

Named Entity Recognition for Medical Records of Heart Failure Using a Pre-trained BERT Model

Mikael Triartama Manurung ^{1*}, I Gusti Ngurah Lanang Wijayakusuma ^{2*}, I Putu Winada Gautama ^{3*}

* Matematika, Universitas Udayana

mikaeltriartamanurung@gmail.com ¹, lanang_wijaya@unud.ac.id ², winadagautama@unud.ac.id ³

Article Info

Article history:

Received 2025-01-19

Revised 2025-02-25

Accepted 2025-02-26

Keyword:

Named Entity Recognition, BERT, Medical Records, Heart Failure, Transformer, Medical Text Classification.

ABSTRACT

This study aims to develop a Named Entity Recognition (NER) model based on a pre-trained BERT model for medical records of heart failure patients. The focus of this research is to classify essential medical entities from unstructured medical record texts. The classification covers four categories: objective data (patient identity, laboratory test results, and objective examination data), subjective data (patient complaints), prescriptions, and diagnoses (diagnosis codes and descriptions). The methodology employs Natural Language Processing (NLP) techniques using Transformer-based architectures, such as Bidirectional Encoder Representation from Transformers (BERT). The developed model is evaluated based on entity label prediction accuracy and medical entity classification performance. The results indicate that the BERT-based NER model performs well, achieving an entity prediction accuracy of 84.82%. Furthermore, the model effectively classifies medical entities from input texts in alignment with expected medical entities. This research is expected to contribute significantly to medical data management, assist healthcare professionals in clinical decision-making, and serve as a reference for the development of AI-based healthcare technology in Indonesia.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Kesehatan jantung merupakan salah satu isu yang cukup krusial di Indonesia, karena menjadi salah satu penyakit penyebab utama kematian tertinggi. Data Kementerian Kesehatan RI pada tahun 2020 mencatat bahwa terdapat sebesar 1,5 juta peningkatan kasus penyakit jantung tiap tahunnya. Penyakit gagal jantung umumnya dipengaruhi oleh faktor-faktor risiko seperti pola makan tidak sehat, kurangnya aktivitas fisik, dan peningkatan prevalensi diabetes serta hipertensi yang berkontribusi secara signifikan pada penyakit gagal jantung [1]. Masalah ini membutuhkan suatu pendekatan manajemen rekam medis yang lebih terstruktur dan efisien untuk mengatasi lonjakan data yang terjadi [2].

Teknologi *Artificial Intelligence* (AI), memiliki potensi dalam mengotomatisasi berbagai proses, terutama *Natural Language Processing* (NLP) yang mampu mengekstrak dan menganalisis informasi penting dari teks medis yang tidak terstruktur [3]. Arsitektur Transformer merupakan arsitektur terbaik dari jaringan saraf dalam hal menangani data yang

berurutan seperti teks, karena Transformer memproses data secara paralel, sehingga memungkinkan proses pelatihan model menjadi lebih efisien [4]. Transformer memiliki dua komponen utama, yaitu *encoder* dan *decoder* yang masing-masing memiliki enam lapisan identik yang disusun berurutan. Setiap lapisan encoder terdapat dua sub-lapisan yaitu, mekanisme *multi-head self-attention* dan jaringan saraf *feedforward* yang saling terhubung. Setiap sub-lapisan dilakukan operasi penjumlahan dan normalisasi lapisan. Setiap sub-lapisan tambahan dalam lapisan decoder akan melakukan *multi-head attention* terhadap output dari encoder [5].

Bidirectional Encoder Representation from Transformers (BERT) merupakan salah satu model pengembangan dari arsitektur Transformer yang sering digunakan sebagai model dasar untuk menyematkan karakteristik teks, karena BERT secara efektif dapat menangkap informasi konteks kata yang lebih kaya [6]. Hal tersebut dapat terjadi karena BERT dilatih untuk memahami konteks dari suatu kata dalam kalimat dengan memperhatikan konteks kata secara 2 arah,

berdasarkan kata sebelum dan kata sesudahnya [7]. Arsitektur BERT dibangun berdasarkan pada *encoder* transformer yang menggunakan mekanisme *self-attention*. Mekanisme *multi-head attention* pada BERT memungkinkan model untuk menangkap berbagai pola perhatian dengan memproses perhatian dalam beberapa subruang paralel [8].

Named Entity Recognition (NER) adalah suatu teknik yang digunakan untuk mengidentifikasi entitas tertentu dari teks yang tidak terstruktur [9]. Tugas NER dalam pemrosesan bahasa alami bertujuan untuk mengidentifikasi dan mengklasifikasikan entitas yang memiliki makna tertentu dalam teks, seperti nama orang, organisasi, lokasi, tanggal, dan lainnya [10]. Secara matematis, suatu arsitektur model akan mengubah input berupa kalimat menjadi representasi vektor melalui embedding, yang mencakup embedding kata maupun posisi. Proses ini dilakukan dengan menghitung *attention* terhadap setiap kata dalam konteks kalimat.

Beberapa penelitian terdahulu telah menunjukkan keunggulan arsitektur transformer seperti BERT dalam tugas NER dibandingkan model lain, misalnya penelitian oleh Houssein et al. pada tahun 2024 yang mengekstraksi faktor risiko dari Rekam Medis Elektronik (RME) yang biasanya tidak terstruktur, menggunakan teknik transfer *learning* dengan model seperti BERT, BioBERT, dan RoBERTa. Hasil akurasi menunjukkan angka yang tinggi, yaitu BERT dengan Skor F1 mikro sebesar 93.73%, BioBERT dengan Skor F1 mikro sebesar 93.99%, dan RoBERTa dengan skor F1 mikro tertinggi sebesar 94.27% [4].

Pada penelitian oleh García-Barragán dkk pada tahun 2024 yang berjudul “*GPT For Medical Entity Recognition in Spanish*” dibahas dan dibandingkan dua model pengenalan teks, yaitu: *fine-tuning* BERT dan *In-Context Learning* dengan *Generative Pretrained Transformer* (GPT). Pada penelitian ini diperoleh bahwa model BERT memiliki akurasi yang lebih baik dalam presisi dibandingkan model GPT walaupun membutuhkan waktu komputasi yang lebih lama [11].

Penelitian lain oleh Kusumawardani et al. menerapkan teknologi pengolahan bahasa alami (*Natural Language Processing*) dan kecerdasan buatan dalam konteks layanan konsultasi kesehatan di Indonesia. NER dibangun menggunakan model *Bidirectional Long Short-Term Memory* yang dikombinasikan dengan *Conditional Random Fields* yang memperoleh hasil akurasi sebesar 99,68%. Model lain sebagai pembanding, yaitu model berbasis BERT ('cahya/XLM-RoBERTa-large-indonesian-NER') dengan akurasi sebesar 98,51%. Akurasi dari model XLM-R lebih kecil dibandingkan model BiLSTM, namun nilai presisi dan skor F1 dari model XLM-R lebih besar dibandingkan dengan model BiLSTM, yang mana sangat bagus untuk data dengan kebutuhan presisi yang tinggi seperti data medis [12-15].

Berdasarkan urgensi dan potensi yang telah dipaparkan sebelumnya, penelitian ini bertujuan untuk membangun suatu NER berbasis model BERT standar yang di *fine-tuning* dengan data medis, untuk mengekstrak dan mengklasifikasikan entitas pada rekam medis pasien gagal

jantung dalam bahasa Indonesia. Hasil penelitian ini diharapkan dapat menjadi solusi yang relevan dalam mendukung pengambilan keputusan klinis dan meningkatkan efisiensi pengelolaan data medis.

II. METODE

A. Pengumpulan Data

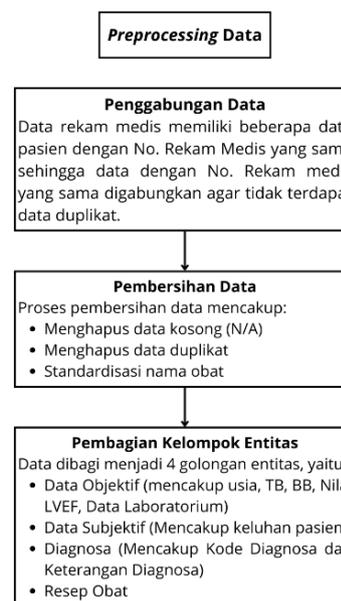
Data yang digunakan dalam penelitian ini merupakan data primer yang berasal dari kumpulan rekam medis pasien penyakit jantung. Data diperoleh dari kumpulan rekam medis penyakit gagal jantung di RSUD pusat Prof.dr.I.G.N.G. Ngoerah. Data mencakup informasi seperti:

- No. Rekam Medis dan Identitas Pasien
- Kondisi Fisik Pasien (usia, tinggi badan, berat badan, dan nilai LVEF)
- Kode Diagnosa dan Keterangan Diagnosa
- Keluhan Pasien
- Data Objektif dan Data Laboratorium
- Resep Obat

Dataset berjumlah 1000 rekam medis yang kemudian dibersihkan dengan menggabungkan konten rekam medis dari data yang memiliki nomor rekam medis yang sama.

B. Preprocessing Data

Data akan dibersihkan dan digabungkan berdasarkan kategori entitas-entitas yang telah ditentukan. Proses preprocessing data digambarkan pada diagram alur berikut:



Gambar 1. Diagram alur proses preprocessing data

C. Labeling Data

Data yang telah dikelompokkan berdasarkan golongan entitas selanjutnya akan di labeli dengan entitas masing-masing dari teks. Misalnya, teks yang berisi identitas pasien dan data hasil pemeriksaan laboratorium yang merupakan

data objektif, akan dilabeli dengan “OBJ”, keluhan pasien yang merupakan bagian dari data subjektif akan dilabeli “SUBJ”, teks yang berisi kode dan keterangan diagnosa akan dilabeli “DIAG”, dan teks yang berisi Resep Obat akan dilabeli “RESEP”. Setelah melalui preprocessing data dan labeling data, terdapat sebesar 724 token kalimat yang terdiri dari 4 jenis entitas, yaitu entitas dengan label OBJ, SUBJ, RESEP, dan DIAG.

D. Split Data

Dataset rekam medis yang telah dilabeli akan dibagi menjadi dua bagian utama, yaitu 80% data untuk pelatihan dan pengujian, dan 20% data untuk validasi. Data pelatihan digunakan untuk melatih model NER berbasis BERT dalam mengenali pola entitas medis yang akan diuji menggunakan data pengujian, sedangkan data validasi berfungsi untuk memantau performa model dalam proses evaluasi dan mencegah terjadinya *overfitting*. Dalam bentuk tabel berikut pembagian data.

TABEL 1.
PEMBAGIAN DATA

Data Train	463
Data Test	116
Data Validasi	145

E. Representasi Teks

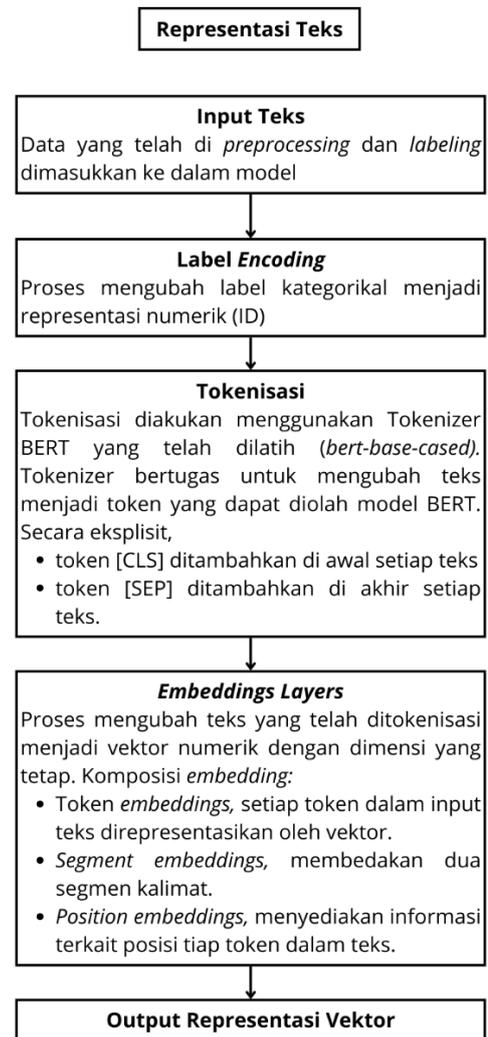
Representasi teks merupakan proses untuk mengubah teks mentah menjadi suatu format yang dapat dimengerti dan diproses oleh *machine learning*. Proses representasi teks dilakukan dengan memanfaatkan *word embedding* berbasis transformer yang secara eksplisit memanfaatkan *library* dari model BERT.

Label encoding adalah teknik yang digunakan untuk mengkonversi label kategorikal menjadi representasi numerik. Cara kerja label encoding:

- Mengidentifikasi kategori dalam kolom yang akan dilakukan encoding. Pada penelitian ini, kategori yang digunakan adalah objektif, subjektif, diagnosa, dan resep obat.
- Setiap kategori diberikan angka unik, misalnya: objektif : 0, subjektif: 1, diagnosa: 2, dan resep obat: 3.
- Kategori dikonversikan menjadi angka yang merepresentasikan kategori tersebut.

Word embedding adalah metode untuk merepresentasikan teks dengan mengonversi data teks menjadi vektor numerik dengan dimensi tertentu. Vektor ini menggambarkan makna kata dalam suatu ruang vektor, sehingga kata-kata yang memiliki arti serupa akan memiliki representasi vektor yang berdekatan. Dalam konteks word embedding, vektor adalah array numerik berdimensi tinggi yang digunakan untuk merepresentasikan setiap kata, frasa, atau kalimat. Setiap elemen dalam vektor mencerminkan karakteristik semantik atau sintaktik tertentu dari teks tersebut. Sebagai contoh, kata “sesak nafas” dan “asma” bisa saja memiliki vektor yang saling berdekatan dibandingkan dengan kata “75 tahun”. Setelah data di lakukan labeling encoding, proses selanjutnya

adalah melakukan tokenisasi dan pembuatan embedding layer.



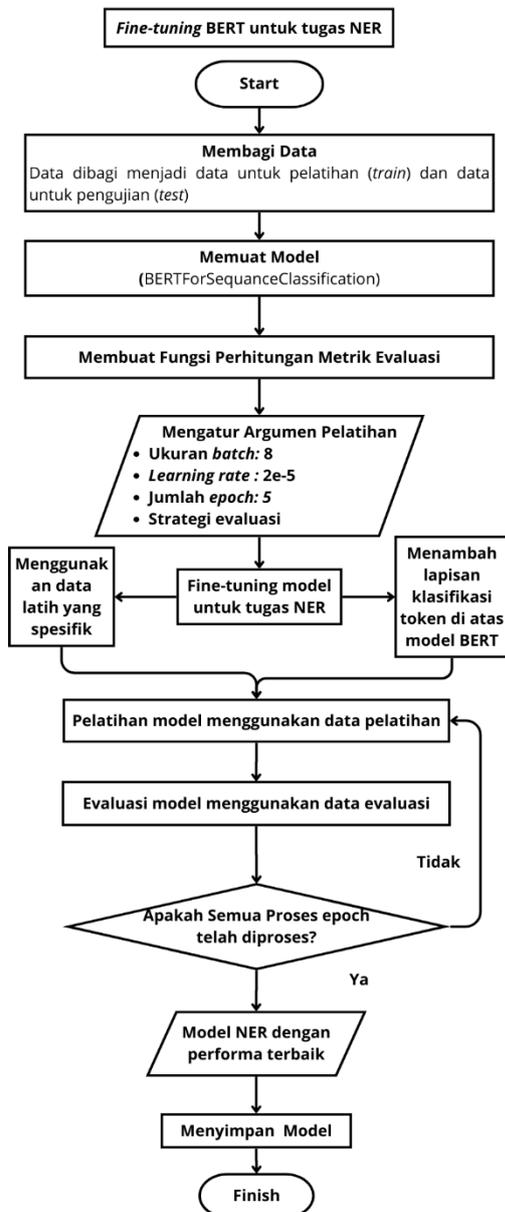
Gambar 2. Diagram alur proses representasi teks

Dalam konteks *word embedding*, vektor adalah *array* numerik berdimensi tinggi yang digunakan untuk merepresentasikan setiap kata, frasa, atau kalimat. Setiap elemen dalam vektor mencerminkan karakteristik semantik atau sintaktik tertentu dari teks tersebut. Sebagai contoh, kata “sesak nafas” dan “asma” bisa saja memiliki vektor yang saling berdekatan dibandingkan dengan kata “75 tahun”. Setelah data di lakukan *labeling encoding*, proses selanjutnya adalah melakukan tokenisasi menggunakan *bert-base-cased* dan pembuatan *embedding layer*.

F. Pemodelan dan Fine Tuning Model BERT

Proses pemodelan dimulai dengan pemisahan data menjadi dua bagian, yaitu data pelatihan (*train*) dan data pengujian (*test*). Model yang digunakan adalah model *pre-trained* BERT dengan fungsi *BertForSequenceClassification*, yang dirancang untuk tugas klasifikasi teks. Proses selanjutnya, mengatur argumen pelatihan, seperti ukuran *batch*

(menentukan banyak contoh yang diproses dalam satu iterasi), laju pembelajaran (mengatur besar langkah setiap iterasi untuk memperbaharui bobot), dan strategi evaluasi yang akan digunakan



Gambar 3. Diagram alur proses pemodelan dan fine-tuning

Gambar 3 menunjukkan diagram alur proses pemodelan dan fine-tuning BERT untuk tugas Named Entity Recognition (NER). Proses ini dimulai dengan inisialisasi fine-tuning BERT untuk tugas NER. Langkah pertama adalah membagi data menjadi dua bagian, yaitu data pelatihan (train) yang digunakan untuk melatih model dan data pengujian (test) yang digunakan untuk menguji performa model setelah pelatihan selesai.

Selanjutnya, model `BERTForSequenceClassification` dimuat untuk menangani tugas klasifikasi urutan. Agar

performa model dapat diukur dengan baik, dibuat fungsi perhitungan metrik evaluasi yang akan digunakan untuk menilai hasil pelatihan. Pada tahap berikutnya, argumen pelatihan ditetapkan, termasuk ukuran batch sebanyak 8, learning rate sebesar $2e-5$, jumlah epoch sebanyak 5, serta strategi evaluasi yang digunakan untuk mengukur kinerja model selama pelatihan.

Setelah parameter pelatihan ditentukan, lapisan klasifikasi token ditambahkan di atas model BERT untuk memungkinkan model mengenali dan mengklasifikasikan entitas dalam teks. Kemudian, proses fine-tuning dilakukan menggunakan data latih yang spesifik untuk tugas NER. Model mulai dilatih menggunakan data pelatihan, dan setelah itu dievaluasi dengan data evaluasi untuk menilai seberapa baik model dalam mengklasifikasikan entitas.

Jika seluruh epoch belum selesai diproses, pelatihan akan terus dilakukan hingga semua epoch selesai. Setelah semua epoch diproses, model terbaik dengan performa tertinggi dipilih berdasarkan hasil evaluasi. Model NER yang memiliki akurasi terbaik kemudian disimpan agar dapat digunakan di masa depan. Dengan demikian, proses pemodelan dan fine-tuning BERT untuk tugas NER selesai, menghasilkan model yang optimal untuk mengklasifikasikan entitas dalam rekam medis pasien gagal jantung.

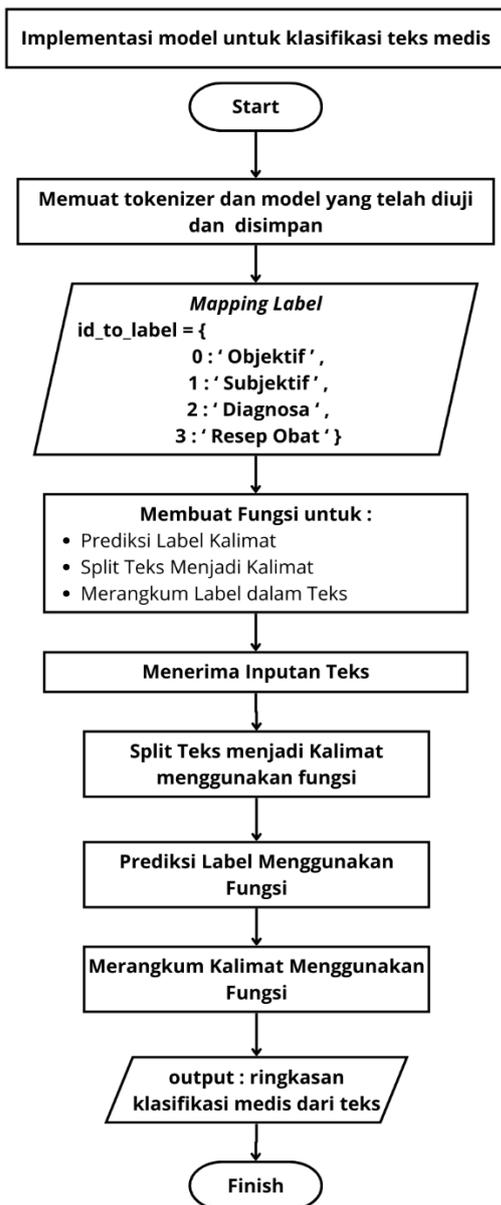
G. Evaluasi Model

Pada tahap evaluasi model, dilakukan pengukuran performa model NER berbasis BERT dengan menggunakan data validasi yang belum pernah digunakan pada proses pelatihan sebelumnya. Evaluasi dilakukan dengan membandingkan hasil prediksi model terhadap label sebenarnya untuk memperoleh nilai akurasi. Proses evaluasi model juga menghasilkan beberapa visualisasi dari evaluasi yang digambarkan dalam bentuk beberapa plot, seperti plot distribusi prediksi dan label sebenarnya, yang disusun untuk memeriksa sebaran hasil klasifikasi model terhadap setiap kelas. Selanjutnya, plot *epoch* dan *loss* digunakan untuk memvisualisasikan perkembangan nilai *training loss* dan *validation loss* pada setiap *epoch*, sehingga memudahkan dalam memantau proses konvergensi model. Selain itu, ditampilkan pula plot *confusion matrix*, yang memberikan representasi grafis atas perbandingan prediksi model terhadap label yang benar, sehingga mempermudah identifikasi kelas yang mengalami kesalahan klasifikasi. Proses evaluasi ini bertujuan untuk memastikan bahwa model tidak hanya memiliki akurasi yang tinggi, namun juga mampu memberikan hasil klasifikasi yang seimbang dan konsisten pada setiap kelas entitas.

H. Implementasi NER untuk Klasifikasi Teks Rekam Medis

Proses Implementasi NER diperlukan untuk mengadaptasi model yang telah dilatih dan disesuaikan untuk mengklasifikasikan entitas khusus dalam teks medis. Melalui proses *mapping label* yang sesuai, seperti objektif, subjektif, diagnosa, dan resep obat, model menjadi lebih efisien dalam mengenali konteks kalimat yang relevan.

Proses dimulai dengan memuat model BERT untuk data rekam medis yang telah dilatih sebelumnya bersamaan dengan *tokenizer*. Proses selanjutnya adalah melakukan *mapping label* untuk mengonversi output model menjadi format yang lebih dapat dipahami. Setelah proses *mapping label* selesai, beberapa fungsi dibuat untuk memprediksi label kalimat berdasarkan input teks yang diberikan. Fungsi ini akan memproses teks dan menghasilkan output yang sesuai dengan label yang telah ditetapkan sebelumnya. Proses ini digambarkan dengan diagram gambar 4.



Gambar 4. Diagram alur implementasi model

Gambar 4 menunjukkan diagram alur implementasi model untuk klasifikasi teks medis menggunakan Named Entity Recognition (NER). Proses ini dimulai dengan Start, di mana sistem memuat tokenizer dan model yang telah diuji serta

disimpan sebelumnya. Model ini digunakan untuk mengklasifikasikan teks medis ke dalam beberapa kategori tertentu.

Selanjutnya, dilakukan *mapping label* untuk menghubungkan setiap ID ke kategori klasifikasi, yaitu 0 untuk Objektif, 1 untuk Subjektif, 2 untuk Diagnosa, dan 3 untuk Resep Obat. Setelah itu, dibuat beberapa fungsi utama yang digunakan dalam proses klasifikasi, meliputi prediksi label kalimat, pemisahan teks menjadi kalimat, dan perangkuman label dalam teks.

Tahap berikutnya adalah sistem menerima input teks dari pengguna. Teks ini kemudian dipecah menjadi beberapa kalimat menggunakan fungsi yang telah dibuat sebelumnya. Setelah itu, model melakukan prediksi label untuk setiap kalimat berdasarkan kategori yang telah ditetapkan.

Setelah semua kalimat diklasifikasikan, sistem merangkum hasil klasifikasi tersebut dengan fungsi yang telah disiapkan, sehingga dapat menghasilkan ringkasan klasifikasi medis dari teks yang diinput. Proses implementasi ini berakhir dengan langkah Finish, di mana output akhir berupa ringkasan klasifikasi teks medis disajikan kepada pengguna.

III. HASIL DAN PEMBAHASAN

NER menggunakan model BERT standar yang disesuaikan dengan data medis dilatih menggunakan data rekam medis pasien gagal jantung yang telah melewati proses preprocessing, *labeling*, pembagian, dan representasi teks. Hasil dari proses training dan evaluasi mencakup nilai *training loss*, *validation loss*, akurasi model, implementasi untuk klasifikasi teks medis, serta visualisasi berupa grafik *loss*, *confusion matrix*, dan distribusi label asli dan prediksi.

A. Training loss dan Validation loss

Berdasarkan 5 epoch yang telah dilakukan saat proses *training*, diperoleh nilai *training loss* dan *validation loss* sebagai berikut:

TABEL 2.
NILAI LOSS PADA SETIAP EPOCH

Epoch	Training Loss	Validation Loss
1	0.099800	0.029932
2	0.006700	0.003716
3	0.003900	0.002297
4	0.003000	0.001848
5	0.002800	0.001703

Training loss mengalami penurunan yang signifikan, yaitu dari nilai 0.0998 pada epoch pertama menjadi 0.0028 pada epoch kelima. Hal tersebut memperlihatkan bahwa model mampu mempelajari pola data dengan cukup baik. Penurunan yang signifikan pada epoch kedua merupakan suatu hal yang umum, karena pelatihan awal model masih baru memulai untuk menemukan pola dasar pada data yang belum sepenuhnya baik, dan kemudian terus belajar untuk menemukan pola yang baik pada *epoch* berikutnya.

Validation loss juga menurun secara bertahap dari 0.029932 menjadi 0.0017 pada *epoch* kelima. Hal ini merupakan suatu indikasi yang sangat baik karena memperlihatkan bahwa model mampu memprediksi entitas teks dengan baik pada data yang belum pernah dilihat sebelumnya.

B. Performa Model

Pada bagian hasil evaluasi akurasi model, diperoleh nilai akurasi sebesar 0.848 atau 84,8%. Berdasarkan *classification report*, performa model dalam mengenali masing-masing label adalah sebagai berikut:

TABEL 3.
NILAI PERFORMA MODEL DALAM PREDIKSI ENTITAS TEKS

	DIAG	OBJ	RESEP	SUBJ
<i>Precision</i>	0.97	0.69	1.00	0.82
<i>Recall</i>	0.89	1.00	1.00	0.50
<i>F1-score</i>	0.93	0.81	1.00	0.62

- **DIAG:** *Precision* 0.97, *recall* 0.89, dan *F1-score* 0.93 menunjukkan bahwa kemampuan model yang sangat baik dalam mengenali data entitas diagnosis.
- **OBJ:** *Precision* 0.69, *recall* 1.00, *F1-score* 0.81, menunjukkan bahwa model cenderung mengklasifikasikan seluruh data objektif dengan benar meskipun presisinya masih cukup rendah.
- **RESEP:** Skor sempurna (1.00) pada *precision*, *recall*, dan *F1-score* menegaskan bahwa kemampuan model sangat akurat untuk mengenali entitas resep obat.
- **SUBJ:** *Precision* 0,82, *recall* 0,50, dan *F1-score* 0,62 menunjukkan bahwa model masih kesulitan mengenali data subjektif pasien, yang kemungkinan disebabkan oleh variasi bahasa atau konteks yang lebih kompleks.

Secara keseluruhan, nilai rata-rata makro *F1-score* sebesar 0,84 atau sekitar 84% yang menunjukkan keseimbangan performa antar-label. Namun, *recall* yang cukup rendah pada label SUBJ menjadi suatu catatan penting yang dapat diperbaiki melalui peningkatan data pelatihan atau teknik augmentasi data.

C. Perbandingan akurasi dengan penelitian sebelumnya

Pada penelitian ini, model BERT standar yang di *fine-tuning* dengan data medis digunakan untuk tugas *Named Entity Recognition* (NER) pada rekam medis berbahasa Indonesia. Perbandingan dengan beberapa penelitian yang membahas NER dalam dunia medis diperlukan untuk memahami keunggulan dan kekurangan dari model BERT yang di *fine-tuning* dalam dunia medis. Berikut merupakan tabel perbandingan performa berbagai model dalam tugas NER rekam medis.

TABEL 4.
PERBANDINGAN PERFORMA MODEL PADA STATE OF THE ART

Penelitian	Model	Dataset	Metrik	Akurasi
[12]	BiLSTM-CRF	Data konsultasi kesehatan online (Indonesia)	<i>Accuracy</i>	99.68%
			<i>Precision</i>	76.98%
			<i>F1-Score</i>	79.27%
[12]	Fine-tuned XLM-RoBERTa	Data konsultasi kesehatan online (Indonesia)	<i>Accuracy</i>	98.51%
			<i>Precision</i>	89.27%
			<i>Recall</i>	90.42%
			<i>F1-Score</i>	89.99%
[11]	GPT	<i>Health Electronic Record</i> (HER) dalam bahasa spanyol terkait kanker payudara	<i>Precision</i>	92%
			<i>Recall</i>	97%
			<i>F1-Score</i>	95%
	Fine-tuning BERT	<i>Health Electronic Record</i> (HER) dalam bahasa spanyol terkait kanker payudara	<i>Precision</i>	93%
			<i>Recall</i>	94%
			<i>F1-Score</i>	94%

D. Implementasi untuk Klasifikasi teks rekam medis

Pada tahap implementasi, model klasifikasi berbasis NER dengan BERT digunakan untuk memetakan kalimat dari rekam medis ke dalam empat label: Objektif (OBJ), Subjektif (SUBJ), Diagnosa (DIAG), dan Resep Obat (RESEP). Model akan menerima input berupa kumpulan kalimat rekam medis dan akan menghasilkan ringkasan berupa kalimat dengan label entitas yang sesuai berdasarkan label yang terdeteksi. Hasil ringkasan menunjukkan bahwa model mampu mengklasifikasikan kalimat dengan cukup baik. Informasi seperti data pemeriksaan fisik dan laboratorium terdeteksi sebagai Objektif, keluhan pasien sebagai Subjektif, dan diagnosa penyakit sebagai Diagnosa.

Contoh implementasi

Contoh input 1 teks (kumpulan kalimat dengan berbagai entitas)

```
input_text = ""
```

```
Bisoprolol 1 x 2.5 mg, Candesartan 1 x 8 mg,  
Simvastatin 1 x 20 mg, Spironolakton 1 x 25 mg,  
Furosemide tab 1 x 40 mg.
```

```
Pasien Tidak Memiliki keluhan. I50.0 dan D4 yang  
artinya Congestive heart failure dan CHF FC II.
```

```
42.0 tahun dengan berat badan 62 kg dan tinggi badan  
171 cm memiliki nilai Ivesf 32.20 % dengan Cor: S1 S2  
Single reguler Pulmo: ves (+++/+++)
```

```
Abdomen: BU (+) Normal Ext: hangat (++)/+++ Lainnya:
```

```
- murmur (-) ronkhi (---/---) Distensi edema (-/-)
```

```
wheezing (-/-) ireguler
```

dan hasil laboratorium 4.25 Kalium, 141.0 Natrium, 0.81 Kreatinin.

```

"""
# Prediksi dan rangkuman label
summary = summarize_text(input_text)
# Menampilkan hasil
print("Ringkasan Label dalam Teks:\n")
for label, sentences in summary.items():
    print(f"{label}:")
    for sentence in sentences:
        print(f" - {sentence}")
    print()

```

Output implementasi

Ringkasan Label dalam Teks:

1) *Objektif*: berisi data objektif pasien, seperti usia, berat badan, tinggi badan, nilai LVEF, kondisi jantung (Cor), paru-paru (Pulmo), abdomen, dan hasil laboratorium. Informasi ini mencerminkan kondisi fisik pasien yang dapat diukur secara langsung.

→ 42.0 tahun dengan berat badan 62 kg dan tinggi badan 171 cm memiliki nilai lvef 32.20 % dengan Cor: S1 S2 Single reguler Pulmo: ves (+++/++) Abdomen: BU (+) Normal Ext: hangat (++/++) Lainnya: - murmur (-) ronkhi (---/---) Distensi edema (-/-) wheezing (---/---) ireguler dan hasil laboratorium 4.25 Kalium, 141.0 Natrium, 0.81 Kreatinin; 4.26 Kalium, 138.0 Natrium, 0.78 Kreatinin dan data hasil laboratorium 4.25 Kalium, 141.0 Natrium, 0.81 Kreatinin.

2) *Subjektif*: berisi keluhan atau pernyataan pasien, dalam hal ini pasien melaporkan tidak memiliki keluhan.

→ Pasien Tidak Memiliki keluhan.

3) *Diagnosa*: berisi hasil diagnosis yang diberikan oleh tenaga medis, seperti kode I50.0 dan D4, yang mengindikasikan Congestive Heart Failure (CHF) dengan Functional Class II.

→ I50.0 dan D4 yang artinya Congestive heart failure dan CHF FC II.

4) *Resep Obat*: berisi daftar obat yang diresepkan kepada pasien, seperti Bisoprolol, Candesartan, Simvastatin, Spironolakton, dan Furosemide, lengkap dengan dosis dan aturan pemakaian.

→ Bisoprolol 1 x 2.5 mg, Candesartan 1 x 8 mg, Simvastatin 1 x 20 mg, Spironolakton 1 x 25 mg, Furosemide tab 1 x 40 mg.

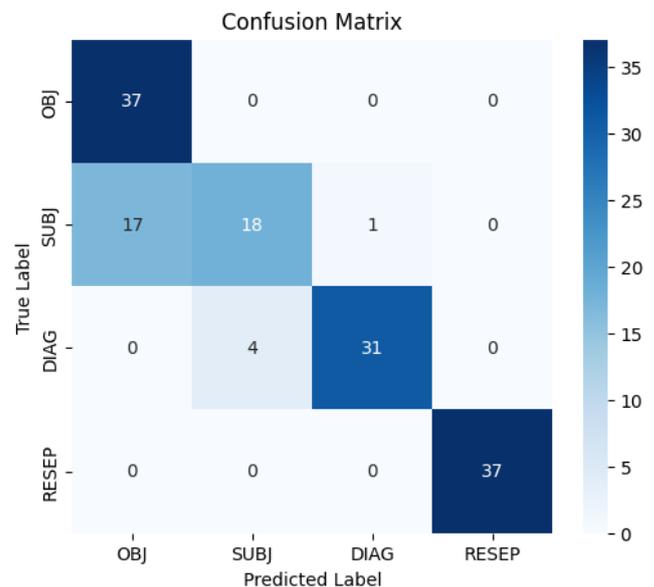
Meskipun secara umum performa model tergolong memuaskan, namun masih terdapat beberapa kekeliruan kecil pada proses implementasi, terutama pada data Subjektif yang terkadang diklasifikasikan sebagai Objektif oleh model. Namun, kesalahan tersebut hanya terjadi pada beberapa kasus dan tidak signifikan secara keseluruhan. Dengan hasil ini, model sudah cukup dapat diandalkan untuk membantu proses klasifikasi otomatis pada rekam medis, meskipun masih perlu

dilakukan perbaikan, terutama pada pengoptimalan pada pemisahan informasi Subjektif dan Objektif.

E. Analisis Visualisasi Hasil Klasifikasi

1) Plot Confusion Matrix

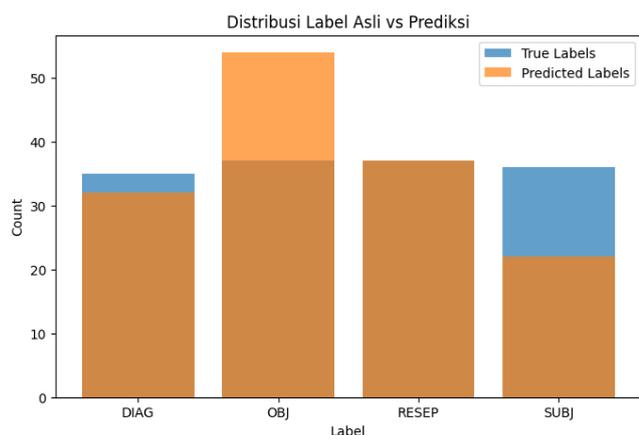
Gambar pertama menunjukkan *Confusion Matrix* dari model klasifikasi, yang menggambarkan bagaimana model mengkategorikan setiap label berdasarkan prediksi dan label sebenarnya. Model memiliki performa yang baik dalam mengklasifikasikan label OBJ, DIAG, dan RESEP, terbukti dari tingginya jumlah prediksi benar di diagonal utama matriks. Namun, terdapat beberapa kesalahan klasifikasi yang cukup mencolok pada label SUBJ, di mana 17 sampel yang seharusnya merupakan SUBJ justru diklasifikasikan sebagai OBJ. Selain itu, terdapat 4 sampel DIAG yang terprediksi sebagai SUBJ serta 1 sampel SUBJ yang terklasifikasi sebagai DIAG



Gambar 5. Plot confusion matrix

2) Plot Distribusi Label Asli VS Prediksi

Gambar 6 kedua merupakan grafik distribusi label asli dan prediksi. Berdasarkan visualisasi ini, terlihat bahwa jumlah prediksi model untuk label OBJ lebih tinggi dibandingkan dengan label aslinya, yang menunjukkan bahwa model cenderung mengklasifikasikan lebih banyak teks ke dalam kategori OBJ, meskipun sebagian sebenarnya merupakan SUBJ. Sebaliknya, label SUBJ memiliki jumlah prediksi yang lebih rendah dibandingkan dengan label aslinya, yang sejalan dengan kesalahan yang ditunjukkan pada Confusion Matrix. Sementara itu, label DIAG dan RESEP memiliki distribusi prediksi yang cukup sesuai dengan data aslinya, yang menunjukkan performa yang lebih baik pada kedua kategori ini.



Gambar 6. Plot distribusi label asli dan prediksi

Berdasarkan hasil visualisasi ini, dapat disimpulkan bahwa model sudah cukup baik dalam mengenali label OBJ, DIAG, dan RESEP, tetapi masih perlu perbaikan dalam membedakan antara SUBJ dan OBJ, yang kemungkinan disebabkan oleh kemiripan struktur kalimat atau kurangnya representasi data pada kategori SUBJ.

IV. KESIMPULAN

Berdasarkan hasil implementasi model klasifikasi teks rekam medis berbasis NER dengan BERT, diperoleh akurasi sebesar 84,8%, yang menunjukkan bahwa model memiliki performa yang cukup baik dalam mengkategorikan teks ke dalam empat label: Objektif (OBJ), Subjektif (SUBJ), Diagnosa (DIAG), dan Resep Obat (RESEP).

Classification report, model memiliki precision tertinggi pada label RESEP (1.00), yang menunjukkan bahwa semua prediksi untuk resep obat selalu benar. Label DIAG juga memiliki performa yang sangat baik dengan *F1-score* sebesar 0.93, yang menunjukkan keseimbangan antara *precision* dan *recall* yang tinggi. Namun, model mengalami kesulitan dalam membedakan label SUBJ, dengan *recall* hanya 0.50 dan *F1-score* 0.62, yang mengindikasikan banyaknya data subjektif yang terklasifikasi sebagai kategori lain, terutama OBJ.

Hasil *Confusion Matrix* juga mendukung hal tersebut, yang mana terdapat 17 sampel SUBJ yang salah diklasifikasikan sebagai OBJ. Selain itu, terdapat 4 sampel DIAG yang dikategorikan sebagai SUBJ. Sebaliknya, label RESEP dan OBJ memiliki tingkat klasifikasi yang sangat akurat, dengan sedikit atau bahkan tidak ada kesalahan prediksi.

Proses *training*, model mengalami penurunan *training loss* dan *validation loss* yang signifikan selama 5 epoch, menunjukkan bahwa model berhasil belajar dengan baik tanpa mengalami *overfitting*. Pada epoch terakhir, nilai *training loss* mencapai 0.0028 dan *validation loss* sebesar 0.0017, yang menunjukkan bahwa model telah mencapai konvergensi dengan performa yang stabil.

Secara keseluruhan, model ini telah menunjukkan performa yang baik dalam mengklasifikasikan teks rekam medis, terutama pada label DIAG, OBJ, dan RESEP. Namun, masih terdapat kekurangan dalam mengenali label SUBJ,

sehingga perbaikan dapat dilakukan dengan menyeimbangkan jumlah data subjektif dalam dataset atau melakukan penyempurnaan pada representasi fitur model agar lebih sensitif terhadap perbedaan antara OBJ dan SUBJ.

DAFTAR PUSTAKA

- [1] A. A. Morris, "Looking North to GUIDE Better Care for Heart Failure Is Not Black or White," *JACC: Heart Failure*.
- [2] J. Rangaswami and P. A. McCullough, "Clinical Context of Dyskalemias Across the Heart Failure Spectrum and Their Associated Adverse Outcomes," *JACC Heart Fail*, vol. 7, no. 6, p. 533, Jun. 2019, doi: 10.1016/j.jchf.2019.01.005.
- [3] D. Jurafsky and J. H. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models Third Edition draft Summary of Contents."
- [4] E. H. Houssein, R. E. Mohamed, G. Hu, and A. A. Ali, "Adapting transformer-based language models for heart disease detection and risk factors extraction," *J Big Data*, vol. 11, no. 1, Dec. 2024, doi: 10.1186/s40537-024-00903-y.
- [5] A. Vaswani *et al.*, "Attention Is All You Need," Jun. 2017, [Online]. Available: <http://arxiv.org/abs/1706.03762>
- [6] C. Wang *et al.*, "Named entity recognition (NER) for Chinese agricultural diseases and pests based on discourse topic and attention mechanism," *Evol Intell*, vol. 17, no. 1, pp. 457–466, Feb. 2024, doi: 10.1007/s12065-022-00727-w.
- [7] H. Pooja and M. P. P. Jagadeesh, "A Deep Learning Based Approach for Biomedical Named Entity Recognition Using Multitasking Transfer Learning with BiLSTM, BERT and CRF," *SN Comput Sci*, vol. 5, no. 5, Jun. 2024, doi: 10.1007/s42979-024-02835-z.
- [8] A. Vaswani *et al.*, "Attention Is All You Need," Jun. 2017, [Online]. Available: <http://arxiv.org/abs/1706.03762>
- [9] 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, Dec. 2022, doi: 10.1186/s12911-022-02059-2.
- [10] A. Thukral, S. Dhiman, R. Meher, and P. Bedi, "Knowledge graph enrichment from clinical narratives using NLP, NER, and biomedical ontologies for healthcare applications," *International Journal of Information Technology (Singapore)*, vol. 15, no. 1, pp. 53–65, Jan. 2023, doi: 10.1007/s41870-022-01145-y.
- [11] Á. García-Barragán, A. González Calatayud, O. Solarte-Pabón, M. Provencio, E. Menasalvas, and V. Robles, "GPT for medical entity recognition in Spanish," *Multimed Tools Appl*, 2024, doi: 10.1007/s11042-024-19209-5.
- [12] R. P. Kusumawardani and K. N. Kusumawati, "Named entity recognition in the medical domain for Indonesian language health consultation services using bidirectional-lstmcrf algorithm," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 1146–1156, doi: 10.1016/j.procs.2024.10.344
- [13] N. Liu, Q. Hu, H. Xu, X. Xu and M. Chen, "Med-BERT: A Pretraining Framework for Medical Records Named Entity Recognition," in *IEEE Transactions on Industrial Informatics*, vol. 18, no. 8, pp. 5600-5608, Aug. 2022, doi: 10.1109/TII.2021.3131180.
- [14] U. Naseem, M. Khushi, V. Reddy, S. Rajendran, I. Razzak and J. Kim, "BioALBERT: A Simple and Effective Pre-trained Language Model for Biomedical Named Entity Recognition," 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 2021, pp. 1-7, doi: 10.1109/IJCNN52387.2021.9533884.
- [15] M. U. Javeed, M. S. Ali, A. Iqbal, M. Azhar, S. M. Aslam and I. Shabbir, "Transforming Heart Disease Detection with BERT: Novel Architectures and Fine-Tuning Techniques," 2024 International Conference on Frontiers of Information Technology (FIT), Islamabad, Pakistan, 2024, pp. 1-6, doi: 10.1109/FIT63703.2024.10838424.