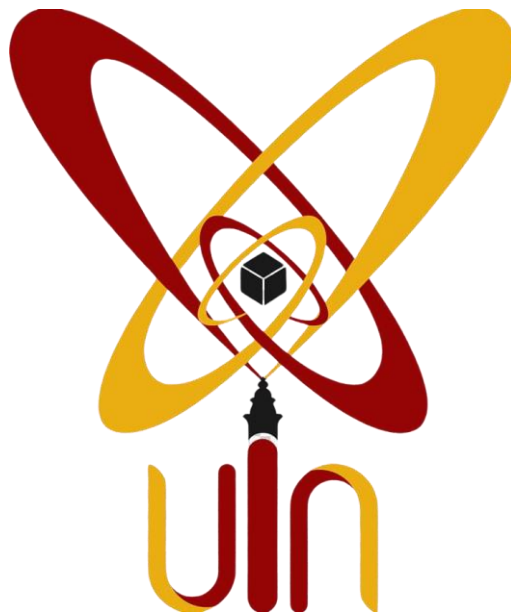


**LAPORAN HASIL PELAKSANAAN PKL/MAGANG**

**PENGEMBANGAN MODEL RELATION EXTRACTION**  
**BERBASIS TRANSFORMER DAN SISTEM VISUALISASI**  
**INFORMASI LITERATUR RUMPUT LAUT BERBASIS WEB**

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



Disusun oleh:  
Nama : Mursalin  
NIM : 231730006

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

## HALAMAN PENGESAHAN

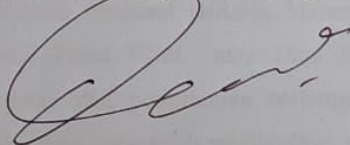
Laporan hasil pelaksanaan PKL/Magang ini telah disusun oleh:

|                   |                                                                                                                                         |
|-------------------|-----------------------------------------------------------------------------------------------------------------------------------------|
| Nama              | : Mursalin                                                                                                                              |
| NIM               | : 231730006                                                                                                                             |
| Program Studi     | : Informatika                                                                                                                           |
| Judul PKL/Magang  | : Pengembangan Model Relation Extraction Berbasis<br>Transformer dan Sistem Visualisasi Informasi Literatur<br>Rumput Laut Berbasis Web |
| Tempat PKL/Magang | : Badan Riset Inovasi Nasional (BRIN) Cisitub Bandung                                                                                   |
| Waktu Pelaksanaan | : 17 Februari 2026- 17 Juli 2026                                                                                                        |

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

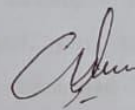
Bandung, 20 Juni 2026

Dosen Pembimbing PKL/Magang,



Ahmad Tabrani, M.T.I  
NIP. 198509172018011002

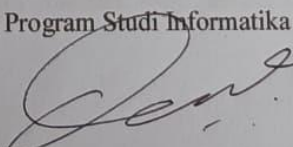
Pembimbing Lapangan,



Dr. Dipl. Ing. (FH) Anne Parlina, M.P.  
NIP. 197602141994032001

Mengetahui,

Ketua Program Studi Informatika



Ahmad Tabrani, M.T.I  
NIP. 198509172018011002

## KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan laporan hasil kegiatan Magang Merdeka Belajar Kampus Merdeka (MBKM) di Badan Riset dan Inovasi Nasional (BRIN) dengan baik. Laporan ini disusun sebagai bentuk pertanggungjawaban atas pelaksanaan kegiatan magang yang telah dilaksanakan.

Selama kegiatan magang, penulis terlibat dalam penelitian yang berfokus pada pengembangan model ekstraksi informasi pada literatur rumput laut serta pengembangan sistem visualisasi informasi dan hubungan entitas berbasis web dalam penelitian yang berjudul “Pengembangan Model Relation Extraction Berbasis Transformer dan Sistem Visualisasi Informasi Literatur Rumput Laut Berbasis Web”.

Penulis menyampaikan terima kasih kepada: Bapak Ahmad Tabrani, M.T.I., selaku Ketua Program Studi Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Maulana Hasanuddin Banten sekaligus dosen pembimbing. Ibu Dr. Dipl. Ing. (FH) Anne Parlina, M.P., selaku pembimbing lapangan dan peneliti di Badan Riset dan Inovasi Nasional (BRIN). Badan Riset dan Inovasi Nasional (BRIN), khususnya pada Kelompok Riset *Information Retrieval* Pusat Riset Sains Data dan Informasi, yang telah memberikan kesempatan dan pengalaman berharga selama pelaksanaan kegiatan magang. Seluruh pihak yang telah memberikan bantuan, dukungan, dan kerja sama selama pelaksanaan kegiatan magang hingga penyusunan laporan ini. Penulis menyadari bahwa laporan ini masih memiliki keterbatasan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan untuk perbaikan di masa mendatang. Semoga laporan ini dapat memberikan manfaat bagi pembaca dan pihak-pihak yang berkepentingan.

Bandung, 17 Juli 2026

Mursalin

## DAFTAR ISI

|                                                       |     |
|-------------------------------------------------------|-----|
| HALAMAN JUDUL.....                                    | i   |
| HALAMAN PENGESAHAN .....                              | ii  |
| KATA PENGANTAR .....                                  | iii |
| DAFTAR ISI.....                                       | iv  |
| RINGKASAN .....                                       | vi  |
| BAB 1 PENDAHULUAN .....                               | 1   |
| 1.1 Latar Belakang Masalah.....                       | 1   |
| 1.2 Rumusan Masalah .....                             | 2   |
| 1.3 Tujuan PKL/Magang.....                            | 3   |
| 1.3.1 Tujuan Umum .....                               | 3   |
| 1.3.2 Tujuan Khusus .....                             | 3   |
| 1.4 Manfaat PKL/Magang.....                           | 3   |
| 1.4.1 Manfaat Teoritis.....                           | 3   |
| 1.4.2 Manfaat Praktis .....                           | 4   |
| BAB II KAJIAN PUSTAKA .....                           | 5   |
| 2.1 Kajian Pustaka.....                               | 5   |
| 2.1.1 Information Extraction (IE).....                | 5   |
| 2.1.2 Named Entity Recognition (NER).....             | 6   |
| 2.1.3 Relation Extraction (RE).....                   | 7   |
| 2.1.4 Visualisasi Informasi dan Knowledge Graph ..... | 8   |
| 2.2 Kerangka Pikir Teoritis .....                     | 8   |
| BAB III METODE PKL/MAGANG .....                       | 10  |
| 3.1 Lokasi PKL/Magang .....                           | 10  |
| 3.2 Desain Pelaksanaan PKL/Magang .....               | 10  |
| 3.3 Penentuan Objek PKL/Magang.....                   | 11  |
| 3.4 Metode Pengumpulan Data.....                      | 11  |
| 3.5 Instrumen PKL/Magang .....                        | 12  |
| 3.6 Teknik Analisis Data.....                         | 12  |

|                                             |    |
|---------------------------------------------|----|
| BAB IV HASIL PKL/MAGANG DAN PEMBAHASAN..... | 13 |
| 4.1 Deskripsi Kegiatan PKL/Magang .....     | 13 |
| 4.2 Permasalahan yang Ditemukan .....       | 15 |
| 4.3 Metode Penyelesaian Masalah .....       | 15 |
| 4.4 Hasil Pelaksanaan PKL/Magang.....       | 16 |
| 4.5 Pembahasan.....                         | 22 |
| 4.6 Pelajaran yang Diperoleh .....          | 24 |
| BAB V KESIMPULAN DAN SARAN.....             | 26 |
| 5.1 Kesimpulan .....                        | 26 |
| 5.2 Keterbatasan.....                       | 26 |
| 5.3 Implikasi.....                          | 27 |
| 5.4 Saran.....                              | 28 |
| DAFTAR PUSTAKA .....                        | 30 |

## **RINGKASAN**

Literatur ilmiah mengenai rumput laut terus mengalami peningkatan dan mencakup berbagai informasi terkait spesies, senyawa bioaktif, manfaat, serta lokasi persebarannya. Informasi tersebut umumnya masih tersimpan dalam bentuk teks tidak terstruktur sehingga identifikasi dan analisis hubungan antar entitas pada berbagai dokumen menjadi kurang efisien. Keterbatasan tersebut menyulitkan proses eksplorasi informasi, terutama dalam mengidentifikasi keterkaitan antara spesies rumput laut, senyawa yang terkandung, manfaat yang dihasilkan, dan lokasi penelitian yang dilaporkan dalam literatur.

Kegiatan magang dilaksanakan di Pusat Riset Sains Data dan Informasi, Badan Riset dan Inovasi Nasional (BRIN), dengan fokus pada pengembangan model ekstraksi informasi dan sistem visualisasi hubungan entitas literatur rumput laut berbasis web. Tahapan kegiatan meliputi studi literatur, perancangan sistem, anotasi dataset menggunakan Label Studio, pengembangan model Relation Extraction (RE) berbasis Transformer, integrasi hasil ekstraksi informasi, serta pengembangan aplikasi menggunakan framework Streamlit. Data yang digunakan berupa literatur ilmiah rumput laut yang telah melalui proses anotasi entitas dan relasi. Hasil ekstraksi kemudian diolah menjadi data terstruktur yang digunakan sebagai sumber data pada sistem visualisasi.

Hasil kegiatan menghasilkan sistem visualisasi informasi berbasis web yang mampu menampilkan hasil ekstraksi entitas dan relasi dalam bentuk tabel, ringkasan informasi, serta visualisasi knowledge graph. Sistem juga menyediakan fitur unggah dokumen, penyaringan relasi, dan ekspor data ke format CSV. Implementasi sistem memungkinkan representasi hubungan antar entitas pada literatur rumput laut disajikan secara lebih terstruktur sehingga mendukung proses eksplorasi, penelusuran, informasi secara lebih efektif

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

Pelaksanaan Magang Merdeka Belajar Kampus Merdeka (MBKM) merupakan salah satu kegiatan pembelajaran yang memberikan kesempatan kepada mahasiswa untuk memperoleh pengalaman melalui keterlibatan langsung dalam kegiatan profesional dan penelitian. Kegiatan ini bertujuan untuk memperluas pemahaman mahasiswa mengenai penerapan ilmu yang diperoleh selama perkuliahan serta meningkatkan kompetensi sesuai dengan bidang keilmuan yang dipelajari. Kegiatan magang dilaksanakan di Badan Riset dan Inovasi Nasional (BRIN), khususnya pada Pusat Riset Sains Data dan Informasi.

Selama pelaksanaan magang, mahasiswa terlibat dalam penelitian yang berjudul “Pengembangan Model Relation Extraction Berbasis Transformer dan Sistem Visualisasi Informasi Literatur Rumput Laut Berbasis Web.” Penelitian tersebut berfokus pada pengembangan model ekstraksi informasi dari literatur rumput laut dan pengembangan sistem visualisasi yang digunakan untuk menyajikan informasi hasil ekstraksi secara terstruktur. Rumput laut merupakan salah satu objek penelitian yang banyak dikaji dalam berbagai bidang, seperti pangan, farmasi, kosmetik, dan bioteknologi.

Peningkatan jumlah penelitian mengenai rumput laut menyebabkan jumlah publikasi ilmiah yang membahas spesies, senyawa bioaktif, manfaat, dan lokasi penelitian terus bertambah. Publikasi tersebut memuat informasi yang dapat digunakan sebagai sumber pengetahuan bagi kegiatan penelitian maupun pengembangan produk berbasis rumput laut. Sebagian besar informasi pada literatur rumput laut masih tersaji dalam bentuk teks tidak terstruktur. Informasi mengenai spesies rumput laut, senyawa bioaktif, manfaat, dan lokasi penelitian tersebar pada berbagai dokumen sehingga proses identifikasi dan penelusuran informasi memerlukan pembacaan terhadap banyak sumber.

Kondisi tersebut menyebabkan proses analisis hubungan antar informasi menjadi kurang efisien, terutama ketika jumlah literatur yang digunakan semakin banyak. Perkembangan Natural Language Processing (NLP) memungkinkan proses ekstraksi informasi dari dokumen teks melalui pendekatan Named Entity Recognition (NER) dan Relation Extraction (RE). Melalui pendekatan tersebut, entitas dan hubungan antar entitas dapat diidentifikasi secara otomatis dan direpresentasikan dalam bentuk data terstruktur. Namun, hasil ekstraksi umumnya disajikan dalam bentuk tabel atau data tekstual sehingga hubungan antar entitas belum dapat diamati secara menyeluruh.

Berdasarkan kondisi tersebut, diperlukan suatu sistem yang mampu menyajikan hasil ekstraksi informasi dalam bentuk visualisasi yang memudahkan pengguna untuk menelusuri hubungan antar entitas. Pengembangan sistem visualisasi informasi dan hubungan entitas berbasis web diharapkan dapat mendukung proses eksplorasi informasi pada literatur rumput laut melalui penyajian data yang lebih terstruktur dan mudah diakses.

## **1.2 Rumusan Masalah**

Berdasarkan kegiatan yang dilaksanakan selama program Magang MBKM di Badan Riset dan Inovasi Nasional (BRIN), rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana proses pembangunan dataset anotasi entitas dan relasi pada literatur rumput laut untuk mendukung pengembangan model ekstraksi informasi?
2. Bagaimana pengembangan dan evaluasi model Relation Extraction dalam mengidentifikasi hubungan antar entitas pada literatur rumput laut?
3. Bagaimana mengintegrasikan hasil ekstraksi informasi ke dalam sistem visualisasi informasi dan hubungan entitas berbasis web?
4. Bagaimana menyajikan informasi hasil ekstraksi dalam bentuk visualisasi yang mendukung eksplorasi hubungan antar entitas pada literatur rumput laut?



### **1.3 Tujuan PKL/Magang**

#### **1.3.1 Tujuan umum**

Melaksanakan kegiatan magang di Badan Riset dan Inovasi Nasional (BRIN) guna memperoleh pengalaman dan pemahaman mengenai proses penelitian serta pengembangan sistem pada bidang ekstraksi informasi dan visualisasi data berbasis web.

#### **1.3.2 Tujuan khusus**

Secara khusus, kegiatan magang yang dilaksanakan di Badan Riset dan Inovasi Nasional (BRIN) bertujuan untuk:

1. Mempelajari proses penelitian yang berkaitan dengan ekstraksi informasi pada literatur rumput laut.
2. Melakukan pembangunan dataset melalui proses anotasi entitas dan relasi pada literatur rumput laut.
3. Mengembangkan dan mengevaluasi model Relation Extraction untuk mengidentifikasi hubungan antar entitas pada literatur rumput laut.
4. Mengintegrasikan hasil ekstraksi informasi ke dalam sistem visualisasi informasi dan hubungan entitas berbasis web.
5. Mengembangkan aplikasi visualisasi berbasis web untuk menyajikan hasil ekstraksi entitas dan relasi secara interaktif.
6. Menyusun laporan hasil kegiatan magang sebagai bentuk dokumentasi dan pertanggungjawaban pelaksanaan kegiatan.

### **1.4 Manfaat PKL/Magang**

#### **1.4.1 Manfaat teoritis**

Kegiatan magang ini diharapkan dapat memberikan kontribusi pada pengembangan pengetahuan di bidang Natural Language Processing (NLP), khususnya terkait ekstraksi informasi, anotasi data, Relation Extraction, dan visualisasi informasi berbasis web. Selain itu, laporan ini dapat menjadi referensi

bagi penelitian selanjutnya yang berkaitan dengan pemanfaatan teknik ekstraksi informasi untuk mengelola dan menyajikan pengetahuan dari literatur ilmiah.

#### **1.4.2 Manfaat praktis**

Pelaksanaan kegiatan magang di Badan Riset dan Inovasi Nasional (BRIN) memberikan manfaat bagi berbagai pihak yang terlibat, baik mahasiswa, instansi mitra, maupun perguruan tinggi. Manfaat tersebut diperoleh melalui keterlibatan langsung dalam kegiatan penelitian yang mencakup pembangunan dataset, pengembangan model ekstraksi informasi, serta pengembangan sistem visualisasi informasi dan hubungan entitas berbasis web.:

1. Bagi Mahasiswa Kegiatan magang memberikan kesempatan untuk menerapkan pengetahuan yang telah diperoleh selama perkuliahan ke dalam kegiatan penelitian secara langsung. Melalui keterlibatan dalam proses anotasi dataset, pengembangan dan evaluasi model Relation Extraction, serta pembangunan sistem berbasis web, mahasiswa memperoleh pemahaman yang lebih komprehensif mengenai penerapan Natural Language Processing dan pengembangan sistem pada lingkungan penelitian.
2. Bagi Badan Riset dan Inovasi Nasional (BRIN) Kegiatan magang memberikan dukungan terhadap pelaksanaan penelitian melalui kontribusi dalam proses anotasi data, pengembangan model ekstraksi informasi, serta implementasi sistem visualisasi informasi dan hubungan entitas pada literatur rumput laut. Hasil yang diperoleh dapat dimanfaatkan sebagai bagian dari pengembangan penelitian dan pengelolaan informasi pada domain yang dikaji.
3. Bagi Universitas Islam Negeri Sultan Maulana Hasanuddin Banten Kegiatan magang menjadi sarana implementasi pembelajaran yang menghubungkan kompetensi akademik dengan kebutuhan dunia penelitian. Selain itu, kegiatan ini dapat memperkuat kerja sama antara perguruan tinggi dan lembaga penelitian serta memberikan pengalaman praktis yang mendukung pengembangan kompetensi mahasiswa sesuai dengan capaian pembelajaran program studi.

## **BAB II**

### **KAJIAN PUSTAKA**

#### **2.1 Kajian Pustaka**

Kajian pustaka merupakan landasan teoritis yang digunakan untuk mendukung pelaksanaan kegiatan magang dan pengembangan penelitian. Melalui kajian pustaka, berbagai konsep, teori, serta hasil penelitian terdahulu yang relevan dapat digunakan sebagai dasar dalam memahami permasalahan, menentukan pendekatan penelitian, dan mendukung proses pengembangan sistem. Pada laporan ini, kajian pustaka mencakup konsep Information Extraction, Named Entity Recognition (NER), Relation Extraction (RE), visualisasi informasi, serta penelitian terdahulu yang berkaitan dengan ekstraksi dan representasi informasi dari dokumen ilmiah. Teori-teori tersebut menjadi dasar dalam proses pembangunan dataset, pengembangan model ekstraksi informasi, dan implementasi sistem visualisasi informasi dan hubungan entitas literatur rumput laut berbasis web.

##### **2.1.1 Information Extraction (IE)**

Information Extraction (IE) merupakan salah satu bidang dalam Natural Language Processing (NLP) yang berfokus pada proses ekstraksi informasi dari data teks tidak terstruktur menjadi data yang terorganisasi dan dapat diolah secara otomatis. Berdasarkan kajian dalam [1] bertujuan untuk mengidentifikasi informasi penting yang terdapat dalam dokumen teks dan merepresentasikannya dalam bentuk yang lebih terstruktur. Tugas yang umum dilakukan dalam Information Extraction meliputi Named Entity Recognition (NER) untuk mengidentifikasi entitas serta Relation Extraction (RE) untuk mengidentifikasi hubungan antar entitas yang ditemukan dalam teks. Dalam dokumen ilmiah, informasi umumnya disajikan dalam bentuk narasi sehingga tidak dapat diolah secara langsung oleh sistem komputer.

Dalam dokumen ilmiah, informasi umumnya disajikan dalam bentuk narasi sehingga tidak dapat diolah secara langsung oleh sistem komputer. Informasi yang terdapat pada dokumen perlu diidentifikasi dan direpresentasikan ke dalam bentuk yang lebih terstruktur agar dapat digunakan pada proses pengolahan dan analisis data. Melalui pendekatan Information Extraction, informasi dalam dokumen dapat direpresentasikan sebagai entitas dan relasi yang menggambarkan keterkaitan antar informasi yang terdapat dalam teks.

Pada literatur rumput laut, informasi yang tersedia mencakup berbagai aspek, seperti spesies rumput laut, kandungan senyawa bioaktif, manfaat, serta lokasi penelitian atau persebaran geografis. Informasi tersebut tersebar pada berbagai publikasi ilmiah dan sebagian besar masih disajikan dalam bentuk teks tidak terstruktur. Proses identifikasi informasi dari sejumlah dokumen memerlukan penelusuran dan pencatatan secara manual sehingga kurang efisien apabila dilakukan pada koleksi literatur yang besar. Penerapan Information Extraction memungkinkan informasi yang terdapat pada literatur rumput laut direpresentasikan dalam bentuk data terstruktur sehingga dapat digunakan sebagai dasar dalam proses analisis dan pengembangan sistem visualisasi informasi

### **2.1.2 Named Entity Recognition (NER)**

Named Entity Recognition (NER) merupakan salah satu tugas dalam Natural Language Processing (NLP) yang bertujuan untuk mengidentifikasi dan mengklasifikasikan entitas yang terdapat dalam suatu teks ke dalam kategori yang telah ditentukan. Dalam kajian NER [2] dijelaskan sebagai tahapan yang digunakan untuk mengenali informasi penting dalam dokumen, seperti nama objek, lokasi, organisasi, maupun kategori lain yang sesuai dengan kebutuhan suatu domain. Hasil dari proses NER berupa entitas yang telah diberi label sehingga dapat digunakan pada tahapan pengolahan informasi berikutnya. Pada penelitian ini, NER digunakan untuk mengidentifikasi empat kategori entitas yang terdapat pada literatur rumput laut, yaitu Seaweed, Compound, Benefit, dan Location. Keempat kategori tersebut dipilih karena merepresentasikan informasi utama yang sering ditemukan dalam publikasi ilmiah mengenai rumput laut.

Proses identifikasi entitas dilakukan untuk memperoleh informasi yang berkaitan dengan spesies rumput laut, kandungan senyawa bioaktif, manfaat yang dilaporkan, serta lokasi penelitian atau persebaran. Penerapan NER pada domain khusus telah banyak digunakan dalam berbagai penelitian ekstraksi informasi. Salah satunya ditunjukkan pada penelitian [2], yang memanfaatkan NER untuk mengidentifikasi entitas spesifik pada literatur tanaman. Pendekatan serupa diterapkan dalam penelitian ini untuk mengidentifikasi entitas yang terdapat pada literatur rumput laut. Hasil identifikasi entitas selanjutnya digunakan sebagai masukan pada proses Relation Extraction guna mengidentifikasi hubungan antar entitas yang terdapat dalam dokumen.

### **2.1.3 Relation Extraction (RE)**

Relation Relation Extraction (RE) merupakan salah satu tugas dalam Natural Language Processing (NLP) yang bertujuan untuk mengidentifikasi hubungan antar entitas dalam suatu teks. Apabila Named Entity Recognition (NER) digunakan untuk mengenali entitas, maka RE digunakan untuk menentukan jenis hubungan yang terdapat di antara entitas tersebut. Berdasarkan penelitian [1] proses RE dilakukan dengan mengidentifikasi relasi antar pasangan entitas yang telah dikenali sebelumnya, baik dalam satu kalimat (intra-sentence) maupun antar kalimat (inter-sentence).

Dalam penelitian ini, RE digunakan untuk mengidentifikasi hubungan antar entitas Seaweed, Compound, Benefit, dan Location yang terdapat pada literatur rumput laut. Hasil identifikasi relasi digunakan untuk membentuk data terstruktur yang merepresentasikan keterkaitan informasi dalam dokumen ilmiah. Pengembangan model RE dilakukan menggunakan model berbasis Transformer, yaitu BioBERT, SciBERT, dan PubMedBERT. Pemanfaatan model pada domain ilmiah sejalan dengan penelitian [3] yang menunjukkan penerapan model bahasa untuk mendukung ekstraksi informasi dari literatur sains. Pada penelitian ini, ketiga model digunakan untuk mengidentifikasi hubungan antar entitas berdasarkan dataset literatur rumput laut yang telah dianotasi.

#### **2.1.4 Visualisasi Informasi dan Knowledge Graph**

Visualisasi informasi merupakan teknik penyajian data dalam bentuk representasi visual untuk membantu proses pemahaman dan eksplorasi informasi. Melalui visualisasi, data dapat disajikan dalam bentuk yang lebih terstruktur sehingga memudahkan pengguna dalam mengamati pola, keterkaitan, dan informasi yang terkandung di dalamnya. Salah satu bentuk visualisasi yang digunakan dalam representasi pengetahuan adalah knowledge graph. Berdasarkan penelitian [4], knowledge graph merepresentasikan entitas sebagai simpul (node) dan hubungan antar entitas sebagai sisi (edge). Representasi tersebut memungkinkan informasi yang diperoleh dari dokumen teks disusun dalam bentuk struktur graf yang menunjukkan keterkaitan antar entitas.

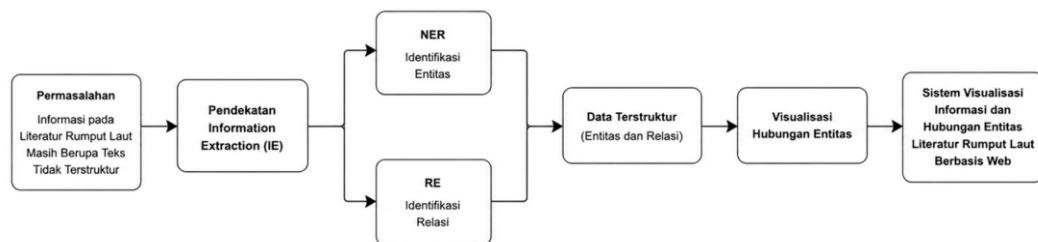
Pemanfaatan visualisasi berbasis graf juga diterapkan dalam penelitian [5], yang memanfaatkan hasil ekstraksi informasi untuk merepresentasikan hubungan antar komponen informasi dalam dokumen ilmiah. Pendekatan tersebut menunjukkan bahwa hasil ekstraksi entitas dan relasi dapat digunakan sebagai dasar dalam pembentukan representasi pengetahuan yang lebih terstruktur. Dalam penelitian ini, hasil ekstraksi entitas dan relasi digunakan sebagai sumber data untuk membangun visualisasi hubungan antar entitas pada literatur rumput laut. Visualisasi tersebut menampilkan keterkaitan antara entitas Seaweed, Compound, Benefit, dan Location dalam bentuk graf sehingga informasi yang diperoleh dari literatur dapat ditampilkan secara lebih terorganisasi.

#### **2.2 Kerangka Pikir Teoritis**

Publikasi ilmiah mengenai rumput laut mengandung berbagai informasi terkait spesies, senyawa bioaktif, manfaat, dan lokasi penelitian yang umumnya disajikan dalam bentuk teks tidak terstruktur. Bentuk penyajian tersebut menyebabkan keterkaitan antar informasi sulit diidentifikasi dan memerlukan penelusuran dokumen secara manual.

Penelitian ini menerapkan pendekatan Information Extraction untuk mengubah data tekstual menjadi data terstruktur. Proses tersebut dilakukan

melalui Named Entity Recognition (NER) untuk mengidentifikasi entitas serta Relation Extraction (RE) untuk menentukan hubungan antar entitas yang terdapat dalam dokumen. Hasil ekstraksi berupa kumpulan entitas dan relasi yang merepresentasikan isi dokumen secara terstruktur. Data yang dihasilkan kemudian dimanfaatkan dalam pengembangan sistem visualisasi berbasis web untuk menampilkan hubungan antar entitas secara lebih terorganisasi. Melalui pendekatan tersebut, informasi yang tersebar pada berbagai publikasi dapat disajikan dalam bentuk visual yang mendukung proses penelusuran dan eksplorasi data.



**Gambar 2.1.** Kerangka Berpikir

## **BAB III**

### **METODE PKL/MAGANG**

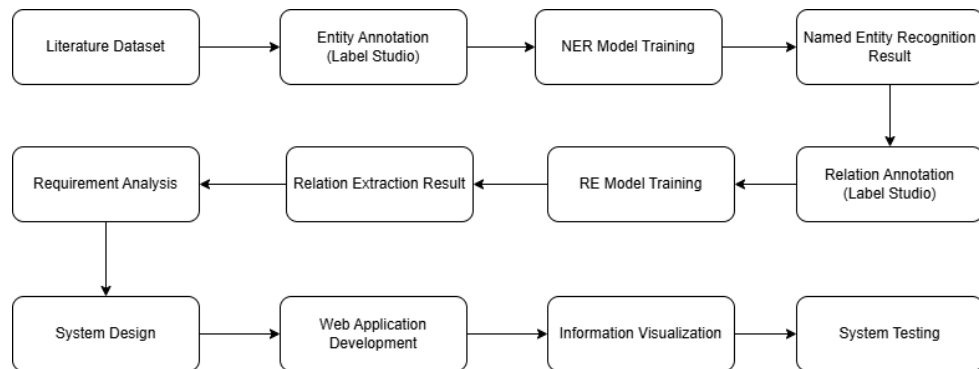
#### **3.1 Lokasi PKL/Magang**

Kegiatan PKL/Magang dilaksanakan di Badan Riset dan Inovasi Nasional (BRIN) pada Kelompok Riset Temu Kembali Informasi (Information Retrieval) yang berada di bawah Pusat Riset Sains Data dan Informasi. Kelompok riset ini berfokus pada pengembangan metode dan sistem temu kembali informasi, pengolahan data, kecerdasan artifisial, serta sains data untuk mendukung kegiatan penelitian. Pelaksanaan magang berlangsung selama periode Februari 2026 hingga Juli 2026. Selama kegiatan magang, penulis terlibat dalam penelitian yang berjudul “Rancang Bangun Sistem Visualisasi Informasi dan Hubungan Entitas Literatur Rumput Laut Berbasis Web”. Kegiatan yang dilakukan mencakup proses anotasi data, pengembangan model ekstraksi informasi, serta pengembangan sistem visualisasi informasi dan hubungan entitas berbasis web.

#### **3.2 Desain Pelaksanaan PKL/Magang**

Pelaksanaan PKL/Magang menggunakan pendekatan deskriptif melalui keterlibatan langsung dalam kegiatan penelitian dan pengembangan sistem. Pendekatan ini dipilih karena memungkinkan penulis memahami tahapan penelitian secara menyeluruh, mulai dari pengolahan data hingga implementasi hasil penelitian ke dalam sistem yang dapat digunakan oleh pengguna. Kegiatan magang dilaksanakan melalui beberapa tahapan, yaitu pengumpulan literatur, anotasi entitas dan relasi, pengembangan model ekstraksi informasi menggunakan pendekatan Named Entity Recognition (NER) dan Relation Extraction (RE), analisis kebutuhan sistem, perancangan sistem, pengembangan aplikasi berbasis web, serta pengujian sistem. Alur pelaksanaan kegiatan magang disajikan pada Gambar 3.1.





**Gambar 3.1.** Alur Pelaksanaan Penelitian

### 3.3 Penentuan Objek PKL/Magang

Objek dalam pelaksanaan PKL/Magang ini adalah literatur ilmiah mengenai rumput laut yang memuat informasi terkait spesies, kandungan senyawa bioaktif, manfaat, dan lokasi penelitian. Fokus kegiatan diarahkan pada proses ekstraksi informasi dari dokumen ilmiah melalui identifikasi entitas dan hubungan antar entitas, serta pemanfaatan hasil ekstraksi tersebut dalam pengembangan sistem visualisasi informasi berbasis web. Entitas yang digunakan dalam penelitian meliputi kategori Seaweed, Compound, Benefit, dan Location. Hubungan antar entitas yang berhasil diidentifikasi selanjutnya digunakan sebagai sumber data untuk membangun visualisasi hubungan informasi pada literatur rumput laut.

### 3.4 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan selama pelaksanaan PKL/Magang meliputi observasi, wawancara dan diskusi, serta dokumentasi.

#### a. Observasi

Observasi dilakukan melalui pengamatan langsung terhadap kegiatan penelitian dan pengembangan sistem yang berlangsung selama program magang. Melalui metode ini, penulis memperoleh pemahaman mengenai proses anotasi data, pengembangan model ekstraksi informasi, dan pembangunan sistem visualisasi berbasis web.

#### b. Wawancara dan Diskusi

Wawancara dan diskusi dilakukan bersama pembimbing lapangan dan anggota tim penelitian untuk memperoleh informasi mengenai kebutuhan penelitian, skema anotasi data, proses pengembangan model, serta kebutuhan sistem yang akan dikembangkan.

c. Dokumentasi

Dokumentasi dilakukan dengan mengumpulkan berbagai dokumen pendukung, seperti dataset literatur rumput laut, pedoman anotasi, hasil ekstraksi informasi, dokumentasi kegiatan magang, serta referensi ilmiah yang digunakan selama penelitian.

### **3.5 Instrumen PKL/Magang**

Instrumen yang digunakan selama pelaksanaan PKL/Magang terdiri atas perangkat keras dan perangkat lunak yang mendukung kegiatan penelitian dan pengembangan sistem. Perangkat keras yang digunakan berupa laptop sebagai sarana utama dalam pengolahan data dan pengembangan aplikasi. Adapun perangkat lunak yang digunakan meliputi Label Studio untuk proses anotasi data, Visual Studio Code sebagai lingkungan pengembangan aplikasi, Python sebagai bahasa pemrograman, Streamlit sebagai framework pengembangan aplikasi web, serta berbagai pustaka pendukung yang digunakan dalam proses ekstraksi informasi dan visualisasi data.

### **3.6 Teknik Analisis Data**

Teknik analisis data yang digunakan dalam kegiatan PKL/Magang adalah analisis deskriptif. Data yang diperoleh melalui observasi, wawancara, diskusi, dan dokumentasi disusun secara sistematis untuk menggambarkan proses pelaksanaan kegiatan magang dan hasil yang diperoleh selama penelitian. Analisis dilakukan terhadap setiap tahapan kegiatan, meliputi proses anotasi data, pengembangan model ekstraksi informasi, serta pengembangan sistem visualisasi berbasis web. Hasil analisis digunakan untuk menjelaskan pelaksanaan kegiatan, mengidentifikasi kendala yang dihadapi, serta mendeskripsikan solusi yang diterapkan selama program PKL/Magang berlangsung.

## **BAB IV**

### **HASIL PKL/MAGANG DAN PEMBAHASAN**

#### **4.1 Deskripsi Kegiatan PKL/Magang**

Kegiatan PKL/Magang dilaksanakan melalui keterlibatan langsung dalam penelitian yang berfokus pada pengembangan sistem visualisasi informasi dan hubungan entitas pada literatur rumput laut berbasis web. Selama pelaksanaan magang, penulis terlibat dalam berbagai tahapan kegiatan yang meliputi pengumpulan dan pengolahan data, proses anotasi entitas dan relasi, pengembangan model ekstraksi informasi, serta pengembangan dan pengujian sistem visualisasi berbasis web.

**Tabel 4.1** Kegiatan PKL/Magang

| <b>No</b> | <b>Tanggal</b>             | <b>Kegiatan</b>                                 | <b>Keterangan</b>                                                                                                                                                                |
|-----------|----------------------------|-------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1         | Februari<br>2025           | Orientasi dan<br>Pengenalan<br>Kegiatan Magang  | Mengikuti pengarahan mengenai kegiatan penelitian, pembagian tugas, serta pengenalan topik penelitian terkait ekstraksi informasi literatur rumput laut.                         |
| 2         | Februari<br>2025           | Studi Literatur                                 | Mempelajari literatur mengenai Information Extraction, Named Entity Recognition (NER), Relation Extraction (RE), dan visualisasi informasi sebagai dasar pelaksanaan penelitian. |
| 3         | Februari–<br>Maret<br>2025 | Pengumpulan<br>Dataset Literatur<br>Rumput Laut | Mengumpulkan dan menyeleksi artikel ilmiah yang digunakan sebagai sumber data penelitian.                                                                                        |
| 4         | Maret<br>2025              | Penyusunan<br>Pedoman Anotasi                   | Mempelajari dan menerapkan pedoman anotasi untuk kategori entitas Seaweed, Compound, Benefit, dan Location serta relasi antar entitas.                                           |

|    |                      |                                    |                                                                                                                            |
|----|----------------------|------------------------------------|----------------------------------------------------------------------------------------------------------------------------|
| 5  | Maret–<br>April 2025 | Anotasi Entitas                    | Melakukan proses pelabelan entitas pada dataset literatur rumput laut menggunakan perangkat anotasi yang telah ditentukan. |
| 6  | April 2025           | Validasi Hasil Anotasi Entitas     | Melakukan pemeriksaan dan perbaikan hasil anotasi untuk menjaga konsistensi data.                                          |
| 7  | April 2025           | Anotasi Relasi                     | Melakukan pelabelan hubungan antar entitas berdasarkan pedoman anotasi yang telah ditetapkan.                              |
| 8  | April–Mei 2025       | Pengembangan Model NER             | Melakukan pelatihan dan evaluasi model Named Entity Recognition untuk mengidentifikasi entitas pada literatur rumput laut. |
| 9  | Mei 2025             | Pengembangan Model RE              | Melakukan pelatihan dan evaluasi model Relation Extraction untuk mengidentifikasi hubungan antar entitas.                  |
| 10 | Mei 2025             | Analisis Kebutuhan Sistem          | Mengidentifikasi kebutuhan fungsional dan kebutuhan data yang diperlukan dalam pengembangan sistem visualisasi.            |
| 11 | Mei 2025             | Perancangan Sistem                 | Menyusun rancangan antarmuka, alur sistem, dan visualisasi hubungan entitas yang akan dikembangkan.                        |
| 12 | Mei–Juni 2025        | Pengembangan Aplikasi Web          | Mengembangkan sistem visualisasi informasi dan hubungan entitas berbasis web menggunakan Streamlit.                        |
| 13 | Juni 2025            | Implementasi Visualisasi Informasi | Mengintegrasikan hasil ekstraksi informasi ke dalam fitur visualisasi hubungan entitas.                                    |
| 14 | Juni 2025            | Pengujian Sistem                   | Melakukan pengujian terhadap fungsi                                                                                        |

|    |           |                           |                                                                                            |
|----|-----------|---------------------------|--------------------------------------------------------------------------------------------|
|    |           |                           | sistem dan tampilan visualisasi untuk memastikan sistem berjalan sesuai kebutuhan.         |
| 15 | Juni 2025 | Penyusunan Laporan Magang | Menyusun laporan kegiatan PKL/Magang berdasarkan seluruh kegiatan yang telah dilaksanakan. |

Sumber: Data diolah oleh penulis, [2026]

#### 4.2 Permasalahan yang Ditemukan

Pelaksanaan PKL/Magang dilakukan pada lingkungan penelitian di Badan Riset dan Inovasi Nasional (BRIN) dengan fokus kegiatan pada pengembangan ekstraksi informasi dan sistem visualisasi hubungan entitas literatur rumput laut. Kegiatan yang dilaksanakan mencakup tahapan pengumpulan data [6], anotasi dataset, pengembangan model ekstraksi informasi, pengembangan sistem berbasis web, hingga pengujian sistem. Seluruh kegiatan tersebut dilaksanakan secara bertahap sesuai dengan kebutuhan penelitian dan pengembangan sistem. Ringkasan kegiatan yang dilakukan selama PKL/Magang disajikan pada

#### 4.3 Metode Penyelesaian Masalah

Metode penyelesaian masalah yang diterapkan selama pelaksanaan magang terdiri atas beberapa tahapan, yaitu studi literatur dan pengumpulan data, anotasi data, pengembangan model Natural Language Processing (NLP), pengembangan sistem visualisasi, serta pengujian sistem. Tahap pertama adalah studi literatur dan pengumpulan data. Pada tahap ini dilakukan penelusuran literatur yang relevan dengan topik penelitian untuk memperoleh pemahaman mengenai ekstraksi informasi pada domain rumput laut. Data yang digunakan berupa 1.122 abstrak publikasi ilmiah mengenai rumput laut yang diperoleh dari basis data Scopus. Data yang telah dikumpulkan kemudian dipersiapkan untuk proses anotasi dan pengembangan model.

Tahap kedua adalah anotasi data. Proses anotasi dilakukan menggunakan perangkat lunak Label Studio. Anotasi mencakup pelabelan empat jenis entitas, yaitu Seaweed, Location, Benefit, dan Compound. Selain anotasi entitas,

dilakukan pula anotasi relasi antarentitas untuk menghasilkan dataset yang digunakan dalam pengembangan model Relation Extraction. Tahap ketiga adalah pengembangan model NLP. Dataset yang telah dianotasi dibagi menjadi data pelatihan, validasi, dan pengujian. Selanjutnya dilakukan proses tokenisasi untuk mengubah teks menjadi representasi yang dapat diproses oleh model Transformer. Pelatihan model Relation Extraction dilakukan menggunakan tiga arsitektur Transformer, yaitu SciBERT, BioBERT, dan PubMedBERT. Model yang dilatih digunakan untuk mengidentifikasi relasi antarentitas pada abstrak literatur rumput laut.

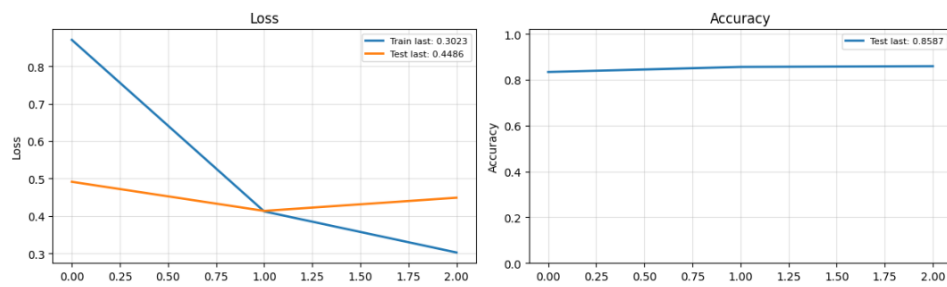
Tahap keempat adalah pengembangan sistem visualisasi berbasis web. Kegiatan pada tahap ini meliputi analisis kebutuhan sistem, perancangan arsitektur sistem, dan implementasi aplikasi menggunakan framework Streamlit. Sistem yang dikembangkan digunakan untuk menampilkan hasil ekstraksi informasi dalam bentuk visualisasi hubungan antarentitas yang diperoleh dari model Relation Extraction. Tahap terakhir adalah pengujian sistem. Pengujian dilakukan menggunakan metode Black Box Testing untuk memverifikasi kesesuaian fungsi sistem dengan kebutuhan yang telah ditetapkan. Pengujian difokuskan pada masukan, keluaran, dan perilaku fungsional sistem tanpa mempertimbangkan implementasi kode program.

#### **4.4 Hasil Pelaksanaan PKL/Magang**

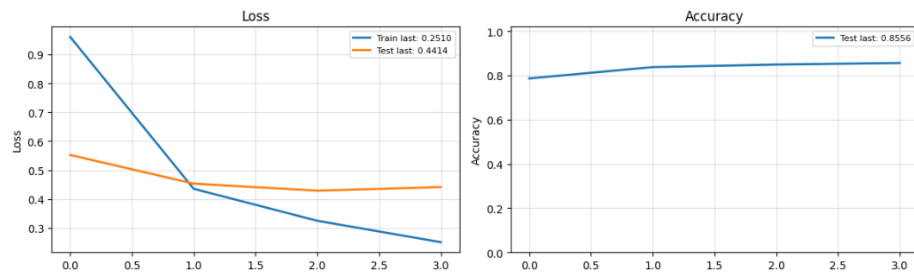
Pelaksanaan kegiatan magang menghasilkan model Relation Extraction untuk mengidentifikasi relasi antarentitas pada literatur rumput laut serta sistem visualisasi berbasis web untuk menampilkan hasil ekstraksi informasi. Hasil yang diperoleh meliputi pengembangan model Relation Extraction, pembangunan sistem visualisasi, penyajian hasil ekstraksi dalam berbagai bentuk visualisasi, dan pengujian fungsionalitas sistem. Pada tahap pengembangan model Relation Extraction, dilakukan pelatihan dan evaluasi menggunakan tiga arsitektur Transformer, yaitu SciBERT, BioBERT, dan PubMedBERT. Evaluasi dilakukan menggunakan data uji yang telah dipisahkan pada tahap persiapan dataset. Hasil evaluasi masing-masing model disajikan pada **Tabel 4.2**.

**Table 4.2** Hasil Evaluasi Model Relation Extraction

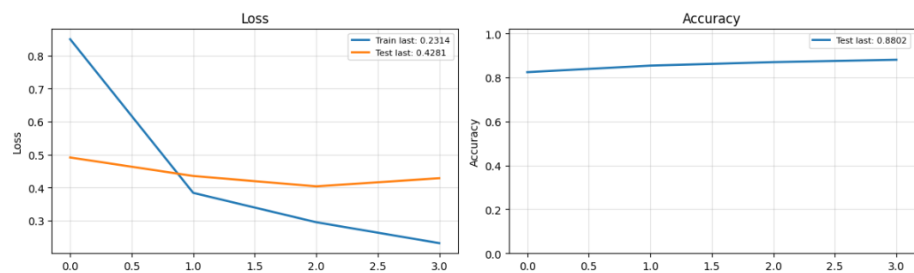
| Model      | Epoch | Training Loss | Validation Loss | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|------------|-------|---------------|-----------------|--------------|---------------|------------|--------------|
| SciBERT    | 2     | 0.412088      | 0.413442        | 85.56        | 86.21         | 85.56      | 85.69        |
| BioBERT    | 3     | 0.324958      | 0.428929        | 84.88        | 86.30         | 84.88      | 85.17        |
| PubMedBERT | 3     | 0.294770      | 0.403423        | 86.92        | 87.64         | 86.92      | 87.02        |



**Gambar 4.1** Grafik Loss dan Accuracy Model SciBERT



**Gambar 4.2** Grafik Loss dan Accuracy Model BioBERT



**Gambar 4.3** Grafik Loss dan Accuracy Model PubmedBERT

| True Label | No Relation  | Seaweed-Loc     | Seaweed-Ben | Seaweed-Comp | Comp-Ben     |          |
|------------|--------------|-----------------|-------------|--------------|--------------|----------|
|            | No Relation  | 731             | 1           | 35           | 46           | 66       |
|            | Seaweed-Loc  | 2               | 85          | 1            | 0            | 0        |
|            | Seaweed-Ben  | 6               | 1           | 156          | 9            | 1        |
|            | Seaweed-Comp | 14              | 0           | 35           | 392          | 0        |
|            | Comp-Ben     | 59              | 0           | 0            | 0            | 271      |
|            |              | Predicted Label |             |              |              |          |
|            |              | No Relation     | Seaweed-Loc | Seaweed-Ben  | Seaweed-Comp | Comp-Ben |

**Gambar 4.4** Confusion Matrix Model SciBERT

| True Label | No Relation  | Seaweed-Loc     | Seaweed-Ben | Seaweed-Comp | Comp-Ben     |          |
|------------|--------------|-----------------|-------------|--------------|--------------|----------|
|            | No Relation  | 716             | 1           | 39           | 33           | 90       |
|            | Seaweed-Loc  | 3               | 79          | 5            | 1            | 0        |
|            | Seaweed-Ben  | 7               | 0           | 156          | 8            | 2        |
|            | Seaweed-Comp | 17              | 0           | 41           | 382          | 1        |
|            | Comp-Ben     | 40              | 0           | 1            | 0            | 289      |
|            |              | Predicted Label |             |              |              |          |
|            |              | No Relation     | Seaweed-Loc | Seaweed-Ben  | Seaweed-Comp | Comp-Ben |

**Gambar 4.5** Confusion Matrix Model BioBERT

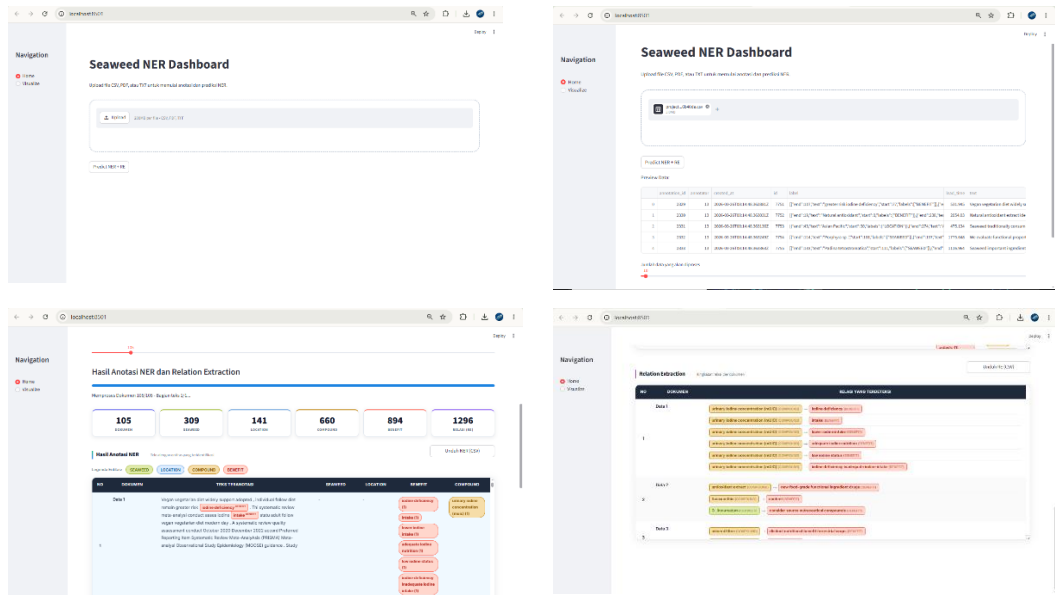
| True Label | No Relation     | 726         | 1           | 34           | 39       | 79  |
|------------|-----------------|-------------|-------------|--------------|----------|-----|
|            | Seaweed-Loc     | 4           | 83          | 1            | 0        | 0   |
|            | Seaweed-Ben     | 6           | 0           | 156          | 10       | 1   |
|            | Seaweed-Comp    | 13          | 1           | 22           | 405      | 0   |
|            | Comp-Ben        | 39          | 0           | 0            | 0        | 291 |
|            | Predicted Label |             |             |              |          |     |
|            | No Relation     | Seaweed-Loc | Seaweed-Ben | Seaweed-Comp | Comp-Ben |     |

**Gambar 4.6** Confusion Matrix Model PubmedBERT

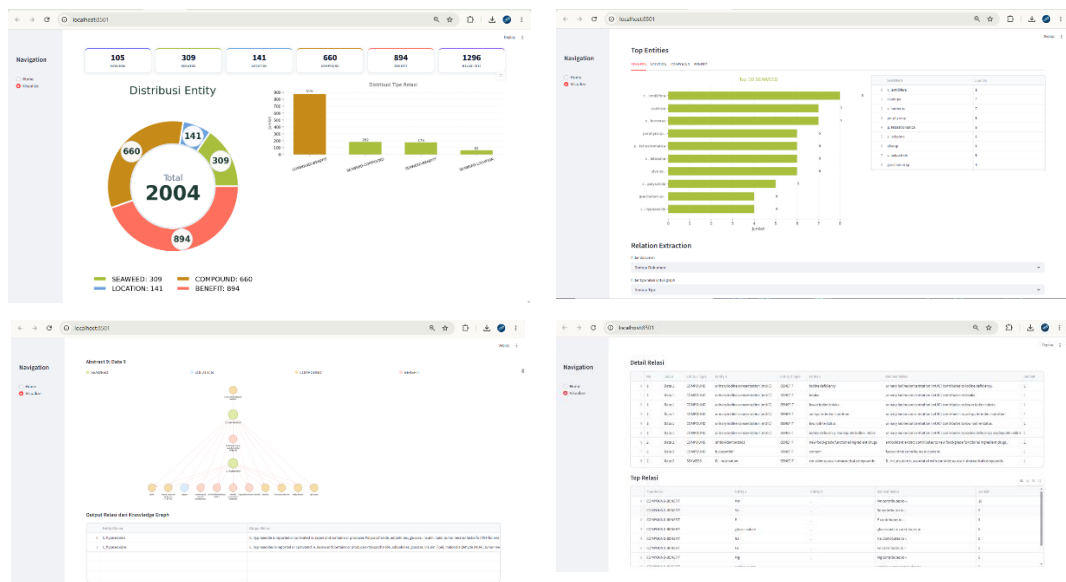
Berdasarkan hasil evaluasi pada Tabel 4.1, model PubMedBERT memperoleh nilai akurasi sebesar 86,92%, precision sebesar 87,64%, recall sebesar 86,92%, dan F1-score sebesar 87,02%. Nilai tersebut merupakan hasil evaluasi yang diperoleh pada data uji setelah proses pelatihan model selesai dilakukan. Selain pengembangan model, dihasilkan sebuah sistem visualisasi



berbasis web yang dibangun menggunakan framework Streamlit. Sistem terdiri atas halaman utama (Home Page) yang digunakan untuk mengunggah dan memproses dokumen, serta halaman visualisasi (Visualize Page) yang digunakan untuk menampilkan hasil ekstraksi informasi.



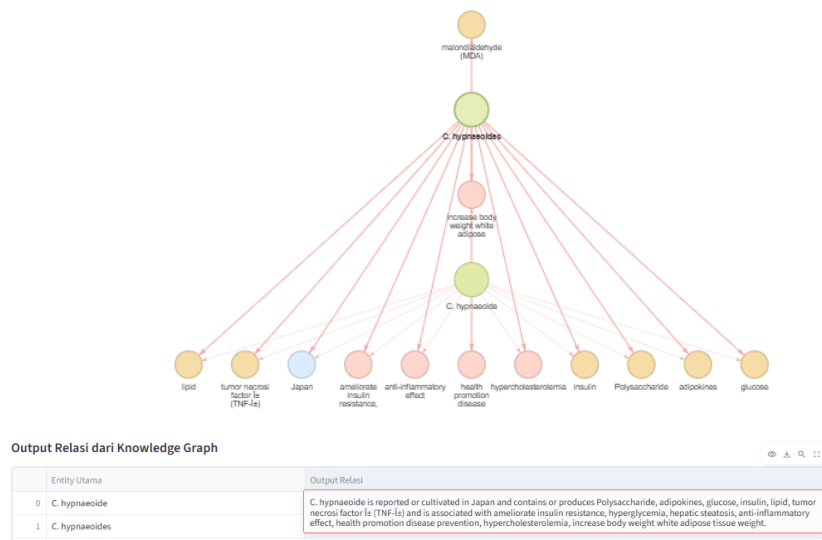
Gambar 4.7 Home Page Sistem



| Location | Entity | Value | Relationship | Value |
|----------|--------|-------|--------------|-------|
| Location | Entity | Value | Relationship | Value |
| Location | Entity | Value | Relationship | Value |
| Location | Entity | Value | Relationship | Value |
| Location | Entity | Value | Relationship | Value |
| Location | Entity | Value | Relationship | Value |
| Location | Entity | Value | Relationship | Value |
| Location | Entity | Value | Relationship | Value |
| Location | Entity | Value | Relationship | Value |
| Location | Entity | Value | Relationship | Value |
| Location | Entity | Value | Relationship | Value |

**Gambar 2.8** Tampilan Visualize Sistem

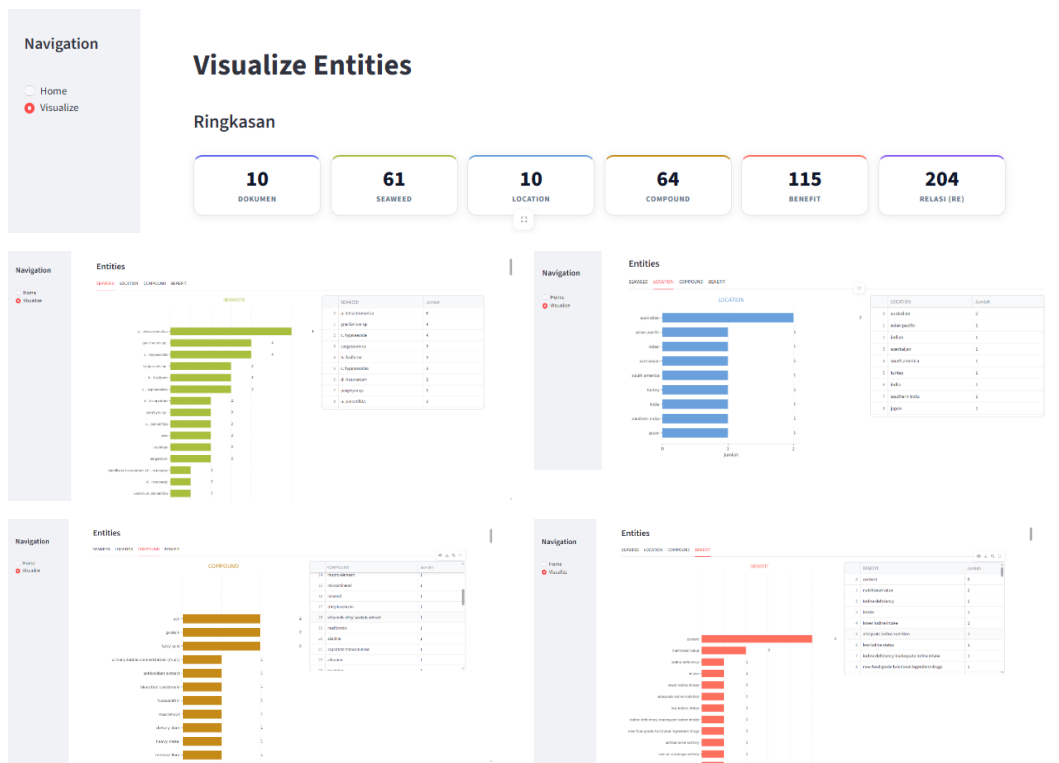
Sistem menampilkan hasil ekstraksi informasi dalam beberapa bentuk representasi, yaitu knowledge graph, tabel relasi, distribusi entitas, dan ringkasan statistik. Visualisasi tersebut digunakan untuk menyajikan hubungan antara entitas Seaweed, Location, Benefit, dan Compound yang diperoleh dari hasil ekstraksi relasi.



**Gambar 4.9** Visualisasi Knowledge Graph

| No  | Judul  | Entity 1 Type | Entity 1                    | Entity 2 Type | Entity 2                                              | Kalimat Relasi                                                                                       | Jumlah |
|-----|--------|---------------|-----------------------------|---------------|-------------------------------------------------------|------------------------------------------------------------------------------------------------------|--------|
| 81  | Data 9 | SEAWEED       | Campylaeaophora hypnaeoides | LOCATION      | Japan                                                 | Campylaeaophora hypnaeoides is reported or cultivated in Japan.                                      | 1      |
| 82  | Data 9 | SEAWEED       | (C. hypnaeoides)            | LOCATION      | Japan                                                 | (C. hypnaeoides) is reported or cultivated in Japan.                                                 | 1      |
| 83  | Data 9 | SEAWEED       | C. hypnaeoides              | LOCATION      | Japan                                                 | C. hypnaeoides is reported or cultivated in Japan.                                                   | 2      |
| 84  | Data 9 | SEAWEED       | C. hypnaeoides              | LOCATION      | Japan                                                 | C. hypnaeoides is reported or cultivated in Japan.                                                   | 2      |
| 85  | Data 9 | SEAWEED       | Campylaeaophora hypnaeoides | BENEFIT       | health promotion disease prevention                   | Campylaeaophora hypnaeoides is associated with health promotion disease prevention.                  | 1      |
| 86  | Data 9 | SEAWEED       | Campylaeaophora hypnaeoides | BENEFIT       | anti-inflammatory effect                              | Campylaeaophora hypnaeoides is associated with anti-inflammatory effect.                             | 1      |
| 87  | Data 9 | SEAWEED       | Campylaeaophora hypnaeoides | BENEFIT       | increase body weight white adipose tissue weight      | Campylaeaophora hypnaeoides is associated with increase body weight white adipose tissue weight.     | 1      |
| 88  | Data 9 | SEAWEED       | Campylaeaophora hypnaeoides | BENEFIT       | ameliorate insulin resistance , hyperglycemia , hepat | Campylaeaophora hypnaeoides is associated with ameliorate insulin resistance , hyperglycemia , hepat | 1      |
| 89  | Data 9 | SEAWEED       | Campylaeaophora hypnaeoides | BENEFIT       | hypercholesterolemia                                  | Campylaeaophora hypnaeoides is associated with hypercholesterolemia.                                 | 1      |
| 90  | Data 9 | SEAWEED       | (C. hypnaeoides)            | BENEFIT       | health promotion disease prevention                   | (C. hypnaeoides) is associated with health promotion disease prevention.                             | 1      |
| 91  | Data 9 | SEAWEED       | (C. hypnaeoides)            | BENEFIT       | anti-inflammatory effect                              | (C. hypnaeoides) is associated with anti-inflammatory effect.                                        | 1      |
| 92  | Data 9 | SEAWEED       | (C. hypnaeoides)            | BENEFIT       | increase body weight white adipose tissue weight      | (C. hypnaeoides) is associated with increase body weight white adipose tissue weight.                | 1      |
| 93  | Data 9 | SEAWEED       | (C. hypnaeoides)            | BENEFIT       | ameliorate insulin resistance , hyperglycemia , hepat | (C. hypnaeoides) is associated with ameliorate insulin resistance , hyperglycemia , hepat            | 1      |
| 94  | Data 9 | SEAWEED       | (C. hypnaeoides)            | BENEFIT       | hypercholesterolemia                                  | (C. hypnaeoides) is associated with hypercholesterolemia.                                            | 1      |
| 95  | Data 9 | SEAWEED       | C. hypnaeoides              | BENEFIT       | health promotion disease prevention                   | C. hypnaeoides is associated with health promotion disease prevention.                               | 2      |
| 96  | Data 9 | SEAWEED       | C. hypnaeoides              | BENEFIT       | anti-inflammatory effect                              | C. hypnaeoides is associated with anti-inflammatory effect.                                          | 2      |
| 97  | Data 9 | SEAWEED       | C. hypnaeoides              | BENEFIT       | increase body weight white adipose tissue weight      | C. hypnaeoides is associated with increase body weight white adipose tissue weight.                  | 2      |
| 98  | Data 9 | SEAWEED       | C. hypnaeoides              | BENEFIT       | ameliorate insulin resistance , hyperglycemia , hepat | C. hypnaeoides is associated with ameliorate insulin resistance , hyperglycemia , hepat              | 2      |
| 99  | Data 9 | SEAWEED       | C. hypnaeoides              | BENEFIT       | hypercholesterolemia                                  | C. hypnaeoides is associated with hypercholesterolemia.                                              | 2      |
| 100 | Data 9 | SEAWEED       | C. hypnaeoides              | BENEFIT       | health promotion disease prevention                   | C. hypnaeoides is associated with health promotion disease prevention.                               | 2      |
| 101 | Data 9 | SEAWEED       | C. hypnaeoides              | BENEFIT       | anti-inflammatory effect                              | C. hypnaeoides is associated with anti-inflammatory effect.                                          | 2      |

Gambar 4.10 Tabel Relasi Antarentitas



Gambar 4.11 Distribusi Entitas dan Ringkasan Statistik

Fungsionalitas sistem diuji menggunakan metode Black Box Testing. Pengujian dilakukan terhadap sembilan skenario yang mencakup proses unggah dokumen, pemrosesan dokumen, visualisasi hasil ekstraksi, penerapan filter relasi, dan ekspor data ke format CSV. Hasil pengujian disajikan pada Tabel 4.3

**Table 4.3** Hasil Black Box Testing

| No | Fitur               | Skenario Pengujian                 | Hasil yang Diharapkan               | Hasil Pengujian |
|----|---------------------|------------------------------------|-------------------------------------|-----------------|
| 1  | Upload Dokumen      | Mengunggah file CSV, PDF, atau TXT | Dokumen berhasil dimuat oleh sistem | Berhasil        |
| 2  | Entity Annotation   | Membuka hasil anotasi entitas      | Data entitas ditampilkan            | Berhasil        |
| 3  | Relation Annotation | Membuka hasil anotasi relasi       | Data relasi ditampilkan             | Berhasil        |
| 4  | RE Result           | Menampilkan hasil ekstraksi relasi | Hasil ekstraksi ditampilkan         | Berhasil        |
| 5  | Export CSV          | Mengekspor hasil ekstraksi         | File CSV berhasil diunduh           | Berhasil        |
| 6  | Statistics Summary  | Membuka halaman visualisasi        | Statistik ditampilkan               | Berhasil        |
| 7  | Knowledge Graph     | Menampilkan knowledge graph        | Graph berhasil ditampilkan          | Berhasil        |
| 8  | Relation Filtering  | Melakukan filter relasi            | Data sesuai filter ditampilkan      | Berhasil        |
| 9  | Relation Table      | Menampilkan tabel relasi           | Tabel relasi ditampilkan            | Berhasil        |

#### 4.5 Pembahasan

Pada penelitian ini, pengembangan model Relation Extraction dilakukan menggunakan tiga arsitektur Transformer, yaitu SciBERT, BioBERT, dan

PubMedBERT. Berdasarkan hasil evaluasi pada Tabel 4.2, model PubMedBERT memperoleh nilai akurasi sebesar 86,92% dan F1-score sebesar 87,02%. SciBERT memperoleh akurasi sebesar 85,56% dan F1-score sebesar 85,69%, sedangkan BioBERT memperoleh akurasi sebesar 84,88% dan F1-score sebesar 85,17%. Perbedaan nilai evaluasi antar model menunjukkan adanya variasi performa dalam tugas klasifikasi relasi antarentitas pada dataset literatur rumput laut yang digunakan dalam eksperimen.

Variasi tersebut berkaitan dengan perbedaan korpus pre-training masing-masing model. SciBERT dilatih menggunakan korpus ilmiah lintas domain yang mencakup berbagai bidang penelitian. BioBERT dilatih menggunakan korpus biomedis yang bersumber dari PubMed abstracts dan PubMed Central (PMC). PubMedBERT dilatih dari awal menggunakan korpus PubMed tanpa tambahan data PMC, sehingga distribusi data pre-training lebih terfokus pada abstrak biomedis. Selain metrik evaluasi, hasil pelatihan menunjukkan bahwa PubMedBERT menghasilkan training loss sebesar 0,294770 dan validation loss sebesar 0,403423. SciBERT memiliki training loss sebesar 0,412088 dan validation loss sebesar 0,413442, sedangkan BioBERT memiliki training loss sebesar 0,324958 dan validation loss sebesar 0,428929. Nilai-nilai tersebut merupakan hasil keluaran proses pelatihan pada konfigurasi eksperimen yang digunakan.

Model yang telah dilatih kemudian diintegrasikan ke dalam sistem visualisasi berbasis web menggunakan framework Streamlit. Integrasi ini memungkinkan proses inferensi model dilakukan pada dokumen yang diunggah pengguna. Hasil ekstraksi relasi kemudian ditampilkan dalam beberapa bentuk representasi. Representasi pertama adalah knowledge graph yang menggambarkan relasi antarentitas dalam bentuk graf. Setiap node merepresentasikan entitas seperti Seaweed, Location, Benefit, dan Compound, sedangkan edge merepresentasikan relasi yang dihasilkan oleh model Relation Extraction. Representasi lainnya mencakup tabel relasi, distribusi entitas, dan ringkasan statistik. Tabel relasi digunakan untuk menyajikan hasil ekstraksi dalam bentuk terstruktur.

Distribusi entitas menunjukkan jumlah kemunculan setiap jenis entitas pada dataset yang diproses. Ringkasan statistik menyajikan informasi agregat dari hasil ekstraksi yang dihasilkan sistem. Pengujian sistem dilakukan menggunakan metode Black Box Testing dengan sembilan skenario pengujian. Skenario tersebut mencakup proses unggah dokumen, pemrosesan teks, visualisasi hasil ekstraksi, penerapan filter relasi, dan ekspor data ke format CSV. Seluruh skenario pengujian menghasilkan keluaran sesuai dengan spesifikasi fungsional sistem. Secara keseluruhan, hasil eksperimen menunjukkan bahwa sistem mampu melakukan proses ekstraksi relasi dan visualisasi informasi berdasarkan model Relation Extraction yang digunakan, sesuai dengan desain sistem yang diimplementasikan. Kualitas hasil visualisasi bergantung pada keluaran model yang digunakan dalam proses inferensi.

#### **4.6 Pelajaran yang Diperoleh**

Pelaksanaan kegiatan MBKM di Badan Riset dan Inovasi Nasional (BRIN) memberikan pengalaman langsung dalam alur kerja riset berbasis Data Science dan Natural Language Processing (NLP) untuk domain literatur rumput laut. Melalui keterlibatan dalam tahapan anotasi entitas dan relasi menggunakan platform Label Studio, diperoleh pemahaman praktis mengenai sensitivitas kualitas data terhadap performa model ekstraksi informasi. Selain itu, proses eksperimen training dan evaluasi pada arsitektur Transformer—khususnya varian SciBERT, BioBERT, dan PubmedBERT—memberikan data komparatif mengenai kemampuan masing-masing model bahasa dalam mengenali entitas ilmiah serta mengekstraksi hubungan antar-entitas secara kuantitatif. Pengetahuan teknis tersebut kemudian diimplementasikan pada pengembangan sistem web berbasis Streamlit untuk menyajikan visualisasi data hasil ekstraksi dalam bentuk Knowledge Graph, distribusi entitas, serta tabel relasi yang dapat difilter sesuai kebutuhan analisis sistem.

Secara non-teknis, partisipasi dalam kelompok penelitian ini membentuk pola kerja yang terstruktur dan menuntut manajemen waktu yang ketat guna memenuhi target yang telah ditetapkan dalam timeline riset. Interaksi berkala

bersama pembimbing dan rekan tim selama proses bimbingan, pemaparan progres, serta fase pengujian sistem berfungsi sebagai sarana untuk melatih komunikasi ilmiah dan kemampuan menyelesaikan kendala teknis secara kolaboratif. Dinamika lingkungan riset nasional ini juga mendorong kemandirian belajar dalam mengidentifikasi serta menyelesaikan galat (error) teknis yang muncul selama proses pengkodean dan pelatihan model, yang secara langsung berkontribusi pada kesiapan sikap kerja di bidang penelitian dan pengembangan teknologi.

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Berdasarkan hasil penelitian dan pembahasan, kesimpulan yang diperoleh untuk menjawab rumusan masalah mencakup tiga capaian. Pertama, dataset anotasi entitas dan relasi dari literatur rumput laut telah dibangun untuk menjadi sumber data terstruktur pada tahap ekstraksi informasi. Kedua, pengembangan model Relation Extraction (RE) telah dieksekusi menggunakan arsitektur Transformer, di mana hasil pengujian kuantitatif menunjukkan bahwa model PubMedBERT menghasilkan tingkat akurasi sebesar 86,92% dan F1-Score sebesar 87,02% pada data pengujian. Model dengan matriks performa tersebut selanjutnya diintegrasikan ke dalam sistem ekstraksi informasi berbasis web. Ketiga, data hasil ekstraksi disajikan menggunakan sistem visualisasi yang mencakup representasi knowledge graph, tabel relasi, ringkasan statistik, dan distribusi entitas untuk memetakan hubungan spesifik antar entitas pada dokumen literatur rumput laut.

#### **5.2 Keterbatasan**

Pelaksanaan program MBKM di Badan Riset dan Inovasi Nasional (BRIN) menghadapi beberapa keterbatasan yang berkaitan dengan ketersediaan data, sumber daya komputasi, dan pengembangan sistem. Pada tahap pengumpulan data, jumlah literatur ilmiah mengenai rumput laut yang tersedia dalam format teks dan dapat digunakan untuk proses ekstraksi informasi memengaruhi jumlah data yang berhasil dikumpulkan. Selain itu, durasi program yang terbatas juga berpengaruh terhadap jumlah data yang dapat dianotasi untuk kebutuhan Named Entity Recognition (NER) dan Relation Extraction (RE) menggunakan platform Label Studio. Akibatnya, distribusi data hasil anotasi yang digunakan pada tahap pelatihan model belum sepenuhnya merata pada seluruh jenis entitas dan relasi yang ditetapkan dalam penelitian.



Pada tahap pengembangan model, eksperimen dilakukan menggunakan tiga arsitektur bahasa, yaitu SciBERT, BioBERT, dan PubMedBERT. Proses pelatihan model bergantung pada sumber daya komputasi yang tersedia selama kegiatan berlangsung, khususnya kapasitas memori dan unit pemrosesan grafis (GPU). Kondisi tersebut memengaruhi konfigurasi eksperimen yang dapat diterapkan, termasuk penentuan ukuran batch (batch size) dan jumlah epoch selama proses pelatihan. Dari sisi pengembangan sistem, aplikasi web berbasis Streamlit yang dibangun dalam kegiatan ini difokuskan pada penyajian hasil ekstraksi informasi dan visualisasi data.

Fitur yang diimplementasikan meliputi Statistics Summary, Entity Distribution, Knowledge Graph, dan Relation Table. Meskipun demikian, sistem belum mendukung alur pemrosesan terintegrasi yang memungkinkan pengguna mengunggah dokumen PDF dan memperoleh hasil ekstraksi secara otomatis melalui model RE. Selain kegiatan penelitian dan pengembangan sistem, pada tahap awal program juga dilakukan proses adaptasi terhadap perangkat anotasi yang digunakan serta pemahaman mengenai implementasi visualisasi hubungan antarentitas dalam bentuk Knowledge Graph. Kegiatan tersebut merupakan bagian dari tahapan persiapan yang diperlukan sebelum memasuki proses anotasi data, pelatihan model, dan pengujian sistem.

### **5.3 Implikasi**

Pelaksanaan kegiatan MBKM ini memberikan beberapa implikasi teknis maupun operasional bagi mahasiswa, instansi mitra, dan universitas. Bagi mahasiswa, keterlibatan dalam alur riset ini berimplikasi pada penguasaan praktis mengenai siklus pemrosesan data teks (Natural Language Processing), yang mencakup tahapan anotasi data menggunakan Label Studio, pelatihan model bahasa berbasis Transformer (SciBERT, BioBERT, dan PubmedBERT), hingga integrasi model ke dalam antarmuka web berbasis Streamlit. Aktivitas tersebut memberikan pemahaman metodologis mengenai pengaruh kualitas data terhadap performa model serta pemecahan kendala teknis dalam pengembangan sistem pengekstrakan informasi.

Bagi instansi mitra (BRIN), kontribusi berupa dataset hasil anotasi entitas (NER dataset) dan anotasi relasi menyediakan aset data latih tambahan yang terstruktur untuk domain literatur ilmiah rumput laut. Selain itu, purwarupa sistem visualisasi berbasis web yang mengintegrasikan komponen Knowledge Graph dan Relation Table dapat difungsikan sebagai infrastruktur awal atau pembanding bagi penelitian lanjutan terkait otomatisasi ekstraksi informasi di kelompok riset terkait. Bagi universitas, kegiatan ini mengimplikasikan pemenuhan capaian pembelajaran mata kuliah melalui mekanisme konversi program MBKM serta menyediakan dokumentasi studi kasus riset terapan yang dapat dijadikan referensi akademik bagi penelitian mahasiswa selanjutnya di bidang serupa.

#### **5.4 Saran**

Berdasarkan analisis terhadap proses riset dan keterbatasan sistem yang diidentifikasi selama pelaksanaan kegiatan, beberapa rekomendasi teknis diajukan untuk pengembangan berkelanjutan bagi mahasiswa, instansi mitra, dan pihak universitas. Bagi mahasiswa yang akan melakukan penelitian sejenis, disarankan untuk mengalokasikan waktu pemahaman sintaksis pengkodean visualisasi grafik hubungan (Knowledge Graph) dan pengenalan perangkat lunak anotasi sebelum linimasa riset inti dimulai, guna mengoptimalkan efisiensi waktu eksekusi pada tahapan pemodelan dan pengujian sistem.

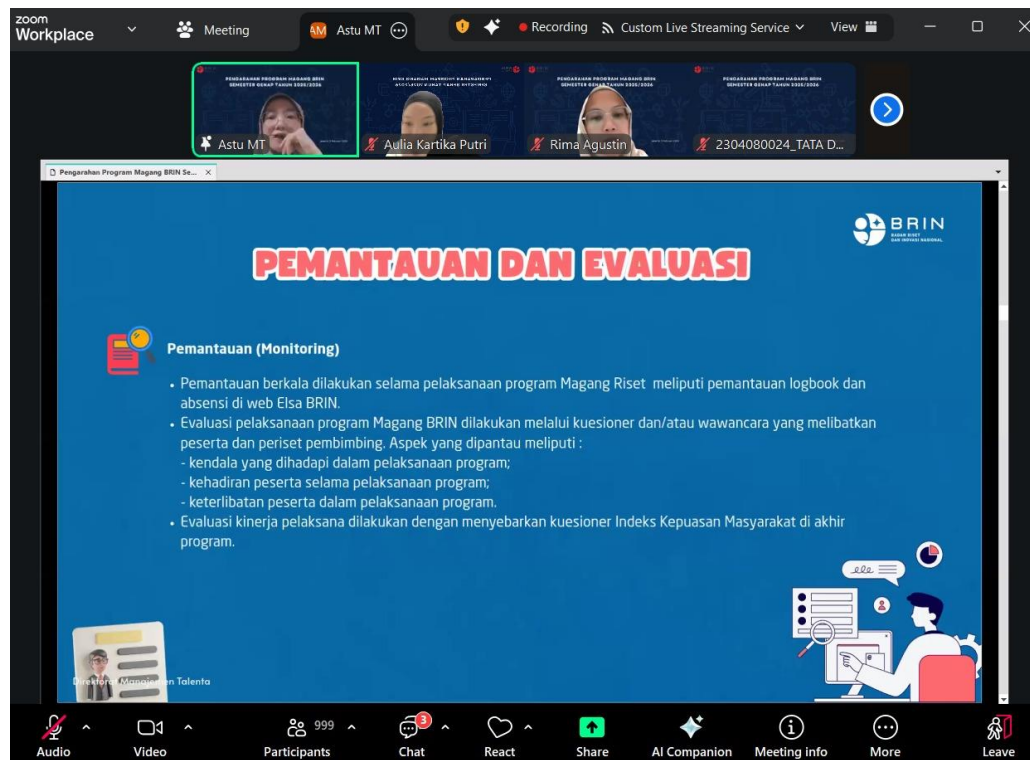
Bagi instansi mitra (BRIN), pengembangan riset selanjutnya disarankan untuk memperluas volume dataset melalui penambahan literatur ilmiah rumput laut secara berkala guna meningkatkan variasi kluster entitas pada fase pelatihan model Relation Extraction (RE). Selain itu, eksperimen pemodelan disarankan menggunakan infrastruktur komputasi dengan alokasi memori dan spesifikasi GPU yang lebih tinggi demi memungkinkan pengujian nilai hyperparameter pada ukuran batch (batch size) serta jumlah iterasi (epoch) yang lebih besar. Dari segi fungsionalitas aplikasi web Streamlit, direkomendasikan untuk beralih dari visualisasi data statis menuju integrasi sistem pemrosesan langsung (end-to-end pipelining) yang mampu menerima input dokumen PDF mentah dan mengekstraksinya secara real-time di latar belakang.

Bagi pihak universitas, direkomendasikan untuk mempertahankan dan memperluas jaringan kerja sama sela akademik dengan lembaga riset nasional seperti BRIN guna menyediakan lebih banyak variasi topik riset terapan bagi mahasiswa. Selain itu, universitas disarankan untuk menyusun dokumentasi repositori khusus yang mengintegrasikan hasil riset laporan MBKM berbasis Data Science ini sebagai referensi metodologis dan studi kasus akademik bagi penelitian mahasiswa pada masa mendatang.

## DAFTAR PUSTAKA

- [1] H. Zheng, S. Wang, and L. Huang, “A Comprehensive Survey on Document-Level Information Extraction,” pp. 58–72, 2024.
- [2] E. B. Boztepe, B. Karakaya, K. Bahadır, and İ. Ünlü, “An Approach for Audio-Visual Content Understanding of Video using Multimodal Deep Learning Methodology,” vol. 5, no. 2, 2022, doi: 10.35377/saucis.05.02.1139765.
- [3] X. Wang and R. E. Mercer, “Incorporating Figure Captions and Descriptive Text in MeSH Term Indexing,” pp. 165–175, 2019.
- [4] S. Zhao, R. Meng, D. He, S. Andi, and P. Bambang, “Integrating Transformer and Paraphrase Rules for Sentence Simplification,” pp. 3164–3173, 2018.
- [5] S. Jain, M. Van Zuylen, H. Hajishirzi, and I. Beltagy, “S CI REX : A Challenge Dataset for Document-Level Information Extraction,” pp. 7506–7516, 2020.
- [6] “BRIN. Badan Riset dan Inovasi Nasional. Diakses 10 Juni 2026. <https://www.brin.go.id/>.”

## LAMPIRAN



**Gambar 1.** Pertemuan Pengantar MBKM



**Gambar 2.** Serah Terima Mahasiswa MBKM



**Gambar 3.** Pertemuan Kelompok Riset Information Retrieval



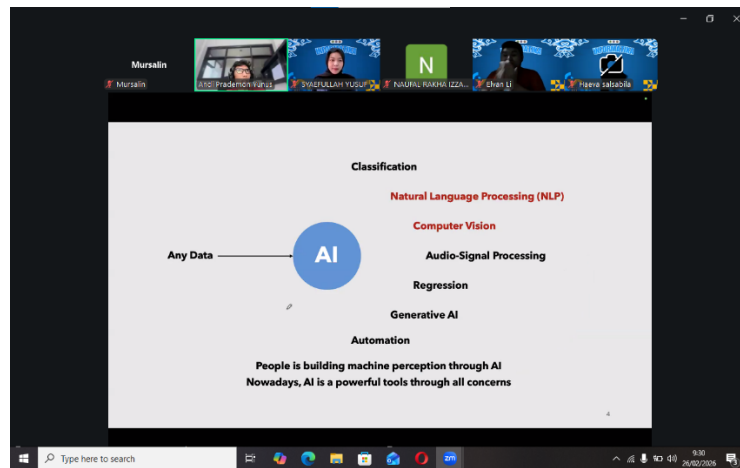
**Gambar 4.** Pertemuan Mingguan dengan Pembimbing



**Gambar 5.** Pertemuan Pekan Kelompok Riset Retrieval Information



**Gambar 6.** Mengikuti Kegiatan Webinar DESAIN #3 Dialog Eksplorasi Sains Data dan Informasi



**Gambar 7.** Mengikuti Kegiatan Webinar



**Gambar 8.** Pembelajaran Tentang *GIT*





**Gambar 9.** Pertemuan Mingguan dengan Pembimbing



**Gambar 10.** Laporan Progres Mingguan



**Gambar 11.** Monitoring Pihak Kampus ke BRIN