 25.
- [20] L. N. Smith, “Cyclical Learning Rates for Training Neural Networks,” in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Santa Rosa, CA, USA: IEEE, Mar. 2017, pp. 464–472. doi: 10.1109/WACV.2017.58.
- [21] A. Mascio *et al.*, “Comparative Analysis of Text Classification Approaches in Electronic Health Records,” May 08, 2020, *arXiv*: arXiv:2005.06624. doi: 10.48550/arXiv.2005.06624.
- [22] Y. Munarko, M. S. Sutrisno, W. A. I. Mahardika, I. Nuryasin, and Y. Azhar, “Named entity recognition model for Indonesian tweet using CRF classifier,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 403, p. 012067, Oct. 2018, doi: 10.1088/1757-899X/403/1/012067.