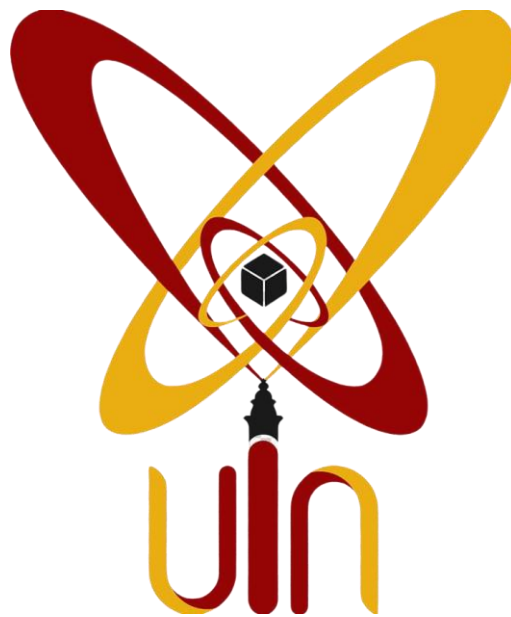


**ANALISIS KOMPARATIF SCIBERT, BIOBERT, DAN
PUBMEDBERT UNTUK NAMED ENTITY RECOGNATION
PADA LITERATUR RUMPUT LAUT**

Disusun untuk memenuhi salah satu syarat pelaksanaan Magang
pada Program Studi Informatika



Disusun oleh:

Nama : Nabila Nurasyifa

NIM : 231730020

**PROGRAM STUDI INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI
SULTAN MAULANA HASANUDDIN BANTEN
2026**

HALAMAN PENGESAHAN

Laporan hasil pelaksanaan PKL/Magang ini telah disusun oleh:

Nama : Nabila Nurasyifa
NIM : 231730020
Program Studi : Informatika
Judul Magang : Analisis Komparatif SciBERT, BioBERT, dan PubMedBERT untuk Named Entity Recognition pada Literatur Rumput Laut
Tempat Magang : Badan Riset dan Inovasi Nasional (BRIN)
Waktu Pelaksanaan : 18 Februari 2026 - 31 Juli 2026

Laporan ini telah diperiksa dan disahkan sebagai laporan hasil pelaksanaan Magang.

Bandung, 24 Juni 2026

Dosen Pembimbing Magang,

Pembimbing Lapangan,

Ahmad Tabrani, M.T.I

Dr. Dipl. Ing. (FH) Anne Parlina M.P.

NIP. 198509172018011002

NIP 197602141994032001

Mengetahui,

Ketua Program Studi Informatika

Ahmad Tabrani, M.T.I

NIP. 198509172018011002

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas rahmat dan karunia-Nya, penulis dapat menyelesaikan laporan hasil pelaksanaan Magang yang berjudul “Analisis Komparatif SciBERT, BioBERT, dan PubMedBERT untuk *Named Entity Recognition* pada Literatur Rumput Laut” dengan baik.

Laporan ini disusun sebagai salah satu bentuk pertanggungjawaban atas kegiatan Magang yang telah dilaksanakan di Badan Riset dan Inovasi Nasional (BRIN) Melalui kegiatan ini, penulis memperoleh pengalaman, pengetahuan, serta pemahaman secara langsung mengenai dunia kerja, khususnya dalam bidang *kecerdasan buatan (Artificial Intelligence)*, *pengolahan bahasa alami (Natural Language Processing)*, dan *data science*.

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan bantuan, bimbingan, dan dukungan selama pelaksanaan Magang hingga penyusunan laporan ini, terutama kepada:

1. Bapak Ahmad Tabrani, M.T.I
2. Ibu Dr. Dipl. Ing. (FH) Anne Parlina M.P.
4. Badan Riset Inovasi Nasional (BRIN)
5. Semua pihak yang telah membantu dalam pelaksanaan Magang.

Penulis menyadari bahwa laporan ini masih memiliki kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan agar laporan ini dapat menjadi lebih baik. Semoga laporan ini dapat memberikan manfaat bagi pembaca dan pihak-pihak yang membutuhkan.

Bandung, 24 Juni 2026

Nabila Nurasyifa

DAFTAR ISI

HALAMAN PENGESAHAN.....	i
KATA PENGANTAR	ii
DAFTAR ISI.....	iii
RINGKASAN	v
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	2
1.3 Tujuan Magang	3
1.3.1 Tujuan umum	3
1.4 Manfaat PKL/Magang.....	3
1.4.1 Manfaat teoritis	3
BAB II KAJIAN PUSTAKA	5
2.1 Kajian Pustaka.....	5
2.1.1 Pengertian NER.....	5
2.1.2 Konsep SciBERT, BioBERT, PubMedBERT	5
2.1.3 Penelitian atau referensi terkait.....	6
2.2 Kerangka Pikir Teoritis	7
BAB III METODE PKL/MAGANG	8
3.1 Lokasi Magang.....	8
3.2 Desain Pelaksanaan Magang.....	8
3.3 Penentuan Objek Magang	9
3.4 Metode Pengumpulan Data	9
3.5 Instrumen PKL/Magang.....	10

3.6 Teknik Analisis Data.....	10
BAB IV HASIL PKL/MAGANG DAN PEMBAHASAN.....	11
4.1 Deskripsi Kegiatan PKL/Magang	11
4.2 Permasalahan yang Ditemukan	11
4.3 Metode Penyelesaian Masalah	12
4.4 Hasil Pelaksanaan PKL/Magang.....	12
4.5 Pembahasan.....	13
4.6 Pelajaran yang Diperoleh	14
BAB V KESIMPULAN DAN SARAN.....	15
5.1 Kesimpulan	15
5.2 Keterbatasan.....	15
5.3 Implikasi.....	15
5.4 Saran.....	15
DAFTAR PUSTAKA	17
LAMPIRAN.....	18

RINGKASAN

Selama pelaksanaan Magang di Badan Riset dan Inovasi Nasional (BRIN), permasalahan utama yang ditemukan adalah proses anotasi dataset yang masih dilakukan secara manual serta pelatihan model *Named Entity Recognition* (NER) yang memerlukan beberapa kali percobaan untuk memperoleh performa terbaik. Kondisi tersebut menyebabkan proses penelitian menjadi lebih lama, padahal diharapkan tersedia dataset yang berkualitas dan proses pengembangan model yang lebih efisien. Permasalahan ini penting karena kualitas data dan model sangat memengaruhi keberhasilan ekstraksi informasi dari literatur ilmiah.

Magang dilaksanakan di BRIN, Organisasi Riset Elektronika dan Informatika (OREI), Pusat Riset Sains Data dan Informasi (PRSDI), dengan objek pengembangan model NER menggunakan SciBERT, BioBERT, dan PubMedBERT. Metode yang digunakan meliputi observasi, wawancara, dokumentasi, dan praktik langsung. Instrumen yang digunakan antara lain Python, Google Colab, Label Studio, Kaggle, dan GitHub, sedangkan data dianalisis secara deskriptif.

Hasil Magang menunjukkan bahwa seluruh tahapan penelitian, mulai dari studi literatur, anotasi dataset, pelatihan, hingga evaluasi model, berhasil dilaksanakan dengan baik. Kegiatan ini menghasilkan model NER berbasis Transformer, laporan hasil magang, dan paper penelitian, serta meningkatkan kemampuan teknis di bidang *Natural Language Processing* (NLP), pemecahan masalah, komunikasi, dan kerja sama tim. Pengalaman tersebut menjadi bekal yang bermanfaat bagi pengembangan kompetensi akademik maupun profesional.

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Pelaksanaan Magang merupakan salah satu kegiatan akademik yang bertujuan untuk memberikan pengalaman langsung kepada mahasiswa dalam dunia kerja. Melalui kegiatan ini, mahasiswa dapat menerapkan ilmu yang telah diperoleh selama perkuliahan serta memahami kondisi nyata di lingkungan kerja. Selain itu, kegiatan Magang menjadi sarana bagi mahasiswa untuk mengembangkan kompetensi teknis maupun nonteknis sehingga memiliki kesiapan yang lebih baik dalam menghadapi dunia profesional.

Perkembangan teknologi informasi, khususnya di bidang kecerdasan buatan (*Artificial Intelligence*), analisis data, dan *Natural Language Processing* (NLP), telah mendorong berbagai institusi untuk memanfaatkan teknologi dalam mendukung proses penelitian maupun pengambilan keputusan. Oleh karena itu, mahasiswa dituntut tidak hanya menguasai konsep teoritis, tetapi juga mampu mengimplementasikan teknologi tersebut untuk menyelesaikan permasalahan nyata sesuai dengan kebutuhan di lingkungan kerja.

Selama pelaksanaan PKL/Magang, berbagai kegiatan yang berkaitan dengan penelitian dan pengembangan sistem berbasis *Natural Language Processing* (NLP) dilaksanakan, khususnya penerapan metode *Named Entity Recognition* (NER) untuk melakukan ekstraksi informasi dari literatur ilmiah. Kegiatan yang dilakukan meliputi studi literatur, pengolahan dan anotasi data, proses *preprocessing*, pembangunan model, hingga evaluasi hasil menggunakan berbagai metrik performa. Keterlibatan secara langsung dalam rangkaian kegiatan tersebut memberikan pengalaman mengenai bagaimana suatu penelitian atau pengembangan sistem dilakukan secara sistematis dan sesuai dengan kaidah ilmiah.

Kondisi ideal yang diharapkan dalam pelaksanaan Magang adalah terciptanya keselarasan antara kompetensi yang diperoleh selama perkuliahan dengan kebutuhan dunia kerja. Mahasiswa diharapkan mampu menerapkan

pengetahuan yang dimiliki untuk menyelesaikan permasalahan nyata, beradaptasi dengan lingkungan profesional, serta menghasilkan kontribusi yang bermanfaat bagi instansi tempat magang. Namun demikian, pada praktiknya masih terdapat kesenjangan antara pembelajaran akademik yang bersifat teoritis dengan penerapan teknologi pada kasus nyata, sehingga diperlukan pengalaman langsung melalui kegiatan Magang.

Berdasarkan kondisi tersebut, pelaksanaan Magang menjadi penting karena memberikan kesempatan kepada mahasiswa untuk memahami proses kerja secara nyata, meningkatkan kemampuan berpikir kritis, pemecahan masalah, komunikasi, kerja sama tim, serta penguasaan teknologi yang relevan dengan bidang keilmuan. Pengalaman tersebut tidak hanya mendukung pencapaian kompetensi akademik, tetapi juga menjadi bekal dalam menghadapi tantangan di dunia kerja maupun penelitian.

Oleh karena itu, laporan Magang ini disusun sebagai bentuk pertanggungjawaban atas seluruh kegiatan yang telah dilaksanakan selama masa magang. Pembahasan dalam laporan ini diharapkan dapat memberikan gambaran mengenai penerapan ilmu pengetahuan di lingkungan kerja sekaligus menjadi dokumentasi atas pelaksanaan kegiatan yang telah dilakukan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam laporan Magang ini adalah sebagai berikut.

1. Bagaimana pelaksanaan kegiatan Magang di Badan Riset dan Inovasi Nasional?
2. Apa saja permasalahan yang ditemukan selama pelaksanaan Magang?
3. Bagaimana upaya penyelesaian terhadap permasalahan yang ditemukan selama Magang?

1.3 Tujuan Magang

1.3.1 Tujuan umum

Pelaksanaan Magang bertujuan untuk memberikan pengalaman kerja secara langsung kepada mahasiswa sehingga mampu menerapkan pengetahuan dan keterampilan yang diperoleh selama perkuliahan, mengembangkan kompetensi sesuai bidang keilmuan, serta memahami kondisi dan budaya kerja di lingkungan instansi/perusahaan.

1.3.2 Tujuan khusus

1. Mengetahui proses kerja dan pelaksanaan kegiatan di Badan Riset dan Inovasi Nasional
2. Mengidentifikasi permasalahan atau kendala yang ditemukan selama pelaksanaan Magang.
3. Menganalisis upaya penyelesaian terhadap permasalahan atau kendala yang ditemukan selama pelaksanaan Magang.
4. Menyusun laporan hasil pelaksanaan Magang secara sistematis sesuai dengan ketentuan.

1.4 Manfaat PKL/Magang

1.4.1 Manfaat teoritis

Pelaksanaan Magang diharapkan dapat memberikan kontribusi terhadap pengembangan ilmu pengetahuan, khususnya yang berkaitan dengan penerapan teori sesuai bidang keilmuan dalam lingkungan kerja. Selain itu, laporan ini diharapkan dapat menjadi referensi akademik bagi mahasiswa maupun pihak lain yang ingin mempelajari proses pelaksanaan Magang serta penerapan ilmu pengetahuan.

1.4.2 Manfaat praktis

Kegiatan Magang memberikan manfaat bagi berbagai pihak. Bagi mahasiswa, kegiatan ini menjadi sarana untuk menambah pengalaman dan

wawasan mengenai dunia kerja, menerapkan ilmu yang telah dipelajari selama perkuliahan, serta mengembangkan kemampuan berpikir kritis, komunikasi, kerja sama tim, dan penyelesaian masalah. Bagi instansi atau perusahaan, kegiatan ini dapat mempererat kerja sama dengan perguruan tinggi, memberikan kontribusi dalam pelaksanaan tugas dan kegiatan, serta menjadi bahan masukan untuk pengembangan program magang di masa mendatang. Sementara itu, bagi perguruan tinggi, kegiatan ini bermanfaat dalam memperkuat kemitraan dengan instansi atau perusahaan, menjadi bahan evaluasi untuk pengembangan kurikulum agar sesuai dengan kebutuhan dunia kerja.

BAB II

KAJIAN PUSTAKA

2.1 Kajian Pustaka

Kajian pustaka merupakan landasan teoritis yang digunakan untuk memahami konsep, metode, dan penelitian terdahulu yang berkaitan dengan pengembangan sistem *Named Entity Recognition* (NER). Pada penelitian ini, teori yang digunakan meliputi konsep *Named Entity Recognition*, *Natural Language Processing* (NLP), serta model *domain-spesifik* yaitu SciBERT, BioBERT, dan PubMedBERT.

2.1.1 Pengertian NER

Named Entity Recognition (NER) merupakan salah satu tugas utama dalam bidang *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi serta mengklasifikasikan entitas yang terdapat di dalam suatu teks ke dalam kategori tertentu, seperti nama orang, lokasi, organisasi, senyawa kimia, maupun kategori lain sesuai kebutuhan penelitian. Melalui proses ini, informasi penting yang sebelumnya berada dalam bentuk teks tidak terstruktur dapat diubah menjadi data terstruktur sehingga lebih mudah dianalisis.

Dalam penelitian ini, NER digunakan untuk mengenali empat jenis entitas pada abstrak ilmiah mengenai rumput laut, yaitu *Seaweed*, *Location*, *Compound*, dan *Benefit*. Pendekatan tersebut memungkinkan proses ekstraksi informasi dilakukan secara otomatis sehingga mempercepat analisis literatur ilmiah yang jumlahnya terus meningkat.

2.1.2 Konsep SciBERT, BioBERT, PubMedBERT

SciBERT merupakan pengembangan BERT yang dilatih menggunakan jutaan artikel ilmiah dari berbagai bidang, khususnya ilmu komputer dan biomedis. Model ini dirancang agar mampu memahami istilah ilmiah secara lebih baik dibandingkan BERT umum sehingga sesuai digunakan pada dokumen penelitian.

BioBERT merupakan model BERT yang memperoleh pelatihan lanjutan menggunakan abstrak PubMed dan artikel penuh PubMed Central. Model ini dikembangkan untuk meningkatkan performa berbagai tugas *Natural Language Processing* (NLP) pada domain biomedis, termasuk *Named Entity Recognition*.

PubMedBERT merupakan model *Transformer* yang dilatih sejak awal (*from scratch*) menggunakan *corpus* PubMed. Berbeda dengan BioBERT yang melakukan *pre-training* lanjutan, PubMedBERT membangun kosakata khusus domain biomedis sejak awal sehingga mampu menghasilkan representasi bahasa yang lebih sesuai untuk tugas ekstraksi informasi pada literatur ilmiah.

2.1.3 Penelitian atau referensi terkait

Beberapa penelitian menunjukkan bahwa model berbasis *Transformer* mampu meningkatkan kinerja *Named Entity Recognition* (NER), terutama pada dokumen ilmiah dan biomedis. SciBERT dikembangkan menggunakan korpus artikel ilmiah dan terbukti memberikan performa yang lebih baik pada berbagai tugas *Natural Language Processing* (NLP), termasuk NER, dibandingkan model BERT umum [1].

Pada domain biomedis, BioBERT meningkatkan kemampuan model melalui proses *pre-training* menggunakan data PubMed dan PubMed Central sehingga menghasilkan akurasi yang lebih tinggi pada tugas NER dan NLP biomedis lainnya [2]. Selanjutnya, PubMedBERT yang dilatih sepenuhnya menggunakan korpus PubMed menunjukkan performa yang lebih baik dibandingkan model BERT umum maupun model hasil *domain adaptation* pada berbagai tugas NLP biomedis [3].

Selain itu, penelitian mengenai arsitektur NER berbasis *deep learning* menunjukkan bahwa model BiLSTM-CRF mampu mengenali entitas secara efektif tanpa memerlukan rekayasa fitur yang kompleks [4]. Berdasarkan penelitian tersebut, penelitian ini membandingkan performa SciBERT, BioBERT, dan PubMedBERT dalam mengenali entitas pada abstrak ilmiah

mengenai rumput laut untuk menentukan model yang memberikan hasil terbaik.

2.2 Kerangka Pikir Teoritis

Penelitian diawali dari permasalahan banyaknya publikasi ilmiah mengenai rumput laut yang menyebabkan proses ekstraksi informasi secara manual menjadi tidak efisien. Untuk mengatasi permasalahan tersebut digunakan pendekatan *Natural Language Processing* melalui metode *Named Entity Recognition*.

Dataset berupa abstrak ilmiah diperoleh dari Scopus kemudian dilakukan anotasi menggunakan Label Studio berdasarkan empat kategori entitas, yaitu *Seaweed*, *Location*, *Compound*, dan *Benefit*. Selanjutnya data dikonversi ke format *BIO*, diproses melalui tahap tokenisasi dan *preprocessing*, kemudian dibagi menjadi data latih, validasi, dan pengujian.

Tahap berikutnya adalah proses *fine-tuning* menggunakan tiga model *Transformer*, yaitu SciBERT, BioBERT, dan PubMedBERT. Seluruh model dievaluasi menggunakan metrik *precision*, *recall*, *F1-score*, dan *accuracy* untuk mengetahui model yang memiliki performa terbaik. Hasil evaluasi kemudian digunakan sebagai dasar analisis dalam menentukan model yang paling sesuai untuk tugas *Named Entity Recognition* pada literatur ilmiah rumput laut.

BAB III

METODE PKL/MAGANG

3.1 Lokasi Magang

Kegiatan Magang dilaksanakan di Badan Riset dan Inovasi Nasional (BRIN), tepatnya pada Organisasi Riset Elektronika dan Informatika (OREI), Pusat Riset Sains Data dan Informasi (PRSDI). Pusat riset ini memiliki bidang penelitian yang berfokus pada sains data, kecerdasan buatan (*Artificial Intelligence*), pembelajaran mesin (*Machine Learning*), pemrosesan bahasa alami (*Natural Language Processing*), serta temu kembali informasi (*Information Retrieval*).

Alamat instansi berada di Jl. Tamansari No. 71, Lebak Siliwangi, Kecamatan Coblong, Kota Bandung, Jawa Barat 40132, Indonesia. Pelaksanaan Magang dilakukan pada Kelompok Riset Temu Kembali Informasi (*Information Retrieval Research Group*) di bawah Pusat Riset Sains Data dan Informasi (PRSDI) kemudian dilaksanakan selama kurang lebih lima bulan, yaitu mulai 18 Februari 2026 hingga 31 Juli 2026.

3.2 Desain Pelaksanaan Magang

Pelaksanaan Magang menggunakan pendekatan deskriptif dengan metode observasi dan praktik langsung. Pendekatan deskriptif digunakan untuk menggambarkan proses pelaksanaan kegiatan serta aktivitas yang dilakukan pada Kelompok Riset Temu Kembali Informasi (*Information Retrieval Research Group*), Pusat Riset Sains Data dan Informasi (PRSDI), BRIN.

Observasi dilakukan untuk memahami alur kerja, penggunaan teknologi, serta proses penelitian yang diterapkan di lingkungan kerja. Selain itu, praktik langsung dilakukan melalui keterlibatan dalam kegiatan penelitian, mulai dari pengolahan data, implementasi model *Named Entity Recognition* (NER) berbasis *Transformer*, hingga evaluasi hasil model sesuai dengan arahan pembimbing.

3.3 Penentuan Objek Magang

Objek dalam pelaksanaan Magang ini adalah pengembangan dan evaluasi sistem *Named Entity Recognition* (NER) untuk mengekstraksi informasi dari abstrak ilmiah mengenai rumput laut. Sistem tersebut memanfaatkan model berbasis *Transformer*, yaitu SciBERT, BioBERT, dan PubMedBERT, untuk mengenali empat jenis entitas, yaitu *Seaweed*, *Location*, *Compound*, dan *Benefit*.

Kegiatan yang dilakukan meliputi pengumpulan dan pengolahan data, pelabelan (*annotation*) entitas, pelatihan model, serta evaluasi performa masing-masing model menggunakan metrik evaluasi yang sesuai. Objek tersebut dipilih karena mendukung proses ekstraksi informasi secara otomatis dan memungkinkan perbandingan kinerja ketiga model dalam mengenali entitas pada literatur ilmiah.

3.4 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan selama pelaksanaan bertujuan untuk memperoleh informasi yang relevan mengenai proses penelitian, alur kerja, serta implementasi model *Named Entity Recognition* (NER) pada data abstrak ilmiah rumput laut. Pemilihan metode ini dilakukan agar data yang diperoleh bersifat faktual, lengkap, dan dapat mendukung pelaksanaan kegiatan Magang. Adapun metode yang digunakan meliputi:

- a. Observasi dilakukan dengan mengamati secara langsung proses pelaksanaan kegiatan penelitian, mulai dari pengolahan data hingga evaluasi model. Metode ini dipilih untuk memahami alur kerja selama Magang.
- b. Wawancara dilakukan melalui diskusi dengan pembimbing lapangan dan anggota tim riset untuk memperoleh informasi mengenai tahapan penelitian serta penggunaan model *Transformer-based* NER. Metode ini dipilih untuk melengkapi informasi yang diperoleh dari hasil observasi.
- c. Dokumentasi dilakukan dengan mengumpulkan dokumen, arsip, catatan, serta hasil kegiatan yang berkaitan dengan pelaksanaan. Metode ini dipilih sebagai bukti pelaksanaan kegiatan dan pendukung penyusunan laporan.

3.5 Instrumen PKL/Magang

Instrumen yang digunakan selama kegiatan meliputi laptop sebagai perangkat utama, buku catatan untuk mencatat hasil diskusi, serta beberapa perangkat lunak pendukung, yaitu *Python* sebagai bahasa pemrograman, *Google Colab* sebagai lingkungan pengembangan, *Kaggle* sebagai sumber *dataset*, *GitHub* sebagai media referensi implementasi, dan *Label Studio* sebagai alat anotasi data dalam proses pengembangan model *Named Entity Recognition* (NER).

3.6 Teknik Analisis Data

Teknik analisis data yang digunakan dalam adalah analisis deskriptif. Data yang diperoleh melalui observasi, wawancara, dan dokumentasi disusun secara sistematis, kemudian dianalisis dan diinterpretasikan untuk memperoleh pemahaman mengenai proses pelaksanaan penelitian serta pengembangan model *Named Entity Recognition* (NER). Hasil analisis tersebut digunakan sebagai dasar dalam penyusunan laporan Magang.

BAB IV

HASIL PKL/MAGANG DAN PEMBAHASAN

4.1 Deskripsi Kegiatan PKL/Magang

Tabel 1. Kegiatan PKL/Magang

No	Tanggal	Kegiatan	Keterangan
1	Februari – Maret 2026	<i>Research & System Design</i>	Melakukan studi literatur terkait <i>Named Entity Recognition</i> (NER), menentukan skema entitas, serta menyusun rancangan sistem yang akan dikembangkan.
2	Maret – April 2026	<i>Dataset & Annotation</i>	Melakukan pembersihan (<i>cleaning</i>) dataset dan proses anotasi data menggunakan Label Studio sesuai skema entitas yang telah ditentukan.
3	April – Juni 2026	<i>NER Model Development</i>	Melakukan pelatihan (<i>training</i>), pengujian, dan evaluasi model NER menggunakan pendekatan <i>Transformer-based</i> .
4	Juni – Juli 2026	Dokumentasi	Menyusun laporan PKL, menyelesaikan penulisan <i>paper</i> penelitian (<i>paper finalization</i>), serta mendokumentasikan perangkat lunak yang dikembangkan.

Sumber: Data diolah oleh penulis, 2026

4.2 Permasalahan yang Ditemukan

Selama pelaksanaan, terdapat beberapa permasalahan yang ditemukan dalam proses pengembangan model *Named Entity Recognition* (NER). Salah satu kendala utama adalah *dataset* yang masih dilakukan anotasi secara manual agar sesuai dengan skema entitas yang telah ditentukan. Selain itu, proses pelatihan model memerlukan beberapa kali percobaan untuk memperoleh konfigurasi parameter yang menghasilkan performa terbaik.

Permasalahan tersebut menyebabkan proses pengembangan model membutuhkan waktu yang lebih lama karena diperlukan validasi data, perbaikan hasil anotasi, serta evaluasi model secara berulang untuk meningkatkan akurasi.

4.3 Metode Penyelesaian Masalah

Permasalahan yang ditemukan diselesaikan melalui beberapa langkah, yaitu melakukan observasi terhadap alur penelitian, berdiskusi secara berkala dengan pembimbing mengenai kualitas data dan hasil pengembangan model, serta melakukan proses anotasi menggunakan *Label Studio*. Selain itu, dilakukan beberapa kali pelatihan dan evaluasi model dengan variasi parameter menggunakan *Python* dan *Google Colab* hingga diperoleh model dengan kinerja yang optimal.

4.4 Hasil Pelaksanaan PKL/Magang

Hasil yang diperoleh selama pelaksanaan Magang meliputi pengembangan model *Named Entity Recognition* (NER) berbasis *Transformer* yang telah melalui tahapan persiapan *dataset*, anotasi data, pelatihan, dan evaluasi model. Selain itu, dihasilkan dokumentasi, laporan hasil Magang, serta *paper* penelitian sebagai luaran dari kegiatan yang dilaksanakan.



Gambar 1. Dokumentasi Kegiatan Magang
Sumber: Dokumentasi Penulis, 2026

4.5 Pembahasan

Hasil pelaksanaan Magang menunjukkan bahwa pengembangan model *Named Entity Recognition* (NER) berbasis *Transformer* dapat diterapkan untuk mengekstraksi informasi dari abstrak ilmiah mengenai rumput laut. Tahapan yang dilakukan, mulai dari studi literatur, persiapan *dataset*, anotasi data, pelatihan, hingga evaluasi model, telah sesuai dengan konsep pengembangan sistem NER yang dijelaskan pada BAB II.

Penggunaan model SciBERT, BioBERT, dan PubMedBERT juga sesuai dengan teori mengenai model *Transformer* yang telah dipra-latih (*pre-trained*) pada domain ilmiah dan biomedis sehingga mampu memahami konteks teks

secara lebih baik dibandingkan model umum. Selain itu, proses anotasi data menggunakan *Label Studio* mendukung pembentukan *dataset* yang berkualitas, yang merupakan salah satu faktor penting dalam meningkatkan kinerja model NER.

Secara keseluruhan, hasil yang diperoleh selama Magang telah sesuai dengan teori yang dibahas pada BAB II. Meskipun demikian, kualitas *dataset* dan proses anotasi tetap menjadi faktor yang memengaruhi performa model, sehingga diperlukan evaluasi secara berulang untuk memperoleh hasil yang optimal.

4.6 Pelajaran yang Diperoleh

Pelaksanaan Magang yang dilaksanakan di BRIN memberikan pengetahuan dan pengalaman mengenai proses penelitian di bidang *Natural Language Processing* (NLP), khususnya pengembangan model *Named Entity Recognition* (NER). Selain itu, diperoleh keterampilan dalam pengolahan data, anotasi menggunakan *Label Studio*, pengembangan model menggunakan *Python* dan *Google Colab*, serta pemanfaatan *GitHub* sebagai media kolaborasi dan referensi implementasi.

Dari sisi sikap kerja, kegiatan Magang ini meningkatkan kedisiplinan, tanggung jawab, kemampuan berkomunikasi, serta kerja sama dalam tim riset. Pengalaman berdiskusi dengan pembimbing dan anggota tim juga memberikan wawasan mengenai penyelesaian masalah secara sistematis serta pentingnya kolaborasi dalam pelaksanaan penelitian.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Pelaksanaan Magang ini memberikan pengalaman dalam pengembangan model *Named Entity Recognition* (NER) berbasis *Transformer* untuk ekstraksi informasi pada abstrak ilmiah rumput laut. Seluruh tahapan penelitian, mulai dari studi literatur, persiapan *dataset*, anotasi data, pelatihan, hingga evaluasi model, telah dilaksanakan sesuai dengan tujuan Magang. Selain menghasilkan model NER, kegiatan ini juga meningkatkan pemahaman mengenai penerapan *Natural Language Processing* (NLP) dalam bidang penelitian.

5.2 Keterbatasan

Selama pelaksanaan Magang, terdapat beberapa keterbatasan, seperti waktu pelaksanaan yang terbatas, proses anotasi data yang membutuhkan ketelitian dan waktu yang cukup lama, serta perlunya beberapa kali pelatihan model untuk memperoleh performa yang optimal. Selain itu, kualitas dan jumlah data juga memengaruhi hasil pengembangan model.

5.3 Implikasi

Hasil Magang ini memberikan manfaat dalam penerapan teknologi *Natural Language Processing* (NLP), khususnya *Named Entity Recognition* (NER), untuk membantu proses ekstraksi informasi dari literatur ilmiah secara otomatis. Selain itu, kegiatan ini juga memberikan pengalaman praktis yang dapat mendukung pengembangan kompetensi akademik dan profesional di bidang sains data dan kecerdasan buatan.

5.4 Saran

Bagi instansi, diharapkan dapat terus memberikan bimbingan dan kesempatan bagi mahasiswa untuk terlibat dalam kegiatan penelitian. Selain itu, penyampaian arahan, informasi, maupun masukan secara lebih terstruktur dan berkesinambungan diharapkan dapat semakin mendukung kelancaran proses

Magang serta membantu mahasiswa dalam memahami setiap tahapan kegiatan dengan lebih optimal.

Bagi pihak kampus, diharapkan dapat terus memperkuat kerja sama dengan instansi penelitian serta meningkatkan koordinasi dan komunikasi terkait pelaksanaan Magang. Penyampaian informasi yang lebih terstruktur dan tepat waktu diharapkan dapat meminimalkan perubahan jadwal atau informasi yang bersifat mendadak, sehingga proses persiapan dan pelaksanaan Magang dapat berjalan lebih efektif dan optimal.

DAFTAR PUSTAKA

- [1] Beltagy, K. Lo, and A. Cohan, "SciBERT: A Pretrained Language Model for Scientific Text," in Proc. EMNLP-IJCNLP, 2019, pp. 3615–3620.
- [2] J. Lee et al., "BioBERT: A Pre-trained Biomedical Language Representation Model for Biomedical Text Mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, 2020.
- [3] Y. Gu et al., "Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing," *ACM Trans. Comput. Healthcare*, vol. 3, no. 1, pp. 1–23, 2021.
- [4] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural Architectures for Named Entity Recognition," in Proc. NAACL-HLT, 2016, pp. 260–270.

LAMPIRAN





