

Terbit online pada laman : <http://teknosi.fti.unand.ac.id/>

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |



Pra-Anotasi pada Model BioBERT untuk Pengenalan Entitas Bernama Laporan Radiologi Klinis Bahasa Indonesia

Arum Maulia ^{a,*}, Danang Wahyu Utomo ^b

^a Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro

^b Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 04 November 2025

Revisi Akhir: 18 Desember 2025

Diterbitkan Online: 15 Januari 2026

KATA KUNCI

*Pengenalan Entitas Bernama Klinis (NER),
Laporan Radiologi,
Ontologi RadLex,
Pra-anotasi, Validasi Manusia,
BioBERT.*

KORESPONDENSI

E-mail: 111202214108@mhs.dinus.ac.id

ABSTRACT

Penelitian ini menekankan pada teknik pra-anotasi untuk pengembangan model Named Entity Recognition (NER) dalam domain medis, khususnya pada laporan radiologi hati. Salah satu kendala utama dalam pengembangan model Named Entity Recognition (NER) adalah terbatasnya ketersediaan data beranotasi. Tantangan ini semakin kompleks ketika di hadapkan dengan skema anotasi yang di implementasikan pada data privat. Untuk menangani hal tersebut, penelitian ini mengusulkan metode pendekatan penyusunan dataset Named Entity Recognition (NER) yang berkualitas guna meningkatkan performa model. Proses penelitian di mulai dengan pembuatan silver corpus melalui proses pra-anotasi semi otomatis. Pada tahap ini dimulai dengan pembersihan data laporan radiologi, kemudian tahap selanjutnya memasukkan data yang telah di bersihkan ke dalam kamus terminologi medis (Radlex) untuk memetakan dan memberi label pada entitas medis yang relevan. Hasil anotasi terminologi medis tersebut kemudian divalidasi untuk menghasilkan gold corpus sebagai data anotasi final. Untuk mengukur kinerja pendekatan ini, dilakukan perbandingan anatara performa model BioBERT yang telah di fine-tune menggunakan hasil anotasi melalui tahap dari skema yang diusulkan dan model yang dilatih menggunakan data hasil anotasi manual melalui Label Studio. Hasil eksperimen menunjukkan bahwa model yang dilatih menggunakan korpus pre-anotasi Radlex memberikan peningkatan performa 45% dibandingkan dengan anotasi manual, sekaligus menghasilkan efisiensi waktu dalam proses pengerjaan.

1. PENDAHULUAN

Perkembangan pesat di bidang kecerdasan buatan (AI) dan pemrosesan bahasa alami (NLP) telah memberikan peran penting dalam bidang medis, khususnya dalam analisis teks klinis [1]. Disisi lain, penerapan AI telah banyak digunakan dalam konteks analisis citra medis seperti penelitian yang dilakukan oleh Zacky et al. (2025) yang memanfaatkan arsitektur U-NET dengan penggabungan modul residual dan mekanisme attention untuk tugas segmentasi tumor hati citra CT Scan. Pendekatan ini terbukti mampu untuk mengoptimalkan kemampuan model dalam mengenali struktur hati. Penelitian ini menunjukkan bahwa penerapan AI dibidang medis terus berkembang baik domain citra maupun teks [2]. Salah satu bidang penelitian yang sedang

berkembang adalah *Named Entity Recognition* (NER), proses mengidentifikasi dan pelabelan entitas penting dalam teks. Dalam *Natural Language Processing* (NLP), proses pengembangan model berbasis pembelajaran mesin sangat bergantung pada ketersediaan korpus beranotasi berkualitas tinggi.

Korpus ini berfungsi sebagai dasar untuk melatih dan mengevaluasi sistem, seperti sebagaimana diterapkan dalam tugas *Named Entity Recognition* (NER) di bidang medis [3]. Namun, membuat korpus beranotasi secara manual merupakan proses yang sangat memakan waktu, membutuhkan keahlian khusus, serta biaya yang tinggi [4]. Dalam lingkup yang lebih luas ketersediaan korpus beranotasi berbahasa Indonesia masih terbatas sebagai contoh *Spell Error Corpus for Minang Language* (SPEML) sebuah korpus Bahasa Minang dikembangkan oleh

Dewi et al. (2023) dibuat dengan tujuan menangani kesalahan ejaan Bahasa Minang dan di klasifikasikan ke dalam tujuh jenis kesalahan. Proses pengembangan SPEML menghadapi kendala seperti minimnya sumber daya linguistik, tidak adanya korpus kesalahan ejaan khusus, dan proses anotasi yang kompleks [5].

Tantangan tersebut semakin kompleks ketika berhadapan dengan teks medis. Dalam data klinis, belum tersedianya anotasi menyebabkan ambiguitas dalam proses pelabelan, terutama ketika berhadapan teks medis seperti laporan radiologi yang mengandung istilah teknis, struktur kalimat yang tidak baku, dan konteks klinis yang kompleks sehingga sulit dipahami oleh anotator non-medis [6][8]. Misalnya, istilah seperti lesi hipodens atau nodul solid memerlukan pemahaman medis untuk menentukan apakah kata tersebut termasuk kategori “temuan”, “diagnosis”, atau “organ tubuh” [7]. Keterbatasan anotator yang memiliki latar belakang medis menyebabkan proses anotasi manual menjadi tidak efisien, serta berpotensi menghasilkan inkonsistensi dalam pelabelan [8]. Kendala serupa muncul pada penelitian yang fokus terhadap pengkodean (ICD-10) dalam rekam medis Indonesia. Menurut Parjono (2023) pengkodean manual oleh petugas rekam medis sangat bergantung pada interpretasi individu, memakan waktu, serta rentan kesalahan. Diagnosis dokter yang dituliskan dalam bentuk *free text* memiliki struktur dan terminologi yang berbeda, sehingga sulit untuk diolah. Oleh karena itu, penerapan text mining pada data medis penting dalam membantu proses pengkodean penyakit secara konsisten dan akurat [9].

Selain itu, penelitian oleh Singh et al. (2021) menunjukkan bahwa pada dataset umum seperti OntoNotes 5.0 yang digunakan secara luas dalam Named Entity Recognition (NER), ditemukan berbagai kesalahan anotasi yang mempengaruhi kinerja model. Peneliti mengidentifikasi serta memperbaiki kesalahan sekitar 10% kalimat dalam dataset tersebut, dan membuktikan bahwa kesalahan anotasi secara signifikan menurunkan akurasi model dan menghambat pemahaman terhadap performa sistem NER [10]. Temuan ini menegaskan bahwa anotasi manual, meskipun dilakukan oleh ahli, tetap rentan terhadap kesalahan dan inkonsistensi, yang pada akhirnya berdampak pada sistem NLP yang dibangun di atasnya. Untuk mengatasi kendala tersebut, salah satu pendekatan yang telah banyak dikembangkan adalah *Pre-annotation*. *Pre-annotation* merupakan proses pemberian label awal secara otomatis sebelum anotasi manual dilakukan, dengan menggunakan metode berbasis kamus medis (*dictionary-based*), aturan linguistik (*rule-based matching*), atau model pembelajaran mesin yang telah dilatih sebelumnya [11]. Tujuan utama pendekatan ini adalah untuk mengurangi waktu dan beban kognitif anotator dengan menyediakan anotasi awal yang dapat diperiksa dan diperbaiki. Penelitian yang dilakukan oleh Liu et al. (2023) menunjukkan bahwa penggunaan *Pre-annotation* dalam proses anotasi manual pada teks medis format bebas dapat meningkatkan efisiensi secara signifikan, yaitu dengan mengurangi waktu hingga 43% tanpa mengurangi, atau bahkan meningkatkan, kualitas hasil anotasi [11].

Dalam penelitian ini, *Pre-annotation* difasilitasi dengan menggunakan RadLex, sebuah ontologi radiologi yang dikembangkan oleh *Radiological Society of North America* (RSNA). RadLex berfungsi sebagai sumber terminologi standar yang mencakup berbagai konsep dalam radiologi, mulai dari anatomi, deskripsi lesi, prosedur pencitraan, hingga hasil observasi. Dengan mengintegrasikan RadLex ke dalam proses anotasi, peneliti dapat menjaga konsistensi pelabelan antar dokumen dan antar anotator [12]. Selain itu, RadLex memungkinkan proses anotasi yang lebih objektif, karena istilah medis yang ambigu dapat dipetakan ke konsep yang telah distandarisasi dalam sistem ontologi. Penelitian yang dilakukan oleh Tsuji et al. (2021) menunjukkan bahwa penggunaan RadLex dalam sistem Named Entity Recognition (NER) untuk laporan radiologi dapat meningkatkan konektivitas dengan istilah stem tern dalam *Foundational Model of Anatomy* (FMA) sebesar 20,33% dibandingkan dengan penggunaan terminologi default seperti SNOMED-CT dalam sistem cTAKES [13]. Dengan demikian, penggunaan RadLex pada tahap *Pre-annotation* tidak hanya mempercepat proses anotasi, tetapi juga memastikan keseragaman terminologi medis yang digunakan dalam korpus radiologi. Dengan pertimbangan latar belakang tersebut, penelitian ini bertujuan untuk membuat anotasi dengan menerapkan strategi *Pre-annotation* berbasis RadLex pada proses anotasi laporan radiologi berbahasa Indonesia. Dengan demikian, penelitian ini diharapkan dapat memberikan peran penting dalam pengembangan metodologi anotasi semi-otomatis yang efisien, konsisten, dan adaptif untuk ranah medis.

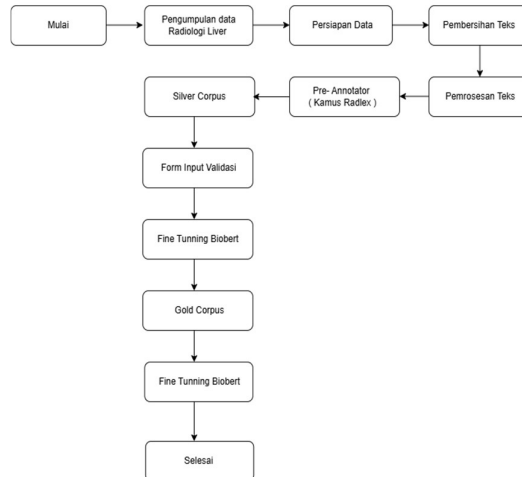
2. METODE

Alur penelitian ini disusun sebagaimana ditunjukkan pada Gambar 1 memperlihatkan tahapan utama dalam pelaksanaan penelitian. Adapun penjabaran mengenai setiap tahapan penelitian yang dilakukan untuk mencapai tujuan penelitian dijelaskan pada Gambar 1.

2.1. Pengumpulan Data

Penelitian ini menggunakan data privat laporan radiologi mengenai liver salah satu rumah sakit di Jawa Tengah sebagaimana yang ditunjukkan pada Tabel 1. Data tersebut didapatkan dari dokumen medis yang telah melalui proses de-identifikasi untuk memastikan perlindungan terhadap privasi informasi pasien yang bersangkutan. Proses de-identifikasi dilakukan dengan cara menghapus dan menyamarkan seluruh informasi yang berpotensi memperlihatkan identitas individu, seperti nama pasien, nomor rekam medis, nama dokter, tanggal pemeriksaan, serta informasi pribadi lainnya.

Dengan demikian, data yang digunakan dalam penelitian ini telah dilindungi sepenuhnya dan tidak dapat ditelusuri kembali kepada individu tertentu. Hasil dari tahapan ini berupa kumpulan teks mentah (*raw text corpus*) yang telah bebas dari informasi identitas pribadi pasien dan siap digunakan sebagai dasar dalam proses anotasi serta pelatihan model dengan metode BioBERT untuk tugas Named Entity Recognition (NER).



Gambar 1. Diagram Alur Penelitian

Tabel 1. Contoh Dataset

ID	Konten_Pemeriksaan	Kesimpulan
1	MSCT ABDOMEN DENGAN KONTRAS\nKLINIS: HCC\nHepar ukuran tampak membesar...	KESAN:\nHepatomegali disertai multipel nodul solid dengan...
2	MSCT ABDOMEN DENGAN KONTRAS\nKLINIS TUMOR INTRAABDOMEN\nTa mpak penebalan circumferential...	KESAN:\nMassa solid pada colon descendens (tebal maksimal ± 4.37 cm...
3	MSCT ABDOMEN DENGAN KONTRAS\nKLINIS : HCC\nHepar ukuran (± 15 cm) membesar...	KESAN:\nHepatomegaly dengan parenkim inhomogen disertai multiple...
4	MSCT ABDOMEN DENGAN KONTRAS\nKLINIS : HCC\nHepar	KESAN:\nHepatomegaly dengan multiple lesi solid difus

2.2. Persiapan Data

2.2.1. Preprocessing

Preprocessing memiliki peran penting dalam proses pengolahan dataset berbasis teks yang dengan tujuan untuk membersihkan, merapikan, serta meningkatkan kualitas data sebelum digunakan ke tahap pemrosesan selanjutnya. Langkah ini memiliki peran penting untuk memastikan bahwa informasi yang diekstraksi dari teks mempunyai akurasi yang tinggi dan dapat memperkuat peningkatan performa model yang dikembangkan dalam penelitian ini. Langkah awal dilakukan dengan mengonversi seluruh teks menjadi huruf kecil untuk memastikan bahwa model menangani kata dengan kapitalisasi berbeda sebagai entitas yang sama. Langkah ini berfungsi meminimalkan pengulangan kosakata dan meningkatkan efisiensi selama tahap *training* model. Tahap selanjutnya, diterapkan ekspresi reguler (*Regular Expression/Regex*) untuk menghapus tanda baca serta karakter selain huruf yang tidak memberikan kontribusi terhadap pemahaman konteks dalam tugas *Named Entity Recognition* (NER).

Tahapan ini diikuti dengan penerapan proses stopword removal guna mengeliminasi kata-kata umum yang tidak memiliki makna spesifik dalam identifikasi entitas medis. Kata sambung seperti “dan”, “yang”, dan “di” dihapus sebab tidak memberikan informasi penting bagi model dalam mengenali istilah medis. Selanjutnya, proses stemming diterapkan untuk mengembalikan setiap kata ke bentuk dasar atau lemma-nya. Tahapan ini bertujuan untuk menekan perbedaan bentuk morfologis pada kata yang memiliki makna dasar serupa, seperti “pembesaran”, “membesar”, dan “besar”, sehingga ketiganya direduksi menjadi satu bentuk dasar. Tahapan ini membantu model mengenali entitas dengan lebih baik karena setiap kata diwakili oleh bentuk dasarnya yang konsisten. Hasil dari semua proses *preprocessing* ini adalah teks yang telah bersih, konsisten, dan terstandarisasi,

sehingga dapat digunakan pada tahap *Pre-annotation* melalui kamus terminologi medis RadLex.

2.2.2. Pre-Annotator (Kamus Radlex)

Pada tahap ini, proses Pre-annotation dilakukan menggunakan bantuan Radlex, yaitu kamus terminologi radiologi yang berguna untuk mengidentifikasi dan melabeli istilah medis secara otomatis. Data yang berhasil melalui tahap pembersihan dan preprocessing kemudian dimasukkan ke dalam sistem Radlex secara bertahap dari mulai report pertama hingga terakhir. Setiap laporan dianotasi untuk mendapatkan istilah medis yang relevan dengan konteks radiologi.

Hasil anotasi yang dihasilkan oleh Radlex kemudian digunakan sebagai dasar untuk melakukan tahap selanjutnya yaitu pemetaan label semantik. Pada tahap ini, setiap istilah medis yang dihasilkan oleh RadLex dimapping ke dalam lima label utama, yaitu *Anatomy*, *Procedure*, *Measurement*, *Disease*, *Finding*, disesuaikan dengan ID terminologi yang diberikan oleh RadLex. Proses ini menghasilkan data berlabel semantik yang sesuai dan struktur, di mana setiap istilah dalam teks radiologi memiliki label sesuai konteks medisnya. Selanjutnya, hasil dari pemetaan label semantik ini diterapkan secara konsisten ke seluruh data yang digunakan dalam penelitian ini, sehingga menghasilkan sebuah korpus beranotasi awal (*Silver Corpus*) dan siap untuk diproses pada tahap berikutnya yaitu memvalidasi *Silver Corpus*.

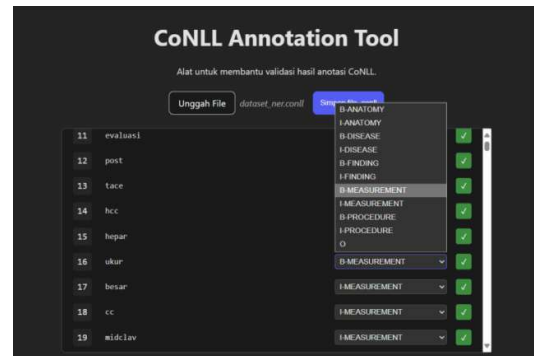
2.2.3. Silver Corpus

Tahap selanjutnya adalah pembentukan *Silver Corpus*, yaitu korpus yang telah teranotasi awal dari hasil proses *Pre-annotation* menggunakan kamus RadLex. Pada tahap ini, hasil anotasi RadLex yang telah dimapping ke dalam lima label semantik diantaranya *Anatomy*, *Procedure*, *Measurement*, *Disease*, dan *Finding* dikompilasi dan disusun menjadi satu kesatuan korpus. *Silver Corpus* berfungsi sebagai data awal yang menggambarkan hasil anotasi otomatis sebelum dilakukan tahap validasi manual oleh para anotator ahli. Meskipun anotasi dihasilkan secara otomatis dengan menggunakan RadLex, masih berpotensi adanya ketidaksesuaian konteks atau klasifikasi istilah medis yang kurang tepat. Oleh karena itu, korpus ini belum dianggap sebagai *Gold Corpus*, melainkan sebagai dasar untuk tahap evaluasi dan penyempurnaan anotasi. Dalam penelitian ini, *Silver Corpus* digunakan sebagai bahan penilaian terhadap kinerja dan kemampuan *Pre-annotation* berbasis RadLex dalam tugas mengidentifikasi entitas medis pada laporan radiologi liver. Hasil dari tahap ini menjadi acuan penting dalam proses validasi dan perbaikan label, sehingga korpus yang dihasilkan pada tahap selanjutnya memiliki tingkat kesesuaian dan konsistensi yang lebih tinggi.

2.2.4. Form Input Validasi

Setelah proses pembuatan *Silver Corpus* selesai, tahap selanjutnya adalah melakukan validasi pada hasil anotasi untuk memastikan kesesuaian label semantik yang dihasilkan dari proses mapping sebelumnya dengan konteks klinis dalam laporan medis. Tahapan ini menjadi penting karena kesalahan pada tahap anotasi dapat menyebabkan masalah terhadap akurasi model yang

akan dilatih. Proses validasi dilakukan menggunakan alat bantu berbasis web (web-based tool) sebagaimana yang ditunjukkan pada gambar 2 dirancang dan dikembangkan khusus untuk mempermudah proses validasi hasil anotasi. Melalui sistem ini, anotator dapat melakukan peninjauan setiap token dan label secara interaktif, memperbaiki kesalahan pelabelan, serta memberikan penyesuaian konteks apabila ditemukan inkonsistensi.



Gambar 2. Conll Annotation Tool

Tool yang digunakan dalam proses validasi diberi nama CoNLL Annotation Tool, yang dirancang untuk membaca file hasil anotasi berformat .conll kemudian akan menampilkan dalam bentuk daftar token beserta label yang berkaitan. Pada antarmuka tersebut, pengguna dapat mengunggah file hasil anotasi format conll (unggah file .conll), memeriksa setiap token yang telah terlabeli, serta mengubah label secara langsung apabila ditemukan ketidaksesuaian. Setiap token dalam laporan ditampilkan secara bersamaan dengan daftar label yang telah ditentukan, seperti *b-anatomy*, *i-anatomy*, *b-disease*, *i-disease*, *b-finding*, *i-finding*, *b-measurement*, *i-measurement*, *b-procedure*, *i-procedure*.

Tahap ini dilakukan secara sistematis untuk memastikan hasil *pre-annotation* dari RadLex sudah sesuai atau memerlukan perbaikan secara manual oleh radiolog. Selain itu, sistem ini dilengkapi dengan tombol validasi di sisi kanan setiap token, yang menandakan bahwa token tersebut telah ditinjau dan divalidasi oleh anotator. Setiap langkah validasi dilakukan secara teliti untuk memastikan bahwa label yang diberikan sesuai dengan konteks medis dan tidak terjadi kesalahan dalam menentukan entitas. Proses validasi ini juga menjaga konsistensi label antar *report*, terutama pada istilah medis yang memiliki kemiripan bentuk atau makna. Setelah seluruh data selesai ditinjau, hasil anotasi yang telah tervalidasi dapat disimpan kembali dalam format .conll menggunakan tombol "Simpan file .conll", sehingga file tersebut dapat langsung digunakan ke tahap pelatihan model *Named Entity Recognition* (NER). Dengan demikian, tool ini tidak hanya dapat mempermudah proses validasi hasil anotasi, tetapi juga meningkatkan efisiensi dan kinerja pembuatan korpus berlabel untuk penelitian selanjutnya.

2.2.5. Fine-tuning Biobert

Tahap fine-tuning BioBERT dilakukan setelah seluruh data hasil anotasi tervalidasi dan disimpan dalam format .conll melalui tool

validasi berbasis web yang dijelaskan pada tahap sebelumnya. Proses ini merupakan lanjutan dari tahap pre-annotation dan validasi, di mana data yang telah di validasi digunakan untuk menyesuaikan model BioBERT guna dapat memahami konteks bahasa medis yang spesifik, terutama pada laporan radiologi. Dalam penelitian ini, model BioBERT yang sebelumnya telah dilatih menggunakan korpus medis umum seperti PubMed dan PMC, diadaptasi kembali (domain adaptation) menggunakan data hasil anotasi dari RadLex pre-annotator dan anotasi manual.

Sebelum dilakukan tahap pelatihan model, data tersebut dikonversi ke dalam format urutan token dan label sesuai ketentuan Name Entity Recognition (NER), yaitu format BIO (*Begin-Inside-Outside*), agar dapat dibaca oleh model BioBERT. Proses training dilakukan dengan mengatur beberapa parameter penting seperti learning rate, batch size, jumlah epoch, dan weight decay guna memastikan kesesuaian model secara optimal. Selama proses training BioBERT, model mempelajari pola linguistik dan semantik dalam teks medis untuk mengetahui entitas seperti *anatomy*, *disease*, *procedure*, *measurement*, dan *finding* berdasarkan hasil anotasi yang telah disiapkan pada tahap sebelumnya.

3. HASIL

3.1. Persiapan Data

Pada tahap awal penelitian ini, dilakukan proses preprocessing pada dataset privat laporan radiologi berupa data yang belum diolah dan tidak memiliki label anotasi. Tahapan ini bertujuan untuk membersihkan dan menstandarkan teks supaya dapat digunakan pada proses anotasi dan pemodelan *Named Entity Recognition* (NER) di tahap selanjutnya. Proses preprocessing data dilakukan melalui beberapa langkah seperti konversi seluruh teks menjadi huruf kecil (*case folding*) untuk menyesuaikan format penulisan, penerapan ekspresi reguler (*regex*) yang bertujuan menghapus tanda baca serta karakter non-alfabet yang tidak relevan, penghapusan *stopword* untuk menghapus kata yang bersifat umum dan tidak mempunyai makna penting dalam proses pengerjaan identifikasi entitas medis, serta menerapkan *stemming* untuk mengembalikan setiap kata ke bentuk dasarnya. Selain itu, dilakukan normalisasi istilah medis dengan tujuan agar ejaan dan penulisan kata tetap konsisten sesuai dengan terminologi medis yang digunakan dalam laporan radiologi. Dengan melalui beberapa tahapan tersebut diharapkan data yang akan digunakan pada langkah selanjutnya dapat terstandarisasi dan terbebas dari noise, sehingga siap digunakan untuk training model BioBERT.

Tabel 2. Hasil *Preprocessing*

Preprocessing Data	Hasil Preprocessing
Teks Asli	MSCT ABDOMEN DENGAN KONTRAS\ nKLINIS : HCC\ nHepar ukuran (± 15 cm) membesar, parenkim inhomogen, tepi iregular, liver tip tumpul.
Case Folding	msct abdomen dengan kontras klinik hcc hepar ukuran (± 15

cm)membesar, parenkim
inhomogen, tepi iregular, liver
tip tumpul

Regular Expression

msct abdomen dengan kontras
klinik hcc hepar ukuran 15 cm
membesar parenkim
inhomogen tepi iregular liver
tip tumpul

Stopword Removal

msct abdomen kontras klinik
hcc hepar ukuran membesar 15
cm parenkim inhomogen tepi
iregular liver tip tumpul

Stemming

msct abdomen kontras klinik
hcc hepar ukuran 15 cm besar
parenkim inhomogen tepi
iregular liver tip tumpul

3.2. Pre-Annotator (Kamus Radlex)

Tahapan selanjutnya setelah teks berhasil melalui proses pembersihan (*preprocessing*) yaitu memasukkan laporan radiologi dari report pertama sampai report terakhir secara bertahap ke dalam kamus terminologi medis RadLex untuk dilakukan proses anotasi otomatis. Tahap ini menerapkan beberapa ketentuan seperti *match longest only*, *match partial words*, dan *include mappings*, dengan ontologi Radlex sebagai utama terminologi medis. Tujuannya adalah untuk memberikan label sebagai istilah medis yang relevan secara konsisten berdasarkan standar terminologi yang diakui dalam domain radiologi. Melalui proses ini, istilah medis dalam teks akan otomatis dikenali dan diberi label awal sesuai dengan entitasnya, sehingga akan mempermudah tahapan validasi manual selanjutnya. Hasil dari proses *pre-annotator* ini dapat dilihat pada Tabel 3 berikut.

Tabel 3. Hasil *Pre-Annotator* Radlex

Teks	ID	Hyperlink
Abdomen	RID56	http://radlex.org/RID/RID56
Multiple	RID5765	http://radlex.org/RID/RID5765
Metastasis	RID5231	http://radlex.org/RID/RID5231
Diameter	RID13432	http://radlex.org/RID/RID13432
Delayed	RID39065	http://radlex.org/RID/RID39065
Phase	RID39065	http://radlex.org/RID/RID39065
Vena	RID39065	http://radlex.org/RID/RID39065
Lienalis	RID1208	http://radlex.org/RID/RID1208
HCC	RID58	http://radlex.org/RID/RID58
Aorta	RID480	http://radlex.org/RID/RID480
Pancreatic	RID176	http://radlex.org/RID/RID176
Duct	RID176	http://radlex.org/RID/RID176

Hasil *pra-anotasi* setelahnya dimapping berdasarkan ID yang dihasilkan dari RadLex. Proses mapping ini dilakukan untuk menghubungkan setiap entitas dengan label semantik yang telah ditetapkan sebelumnya, meliputi *Anatomy*, *Procedure*, *Measurement*, *Disease*, *Finding*. Selanjutnya, hasil dari mapping ini akan digunakan dalam pembentukan *silver corpus*, yaitu

sekumpulan data yang telah dianotasi secara semi-otomatis dan siap untuk proses validasi.

3.3. Silver Corpus

Setelah melalui tahap *pra-anotasi* dan *mapping* label semantik menggunakan RadLex, didapatkan *silver corpus* yang berisi pasangan token serta label entitas dari hasil anotasi semi-otomatis. Korpus ini dibentuk dalam format CoNLL (*Conference on Natural Language Learning*) agar dapat digunakan pada tahap *training* dan evaluasi model *Named Entity Recognition* (NER). Format ini memudahkan sistem dalam membaca hubungan antara token dan label secara terstruktur serta menjaga konsistensi struktur data selama proses *training*. Dengan dibentuknya *silver corpus* yang terstandarisasi, proses *training* model dapat dilakukan secara lebih efisien. Contoh hasil *silver corpus* ditunjukkan pada Tabel 4 berikut.

Tabel 4. Hasil *Silver Corpus*

Teks	Label
msct	O
abdomen	B-Anatomy
klinis	O
evaluasi	O
post	O
pancreas	B-Anatomy
vena	I-Anatomy
cava	O
artery	B-Anatomy
washout	B-finding
venous	O
delayed	B-Procedure
phase	B-Measurement
diameter	I-Measurement
40	O
hcc	B-Disease
metastasis	I-Disease
46-50	O
ct	O
vena	B-Anatomy
lienalis	I-Anatomy
vena	I-Anatomy
porta	I-Anatomy
spondylosis	B-Disease
tumor	I-Disease
struktur	O
retensi	O
segmen	O
8	O
5	O
multiple	O
ilium	B-Anatomy
kanan	O
kiri	O
corpus	B-Anatomy
vertebra	I-Anatomy
th	O
12	O
bone	B-Anatomy
metastasis	B-Disease

3.5. Validasi Silver Corpus

Dari hasil *silver corpus* yang telah dihasilkan, terdapat beberapa entitas yang perlu disesuaikan dengan dilakukan proses validasi

ulang. Beberapa istilah belum sepenuhnya sempurna teridentifikasi ke dalam label semantik yang sesuai. Pada istilah “MSCT” seharusnya dikategorikan sebagai *Procedure*, sementara nilai numerik seperti angka dan ukuran seharusnya masuk dalam label *Measurement*. Selain itu, istilah “CT” semestinya terlabeli sebagai *Procedure*, tetapi pada hasil awal masih terdapat ketidaksesuaian label. Dengan demikian, tahap validasi dilakukan sebagai upaya memperbaiki kesalahan anotasi tersebut agar label menjadi lebih konsisten. Selain itu untuk beberapa entitas lainnya juga perlu dilakukan peninjauan kembali melalui proses validasi guna memastikan kesesuaian label dengan konteks kalimat. Hasil validasi kedua yang kemudian digunakan sebagai dasar pembentukan *gold corpus*.

3.6. Fine-tuning Biobert

Tahap selanjutnya adalah *fine-tuning* dijalankan menggunakan *framework Hugging Face Transformers* untuk menyesuaikan model *pre-trained* BioBERT dengan data laporan radiologi. Proses *training* di kelola melalui kelas *Training Arguments* dengan melibatkan parameter utama, seperti *output_dir* untuk menyimpan hasil model, *eval_strategy="epoch"* agar evaluasi dilakukan setiap *epoch*, *learning_rate=5e-5* untuk menjaga stabilitas pembelajaran, ukuran batch sebesar 16, jumlah *epoch* sebanyak 20, serta *weight_decay* sebesar 0.01 sebagai regularisasi. Model yang paling terbaik akan disimpan berdasarkan nilai *f1-score* tertinggi. Hasil dari *fine-tuning* dapat dilihat pada Tabel 5 sebagai berikut :

Tabel 5. Hasil *Fine-tuning* Biobert

Label	Precision	Recall	F1-score
Anatomy	0,92	0,97	0,94
Disease	0,88	0,95	0,91
Finding	0,92	0,97	0,94
Measurement	1.00	1.00	1.00
Procedure	1.00	1.00	1.00
Micro Avg	0,92	0,97	0,94
Macro Avg	0,94	0,98	0,96
Weighted Average	0,92	0,97	0,94

Hasil evaluasi dapat dilihat pada Tabel 5 merupakan proses *fine-tuning* model BioBERT pada *Name Entity Recognition* (NER) di domain medis. Secara umum, model ini mampu mencapai tingkat efektivitas yang tinggi dengan hasil nilai *f1-score* berada pada rentang 0.91 hingga 1.00 untuk semua label entitas klinis. Label *measurement* dan *procedure* mencapai tingkat ketepatan tinggi, *recall*, dan *f1-score* memperoleh nilai sempurna 1.00, yang menandakan kemampuan model untuk dapat mengenali kedua entitas ini secara konsisten tanpa *false prediction*. Label *anatomy* dan *finding* juga menunjukkan hasil yang tinggi dengan *f1-score* 0.94 menggambarkan model mampu mendeteksi entitas *anatomy* dan *finding* dengan tingkat akurasi yang tinggi.

Berbeda dengan label yang lain, *disease* mendapat nilai *f1-score* terendah 0.91 walaupun nilai *recall* tetap optimal 0.95. Hal ini menunjukkan masih terdapat kesalahan prediksi positif pada entitas ini, yang kemungkinan disebabkan oleh *overlap* semantik antara *disease* dengan entitas lain atau variasi konteks kalimat

pada data *training*. Secara keseluruhan, nilai *micro* dan *weighted average* menghasilkan *f1-score* sebesar 0,94, sedangkan *macro average* mencapai 0.96. Perbedaan ini tidak terlalu signifikan menunjukkan bahwa model tidak hanya unggul pada kelas mayoritas, tetapi juga memiliki performa yang stabil pada semua entitas, sehingga mendukung konsistensi model dalam tugas NER domain medis.

4. PEMBAHASAN

4.1. Fine-tuning Anotasi Manual (Label Studio)

4.1.1. Fine-tuning Anotasi Manual

Tahapan anotasi data dilakukan secara manual menggunakan Label Studio dengan lima label utama, yaitu *anatomy*, *disease*, *finding*, *measurement*, dan *procedure*. Data yang berhasil dianotasi kemudian digunakan untuk bahan dilakukan *fine-tuning* model BioBERT. Proses *training* dikendalikan melalui kelas *TrainingArguments* dengan beberapa parameter utama, yaitu *output_dir* sebagai lokasi penyimpanan hasil model, *eval_strategy="epoch"* agar evaluasi dilakukan di setiap *epoch*, serta *learning rate* sebesar $5e-5$ untuk menjaga stabilitas pada proses *training*. Ukuran *batch* ditetapkan sebesar 16, jumlah *epoch* sebanyak 20, dan *weight decay* sebesar 0.01 digunakan sebagai bentuk mekanisme pengendalian untuk mencegah *overfitting*. Model terbaik secara otomatis disimpan berdasarkan hasil dari nilai *f1-score* tertinggi yang diperoleh selama *training*. Hasil dari proses *fine-tuning* tersebut dapat dilihat pada Tabel 6 sebagai berikut :

Tabel 6. Hasil *Fine-tuning* Biobert (Anotasi Manual)

Label	Precision	Recall	F1-Score
Anatomy	0,61	0,62	0,62
Disease	0,73	0,57	0,64
Finding	0,67	0,73	0,70
Measurement	0,55	0,53	0,54
Procedure	0.63	0,74	0,68

Micro Avg	0,63	0,66	0,64
Macro Avg	0,64	0,64	0,64
Weighted Avg	0,63	0,66	0,64

Hasil *fine-tuning* model BioBERT menggunakan data hasil anotasi manual seperti yang ditunjukkan pada Tabel 6. Secara umum, model menunjukkan performa cukup baik dengan memperoleh nilai rata-rata *f1-score* sebesar 0,64 pada keseluruhan pengujian, baik untuk *micro average*, *macro average*, ataupun *weighted average*. Nilai tersebut menunjukkan bahwa model baik dalam mengenali entitas medis secara stabil dengan keseimbangan antara *precision* dan *recall* di seluruh label. Secara lebih rinci, label *finding* memperoleh performa terbaik dengan *f1-score* sebesar 0.70 *precision* sebesar 0.67, dan *recall* sebesar 0.73. Hasil ini menunjukkan bahwa model mempunyai kemampuan yang baik dalam mengenali entitas *finding* dengan tingkat ketidaksesuaian klasifikasi yang relatif rendah. Kemungkinan besar, hal ini disebabkan oleh munculnya entitas *finding* yang lebih dominan dalam data *training*. Label *disease* menunjukkan nilai *precision* paling tinggi sebesar 0.73, sedangkan nilai *recall* memperoleh nilai sebesar 0.57, dan menghasilkan *f1-score* 0.64. Hasil tersebut membuktikan bahwa model memiliki tingkat ketepatan dalam mengenali istilah *disease*, tetapi masih ditemui sejumlah entitas yang tidak dapat terdeteksi. Sementara itu, label *procedure* memperlihatkan kinerja yang tergolong baik dengan memperoleh nilai *recall* 0.74 dan *f1-score* 0.68, hal ini menandakan kemampuan model yang relatif optimal dalam mengenali tindakan medis walaupun masih terdapat ketidaksesuaian prediksi positif. Pada label *anatomy*, nilai *f1-score* sebesar 0.62 mengindikasikan bahwa model dapat mengenali label *anatomy* secara tepat dan seimbang, walaupun masih terdapat peluang untuk meningkatkan akurasi. Sebaliknya, label *measurement* menunjukkan hasil paling rendah dengan *f1-score* 0.54, hal tersebut menunjukkan bahwa model masih menghadapi kendala dalam mengenali entitas pengukuran seperti angka, satuan, dan simbol medis.

4.2. Hasil Training Model BioBERT

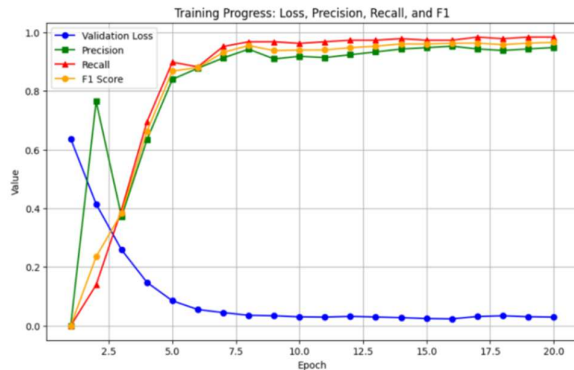


Gambar 3. Training Curve Anotasi Manual

Gambar 3 menunjukkan perkembangan nilai *validation loss*, *precision*, *recall*, dan *f1-score* selama proses *training* model sebanyak 20 *epoch*. Berdasarkan grafik diatas, *validation loss* mengalami penurunan yang cukup signifikan pada proses awal *training*, dari nilai awal sekitar 1.8 menjadi 0.9 hingga mencapai

epoch ke-10, kemudian menunjukkan pola yang relatif stabil sampai akhir proses *training*. Pola ini menandakan bahwa model berhasil mempelajari pola data secara bertahap sampai mencapai kestabilan meskipun belum sepenuhnya optimal. Sementara itu, metrik evaluasi seperti *precision*, *recall*, *f1-score* menunjukkan

peningkatan yang konsisten sampai mencapai nilai dengan rentang 0.55 - 0.6 setelah *epoch* ke-6, dan kemudian stabil. Nilai ini menggambarkan bahwa model telah mampu melakukan prediksi sehingga menghasilkan tingkat akurasi yang cukup baik, meskipun masih mempunyai peluang untuk melakukan perbaikan terutama dalam mendeteksi entitas atau kelas positif secara menyeluruh. Secara keseluruhan, hasil ini menunjukkan bahwa model sudah melalui proses *training* yang stabil tanpa adanya indikasi *overfitting*, namun masih memerlukan optimasi lebih lanjut seperti *learning rate*, *batch size*, jumlah *epoch*, atau peningkatan kualitas data *training* untuk memperoleh performa lebih optimal.

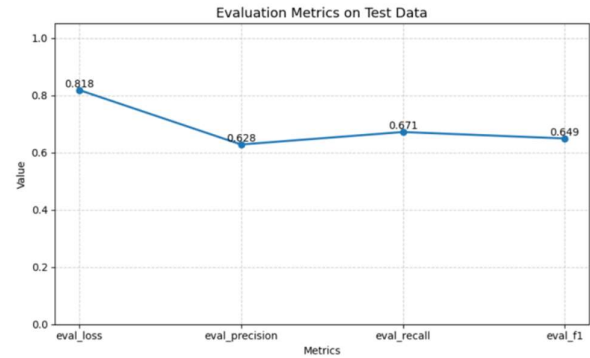


Gambar 4. Training Curve Anotasi (Pre- Annotator)

Gambar 4 menunjukkan perkembangan metrik evaluasi model selama proses *training* berlangsung sebanyak 20 *epoch* yang meliputi *validation loss*, *precision*, *recall*, dan *f1-score*. Berdasarkan grafik tersebut, terlihat bahwa model mengalami proses konvergensi yang stabil setelah melewati beberapa *epoch* awal. Nilai *validation loss* mengalami penurunan signifikan pada lima *epoch* pertama, kemudian menurun secara bertahap hingga mencapai nilai mendekati nol dan relatif stabil setelah *epoch* ke-6. Temuan tersebut menunjukkan bahwa model mampu mengenali pola data secara efektif dan tidak mengalami *overfitting* yang ditandai dengan tidak adanya peningkatan loss pada tahap akhir proses *training*.

Sementara itu, ketiga metrik evaluasi yaitu *precision* (garis hijau), *recall* (garis merah), dan *f1-score* (garis oranye), menunjukkan peningkatan tajam sejak *epoch* pertama hingga mencapai nilai di atas 0.9 setelah *epoch* ke-5. Nilai yang didapatkan kemudian mengalami kestabilan hingga akhir proses *training*. Hal ini menandakan bahwa model mempunyai performa klasifikasi yang sangat baik, disertai dengan tingkat kesalahan prediksi yang rendah (*high precision*), kemampuan mendeteksi entitas yang sesuai dengan sangat baik (*high recall*), serta keseimbangan antara keduanya yang optimal dengan proses *training* yang efisien. Penurunan *validation loss* yang stabil disertai peningkatan nilai *precision*, *recall*, dan *f1-score* secara konsisten menandakan bahwa model memiliki kemampuan untuk menerapkan pola yang telah melalui proses *training* yang baik terhadap data validasi.

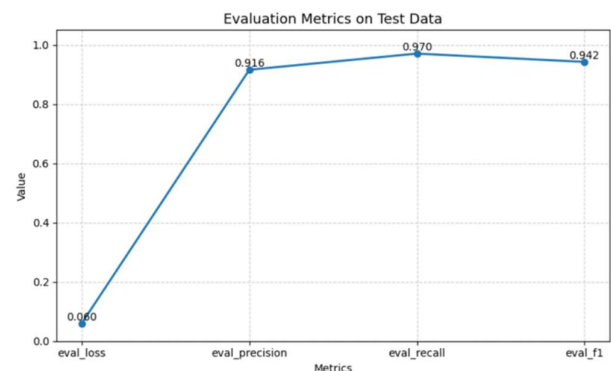
4.3. Hasil Evaluasi Model BioBERT



Gambar 5. Hasil Evaluasi Anotasi manual

Gambar 5 menampilkan hasil dari evaluasi performa model BioBERT pada data uji setelah dilakukan proses *fine-tuning* dengan menggunakan data anotasi manual. Evaluasi dilakukan dengan menggunakan empat metrik utama, yaitu *loss*, *precision*, *recall*, dan *f1-score*, sebagai bahan penilaian sejauh mana model mampu mengenali entitas medis secara akurat. Nilai *evaluation loss* sebesar 0.818 menunjukkan bahwa tingkat kesalahan prediksi model masih berada pada kategori sedang. Meskipun model telah mampu mempelajari pola dari entitas dengan baik, nilai ini menandakan bahwa masih terdapat untuk melakukan optimasi lebih lanjut. Berdasarkan metrik performa, nilai *precision* sebesar 0.628 menunjukkan bahwa sekitar 62,8% dari entitas yang diprediksi oleh model sepenuhnya relevan, menandakan kemampuan model yang cukup baik dalam menghindari *false positive*. Nilai *recall* sebesar 0.671 mengindikasikan bahwa model dapat mengenali sekitar 67,1% dari seluruh entitas medis yang seharusnya terdeteksi, sehingga performa yang ditunjukkan cukup efektif dalam mengenali *true positive*.

Nilai *f1-score* mencapai nilai 0.649 yang merupakan rata - rata antara *precision* dan *recall*, membuktikan bahwa model memiliki performa yang cukup seimbang antara ketepatan dan kelengkapan dalam mengenali entitas. Secara keseluruhan, hasil dari evaluasi ini menunjukkan bahwa model BioBERT proses *fine-tuning* mampu melakukan ekstraksi entitas medis dengan tingkat akurasi yang stabil dan akurat, walaupun masih mempunyai kemungkinan mengalami peningkatan performa melalui cara penyempurnaan parameter *training* dan perluasan korpus anotasi.



Gambar 6. Hasil Evaluasi Anotasi (*Pre-Annotator*)

Berdasarkan Gambar 6 menampilkan hasil metrik evaluasi dari performa model BioBERT pada data uji setelah dilakukan proses *fine-tuning* dengan menggunakan data anotasi yang dihasilkan melalui kamus terminologi medis RadLex sebagai *pre-annotator*. Proses evaluasi dilakukan dengan menggunakan empat metrik utama, yaitu *loss*, *precision*, *recall*, dan *f1-score*, secara keseluruhan memberikan seluruh gambaran mengenai kemampuan model dalam melakukan ekstraksi entitas medis secara akurat. Berdasarkan hasil yang ditampilkan pada grafik diatas, nilai *evaluation loss* sebesar 0.060 menunjukkan bahwa tingkat kesalahan dalam prediksi model berada pada level sangat rendah. Hal ini mengindikasikan bahwa proses *training* model telah berjalan secara stabil, sehingga model mampu mempelajari pola dan representasi label semantik dari data medis dengan baik tanpa adanya indikasi *overfitting*. Nilai *precision* yang sangat tinggi yaitu 0.916 membuktikan bahwa model telah memiliki kemampuan yang unggul dalam menghasilkan prediksi benar dengan tingkat kesalahan minimal pada deteksi entitas medis. Sementara itu, hasil dari *recall* mencapai 0.970 menunjukkan bahwa model hampir sepenuhnya mampu mengenali seluruh

entitas medis yang relevan dalam data *training*. Kombinasi dari kedua metrik ini menghasilkan nilai *f1-score* sebesar 0.942, yang menggambarkan keseimbangan optimal antara ketepatan dan kelengkapan dalam proses ekstraksi entitas.

Performa yang sangat tinggi ini membuktikan bahwa penggunaan kamus RadLex sebagai *pre-annotator* dapat memberikan kontribusi yang signifikan terhadap peningkatan kualitas data *training*. RadLex yang merupakan kamus terminologi medis terstandarisasi dan kaya akan representasi konseptual, berfungsi sebagai label *guidance* membantu proses anotasi awal menjadi lebih efisien dan menjaga konsistensi dalam pelabelan. Dengan adanya bantuan terminologi yang sudah tervalidasi secara medis, model memperoleh *contextual understanding* yang lebih kuat terhadap istilah dan struktur semantik dalam domain medis radiologi. Hal ini memungkinkan model tidak hanya mengenali secara leksikal, tetapi juga memahami konteks penggunaannya dalam kalimat medis yang kompleks.

4.4. Analisis Hasil dan Perbandingan

Tabel 7. Perbandingan Anotasi

Anotasi	Eval Loss	Precision	Recall	F1-Score
Manual	0.818	0.628	0.671	0.649
Pre-Annotator (Radlex)	0.060	0.916	0.970	0.942

Pada tabel 7 di atas terlihat bahwa model BioBERT yang telah dilatih menggunakan anotasi hasil *pre-annotator* RadLex menunjukkan adanya peningkatan performa yang signifikan dibandingkan dengan model yang menggunakan anotasi manual. Berdasarkan hasil yang didapatkan nilai *loss* menurun tajam dari 0.818 menjadi 0.060, yang berarti tingkat kesalahan pada prediksi model berkurang signifikan. Selain itu, nilai *precision* meningkat dari angka 0.628 menjadi 0.916, hal ini menunjukkan bahwa model yang telah dilatih dengan data *pre-annotator* RadLex terbukti lebih mampu menghindari kesalahan dalam mengenali entitas positif yang tidak sesuai (*false positive*). Nilai *recall* mengalami peningkatan dari 0.671 menjadi 0.970, menandakan kemampuan model dalam mengenali hampir seluruh entitas medis yang relevan dalam teks. Peningkatan ini diperkuat oleh nilai *f1-score* yang naik dari 0.649 menjadi 0.942, menunjukkan keseimbangan yang lebih baik diantara *precision* dan *recall*.

Perbedaan hasil tersebut dapat dijelaskan oleh kualitas dan konsistensi label yang digunakan selama proses *training*. Pada anotasi manual, walaupun mampu memahami konteks secara lebih mendalam, hasilnya akan sangat bergantung terhadap pemahaman dan ketelitian anotator. Hal ini dapat menimbulkan perbedaan pemahaman antar anotator serta ketidaksesuaian pelabelan, terutama pada istilah medis yang memiliki makna ganda atau jarang muncul. Sementara itu, anotasi otomatis menggunakan *pre-annotator* RadLex dengan memanfaatkan kamus terminologi medis standar yang mencakup ribuan istilah anatomi, prosedur, penyakit, dan temuan klinis. Pendekatan ini menghasilkan label konsisten dan terstruktur, sehingga

mengurangi adanya kemungkinan ketidaksesuaian anotasi (*label noise*) serta membantu model dalam memahami istilah medis secara lebih umum.

5. KESIMPULAN

Penelitian ini berhasil mengusulkan dan menerapkan pendekatan *pra-anotasi* berbasis RadLex dalam pengembangan model Named Entity Recognition (NER) pada laporan radiologi klinis berbahasa Indonesia, khususnya pada organ hati (liver). Melalui tahapan sistematis yang mencakup *preprocessing*, *pre-annotation*, pembentukan *silver corpus*, validasi anotasi, hingga *fine-tuning* model BioBERT, penelitian ini menunjukkan bahwa penggunaan RadLex sebagai kamus terminologi medis mampu meningkatkan efisiensi dan konsistensi proses anotasi. Hasil eksperimen menunjukkan bahwa model BioBERT yang dilatih menggunakan data hasil pra-anotasi RadLex (*f1-score* = 0.942) jauh melampaui performa model yang dilatih menggunakan anotasi manual (*f1-score* = 0.649). Hal ini membuktikan bahwa pendekatan pra-anotasi mampu mengurangi kesalahan anotasi meningkatkan akurasi model, serta mempercepat dalam proses pengolahan data medis.

Untuk penelitian selanjutnya, direncanakan dilakukan evaluasi pengukuran nilai IAA (*inter-annotator agreement*) dengan melibatkan minimal 2 anotator. Tujuannya adalah untuk menilai dan mengevaluasi apakah label pada anotasi sebelumnya sesuai. Hasil nilai IAA dapat digunakan sebagai bahan untuk menempatkan label dalam gold standar serta mengukur sejauh

mana sistem pra-anotasi dapat membantu mengurangi perbedaan persepsi antar anotator dalam memahami konteks klinis pada laporan radiologi.

DAFTAR PUSTAKA

- [1] D. Fraile Navarro *et al.*, “Clinical named entity recognition and relation extraction using natural language processing of medical free text: A systematic review,” *International Journal of Medical Informatics*, vol. 177. Elsevier Ireland Ltd, Sep. 01, 2023. doi: [10.1016/j.ijmedinf.2023.105122](https://doi.org/10.1016/j.ijmedinf.2023.105122).
- [2] Z. D. Eka Putra and D. W. Utomo, “Penerapan Deep Learning dengan Mekanisme Attention untuk Meningkatkan Performa Segmentasi Liver dan Tumor pada Citra CT Menggunakan ResUnet,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 10, no. 3, pp. 231–239, Jan. 2025, doi: [10.25077/teknosi.v10i3.2024.231-239](https://doi.org/10.25077/teknosi.v10i3.2024.231-239).
- [3] G. H. B. Andrade, S. Yada, and E. Aramaki, “Is Boundary Annotation Necessary? Evaluating Boundary-Free Approaches to Improve Clinical Named Entity Annotation Efficiency: Case Study,” *JMIR Medical Informatics*, vol. 12, 2024, doi: [10.2196/59680](https://doi.org/10.2196/59680).
- [4] S. Silvestri, F. Gargiulo, and M. Ciampi, “Iterative Annotation of Biomedical NER Corpora with Deep Neural Networks and Knowledge Bases,” *Applied Sciences (Switzerland)*, vol. 12, no. 12, Jun. 2022, doi: [10.3390/app12125775](https://doi.org/10.3390/app12125775).
- [5] A. Penelitian, D. Soyusiawaty, A. Fadlil, A. Dahlan, and R. Selatan, “Jurnal Nasional Teknologi dan Sistem Informasi Attribution-ShareAlike 4.0 International Some rights reserved Pengembangan Korpus Bahasa Minang pada Spell Error Corpus for Minang Language (SPEML)”, doi: [10.25077/TEKNOSI.v11i1.2025.17-26](https://doi.org/10.25077/TEKNOSI.v11i1.2025.17-26).
- [6] P. H. Brekke, T. Rama, I. Pilán, Ø. Nytrø, and L. Øvrelid, “Synthetic data for annotation and extraction of family history information from clinical text,” *Journal of Biomedical Semantics*, vol. 12, no. 1, Dec. 2021, doi: [10.1186/s13326-021-00244-2](https://doi.org/10.1186/s13326-021-00244-2).
- [7] W. Lau, K. Lybarger, M. L. Gunn, M. B. Chb, and M. Yetisgen, “Event-based clinical findings extraction from radiology reports with pre-trained language model.” [Online]. Available: <https://github.com/wilsonlau-uw/BERT-EE>
- [8] A. Sylolypavan, D. Sleeman, H. Wu, and M. Sim, “The impact of inconsistent human annotations on AI driven clinical decision making,” *npj Digital Medicine*, vol. 6, no. 1, Dec. 2023, doi: [10.1038/s41746-023-00773-3](https://doi.org/10.1038/s41746-023-00773-3)
- [9] P. Parjono and S. Kusumadewi, “Pemodelan Text Mining dalam Pengkodean Penyakit Pasien Berdasar Kode ICD 10,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 2, pp. 200–207, Sep. 2023, doi: [10.25077/teknosi.v9i2.2023.200-207](https://doi.org/10.25077/teknosi.v9i2.2023.200-207).
- [10] G. Bernier-Colborne and S. Vajjala, “Annotation Errors and NER: A Study with OntoNotes 5.0,” Jun. 2024, [Online]. Available: [http://arxiv.org/abs/2406.19172](https://arxiv.org/abs/2406.19172)
- [11] L. Liu *et al.*, “Web-Based Application Based on Human-in-the-Loop Deep Learning for Deidentifying Free-Text Data in Electronic Medical Records: Development and Usability Study,” *Interactive Journal of Medical Research*, vol. 12, p. e46322, Aug. 2023, doi: [10.2196/46322](https://doi.org/10.2196/46322).
- [12] S. Datta, J. Godfrey-Stovall, and K. Roberts, “RadLex Normalization in Radiology Reports.” [Online]. Available: <http://alt.qcri.org/semeval2014/task7/>
- [13] S. Tsuji, A. Wen, N. Takahashi, H. Zhang, K. Ogasawara, and G. Jiang, “Developing a RadLex-based named entity recognition tool for mining textual radiology reports: development and performance evaluation study,” *Journal of Medical Internet Research*, vol. 23, no. 10, Oct. 2021, doi: [10.2196/25378](https://doi.org/10.2196/25378).
- [14] A. Miranda-Escalada, A. Gonzalez-Agirre, J. Armengol-Estapé, and M. Krallinger, “Overview of automatic clinical coding: annotations, guidelines, and solutions for non-English clinical cases at CodiEsp track of CLEF eHealth 2020.” [Online] Available: <https://www.who.int/classifications/icd/icdonlineversions/en/>
- [15] M. Marquis, I. Bossenko, and P. Ross, “RadLex and SNOMED CT integration: a pilot study for standardising radiology classification,” *Insights into Imaging*, vol. 16, no. 1, Dec. 2025, doi: [10.1186/s13244-025-01935-5](https://doi.org/10.1186/s13244-025-01935-5).
- [16] D. Li *et al.*, “Performance and Agreement When Annotating Chest X-ray Text Reports—A Preliminary Step in the Development of a Deep Learning-Based Prioritization and Detection System,” *Diagnostics*, vol. 13, no. 6, Mar. 2023, doi: [10.3390/diagnostics13061070](https://doi.org/10.3390/diagnostics13061070).
- [17] J.-B. Delbrouck *et al.*, “RadGraph-XL: A Large-Scale Expert-Annotated Dataset for Entity and Relation Extraction from Radiology Reports.” [Online]. Available: <https://github.com/Stanford-AIMI/radgraph-XL>
- [18] Y. Zhu, Y. Ye, M. Li, J. Zhang, and O. Wu, “Investigating annotation noise for named entity recognition,” *Neural Computing and Applications*, vol. 35, no. 1, pp. 993–1007, 2023, doi: [10.1007/s00521-022-07733-0](https://doi.org/10.1007/s00521-022-07733-0).
- [19] M. Mikulová, M. Straka, J. Štěpánek, B. Štěpánková, and J. Hajič, “Quality and Efficiency of Manual Annotation: Pre-annotation Bias,” Jun. 2023, [Online]. Available: [http://arxiv.org/abs/2306.09307](https://arxiv.org/abs/2306.09307)
- [20] H. A. Xu, V. Loftsson, B. Kulynych, B. Kaabachi, and J. L. Raisaro, “Accelerating Clinical Text Annotation in Underrepresented Languages: A Case Study on Text De-Identification,” in *Studies in Health Technology and Informatics*, IOS Press BV, Aug. 2024, pp. 853–857. doi: [10.3233/SHTI240546](https://doi.org/10.3233/SHTI240546).

- [21] M. Syed et al., “DeIDNER corpus: Annotation of clinical discharge summary notes for named entity recognition using BRAT tool,” in *Public Health and Informatics: Proceedings of MIE 2021*, IOS Press, 2021, pp. 432–436. doi: [10.3233/SHTI210195](https://doi.org/10.3233/SHTI210195).
- [22] Z. Bao, Y. Wang, and W. Zhang, “Low-Resource Named Entity Recognition without Human Annotation,” *Academy and Industry Research Collaboration Center (AIRCC)*, Sep. 2021, pp. 327–335. doi: [10.5121/csit.2021.111427](https://doi.org/10.5121/csit.2021.111427).
- [23] Q. Zeng, M. Yu, W. Yu, T. Jiang, and M. Jiang, “Validating Label Consistency in NER Data Annotation,” Sep. 2021, [Online]. Available: <http://arxiv.org/abs/2101.08698>

BIODATA PENULIS



Arum Maulia

Penulis adalah mahasiswa Program Studi Teknik Informatika Universitas Dian Nuswantoro yang memiliki minat penelitian di bidang Natural Language Processing.



Danang Wahyu Utomo, M.Kom

Dosen Fakultas Ilmu Komputer Program Studi Teknik Informatika Universitas Dian Nuswantoro Semarang. Aktif dalam penelitian Natural Language Processing (NLP) untuk domain medis.