

LAPORAN PELAKSANAAN
MERDEKA BELAJAR KAMPUS MERDEKA (MBKM)
BADAN RISET DAN INOVASI NASIONAL
SEMESTER GENAP 2025/2026

RANCANG BANGUN SISTEM VISUALISASI
INFORMASI DAN HUBUNGAN ENTITAS
LITERATUR RUMPUT LAUT BERBASIS WEB

PENELITIAN/RISET

Disusun oleh:

Nama : Mursalin
NIM : 231730006
Prodi/Jurusan : Informatika
Fakultas : Fakultas Sains dan Teknologi
Universitas : UIN Sultan Maulana Hasanuddin Banten
Pembimbing : 1. Dr. Dipl. Ing. (FH) Anne Parlina, M.P
2. Ahmad Tabrani, M.T.I.

PUSAT RISET SAINS DATA DAN INFORMASI
BADAN RISET DAN INOVASI NASIONAL
BANDUNG

2026



BRIN

BADAN RISET
DAN INOVASI NASIONAL



**LEMBAR PENGESAHAN LAPORAN PELAKSANAAN
MERDEKA BELAJAR KAMPUS MERDEKA (MBKM)
BADAN RISET DAN INOVASI NASIONAL
SEMESTER GENAP 2025/2026**

**RANCANG BANGUN SISTEM VISUALISASI
INFORMASI DAN HUBUNGAN ENTITAS
LITERATUR RUMPUT LAUT BERBASIS WEB
PENELITIAN/RISET**

Disusun oleh:

Nama : Mursalin
NIM : 231730006
Prodi/Jurusan : Informatika
Fakultas : Fakultas Sains dan Teknologi
Universitas : UIN Sultan Maulana Hasanudiin Banten
Unit Kerja BRIN : Pusat Riset Sains Data dan Informasi

Bandung, 20 Juni 2026

**Menyetujui,
Pembimbing BRIN**

**Mengetahui,
Pembimbing Perguruan Tinggi**

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

Ahmad Tabrani, M.T.I
NIP. 198509172018011002

DAFTAR ISI

DAFTAR ISI.....	iii
DAFTAR TABEL.....	v
DAFTAR GAMBAR	vi
BAB I.....	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan Kegiatan MBKM	2
1.3 Sasaran Kegiatan MBKM	2
BAB II	3
PROFIL MITRA KEGIATAN MBKM	3
2.1 Sejarah Badan Riset dan Inovasi Nasional	3
2.2 Struktur Organisasi BadanRiset dan Inovasi Nasional	5
2.3 Visi dan Misi Tujuan dan Sasaran StrategisBadan Riset dan Inovasi Nasional	8
2.4 Profil Satuan Kerja Tempat Pelaksanaan MBKM	9
2.5 Profil Kelompok Penelitian/Fungsi Kerja Tempat Pelaksanaan MBKM.....	11
BAB III	13
PELAKSANAAN KEGIATAN MBKM	13
3.1 Uraian Umum Kegiatan MBKM yang Dilakukan	13
3.2 Keterlibatan Penelitian/Riset yang Dilakukan	13
3.2.1 <i>Literature Dataset</i>	15
3.2.2 <i>Entity Annotation</i>	15
3.2.3 <i>Named Entity Recognition (NER) Dataset</i>	16
3.2.4 <i>Relation Annotation</i>	17
3.2.5 <i>Relation Dataset</i>	17
3.2.6 <i>Dataset Splitting</i>	18
3.2.7 <i>Input Representation (Tokenizer)</i>	19
3.2.8 <i>Relation Extraction (RE) Model Training</i>	20
3.2.9 <i>Relation Extraction (RE) Evaluation</i>	20
a.) <i>SciBERT Evaluation</i>	21
b.) <i>BioBERT Evaluation</i>	22
c.) <i>PubmedBERT Evaluation</i>	25
3.2.10 <i>Best Relation Extraction (RE) Model</i>	28
3.2.11 <i>Entity and Relation Extraction Result</i>	28

3.2.12 Requirement Analysis	29
3.2.13 System Design	30
3.2.14 Web Development.....	30
a.) Home Page.....	31
b.) Visualize Page	31
3.2.15 Information Visualization.....	32
a.) Statistics Summary	33
b.) Entity Distribution	34
c.) Knowledge Graph	35
d.) Relation Filtering.....	35
e.) Relation Table.....	36
3.2.16 System Testing.....	36
3.3 Penugasan-Penugasan yang Telah Diselesaikan.....	37
3.4 Pembelajaran Hal Baru (<i>Lessons Learned</i>) yang Diperoleh	40
BAB IV	43
PENUTUP	43
4.1 Kesimpulan	43
4.2 Saran	44
DAFTAR PUSTAKA	45
LAMPIRAN.....	46
Biodata Peserta MBKM	46
Biodata Pembimbing BRIN	46
Biodata Pembimbing Universitas.....	47
Dokumentasi Kegiatan MBKM	48
Luaran Kegiatan MBKM	52

DAFTAR TABEL

Tabel 3.1 Karakteristik Data	15
Tabel 3.2 Kategori Entitas	16
Tabel 3.3 Kategori Relasi yang Digunakan	17
Tabel 3.4 Distribusi Relasi Hasil Anotasi	18
Tabel 3.5 Relasi yang Digunakan untuk Training RE	18
Tabel 3.6 Hasil Training SciBERT	22
Tabel 3.7 Hasil Training BioBERT	24
Tabel 3.8 Hasil Training PubmedBERT	26
Tabel 3.9 Hasil Perbandingan Model	28
Tabel 3.10 Hasil Analisis Kebutuhan Sistem	29
Tabel 3.11 Black Box Testing	36

DAFTAR GAMBAR

Gambar 2.1 Logo Badan Riset dan Inovasi Nasional (BRIN)	3
Gambar 2.2 Struktur Organisasi BRIN	6
Gambar 2.3 Profil Pusat Riset Sains Data dan Informasi	10
Gambar 2.4 Alur Kerja Pusat Riset Sains Data dan Informasi	10
Gambar 2.5 Profil Kelompok Riset Information Retrieval (Pusat Riset Sains Data dan Informasi)	11
Gambar 3.1 Alur Penelitian Pengembangan Model Ekstraksi Informasi dari Literatur Rumput Laut	14
Gambar 3.2 Alur Penelitian Pengembangan Sistem Visualisasi Informasi dan Hubungan Entitas Literatur Rumput Laut Berbasis Web.....	15
Gambar 3.3 Dataset NER.....	16
Gambar 3.4 Arsitektur SciBERT.....	21
Gambar 3.5 Grafik Loss dan Accuracy Model SciBERT	22
Gambar 3.6 Confusion Matrix Model SciBERT	22
Gambar 3.7 Arsitektur BioBERT	23
Gambar 3.8 Grafik Loss dan Accuracy Model BioBERT	24
Gambar 3.9 Confusion Matrix Model BioBERT.....	25
Gambar 3.10 Arsitektur PubmedBERT.....	26
Gambar 3.11 Grafik Loss dan Accuracy Model PubmedBERT	27
Gambar 3.12 Confusion Matrix Model PubmedBERT	27
Gambar 3.13 Dataset Ekstraksi NER dan RE.....	29
Gambar 3.14 Hasil Perancangan Sistem.....	30
Gambar 3.15 Tampilan Home Page	31
Gambar 3.17 Tampilan Visualize Page	32
Gambar 3.18 Statistics Summary	33
Gambar 3.19 Entity Distribution	34
Gambar 3.20 Knowledge Graph.....	35
Gambar 3.21 Relation Filtering.....	35
Gambar 3.22 Relation Table	36
Gambar 3.23 Sertifikat Pelatihan Pengantar MBKM BRIN	38
Gambar 3.24 Daftar Materi Pelatihan Pengantar MBKM BRIN	38

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penelitian memegang peran krusial dalam kemajuan ilmu pengetahuan dan teknologi serta merupakan aspek tak terpisahkan dalam upaya meningkatkan daya saing dan kemandirian suatu bangsa. Untuk mencapai tujuan tersebut, peningkatan kapasitas penelitian di tingkat nasional menjadi suatu kebutuhan yang mendesak. Dalam hal ini, peningkatan kualitas Sumber Daya Manusia (SDM) di bidang penelitian merupakan faktor utama, karena SDM yang berkualitas berpotensi besar untuk mendorong kemajuan riset serta pengembangan ilmu pengetahuan dan teknologi.

Sebagai pihak yang berperan penting dalam pengembangan riset dan inovasi, pemerintah telah mengambil langkah strategis dengan meluncurkan program Merdeka Belajar Kampus Merdeka (MBKM). Program ini memberikan kesempatan luas bagi mahasiswa untuk mendapatkan pengalaman belajar yang lebih mendalam, serta memperoleh keterampilan dan kompetensi baru yang relevan dengan perkembangan ilmu pengetahuan dan teknologi.

Dalam konteks ini, Badan Riset dan Inovasi Nasional (BRIN) berperan sentral sebagai lembaga pemerintah yang memimpin dan mengkoordinasikan aktivitas riset dan inovasi. BRIN secara aktif mengelola dan mengarahkan kegiatan riset dan inovasi melalui pelaksanaan Program MBKM BRIN. Program ini mencakup dua skema utama, yaitu program magang/praktik kerja serta program penelitian/riset. Dengan melibatkan berbagai unit kerja di bawah naungan BRIN, termasuk unit kerja penelitian dan administrative.

Program MBKM BRIN bertujuan lebih dari sekadar memberikan pengalaman belajar kepada mahasiswa. Tujuan utama dari Program MBKM BRIN adalah mempersiapkan mahasiswa untuk memasuki dunia kerja, khususnya dalam bidang riset dan inovasi. Melalui proses mentoring riset di lingkungan BRIN, program ini bertujuan menghasilkan SDM yang berkualitas yang dapat mendukung perkembangan ekosistem riset dan inovasi di Indonesia. Dengan demikian,

langkah-langkah konkret ini diharapkan dapat menciptakan lingkungan riset yang produktif, mendorong inovasi, dan mendukung perkembangan potensi intelektual bangsa.

1.2 Tujuan Kegiatan MBKM

Program MBKM BRIN adalah program yang dicanangkan oleh Mendikbud dan memiliki beberapa tujuan sebagai berikut.

1. Meningkatkan kompetensi lulusan Perguruan Tinggi baik *soft skills* maupun *hard skills*.
2. Menyiapkan lulusan Perguruan Tinggi sebagai pemimpin masa depan bangsa yang unggul dan berkepribadian.
3. Menyiapkan SDM unggul dan berkualitas dalam rangka menumbuhkan ekosistem riset dan inovasi di Indonesia.
4. Memberikan pengalaman bekerja secara langsung bagi mahasiswa untuk mengembangkan potensinya.

1.3 Sasaran Kegiatan MBKM

Program MBKM BRIN terbuka untuk mahasiswa aktif paling rendah berada pada semester 6 (enam) dari seluruh Perguruan Tinggi, baik Negeri maupun Swasta. Mahasiswa yang lulus verifikasi administrasi oleh Direktorat Manajemen Talenta dan lulus seleksi substansi oleh Kelompok/Fungsi Kegiatan dapat mengikuti program MBKM BRIN.

BAB II

PROFIL MITRA KEGIATAN MBKM

2.1 Sejarah Badan Riset dan Inovasi Nasional

BRIN merupakan lembaga pemerintah yang berada di bawah dan bertanggung jawab kepada Presiden dalam menyelenggarakan, penelitian, pengembangan, dan penerapan, serta invensi dan inovasi, penyelenggaraan ketenaganukliran dan penyelenggaraan keantariksaan yang terintegrasi. Pengertian ini berdasar pada Peraturan Presiden Nomor 78 tahun 2021. Berdasar peraturan ini pula Presiden Joko Widodo mengangkat Laksana Tri Handoko sebagai Kepala Badan Riset dan Inovasi Nasional di Istana Negara pada tanggal 28 April 2021 yang kemudian dituangkan pada Keputusan Presiden Republik Indonesia (Keppres) Nomor 19/M Tahun 2021 tanggal 28 April 2021 tentang Pengangkatan Kepala BRIN. Logo Badan Riset dan Inovasi Nasional ditunjukkan pada **Gambar 2.1**.



Gambar 2.1 Logo Badan Riset dan Inovasi Nasional (BRIN)

Badan Riset dan Inovasi Nasional (BRIN) pertama kali dibentuk oleh Presiden Joko Widodo pada tahun 2019 untuk memenuhi Undang-Undang Nomor 11 tahun 2019 Tentang Sistem Nasional Ilmu Pengetahuan dan Teknologi (UU SISNAS IPTEK). Awalnya Lembaga ini menjadi satu dengan Kementerian Riset dan Teknologi (Kemenristek), namun pada 5 Mei 2021, Presiden Joko Widodo menandatangani Peraturan Presiden Nomor 33 Tahun 2021, yang secara efektif menetapkan BRIN sebagai satu-satunya badan penelitian nasional. Peraturan ini membuat semua badan penelitian nasional Indonesia melebur menjadi satu, badan-badan penelitian itu meliputi Lembaga Ilmu Pengetahuan Indonesia (LIPI), Badan Pengkajian dan Penerapan Teknologi (BPPT), Badan Tenaga Nuklir Nasional (BATAN), dan Lembaga Penerbangan dan Antariksa Nasional (LAPAN), hasil

peleburan badan-badan penelitian yang merupakan BRIN.

Dalam menjalankan tugas yang tertera pada Peraturan Presiden Nomor 78 Tahun 2021 BRIN menyelenggarakan 14 fungsi sebagai berikut.

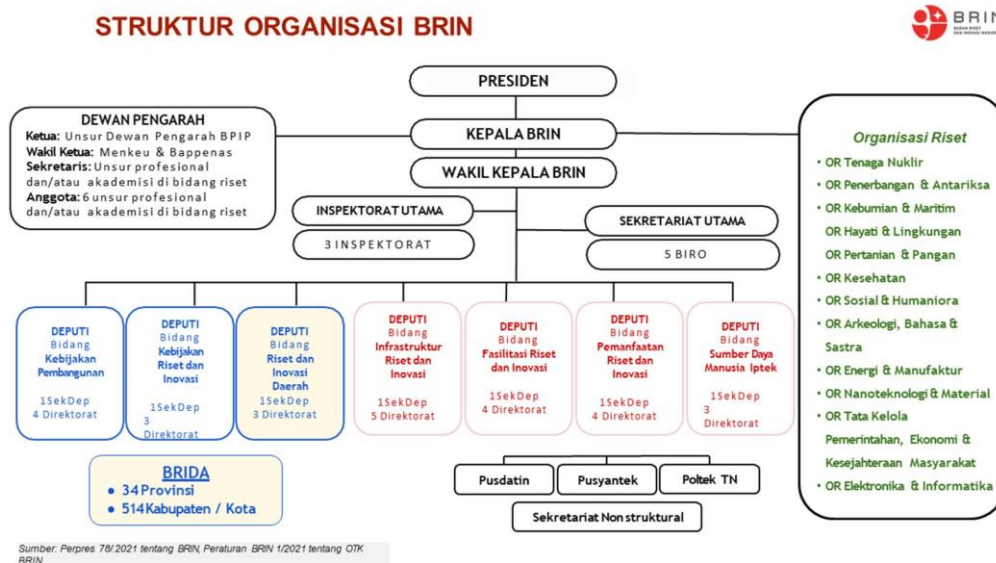
1. Pelaksanaan penelitian, pengembangan, pengkajian, dan penerapan serta invensi dan inovasi dalam rangka penyusunan rekomendasi perencanaan pembangunan nasional berdasarkan hasil kajian ilmiah dengan berpedoman pada nilai Pancasila;
2. Perumusan dan penetapan kebijakan di bidang riset dan inovasi yang meliputi rencana induk pemajuan ilmu pengetahuan dan teknologi, dan peta jalan penelitian, pengembangan, pengkajian, penerapan, serta invensi dan inovasi, penyelenggaraan ketenaganukliran, dan penyelenggaraan keantariksaan;
3. Perumusan, penetapan, dan pelaksanaan kebijakan di bidang pembinaan, pengembangan kompetensi, pengembangan profesi, manajemen talenta, dan pengawasan dan pengendalian sumber daya manusia ilmu pengetahuan dan teknologi, infrastruktur riset dan inovasi, fasilitas riset dan inovasi, dan pemanfaatan riset dan inovasi;
4. Pengintegrasian sistem penyusunan perencanaan, program, anggaran, kelembagaan, dan sumber daya penelitian, pengembangan, pengkajian, dan penerapan, invensi dan inovasi, penyelenggaraan ketenaganukliran, dan penyelenggaraan keantariksaan;
5. Penyelenggaraan penelitian, pengembangan, pengkajian, dan penerapan, serta invensi dan inovasi, penyelenggaraan ketenaganukliran, dan penyelenggaraan keantariksaan;
6. Pengawasan dan pengendalian penelitian, pengembangan, pengkajian, dan penerapan, serta invensi dan inovasi, penyelenggaraan ketenaganukliran, dan penyelenggaraan keantariksaan secara menyeluruh dan berkelanjutan;
7. Pelaksanaan koordinasi pengabdian kepada masyarakat berbasis penelitian, pengembangan, pengkajian, dan penerapan, serta invensi dan inovasi yang dihasilkan oleh kelembagaan ilmu pengetahuan dan teknologi;
8. Pelaksanaan pembangunan, pengelolaan, dan pengembangan sistem informasi penelitian, pengembangan, pengkajian, dan penerapan, serta invensi dan

inovasi, penyelenggaraan ketenaganukliran, dan penyelenggaraan keantariksaan;

9. Pelaksanaan penelitian, pengembangan, invensi, dan inovasi kebijakan yang mengakui, menghormati, mengembangkan dan melestarikan keanekaragaman pengetahuan tradisional, kearifan lokal, sumber daya alam hayati dan nurhayati, serta budaya sebagai bagian dari identitas bangsa;
10. Pemberian fasilitasi, bimbingan teknis, pembinaan, dan supervisi serta pemantauan dan evaluasi di bidang penelitian, pengembangan, pengkajian, dan penerapan, serta invensi dan inovasi, penyelenggaraan ketenaganukliran, dan penyelenggaraan keantariksaan;
11. Pemantauan, pengendalian, dan evaluasi terhadap pelaksanaan tugas dan fungsi BRIDA;
12. Pembinaan dan pemberian dukungan administrasi dan teknis kepada seluruh unsur organisasi di lingkungan BRIN;
13. Pengawasan atas pelaksanaan tugas di lingkungan BRIN; dan
14. Pelaksanaan fungsi lain yang diberikan oleh Presiden.

2.2 Struktur Organisasi Badan Riset dan Inovasi Nasional

Berdasarkan Peraturan Badan Riset dan Inovasi Nasional Nomor 1 Tahun 2021, berikut merupakan struktur organisasi di lingkungan BRIN. Struktur organisasi BRIN dapat dilihat pada **Gambar 2.2**



Gambar 2.2 Struktur Organisasi BRIN

Berdasarkan Pusat Pengembangan Kompetensi BRIN (2021), tugas dan fungsi dari para pelaksana di lingkungan BRIN antara lain :

1. Kepala BRIN
 - a. Memimpin dan bertanggung jawab atas pelaksanaan tugas dan fungsi BRIN.
 - b. Dalam melaksanakan tugas dan fungsi BRIN, kepala memperhatikan arahan dari dewan pengawas.
 - c. Kepala merupakan pengguna anggaran / pengguna barang.
2. Wakil Kepala BRIN
 - a. Membantu kepala dalam perumusan kebijakan dan atau pelaksanaan kebijakan BRIN.
 - b. Membantu kepala dalam mengkoordinasikan pencapaian kebijakan strategis lintas unit organisasi jabatan pimpinan tinggi madya dan Organisasi Riset (OR) di lingkungan BRIN.
3. Sekretariat Utama
 - a. Pengintegrasian sistem penyusunan perencanaan, program, kelembagaan, dan sumber daya penelitian, pengembangan, pengkajian, penerapan, serta

invensi dan inovasi, penyelenggaraan ketenaganukliran dan penyelenggaraan keantariksaan.

- b. Pembinaan dan pemberian dukungan administrasi, ketatausahaan, sumber daya manusia, keuangan kerumahtanggaan, arsip dan dokumentasi BRIN.
 - c. Pengelolaan hubungan masyarakat, keprotokolan, dan pengamanan.
 - d. Pelaksanaan kerjasama dalam rangka optimalisasi tugas dan fungsi BRIN.
 - e. Pembinaan dan penataan organisasi dan tata laksana.
 - f. Pelaksanaan reformasi birokrasi di lingkungan BRIN.
 - g. Pelaksanaan dan pengoordinasian penyusunan produk hukum dan advokasi hukum.
 - h. Penyelenggaraan pengelolaan barang milik/kekayaan negara dan pengelolaan pengadaan barang/jasa.
 - i. Manajemen aparatur sipil negara BRIN.
4. Inspektorat Utama
- Bertugas menyelenggarakan pengawasan intern di lingkungan BRIN.
5. Deputy Bidang Kebijakan Pembangunan
- Bertugas menyelenggarakan pelaksanaan penelitian, pengembangan, pengkajian, penerapan, serta invensi dan inovasi dalam rangka penyusunan rekomendasi kebijakan arah dan prinsip rancangan perencanaan pembangunan nasional dengan berpedoman pada nilai Pancasila.
6. Deputy Bidang Kebijakan Riset dan Inovasi
- Bertugas menyelenggarakan perumusan dan penetapan kebijakan di bidang riset dan inovasi meliputi rencana induk memajukan ilmu pengetahuan dan teknologi, serta peta jalan penelitian, pengembangan, pengkajian, penerapan, serta invensi dan inovasi penyelenggaraan ketenaganukliran dan penyelenggaraan keantariksaan dengan berpedoman pada nilai Pancasila.
7. Deputy Bidang Sumber Daya Manusia Ilmu Pengetahuan dan Teknologi
- Memiliki tugas menyelenggarakan perumusan dan pelaksanaan kebijakan di bidang pembinaan, pengembangan kompetensi, pengembangan profesi, manajemen talenta, serta pengawasan dan pengendalian sumber daya manusia ilmu pengetahuan teknologi.

8. **Deputi Bidang Infrastruktur Riset dan Inovasi**
Bertugas menyelenggarakan perumusan dan pelaksanaan kebijakan di bidang infrastruktur riset dan inovasi.
9. **Deputi Bidang Fasilitasi Riset dan Inovasi**
Memiliki tugas menyelenggarakan perumusan dan pelaksanaan kebijakan di bidang fasilitasi riset dan inovasi.
10. **Deputi Bidang Pemanfaatan Riset dan Inovasi**
Bertugas menyelenggarakan perumusan dan pelaksanaan kebijakan di bidang pemanfaatan riset dan inovasi.
11. **Deputi Bidang Riset dan Inovasi Daerah**
Bertugas menyelenggarakan koordinasi dan pengendalian penelitian, pengembangan, pengkajian, serta invensi dan inovasi BRIDA.
12. **Pusat Pelayanan Teknologi**
Bertugas melaksanakan manajemen pemasaran, manajemen proyek, manajemen kontrak dan lisensi, serta manajemen keuangan dan tata usaha Pusat Pelayanan teknologi.
13. **Pusat Data dan Informasi**
Melaksanakan penyiapan penyusunan kebijakan teknis, pelaksanaan, pemanfaatan, evaluasi, dan pelaporan di bidang data.

2.3 Visi dan Misi Tujuan dan Sasaran Strategis Badan Riset dan Inovasi

Nasional

Badan Riset dan Inovasi Nasional menetapkan visi yaitu “Badan Riset dan Inovasi Nasional yang andal, professional, inovatif dan berintegritas dalam pelayanan kepada Presiden dan Wakil Presiden untuk mewujudkan Visi dan Misi Presiden dan Wakil Presiden : Indonesia Maju yang Berdaulat, Mandiri, dan Berkepribadian berlandaskan Gotong Royong”. Sebagai upaya untuk mewujudkan visi tersebut, misi Badan Riset dan Inovasi Nasional adalah sebagai berikut.

- a. Peningkatan Kualitas Manusia Indonesia,
- b. Peningkatan Struktur Ekonomi yang Produktif, Mandiri, dan Berdaya Saing,
- c. Pembangunan yang Merata dan Berkeadilan,

- d. Mencapai Lingkungan Hidup yang Berkelanjutan,
- e. Kemajuan Budaya yang Mencerminkan Kepribadian Bangsa, dan
- f. Pengelolaan Pemerintahan yang Bersih, Efektif, dan Terpercaya.

Uraian misi tersebut adalah sebagai berikut.

1. Peningkatan Kapabilitas IPTEK, Budaya Riset, dan Penciptaan Inovasi melalui peningkatan Kualitas SDM IPTEK, Penguatan Transformasi Ekonomi, dan Pembangunan Berkelanjutan berlandaskan Budaya Iptek untuk Peningkatan Daya Saing.

2. Peningkatan Pengelolaan Pemerintahan yang Bersih, Efektif, dan Terpercaya.

Misi ini mencakup upaya menjawab permasalahan pembangunan iptek dan pada periode 2020-2024 dalam aspek kebijakan riset dan inovasi, kerja sama pembangunan dan kemitraan, peningkatan penelitian, pengembangan, pengkajian dan penerapan Iptek pada beberapa fokus prioritas riset dan inovasi nasional, serta peningkatan tata kelola pemerintahan yang baik dalam rangka reformasi birokrasi.

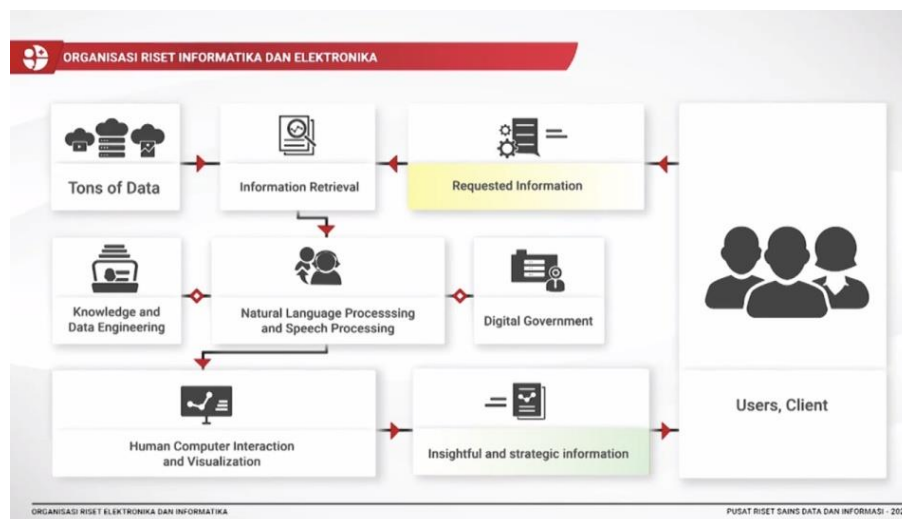
2.4 Profil Satuan Kerja Tempat Pelaksanaan MBKM

Pusat Riset Sains Data dan Informasi berada di bawah Organisasi Riset Elektronika dan Informatika (OREI) yang dipimpin oleh Dr. Esa Prakasa M.T. Terdapat enam Pusat Riset (PR) di bawah OREI antara lain: PR Telekomunikasi, PR Elektronika, PR Sains Data dan Informasi, PR Kecerdasan Artifisial dan Keamanan Siber, PR Komputasi, dan PR Mekatronika Cerdas.

Pusat Riset Sains Data dan Informasi (PR SDI) dibentuk serta disahkan melalui Peraturan Badan Riset dan Inovasi Nasional Republik Indonesia Nomor 10 Tahun 2022 tentang Tugas, Fungsi, Dan Struktur Organisasi Riset Elektronika dan Informatika, pada tanggal 25 Februari 2022. Adapun profil pusat riset sains data dan informasi beserta alur kerjanya dapat dilihat pada **Gambar 2.3.** dan **Gambar 2.4.**



Gambar 2.3 Profil Pusat Riset Sains Data dan Informasi

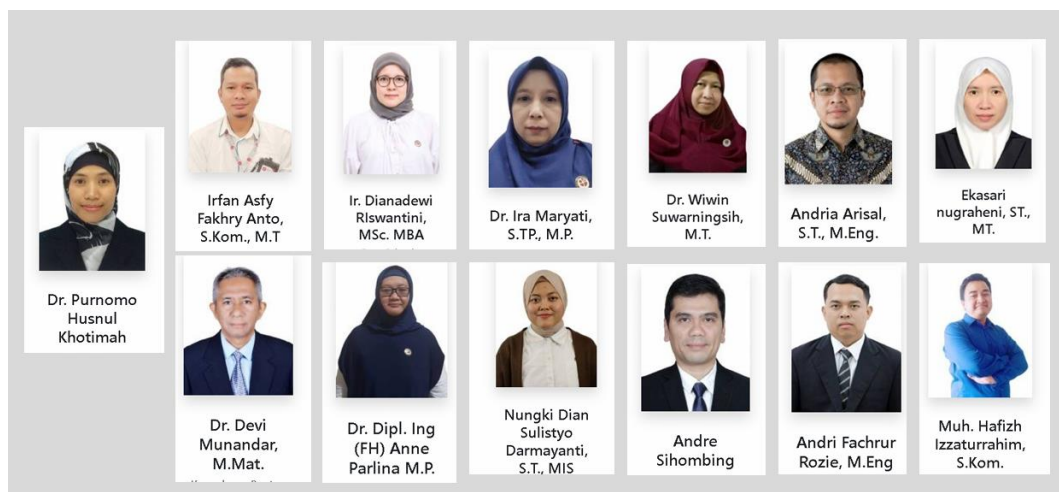


Gambar 2.4 Alur Kerja Pusat Riset Sains Data dan Informasi

Secara khusus, Unit Kerja Pusat Riset Sains Data dan Informasi (PR SDI) mempunyai tugas melaksanakan tugas teknis penelitian, pengembangan, pengkajian, dan penerapan, serta invensi dan inovasi di bidang sains data dan informasi sesuai dengan Pasal 14. Selain itu, dalam melaksanakan tugas sebagaimana dimaksud dalam pasal 14, Pusat Riset Sains Data dan Informasi menyelenggarakan fungsi yang diatur dalam pasal 15 sebagai berikut.

1. Pelaksanaan tugas teknis penelitian, pengembangan, pengkajian, dan penerapan, serta invensi dan inovasi di bidang sains data dan informasi.
2. Penyiapan bahan rekomendasi ilmiah atau tanggapan ilmiah di bidang sains data dan informasi.
3. Pemberian bimbingan teknis dan supervisi di bidang sains data dan informasi.
4. Pelaksanaan kerja sama di bidang sains data dan informasi.
5. Pemantauan, evaluasi, dan pelaporan di bidang sains data dan informasi.

2.5 Profil Kelompok Penelitian/Fungsi Kerja Tempat Pelaksanaan MBKM



Gambar 2.5 Profil Kelompok Riset Information Retrieval (Pusat Riset Sains Data dan Informasi)

Kelompok riset Pengambilan Informasi (*Information Retrieval*) merupakan salah satu kelompok penelitian di Pusat Riset Sains Data dan Informasi. Kelompok penelitian ini diketuai oleh Dr. Purnomo Husnul Khotimah, M.T. dengan tiga belas anggota, yaitu, Dr. Dipl. Ing (FH) Anne Parlina, M.P. ; Irfan Asfy Fakhry Anto, S.Kom., M.T.; Ir. Dianadewi Rlswantini, MSc. MBA; Dr. Ira Maryati, S.TP., M.P.; Dr. Wiwin Suwarningsih, M.T.; Andria Arisal, S.T., M.Eng.; Ekasari Nugraheni, S.T., M.T.; Dr. Devi Munandar, M.Mat.; Nungki Dian Sulisty Darmayanti, S.T., MIS; Andre Sihombing; Andri Fachrur Rozie, M.Eng.; serta Muh. Hafizh Izzaturrahim, S.Kom. Adapun ruang lingkup penelitian adalah sebagai berikut:

1. *Data mining (text extraction, entity recognition, web extraction)*
2. *Data processing (natural language processing, topic/event detection, text/pattern tagging, data integration)*
3. *Data analytics (network graph analysis, natural language analysis, market analysis, medical report analysis)*
4. *Data visualization (information visualization)*

Tentang IRRG (*Information Retrieval Research Group*): Ledakan internet, media sosial, dan teknologi berbasis awan telah mendorong pengembangan model sistem informasi baru dan menciptakan sumber informasi yang sangat besar yang dapat bermanfaat untuk memberikan solusi terhadap permasalahan kemanusiaan dan lingkungan yang muncul. IRRG berperan penting dalam menyediakan kerangka kerja untuk memahami dan mengatasi tantangan informasi ini. Proses menemukan, mengambil, dan mengorganisir informasi yang relevan dari sumber informasi baru ini adalah fokus utama penelitian dan kegiatan IRRG. Ini melibatkan identifikasi data yang diperlukan, pencarian untuk mendapatkannya, dan kemudian mengekstrak serta mengorganisirnya sehingga menjadi bermanfaat, berarti, dan dapat diambil tindakan bagi para pemangku kepentingan. Dengan demikian, IRRG berperan sebagai garda terdepan dalam mengembangkan solusi informatif yang dapat membantu dalam peningkatan kualitas kehidupan manusia dan menjaga keberlanjutan lingkungan.

BAB III

PELAKSANAAN KEGIATAN MBKM

3.1 Uraian Umum Kegiatan MBKM yang Dilakukan

Program MBKM Penelitian/Riset yang dilakukan mahasiswa meliputi kegiatan penelitian, pengembangan, pengkajian, dan penerapan ilmu pengetahuan dan teknologi. Program ini diselenggarakan oleh Direktorat Manajemen Talenta Badan Riset dan Inovasi Nasional yang berlangsung selama 6 (enam) bulan. Kegiatan ini dimulai pada tanggal 18 Februari 2026 hingga 30 Juli 2026 serta dibagi menjadi 2 tahap kegiatan yaitu pelatihan pengantar selama 1 (satu) minggu dan proses pembimbingan selama 24 minggu.

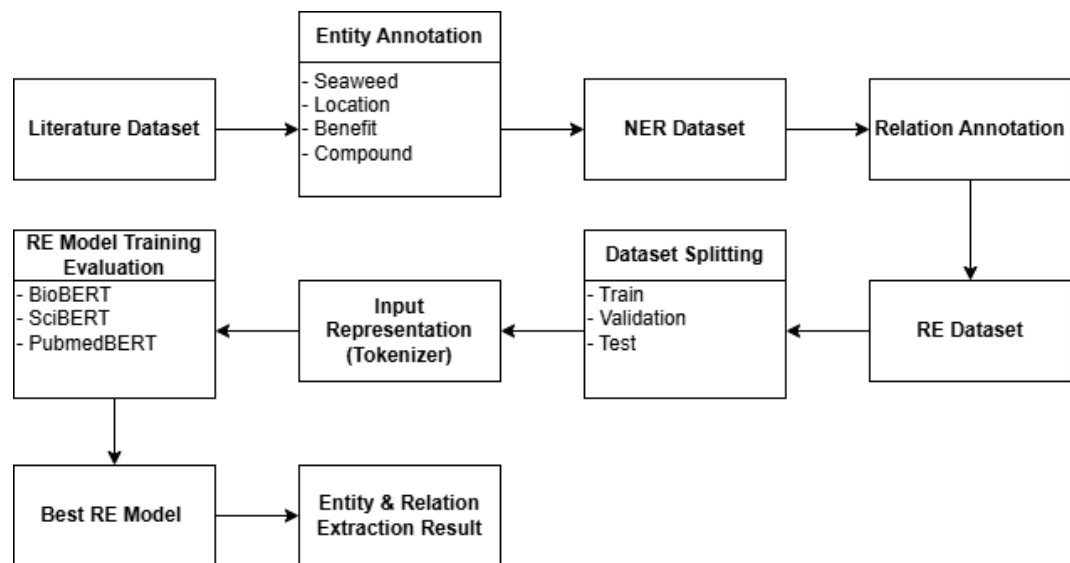
Pelatihan pengantar MBKM bertujuan agar mahasiswa memiliki pengetahuan mengenai *building learning commitment*, BRIN sebagai lembaga riset berkelas dunia, pengantar *research integrity*, membangun tim efektif, komunikasi persuasive, dan teknik presentasi yang baik. Setelah mengikuti pelatihan pengantar MBKM, mahasiswa mengikuti kegiatan pembimbingan berupa aktivitas sesuai bentuk kegiatan pembelajaran yang dipilih, pada lokasi, dan unit kerja yang telah ditetapkan, serta sesuai dengan tugas dan arahan pembimbing. Mahasiswa berkewajiban mengisi daftar hadir, mengisi laporan harian (*logbook*), melaksanakan program MBKM dengan bertanggung jawab, memelihara barang milik negara yang digunakan, dan mengikuti tata tertib yang berlaku di lingkungan BRIN. Disisi lain, mahasiswa berhak menggunakan peralatan kerja yang disediakan, mendapat pelatihan pengantar dan pembimbingan, mendapatkan penilaian program MBKM dari pembimbing, dan mendapatkan sertifikat setelah menyelesaikan Program MBKM.

3.2 Keterlibatan Penelitian/Riset yang Dilakukan

Kegiatan penelitian/riset yang saya lakukan selama program Magang MBKM di Badan Riset dan Inovasi Nasional (BRIN) terbagi menjadi dua fokus utama, yaitu pengembangan model ekstraksi informasi pada literatur rumput laut dan pengembangan sistem visualisasi informasi serta hubungan entitas berbasis web.

Kedua kegiatan tersebut merupakan bagian dari penelitian yang berjudul “Rancang Bangun Sistem Visualisasi Informasi dan Hubungan Entitas Literatur Rumput Laut Berbasis Web.”

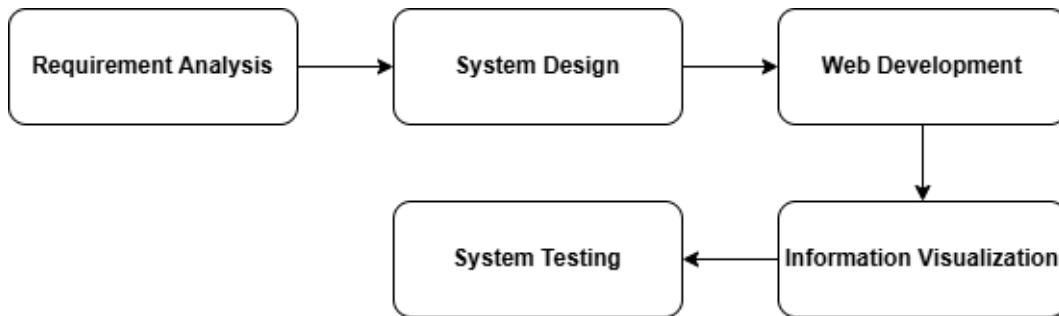
Fokus penelitian pertama adalah pengembangan model ekstraksi informasi dari literatur rumput laut. Pada kegiatan ini saya terlibat dalam proses anotasi entitas dan anotasi relasi menggunakan Label Studio. Entitas yang digunakan dalam penelitian terdiri atas *Seaweed*, *Location*, *Benefit*, dan *Compound*. Dataset hasil anotasi kemudian digunakan untuk pengembangan model *Named Entity Recognition (NER)* dan *Relation Extraction (RE)* berbasis transformer guna mengidentifikasi entitas serta hubungan antar entitas yang terdapat dalam literatur rumput laut. Tahapan penelitian yang dilakukan ditunjukkan pada **Gambar 3.1**.



Gambar 3.1 Alur Penelitian Pengembangan Model Ekstraksi Informasi dari Literatur Rumput Laut

Kegiatan Penelitian/riset kedua berfokus pada pengembangan sistem visualisasi informasi dan hubungan entitas literatur rumput laut berbasis web. Penelitian ini bertujuan untuk menyajikan hasil ekstraksi informasi dalam bentuk visual yang interaktif sehingga memudahkan pengguna dalam memahami hubungan antarentitas yang terdapat pada literatur rumput laut. Tahapan penelitian meliputi analisis kebutuhan sistem, perancangan sistem, implementasi aplikasi berbasis web menggunakan Streamlit, serta pengujian sistem. Alur penelitian ini

ditunjukkan pada **Gambar 3.2**.



Gambar 3.2 Alur Penelitian Pengembangan Sistem Visualisasi Informasi dan Hubungan Entitas Literatur Rumput Laut Berbasis Web.

3.2.1 Literature Dataset

Penelitian ini menggunakan dataset literatur rumput laut yang diperoleh dari Badan Riset dan Inovasi Nasional (BRIN), Pusat Riset Sains Data dan Informasi. Dataset tersebut merupakan kumpulan literatur ilmiah yang bersumber dari basis data Scopus dan digunakan sebagai sumber data dalam proses ekstraksi informasi. Data yang diperoleh menjadi dasar dalam identifikasi entitas dan relasi untuk pembentukan dataset *Named Entity Recognition (NER)* dan *Relation Extraction (RE)*.

Tabel 3.1 Karakteristik Data

Karakteristik	Keterangan
Sumber Data	Badan Riset dan Inovasi Nasional (BRIN), Pusat Riset Sains Data dan Informasi
Basis Data	Scopus
Domain	Literatur Rumput Laut
Jumlah Dokumen	1.122 Dokumen
Format Data	CSV
Bahasa	Bahasa Inggris
Konten Data	Metadata dan Abstrak Artikel
Tujuan Penggunaan	Named Entity Recognition (NER) dan Relation Extraction (RE)

3.2.2 Entity Annotation

Pada tahap ini dilakukan proses anotasi entitas pada dataset literatur rumput laut. Anotasi bertujuan untuk memberikan label pada informasi penting yang terdapat dalam teks sesuai dengan kategori entitas yang telah ditentukan. Dalam penelitian ini, entitas yang digunakan meliputi *Seaweed*, *Location*, *Benefit*, dan *Compound*. Hasil anotasi digunakan sebagai dasar dalam pembentukan dataset

Named Entity Recognition (NER).

Table 3.2 Kategori Entitas

Entitas	Deskripsi
Seaweed	Nama spesies atau jenis rumput laut
Location	Lokasi penelitian atau asal rumput laut
Benefit	Manfaat atau aktivitas biologis rumput laut
Compound	Senyawa bioaktif yang terkandung dalam rumput laut

3.2.3 *Named Entity Recognition (NER) Dataset*

Beberapa penelitian terdahulu telah mencoba mengintegrasikan ekstraksi informasi dari media video. Sebagai contoh, penelitian oleh Boztepe et al., (2022) menggunakan *Named Entity Recognition (NER)* untuk memetakan kategori emosi dari teks transkrip video. Namun, fokus utama mereka masih terbatas pada konotasi kategori lingkungan, sehingga terdapat ruang pengembangan yang luas untuk menerapkan metode *Relation Extraction (RE)* yang lebih mendalam pada interaksi antar-entitas di bidang spesifik lainnya.

Untuk mengatasi keterbatasan cakupan domain yang ada, penelitian ini mengembangkan dataset *Named Entity Recognition (NER)* khusus untuk literatur rumput laut yang mencakup empat kategori entitas, yaitu Seaweed, Location, Benefit, dan Compound. Proses anotasi yang dilakukan memungkinkan transformasi informasi dari teks tidak terstruktur menjadi data terstruktur. Dataset yang dihasilkan berperan sebagai dasar dalam identifikasi entitas sekaligus mendukung pengembangan model *Relation Extraction (RE)* untuk mengungkap hubungan antar entitas pada literatur rumput laut. **Gambar 3.3.**

O	= 184180
B-BENEFIT	= 5571
B-COMPOUND	= 6583
B-LOCATION	= 1118
B-SEAWEED	= 3293
I-BENEFIT	= 17355
I-COMPOUND	= 4289
I-LOCATION	= 241
I-SEAWEED	= 3555

Gambar 3.3 Dataset NER

3.2.4 Relation Annotation

Proses anotasi relasi dilakukan untuk mengidentifikasi hubungan semantik antarentitas yang terdapat dalam literatur rumput laut. Anotasi dilakukan terhadap pasangan entitas yang telah dihasilkan pada tahap *Named Entity Recognition (NER)* sehingga hubungan antarentitas dapat direpresentasikan secara terstruktur.

Penerapan pemrosesan bertahap dari pengenalan entitas menuju ekstraksi relasi informasi ini mengadopsi pendekatan sekuensial yang serupa dengan penelitian oleh Do et al., (2024), di mana lokalisasi informasi kunci baru dilakukan setelah entitas baris teks dikenali.

Dalam penelitian ini, relasi yang dianotasi meliputi hubungan antara *Seaweed* dan *Location*, *Seaweed* dan *Benefit*, *Seaweed* dan *Compound*, serta *Compound* dan *Benefit*. Hasil anotasi relasi digunakan sebagai dasar dalam pembentukan dataset *Relation Extraction (RE)* untuk proses pelatihan dan evaluasi model.

Table 3.3 Kategori Relasi yang Digunakan

Relasi	Deskripsi
Seaweed–Location	Menunjukkan lokasi asal, habitat, atau lokasi penelitian suatu rumput laut
Seaweed–Benefit	Menunjukkan manfaat atau aktivitas biologis yang dimiliki oleh rumput laut
Seaweed–Compound	Menunjukkan senyawa yang terkandung dalam rumput laut
Compound–Benefit	Menunjukkan manfaat atau aktivitas biologis yang dihasilkan oleh suatu senyawa

3.2.5 Relation Dataset

Dataset *Relation Extraction (RE)* disusun berdasarkan hasil anotasi relasi yang dilakukan pada literatur rumput laut. Dataset ini memuat pasangan entitas beserta label relasi yang merepresentasikan hubungan antarentitas dalam teks. Hasil proses anotasi menghasilkan berbagai kategori relasi dengan distribusi yang bervariasi, sebagaimana disajikan pada **Table 3.4**.

Table 3.4 Distribusi Relasi Hasil Anotasi

Relasi	Jumlah
Seaweed-Compound	4100
Compound-Benefit	3213
Seaweed-Benefit	2068
Seaweed-Location	786
Compound-Compound	40
Benefit-Compound	38
Seaweed-Seaweed	18
Benefit-Benefit	12
Compound-Seaweed	8
Benefit-Location	1
Total Relasi	10284

Meskipun proses anotasi menghasilkan berbagai kategori relasi, penelitian ini hanya memanfaatkan relasi yang sesuai dengan tujuan ekstraksi informasi yang telah ditetapkan. Oleh karena itu, pembentukan dataset RE difokuskan pada empat kategori relasi utama, yaitu hubungan antara *Seaweed* dan *Location*, *Seaweed* dan *Benefit*, *Seaweed* dan *Compound*, serta *Compound* dan *Benefit*. Keempat kategori relasi tersebut dipilih karena merepresentasikan informasi yang paling relevan dalam kajian literatur rumput laut.

Table 3.5 Relasi yang Digunakan untuk Training RE

Relasi	Jumlah
Seaweed-Location	786
Seaweed-Benefit	2068
Seaweed-Compound	4100
Compound-Benefit	3213
Total Relasi Digunakan	10167

3.2.6 Dataset Splitting

Dataset splitting merupakan tahap pembagian dataset *Relation Extraction (RE)* menjadi beberapa subset agar proses pelatihan dan evaluasi model dapat dilakukan secara terpisah dan objektif. Tahap ini dilakukan setelah dataset RE terbentuk dan sebelum proses tokenisasi. Pada penelitian ini, dataset dibagi menggunakan metode *random 2-way split* dengan fungsi *train_test_split* dari pustaka *Hugging Face Datasets*. Pembagian dilakukan dengan rasio 90:10, yaitu 90% untuk data pelatihan (*training set*) dan 10% untuk data pengujian (*test/evaluation set*). Data pelatihan

berjumlah 17.195 sampel dan digunakan untuk melatih model dalam mempelajari pola hubungan antara entitas [E1] dan [E2] berdasarkan 5 kategori kelas relasi yang telah ditentukan. Sementara itu, data pengujian berjumlah 1.911 sampel dan digunakan untuk mengevaluasi performa model secara objektif, tanpa ikut serta dalam proses pembaruan bobot model. Dengan adanya pembagian ini, model dapat diuji pada data yang belum pernah dilihat sebelumnya sehingga mampu mengukur kemampuan generalisasi model secara lebih akurat serta menghindari terjadinya overfitting. Tahap ini menghasilkan dua subset data, yaitu *train_dataset* dan *test_dataset*, yang kemudian digunakan pada tahap berikutnya yaitu *Input Representation (Tokenizer)*.

3.2.7 Input Representation (Tokenizer)

Tahap *Input Representation (tokenizer)* merupakan proses konversi data teks menjadi representasi numerik yang dapat diproses oleh model berbasis transformer seperti SciBERT, BioBERT, dan PubMedBERT. Pada tahap ini, kalimat yang telah diberi penanda entitas [E1] dan [E2] diubah menjadi bentuk token dan selanjutnya dipetakan ke dalam indeks numerik berdasarkan vocabulary masing-masing model. Proses tokenisasi menghasilkan tiga komponen utama, yaitu *input_ids* yang merepresentasikan urutan token dalam bentuk angka, *attention_mask* yang menunjukkan token aktif (nilai 1) dan token padding (nilai 0), serta token khusus model transformer seperti [CLS] sebagai representasi awal urutan untuk tugas klasifikasi dan [SEP] sebagai penanda batas kalimat atau segmen.

Tokenisasi diterapkan pada dataset hasil pembagian (dataset splitting), yaitu 17.195 data *training* dan 1.911 data testing, dengan konfigurasi parameter *max_length* = 256, *truncation* = *True*, dan *padding* = "*max_length*". Parameter *max_length* membatasi panjang urutan token, *truncation* digunakan untuk memotong data yang melebihi batas tersebut, sedangkan *padding* memastikan seluruh input memiliki panjang yang seragam. Hasil dari tahap ini adalah dataset dalam bentuk tensor numerik yang telah terstandarisasi dimensinya, sehingga dapat digunakan langsung sebagai input pada tahap *RE Model Training* dan *Evaluation*.

3.2.8 Relation Extraction (RE) Model Training

Tahap pelatihan model *Relation Extraction (RE)* merupakan proses untuk mengoptimalkan arsitektur berbasis transformer domain spesifik, yaitu SciBERT, BioBERT, dan PubMedBERT. Tujuan tahap ini adalah mempelajari pola hubungan semantik antar-entitas berdasarkan representasi tensor hasil tokenisasi, yaitu *input_ids* dan *attention_mask*. Proses transformasi teks mentah menjadi representasi indeks sparse beserta pengelolaan seluruh arsitektur dasar model ini mengadopsi kerangka kerja standar dari Wolf et al., (2020).

Pelatihan menggunakan pendekatan *supervised learning* dengan konfigurasi hyperparameter sebagai berikut: ukuran batch (batch size) 16, *learning rate* sebesar 2×10^{-5} , *weight decay* sebesar 0.15, serta jumlah epoch maksimum sebanyak 5. Model dilatih menggunakan data training sebanyak 17.195 sampel. Pada setiap epoch, data diproses melalui forward propagation untuk menghasilkan prediksi kelas relasi, kemudian dihitung nilai loss berdasarkan perbedaan dengan label aktual. Selanjutnya dilakukan backward propagation untuk memperbaiki bobot (*weights*) model menggunakan algoritma optimasi. Untuk mengurangi risiko overfitting, digunakan mekanisme *early stopping* dengan parameter *patience* = 1. Pada setiap akhir epoch, model dievaluasi menggunakan data sebanyak 1.911 sampel untuk menghitung nilai *validation loss*. Apabila terjadi peningkatan *validation loss*, proses pelatihan dihentikan dan bobot terbaik model dipulihkan melalui parameter *load_best_model_at_end=True*.

3.2.9 Relation Extraction (RE) Evaluation

Tahap evaluasi model *Relation Extraction (RE)* dilakukan untuk mengukur kinerja dan kemampuan generalisasi model SciBERT, BioBERT, dan PubMedBERT dalam menyelesaikan tugas klasifikasi multikelas (5 kelas). Evaluasi menggunakan data pengujian (*test set*) sebanyak 1.911 sampel yang tidak digunakan selama proses pelatihan. Kinerja model diukur secara kuantitatif menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-score (weighted average)* untuk menilai kemampuan klasifikasi hubungan antar-entitas. Selain itu, digunakan *confusion matrix* berukuran 5x5 untuk menganalisis distribusi hasil prediksi secara detail, yang mencakup tingkat ketepatan prediksi pada setiap

kategori (diagonal utama matriks) serta pola kesalahan klasifikasi (misclassification) antar-kelas relasi yang diuji.

a.) *SciBERT Evaluation*

Model SciBERT (*allenai/scibert_scivocab_uncased*) yang digunakan dikonfigurasi menggunakan arsitektur *BertForSequenceClassification* untuk tugas klasifikasi multikelas (5 kelas). Seluruh parameter model bersifat trainable dengan total sebanyak 109.922.309 parameter.

```
[INFO] Arsitektur SciBERT:
=====
Layer (type:depth-idx)                   Output Shape           Param #
=====
BertForSequenceClassification            [8, 5]                 --
├─BertModel: 1-1                         [8, 768]               --
│   ├─BertEmbeddings: 2-1                [8, 256, 768]          --
│   │   └─Embedding: 3-1                 [8, 256, 768]          23,877,120
│   │   └─Embedding: 3-2                 [8, 256, 768]          1,536
│   │   └─Embedding: 3-3                 [1, 256, 768]          393,216
│   │   └─LayerNorm: 3-4                 [8, 256, 768]          1,536
│   │   └─Dropout: 3-5                  [8, 256, 768]          --
│   └─BertEncoder: 2-2                  [8, 256, 768]          --
│       └─ModuleList: 3-6                --
│           └─BertPooler: 2-3            [8, 768]               --
│               └─Linear: 3-7             [8, 768]               590,592
│                   └─Tanh: 3-8           [8, 768]               --
├─Dropout: 1-2                           [8, 768]               --
└─Linear: 1-3                             [8, 5]                 3,845
=====
Total params: 109,922,309
Trainable params: 109,922,309
Non-trainable params: 0
Total mult-adds (Units.MEGABYTES): 876.63
=====
Input size (MB): 0.02
Forward/backward pass size (MB): 1700.32
Params size (MB): 439.69
Estimated Total Size (MB): 2140.02
=====
```

Gambar 3.4 Arsitektur SciBERT

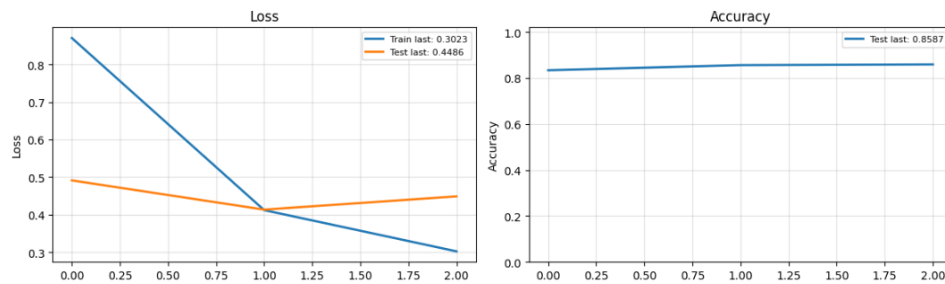
Evaluasi model dilakukan menggunakan 1.911 sampel data uji yang tidak digunakan selama proses pelatihan. Hasil pengujian menunjukkan performa model berdasarkan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-score* sebagaimana ditunjukkan pada Tabel 3.6.

Table 3.6 Hasil Training SciBERT

Epoch	Training Loss	Validation Loss	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	0.870326	0.491176	83.36	83.64	83.36	83.43
2	0.412088	0.413442	85.56	86.21	85.56	85.69
3	0.302321	0.448630	85.87	87.29	85.87	86.08

Berdasarkan Tabel 3.6, nilai *training loss* mengalami penurunan secara bertahap dari 0.870326 pada epoch pertama menjadi 0.302321 pada epoch ketiga. Penurunan tersebut menunjukkan bahwa model semakin mampu mempelajari pola hubungan antar entitas yang terdapat pada data pelatihan. Sementara itu, nilai

validation loss mencapai titik minimum pada epoch kedua sebesar 0.413442 sebelum meningkat menjadi 0.448630 pada epoch ketiga. Pola ini mengindikasikan bahwa peningkatan kemampuan model terhadap data pelatihan tidak lagi diikuti oleh peningkatan performa pada data validasi. Dengan demikian, epoch kedua dapat dianggap sebagai titik optimal pembelajaran model karena menghasilkan nilai *validation loss* terendah yang mencerminkan kemampuan generalisasi terbaik terhadap data yang tidak digunakan selama proses pelatihan.



Gambar 3.5 Grafik Loss dan Accuracy Model SciBERT

True Label	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 3.6 Confusion Matrix Model SciBERT

Berdasarkan nilai *validation loss*, model terbaik diperoleh pada epoch kedua dan dipilih sebagai model akhir melalui mekanisme *early stopping*. Model tersebut menghasilkan *accuracy* sebesar 85,56%, *precision* sebesar 86,21%, *recall* sebesar 85,56%, dan F1-score sebesar 85,69%. Hasil *confusion matrix* menunjukkan bahwa mayoritas sampel berhasil diprediksi pada kelas yang benar, yang ditunjukkan oleh dominasi nilai pada diagonal utama matriks. Temuan ini menunjukkan bahwa

model mampu mengenali dan membedakan lima kelas relasi yang digunakan dalam penelitian dengan tingkat ketepatan yang baik.

b.) *BioBERT Evaluation*

Model BioBERT (dmis-lab/biobert-base-cased-v1.1) yang digunakan dikonfigurasi menggunakan arsitektur *BertForSequenceClassification* untuk tugas klasifikasi multikelas (5 kelas). Seluruh parameter model bersifat trainable dengan total sebanyak 108.314.117 parameter. Perbedaan jumlah parameter dibandingkan SciBERT disebabkan oleh variasi ukuran vocabulary pada masing-masing model. BioBERT menggunakan struktur vocabulary berbasis *BERT-base-cased* yang bersifat sensitif terhadap huruf besar dan kecil (cased), sehingga menghasilkan perbedaan dimensi pada matriks embedding.

Layer (type:depth-idx)	Output Shape	Param #
BertForSequenceClassification	[8, 5]	--
└BertModel: 1-1	[8, 768]	--
└┬BertEmbeddings: 2-1	[8, 256, 768]	--
└└┬Embedding: 3-1	[8, 256, 768]	22,268,928
└└┬Embedding: 3-2	[8, 256, 768]	1,536
└└┬Embedding: 3-3	[1, 256, 768]	393,216
└└┬LayerNorm: 3-4	[8, 256, 768]	1,536
└└┬Dropout: 3-5	[8, 256, 768]	--
└┬BertEncoder: 2-2	[8, 256, 768]	--
└└┬ModuleList: 3-6	--	85,054,464
└┬BertPooler: 2-3	[8, 768]	--
└└┬Linear: 3-7	[8, 768]	590,592
└└┬Tanh: 3-8	[8, 768]	--
└Dropout: 1-2	[8, 768]	--
└Linear: 1-3	[8, 5]	3,845
=====		
Total params: 108,314,117		
Trainable params: 108,314,117		
Non-trainable params: 0		
Total mult-adds (Units.MEGABYTES): 863.76		
=====		
Input size (MB): 0.02		
Forward/backward pass size (MB): 1700.32		
Params size (MB): 433.26		
Estimated Total Size (MB): 2133.59		

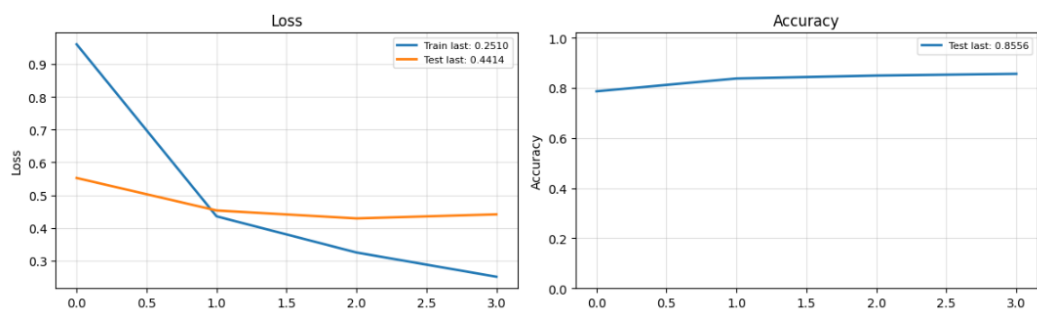
Gambar 3.7 Arsitektur BioBERT

Evaluasi model dilakukan menggunakan 1.911 sampel data uji yang tidak digunakan selama proses pelatihan. Hasil pengujian menunjukkan performa model berdasarkan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-score* sebagaimana ditunjukkan pada Tabel 3.7.

Table 3.7 Hasil Training BioBERT

Epoch	Training Loss	Validation Loss	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	0.960425	0.552405	78.60	79.53	78.60	78.02
2	0.435343	0.453248	83.73	84.86	83.73	83.99
3	0.324958	0.428929	84.88	86.30	84.88	85.17
4	0.250992	0.441361	85.56	86.29	85.56	85.69

Berdasarkan Tabel 3.7, Nilai *training loss* menunjukkan tren penurunan yang konsisten dari epoch pertama hingga epoch keempat, yaitu dari 0.960425 menjadi 0.250992. Hasil tersebut menunjukkan bahwa proses optimasi berhasil meningkatkan kemampuan model dalam mempelajari pola relasi pada data pelatihan. Di sisi lain, *validation loss* mencapai nilai terendah pada epoch ketiga sebesar 0.428929 sebelum meningkat pada epoch keempat menjadi 0.441361. Peningkatan *validation loss* setelah titik minimum tersebut mengindikasikan mulai terjadinya penurunan kemampuan generalisasi model terhadap data validasi. Berdasarkan kriteria pemilihan model menggunakan *validation loss*, bobot model pada epoch ketiga dipilih sebagai model terbaik dan digunakan pada tahap evaluasi. Model tersebut memperoleh *accuracy* sebesar 84,88%, *precision* sebesar 86,30%, *recall* sebesar 84,88%, dan *F1-score* sebesar 85,17%. Nilai tersebut menunjukkan bahwa model mampu mengidentifikasi relasi antar entitas dengan tingkat ketepatan yang relatif baik..

**Gambar 3.8** Grafik Loss dan Accuracy Model BioBERT

True Label	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 3.9 Confusion Matrix Model BioBERT

Berdasarkan kurva evaluasi akhir, model BioBERT mencatatkan nilai *accuracy* sebesar 0.8488 (84,88%) dan *F1-score* sebesar 0.8517, yang merepresentasikan tingkat ketepatan prediksi arsitektur ini terhadap keseluruhan data pengujian. Selanjutnya, visualisasi *confusion matrix* berukuran 5x5 menunjukkan bahwa sebagian besar prediksi terkonsentrasi pada diagonal utama matriks. Pola distribusi ini mengonfirmasi dominasi kapitalisasi prediksi yang benar secara spesifik pada masing-masing dari kelima kelas relasi (*No Relation*, *Seaweed-Loc*, *Seaweed-Ben*, *Seaweed-Comp*, dan *Comp-Ben*) serta meminimalkan akumulasi kesalahan klasifikasi (*misclassification*) antar-kategori relasi yang diuji.

c.) *PubmedBERT Evaluation*

Model PubMedBERT (*microsoft/BiomedNLP-PubMedBERT-base-uncased-abstract-fulltext*) yang digunakan dikonfigurasi menggunakan arsitektur *BertForSequenceClassification* untuk tugas klasifikasi multikelas (5 kelas). Seluruh parameter model bersifat trainable dengan total sebanyak 109.486.085 parameter. Perbedaan jumlah parameter dibandingkan dengan SciBERT dan BioBERT disebabkan oleh variasi ukuran *vocabulary* serta penyesuaian dimensi pada output layer masing-masing arsitektur. PubMedBERT menggunakan struktur *vocabulary* yang di-pretrain dari awal (*from scratch*) memanfaatkan artikel ilmiah PubMed yang bersifat *uncased* (tidak sensitif terhadap huruf besar dan kecil), sehingga menghasilkan perbedaan dimensi pada matriks embedding.

Layer (type:depth-idx)	Output Shape	Param #
BertForSequenceClassification	[8, 5]	--
└BertModel: 1-1	[8, 768]	--
└BertEmbeddings: 2-1	[8, 256, 768]	--
└Embedding: 3-1	[8, 256, 768]	23,440,896
└Embedding: 3-2	[8, 256, 768]	1,536
└Embedding: 3-3	[1, 256, 768]	393,216
└LayerNorm: 3-4	[8, 256, 768]	1,536
└Dropout: 3-5	[8, 256, 768]	--
└BertEncoder: 2-2	[8, 256, 768]	--
└ModuleList: 3-6	--	85,054,464
└BertPooler: 2-3	[8, 768]	--
└Linear: 3-7	[8, 768]	590,592
└Tanh: 3-8	[8, 768]	--
└Dropout: 1-2	[8, 768]	--
└Linear: 1-3	[8, 5]	3,845

Total params: 109,486,085
 Trainable params: 109,486,085
 Non-trainable params: 0
 Total mult-adds (Units.MEGABYTES): 873.14

Input size (MB): 0.02
 Forward/backward pass size (MB): 1700.32
 Params size (MB): 437.94
 Estimated Total Size (MB): 2138.28

Gambar 3.10 Arsitektur PubmedBERT

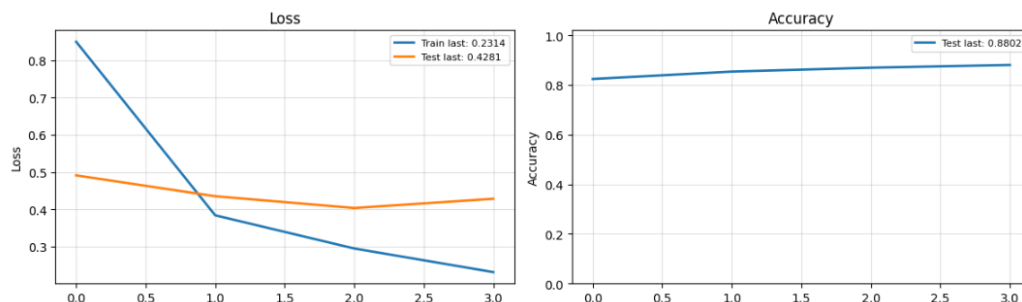
Evaluasi model dilakukan menggunakan 1.911 sampel data uji yang tidak digunakan selama proses pelatihan. Hasil pengujian menunjukkan performa model berdasarkan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-score* sebagaimana ditunjukkan pada Tabel 3.8.

Table 3.8 Hasil Training Model PubmedBERT

Epoch	Training Loss	Validation Loss	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	0.849350	0.490908	82.37	82.38	82.37	81.86
2	0.383749	0.434934	85.35	86.75	85.35	85.56
3	0.294770	0.403423	86.92	87.64	86.92	87.02
4	0.231449	0.428107	88.02	88.46	88.02	88.07

Berdasarkan Tabel 3.8, Berdasarkan hasil pelatihan, nilai training loss PubMedBERT menunjukkan tren penurunan yang konsisten dari epoch 1 hingga epoch 4 (0.849350 menjadi 0.231449). Hal ini merepresentasikan kapabilitas model yang semakin adaptif dalam mengoptimalkan parameter terhadap data pelatihan. Di sisi lain, validation loss mengalami penyusutan hingga mencapai titik minimum pada epoch 3 (0.403423), sebelum akhirnya kembali meningkat pada epoch 4 (0.428107). Peningkatan nilai validation loss pada epoch 4 tersebut menjadi indikasi awal terjadinya overfitting, yang secara otomatis mengaktifasi mekanisme early stopping guna menghentikan iterasi pelatihan. Meskipun kalkulasi metrik

evaluasi (*accuracy* dan *F1-score*) secara matematis mencatatkan angka yang lebih tinggi pada epoch 4, titik balik minimum pada *validation loss* menegaskan bahwa kapabilitas generalisasi model PubMedBERT mencapai titik paling optimal pada epoch 3. Sebagai implikasinya, sistem mempertahankan bobot model terbaik dari epoch 3 sebagai arsitektur final, yang menghasilkan capaian *accuracy* akhir sebesar 0.8692 (86,92%) dan *F1-score* sebesar 0.8702.



Gambar 3.11 Grafik Loss dan Accuracy Model PubmedBERT

True Label	No Relation	Seaweed-Loc	Seaweed-Ben	Seaweed-Comp	Comp-Ben	
	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 3.12 Confusion Matrix Model PubmedBERT

Berdasarkan hasil evaluasi akhir, model PubMedBERT menghasilkan *accuracy* sebesar 86,92%, *precision* sebesar 87,64%, *recall* sebesar 86,92%, dan *F1-score* sebesar 87,02%, sehingga menjadi model dengan performa terbaik dibandingkan BioBERT dan SciBERT. Visualisasi *confusion matrix* menunjukkan bahwa mayoritas sampel berada pada diagonal utama matriks, yang menandakan bahwa sebagian besar data berhasil diklasifikasikan ke dalam kelas yang sesuai. Hasil tersebut menunjukkan bahwa PubMedBERT mampu mempelajari karakteristik masing-masing kategori relasi secara efektif dan memiliki

kemampuan yang baik dalam membedakan kelas *No Relation*, *Seaweed–Location*, *Seaweed–Benefit*, *Seaweed–Compound*, dan *Compound–Benefit* pada dataset literatur rumput laut.

3.2.10 *Best Relation Extraction (RE) Model*

Tahap ini dilakukan untuk menentukan model *Relation Extraction (RE)* yang digunakan pada proses ekstraksi relasi. Penentuan model dilakukan dengan membandingkan kinerja tiga model yang telah dievaluasi, yaitu SciBERT, BioBERT, dan PubMedBERT. Perbandingan dilakukan menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score* pada data pengujian yang sama.

Table 3.9 Hasil Perbandingan Model

No	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	BioBERT	84.88	86.30	84.88	85.17
2	SciBERT	85.56	86.21	85.56	85.69
3	PubMedBERT	86.92	87.64	86.92	87.02

Berdasarkan **Tabel 3.9**, Berdasarkan perbandingan kinerja ketiga model, PubMedBERT menghasilkan nilai tertinggi pada seluruh metrik evaluasi, dengan *Accuracy* sebesar 86,92%, *Precision* sebesar 87,64%, *Recall* sebesar 86,92%, dan *F1-Score* sebesar 87,02%. Temuan ini menunjukkan bahwa PubMedBERT memiliki kemampuan generalisasi yang lebih baik dalam mengenali pola relasi antar entitas dibandingkan SciBERT dan BioBERT. Dengan demikian, PubMedBERT dipilih sebagai model terbaik dan selanjutnya diimplementasikan pada sistem *Relation Extraction* untuk mendukung pembangunan *knowledge graph* literatur rumput laut.

3.2.11 *Entity and Relation Extraction Result*

Tahap ekstraksi entitas dan relasi dilakukan untuk menghasilkan informasi terstruktur dari literatur rumput laut berdasarkan model terbaik yang diperoleh pada tahap sebelumnya. Model *Named Entity Recognition (NER)* digunakan untuk mengidentifikasi entitas yang termasuk dalam kategori *Seaweed*, *Compound*, *Benefit*, dan *Location*, sedangkan model *Relation Extraction (RE)* digunakan untuk mengidentifikasi hubungan antar entitas yang terdapat dalam teks. Hasil ekstraksi

berupa kumpulan entitas dan relasi yang merepresentasikan informasi penting dalam literatur rumput laut. Informasi yang semula tersaji dalam bentuk teks tidak terstruktur direpresentasikan ke dalam bentuk data terstruktur yang terdiri atas entitas dan hubungan antar entitas. Representasi tersebut mengikuti konsep knowledge map yang dikembangkan oleh Ngo & Nguyen, (2024), yaitu penyusunan informasi berdasarkan keterhubungan antar entitas yang diekstraksi dari dokumen teks. Data hasil ekstraksi kemudian digunakan sebagai sumber data pada pengembangan sistem visualisasi informasi dan hubungan entitas berbasis web.

	JUDUL	SEAWEED	LOCATION	BENEFIT	COMPOUND	RE
0	Data 1			iodine deficiency (1), intake (1), lower iodine intake (1), adequate iodine nutrition (1), urinary iodine concentration (muic) (1)		COMPOUND 'urinary iodine con
1	Data 2	<i>durvillaea incurvata</i> (d. incurvata) (1), d. incurvata (1), d. incurvatum (2)		new food-grade functional ingredient drugs (1), content (1), nutritional value (1), and antioxidant extract (1), bioactive carotenoid (1), fucoxanthin (1), macromol (1), diets	COMPOUND 'antioxidant extra	
2	Data 3		asian pacifi	distinct nutritional benefit terrestrial crops (1), content (2), food product developer	mineral fibre (1)	COMPOUND 'mineral fibre' BEN
3	Data 4	<i>porphyra</i> sp. (2), undarium pinnatifida (1), saccharina sp. (1), hickium fusiformis (1)		functional property nutritional composition (1), high total crude (1), content (2), rich	protein (1), fatty acid (2), mufa (1), satur (1), pufo (1), macro element (1), microminer	SEAWEED 'Porphyra sp.' BENEF
4	Data 5	<i>padina tetrastrumatica</i> (1), p. tetrastrumatica (6)		functional food (1), nutritional value (1), nutritional antihyperglycemic potential (1)	streptococci (1), ethanolic ethyl acetate extract (1), metformin (1), alanine (1), aspa	COMPOUND 'alanine' BENEFIT
5	Data 6		indian (3), a	treatments (1), biodiversity (1), human nutrition (1), food security (1), curative proper	fucoidan (1), spirulina (1)	COMPOUND 'fucoidan' BENEF
6	Data 7	<i>palmarium palmata</i> (1), palmarium palmatum (1)		improve sensory characteristic (1), texture (1), optimize proces storage condition (1)	compound (three amino acid volatile compound) (1), macronutrient (1), isoluble prot	SEAWEED 'Palmarium palmata
7	Data 8	<i>acrosiphonia orientalis</i> , caulerpa scalpelliformis, ulva fasciata (1), ulva lactuca (1), —		biological activities (1), antioxidants, (1), healthier human and/or animal consumption	moisture (1), ash (1), total sugar (1), total proteins (1), total lipids (1), crude fiber (1), c	SEAWEED 'Acrosiphonia orient
8	Data 9	<i>camptylaphora hydnaroides</i> (1), c. hydnaroides (1), c. hydnaroides (1), c. hydnaroides (1)	japan (1)	health promotion disease prevention (1), anti-inflammatory effect (1), increase body	glucose (1), insulin (1), adipokines (1), lipid (1), polysaccharide (1), malondialdehyde	SEAWEED 'Camptylaphora hyp

Gambar 3.13 Dataset Ekstraksi NER dan RE

3.2.12 Requirement Analysis

Tahap *requirement analysis* dilakukan untuk mengidentifikasi kebutuhan sistem yang akan dikembangkan berdasarkan hasil ekstraksi entitas dan relasi pada literatur rumput laut. Analisis ini bertujuan untuk menentukan kebutuhan fungsional dan non-fungsional yang diperlukan agar sistem dapat mendukung proses pengelolaan, pencarian, dan visualisasi informasi secara efektif. Hasil analisis kebutuhan sistem pada penelitian ini disajikan pada Tabel 3.10.

Table 3.10 Hasil Analisis Kebutuhan Sistem

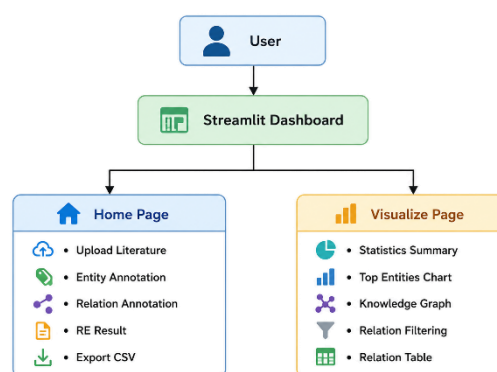
Kategori	Kebutuhan
Fungsional	1. Menerima masukan dokumen dalam format CSV, PDF, dan TXT. 2. Menampilkan hasil ekstraksi entitas Seaweed, Compound, Benefit, dan Location. 3. Menampilkan hasil ekstraksi relasi antar entitas. 4. Menyajikan visualisasi knowledge graph. 5. Menyediakan fitur filtering berdasarkan dokumen dan jenis relasi. 6. Mengekspor hasil ekstraksi ke format CSV.
Non-Fungsional	1. Sistem berbasis web menggunakan Python dan Streamlit.

	2. Mendukung pemrosesan ratusan dokumen. 3. Menyediakan visualisasi yang interaktif dan mudah digunakan.
--	---

Hasil analisis kebutuhan ini menjadi dasar dalam tahap perancangan sistem untuk memastikan bahwa fungsi dan fitur yang dikembangkan sesuai dengan kebutuhan pengelolaan, eksplorasi, dan visualisasi hasil ekstraksi entitas dan relasi.

3.2.13 System Design

Tahap Perancangan sistem dilakukan untuk menggambarkan alur kerja aplikasi serta integrasi antar komponen yang mendukung proses pengolahan, penyimpanan, dan visualisasi hasil ekstraksi entitas dan relasi. Rancangan yang dihasilkan menjadi acuan dalam pengembangan sistem visualisasi informasi dan hubungan entitas berbasis web. Hasil perancangan sistem pada penelitian ini ditunjukkan pada Gambar 3.14.



Gambar 3.14 Hasil Perancangan Sistem

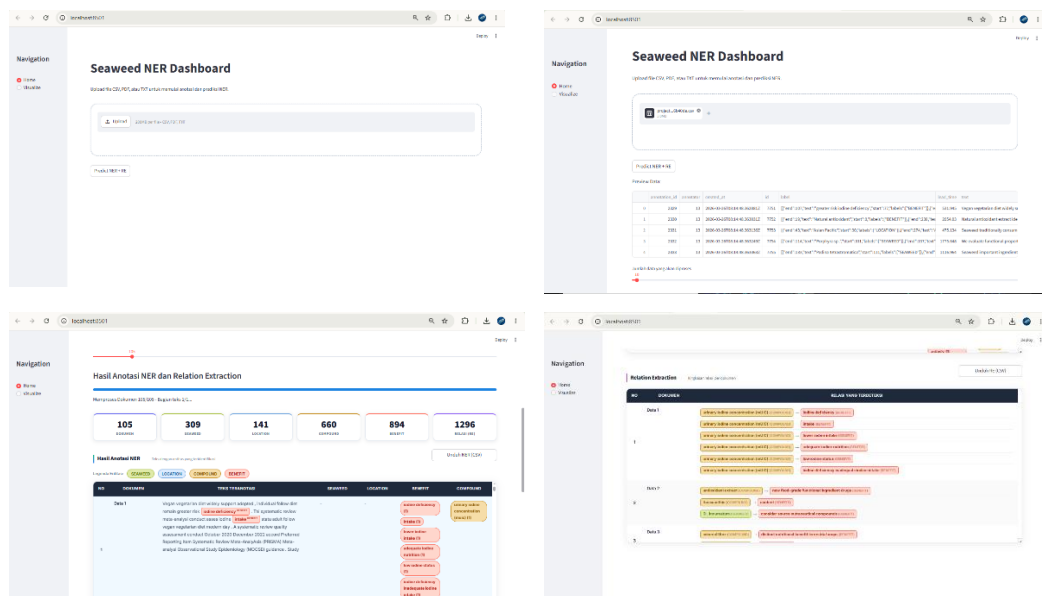
Arsitektur sistem terdiri atas dua halaman utama, yaitu *Home Page* dan *Visualize Page*. *Home Page* digunakan untuk pengelolaan data, meliputi unggah dokumen, anotasi entitas, anotasi relasi, penampilan hasil ekstraksi relasi, dan ekspor data. Sementara itu, *Visualize Page* digunakan untuk menyajikan hasil ekstraksi dalam bentuk statistik, grafik entitas, *knowledge graph*, serta tabel relasi. Melalui arsitektur ini, sistem dapat mendukung pengelolaan data dan visualisasi hasil ekstraksi entitas serta relasi dalam satu aplikasi berbasis web.

3.2.14 Web Development

Tahap pengembangan aplikasi dilakukan untuk mengimplementasikan hasil perancangan sistem ke dalam bentuk aplikasi berbasis web. Implementasi mencakup pengembangan fitur untuk pemrosesan dokumen literatur, penyajian hasil ekstraksi entitas dan relasi, serta visualisasi hubungan antar entitas yang diperoleh dari proses ekstraksi informasi.

a.) Home Page

Home Page merupakan halaman utama aplikasi yang digunakan untuk mengakses fitur-fitur pengolahan literatur dan hasil ekstraksi. Halaman ini menyediakan beberapa fitur utama, yaitu unggah dokumen, anotasi entitas, anotasi relasi, penampilan hasil *Relation Extraction (RE)*, serta ekspor hasil ekstraksi ke format CSV. Tampilan *Home Page* ditunjukkan pada Gambar 3.15.



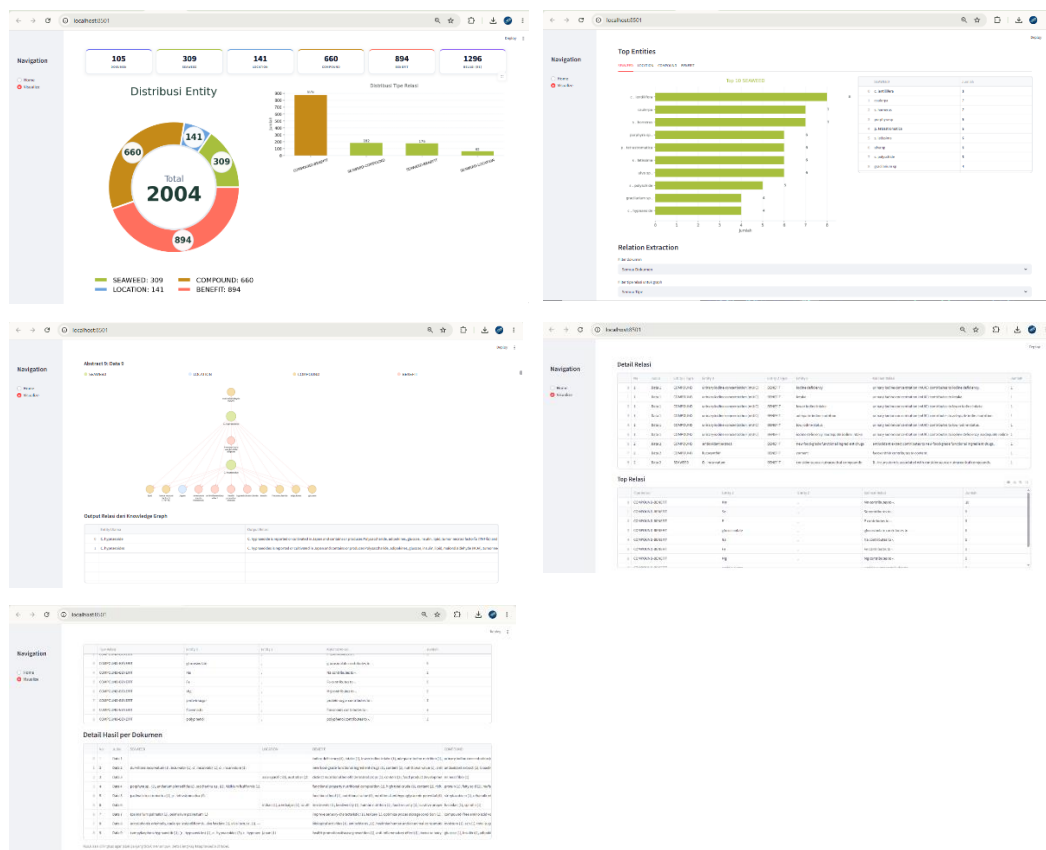
Gambar 3.15 Tampilan Home Page

Pada halaman ini, pengguna dapat mengunggah dokumen literatur yang akan diproses oleh sistem. Selain itu, halaman ini juga menyediakan akses terhadap hasil anotasi entitas, anotasi relasi, hasil ekstraksi relasi, serta fitur ekspor data untuk mendukung proses analisis dan visualisasi pada tahap selanjutnya.

b.) Visualize Page

Visualize Page merupakan halaman yang digunakan untuk menyajikan hasil

ekstraksi entitas dan relasi dalam bentuk visualisasi. Halaman ini menyediakan berbagai komponen visualisasi yang mendukung eksplorasi data, meliputi ringkasan statistik, distribusi entitas, knowledge graph, penyaringan relasi, dan tabel relasi. Tampilan *Visualize Page* ditunjukkan pada Gambar 3.16.



Gambar 3.16 Tampilan Visualize Page

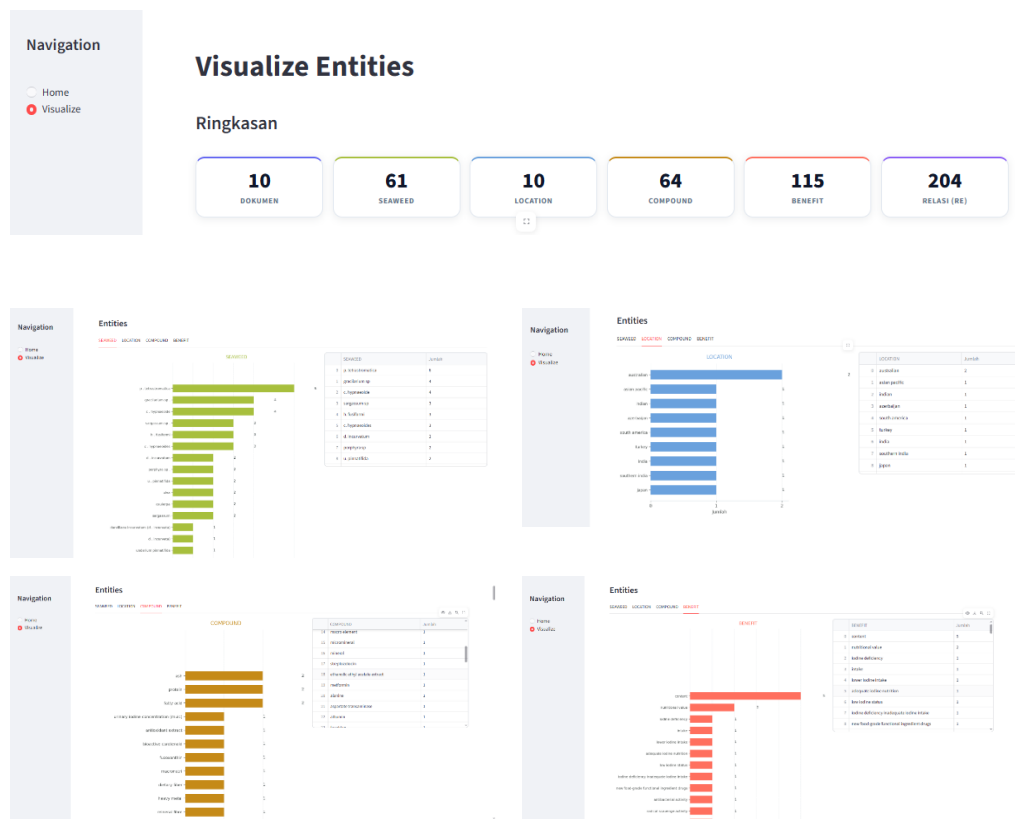
Pada halaman ini, hasil ekstraksi disajikan dalam bentuk visual yang memudahkan pengguna dalam mengidentifikasi pola dan hubungan antar entitas. Selain menampilkan informasi statistik, halaman ini juga menyediakan *knowledge graph* dan fitur penyaringan relasi untuk mendukung proses eksplorasi data secara lebih terstruktur.

3.2.15 Information Visualization

Tahap visualisasi informasi dilakukan untuk menyajikan hasil ekstraksi entitas dan relasi dalam bentuk yang lebih mudah dipahami dan dianalisis. Visualisasi yang dikembangkan mencakup ringkasan statistik, distribusi entitas, knowledge graph, serta tabel relasi yang memungkinkan pengguna mengeksplorasi

hubungan antar entitas pada literatur rumput laut.

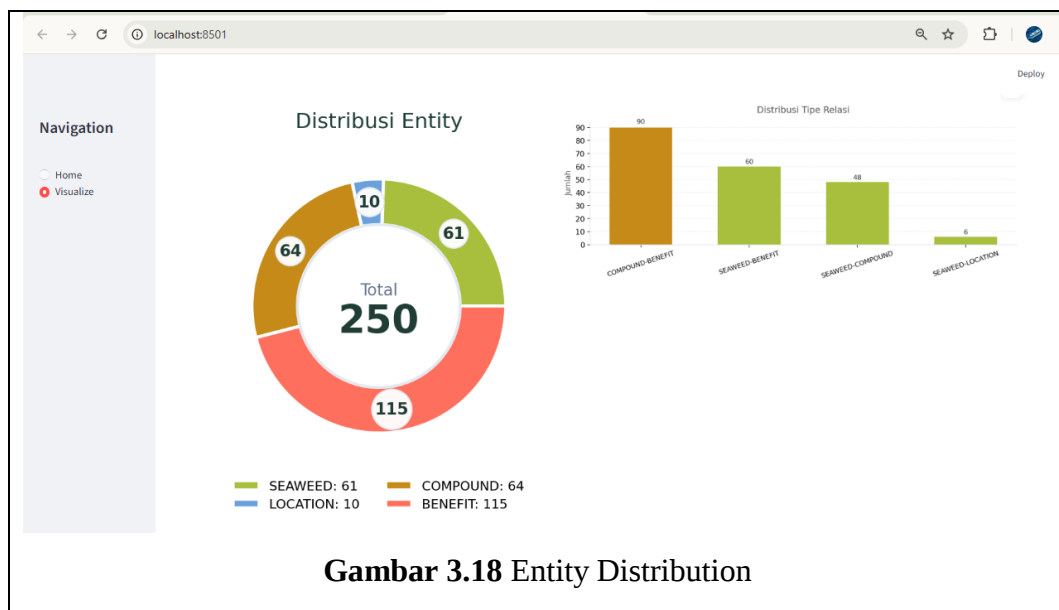
a.) Statistics Summary



Gambar 3.17 Statistics Summary

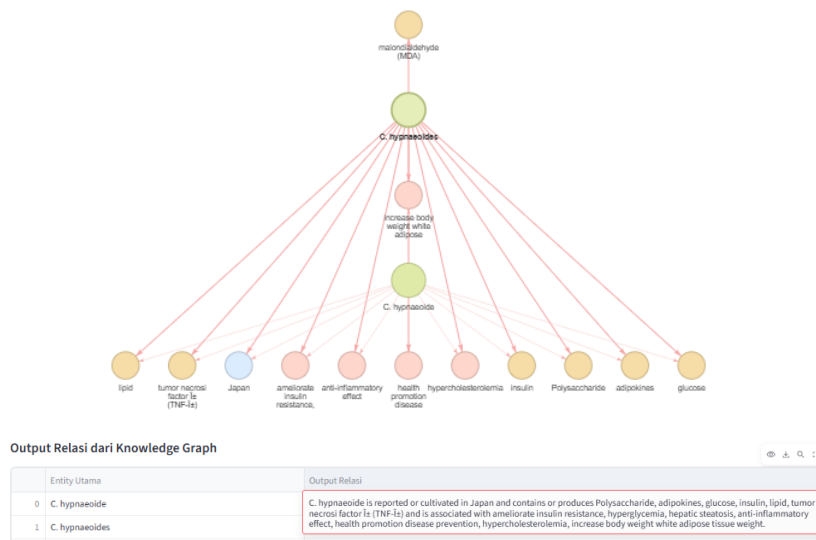
Statistics Summary menyajikan ringkasan informasi hasil ekstraksi dalam bentuk metrik numerik, seperti jumlah dokumen, jumlah entitas, jumlah relasi, dan jumlah jenis entitas. Ringkasan ini memberikan gambaran umum mengenai data yang telah diproses oleh sistem.

b.) Entity Distribution



Visualisasi distribusi entitas digunakan untuk menampilkan frekuensi kemunculan setiap kategori entitas yang berhasil diekstraksi dari literatur rumput laut. Visualisasi ini membantu mengidentifikasi entitas yang paling dominan dalam dataset.

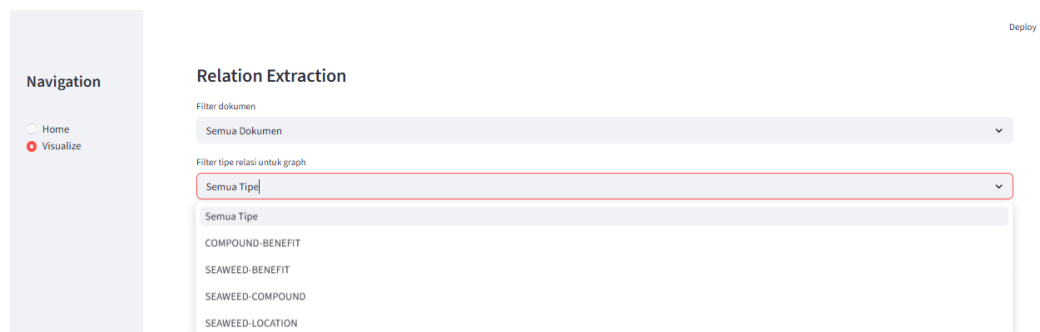
c.) Knowledge Graph



Gambar 3.19 Knowledge Graph

Knowledge graph digunakan untuk merepresentasikan hubungan antar entitas dalam bentuk jaringan. Setiap node merepresentasikan entitas, sedangkan edge merepresentasikan relasi yang terbentuk antar entitas. Visualisasi ini memudahkan identifikasi keterkaitan antara *seaweed*, *compound*, *benefit*, dan *location*.

d.) Relation Filtering



Gambar 3.20 Relation Filtering

Fitur *relation filtering* memungkinkan pengguna melakukan penyaringan data berdasarkan jenis relasi atau dokumen tertentu. Fitur ini digunakan untuk mempermudah eksplorasi hubungan entitas sesuai kebutuhan analisis.

e.) Relation Table

localhost:8501											
No	Judul	Entity 1 Type	Entity 1	Entity 2 Type	Entity 2	Kalimat Relasi	Jumlah				
81	9	Data 9	SEAWEED	Campylaea hypnoides	LOCATION	Japan	Campylaea hypnoides is reported or cultivated in Japan.	1			
82	9	Data 9	SEAWEED	(C. hypnoides)	LOCATION	Japan	(C. hypnoides) is reported or cultivated in Japan.	1			
83	9	Data 9	SEAWEED	C. hypnoides	LOCATION	Japan	C. hypnoides is reported or cultivated in Japan.	2			
84	9	Data 9	SEAWEED	C. hypnoides	LOCATION	Japan	C. hypnoides is reported or cultivated in Japan.	2			
85	9	Data 9	SEAWEED	Campylaea hypnoides	BENEFIT	health promotion disease prevention	Campylaea hypnoides is associated with health promotion disease prevention.	1			
86	9	Data 9	SEAWEED	Campylaea hypnoides	BENEFIT	anti-inflammatory effect	Campylaea hypnoides is associated with anti-inflammatory effect.	1			
87	9	Data 9	SEAWEED	Campylaea hypnoides	BENEFIT	Increase body weight white adipose tissue weight	Campylaea hypnoides is associated with increase body weight white adipose tissue weight.	1			
88	9	Data 9	SEAWEED	Campylaea hypnoides	BENEFIT	ameliorate insulin resistance , hyperglycemia , hepatomegaly	Campylaea hypnoides is associated with ameliorate insulin resistance , hyperglycemia , hepatomegaly.	1			
89	9	Data 9	SEAWEED	Campylaea hypnoides	BENEFIT	hypercholesterolemia	Campylaea hypnoides is associated with hypercholesterolemia.	1			
90	9	Data 9	SEAWEED	(C. hypnoides)	BENEFIT	health promotion disease prevention	(C. hypnoides) is associated with health promotion disease prevention.	1			
91	9	Data 9	SEAWEED	(C. hypnoides)	BENEFIT	anti-inflammatory effect	(C. hypnoides) is associated with anti-inflammatory effect.	1			
92	9	Data 9	SEAWEED	(C. hypnoides)	BENEFIT	Increase body weight white adipose tissue weight	(C. hypnoides) is associated with increase body weight white adipose tissue weight.	1			
93	9	Data 9	SEAWEED	(C. hypnoides)	BENEFIT	ameliorate insulin resistance , hyperglycemia , hepatomegaly	(C. hypnoides) is associated with ameliorate insulin resistance , hyperglycemia , hepatomegaly.	1			
94	9	Data 9	SEAWEED	(C. hypnoides)	BENEFIT	hypercholesterolemia	(C. hypnoides) is associated with hypercholesterolemia.	1			
95	9	Data 9	SEAWEED	C. hypnoides	BENEFIT	health promotion disease prevention	C. hypnoides is associated with health promotion disease prevention.	2			
96	9	Data 9	SEAWEED	C. hypnoides	BENEFIT	anti-inflammatory effect	C. hypnoides is associated with anti-inflammatory effect.	2			
97	9	Data 9	SEAWEED	C. hypnoides	BENEFIT	Increase body weight white adipose tissue weight	C. hypnoides is associated with increase body weight white adipose tissue weight.	2			
98	9	Data 9	SEAWEED	C. hypnoides	BENEFIT	ameliorate insulin resistance , hyperglycemia , hepatomegaly	C. hypnoides is associated with ameliorate insulin resistance , hyperglycemia , hepatomegaly.	2			
99	9	Data 9	SEAWEED	C. hypnoides	BENEFIT	hypercholesterolemia	C. hypnoides is associated with hypercholesterolemia.	2			
100	9	Data 9	SEAWEED	C. hypnoides	BENEFIT	health promotion disease prevention	C. hypnoides is associated with health promotion disease prevention.	2			
101	9	Data 9	SEAWEED	C. hypnoides	BENEFIT	anti-inflammatory effect	C. hypnoides is associated with anti-inflammatory effect.	2			

Gambar 3.21 Relation Table

Visualisasi distribusi entitas digunakan untuk menampilkan frekuensi kemunculan setiap kategori entitas yang berhasil diekstraksi dari literatur rumput laut. Visualisasi ini membantu mengidentifikasi entitas yang paling dominan dalam dataset.

3.2.16 System Testing

Pengujian sistem dilakukan menggunakan metode *Black Box Testing* untuk memverifikasi bahwa seluruh fungsi yang diimplementasikan dapat berjalan sesuai dengan kebutuhan sistem. Pengujian dilakukan dengan memberikan masukan pada setiap fitur dan mengamati keluaran yang dihasilkan tanpa memperhatikan proses internal aplikasi.

Table 1.11 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

3.3 Penugasan-Penugasan yang Telah Diselesaikan

Penugasan yang diberikan meliputi materi pengantar, *review* jurnal, pengolahan data, mengikuti pertemuan pekanan, dan presentasi mingguan. Dengan mengikuti pelatihan pengantar MBKM, mahasiswa diharapkan memiliki pengetahuan mengenai *building learning commitment*, BRIN sebagai lembaga riset berkelas dunia, pengantar *research integrity*, membangun tim efektif, komunikasi persuasif, dan teknis presentasi yang baik seperti pada **Gambar 3.22** dan **3.23**



Gambar 3.22 Sertifikat Pelatihan Pengantar MBKM BRIN

No. 626/STMP/PM/2/2026

Materi Pelatihan Pembekalan Magang BRIN Semester Genap 2025/2026

No.	Materi Pelatihan	Jam
1	Orientation Pelatihan	3
2	Understanding Learning Competency	3
3	Mengenal Badan Riset dan Inovasi Nasional	3
4	Pengantar Research Integrity	3
5	Jurnal Ilmiah & Manajemen Referensi	3
6	Tim Etik	3
7	Teknik Presentasi	4
8	Evaluasi Akademik	3
Jumlah Total		25

Jakarta, 16 Februari 2026
Ketua Tim Fungsi Layanan Learning Management System (LMS)
Pengembangan Kompetensi SDM

TT ELEKTRONIK

Indra Riswadiarta, M.H.

Gambar 3.23 Daftar Materi Pelatihan Pengantar MBKM BRIN

Setelah mengikuti kegiatan orientasi, mahasiswa melaksanakan kegiatan magang di bawah bimbingan peneliti sesuai dengan topik dan ruang lingkup penelitian yang telah ditetapkan. Selama periode magang, mahasiswa melakukan aktivitas penelitian dan pengembangan sistem sesuai dengan arahan pembimbing.

Selain itu, dilakukan presentasi perkembangan secara mingguan sebagai sarana pemantauan kemajuan kegiatan, evaluasi hasil pekerjaan, serta diskusi terkait permasalahan yang ditemui selama proses penelitian dan pengembangan sistem.

Kegiatan magang diawali dengan tahap studi literatur dan perancangan sistem pada bulan pertama. Pada tahap ini dilakukan kajian terhadap berbagai penelitian yang berkaitan dengan literatur rumput laut, ekstraksi informasi, *Named Entity Recognition (NER)*, dan *Relation Extraction (RE)*. Hasil kajian digunakan untuk merumuskan research question serta menyusun rancangan arsitektur sistem yang menjadi dasar pengembangan penelitian.

Pada bulan kedua dilakukan persiapan dataset yang meliputi proses pembersihan data (data cleaning), anotasi entitas dan relasi menggunakan Label Studio, serta pembagian dataset ke dalam data pelatihan, validasi, dan pengujian. Proses anotasi dilakukan untuk menghasilkan data berlabel yang digunakan dalam pengembangan model ekstraksi informasi.

Kegiatan pada bulan ketiga difokuskan pada pengembangan dan evaluasi model *Relation Extraction (RE)*. Proses ini dilakukan dengan memanfaatkan tiga model berbasis Transformer, yaitu SciBERT, BioBERT, dan PubMedBERT. Ketiga model dilatih menggunakan dataset yang telah dipersiapkan dan dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

Hasil evaluasi digunakan untuk menentukan model dengan kinerja terbaik yang selanjutnya diintegrasikan ke dalam sistem. Pada bulan keempat dilakukan pengembangan pipeline ekstraksi informasi secara *end-to-end*. Tahap ini mencakup integrasi model *Relation Extraction* terpilih, pembentukan data terstruktur hasil ekstraksi, serta pengembangan mekanisme identifikasi hubungan antar entitas.

Hasil dari tahap ini berupa representasi informasi terstruktur yang memuat entitas dan relasi yang diekstraksi dari literatur rumput laut. Pada bulan kelima dilakukan pengembangan aplikasi visualisasi berbasis web menggunakan framework Streamlit. Pengembangan mencakup implementasi fitur unggah dokumen, penyorotan entitas (*entity highlighting*), penyajian hasil ekstraksi relasi, visualisasi hubungan entitas, serta ekspor data ke format CSV. Tahap ini

menghasilkan prototipe aplikasi yang dapat digunakan untuk menampilkan dan mengeksplorasi hasil ekstraksi informasi.

Pada bulan keenam dilakukan penyusunan laporan, finalisasi artikel ilmiah, serta dokumentasi perangkat lunak yang telah dikembangkan. Kegiatan ini bertujuan untuk mendokumentasikan proses penelitian, hasil pengembangan model, dan implementasi sistem yang telah dilakukan selama program magang.

Selain kegiatan penelitian dan pengembangan sistem, mahasiswa juga berpartisipasi dalam Pertemuan Pekan Pusat Riset Sains Data dan Informasi BRIN yang diselenggarakan secara berkala. Kegiatan ini menjadi forum ilmiah yang diikuti oleh peneliti, tenaga ahli, dan mahasiswa magang untuk mempresentasikan perkembangan penelitian, memperoleh masukan, serta mendiskusikan berbagai topik penelitian di bidang sains data dan kecerdasan artifisial. Melalui kegiatan tersebut, mahasiswa memperoleh wawasan mengenai berbagai penelitian yang sedang dikembangkan di lingkungan BRIN sekaligus mendapatkan umpan balik yang mendukung pelaksanaan penelitian selama program magang.

3.4 Pembelajaran Hal Baru (*Lessons Learned*) yang Diperoleh

Berdasarkan pengalaman dalam menjalani program ini, terdapat beberapa pembelajaran baru yang saya peroleh:

a. Pentingnya Kolaborasi dalam Kegiatan Penelitian

Selama kegiatan magang, saya memahami bahwa kolaborasi merupakan salah satu faktor penting dalam keberhasilan penelitian. Bimbingan dan arahan dari para peneliti membantu saya memahami berbagai tahapan penelitian, mulai dari perumusan permasalahan, pengolahan data, hingga evaluasi hasil. Selain itu, diskusi yang dilakukan secara rutin memberikan kesempatan untuk memperoleh masukan yang konstruktif serta memperluas pemahaman terhadap topik penelitian yang sedang dikerjakan.

b. Pentingnya Literatur dan Jurnal sebagai Landasan Penelitian

Kajian terhadap jurnal dan artikel ilmiah menjadi salah satu aktivitas penting selama kegiatan magang. Melalui studi literatur, saya memperoleh pemahaman mengenai berbagai metode ekstraksi informasi, perkembangan model berbasis Transformer, serta penerapan NER dan RE pada domain ilmiah. Literatur yang dipelajari menjadi dasar dalam menentukan pendekatan penelitian, pemilihan model, serta evaluasi hasil yang diperoleh.

c. Pentingnya Kualitas Data dalam Pengembangan Model

Salah satu pembelajaran yang saya peroleh adalah bahwa kualitas data memiliki pengaruh yang sangat besar terhadap performa model yang dikembangkan. Proses anotasi entitas dan relasi menggunakan Label Studio menunjukkan bahwa konsistensi pemberian label, ketelitian dalam proses anotasi, serta kesesuaian pedoman anotasi merupakan aspek yang sangat penting untuk menghasilkan dataset yang berkualitas. Dataset yang baik menjadi fondasi utama dalam pengembangan model ekstraksi informasi yang akurat.

d. Pemahaman mengenai Ekstraksi Informasi Berbasis NLP

Melalui keterlibatan dalam penelitian ini, saya memperoleh pemahaman yang lebih mendalam mengenai proses ekstraksi informasi dari literatur ilmiah menggunakan pendekatan *Named Entity Recognition (NER)* dan *Relation Extraction (RE)*. Saya memahami bagaimana informasi yang awalnya tidak terstruktur dapat diubah menjadi data terstruktur melalui identifikasi entitas dan hubungan antarentitas sehingga lebih mudah dianalisis dan dimanfaatkan untuk kebutuhan penelitian lanjutan..

e. Pemanfaatan Tools dan Library dalam Penelitian NLP

Kegiatan magang memberikan pengalaman dalam penggunaan berbagai tools dan library yang mendukung penelitian NLP. Penggunaan Label Studio untuk anotasi data serta berbagai pustaka Python seperti Pandas, NumPy, PyTorch, Transformers, Scikit-learn, dan Matplotlib membantu saya memahami alur pengembangan sistem secara lebih menyeluruh. Selain meningkatkan kemampuan teknis, penggunaan tools tersebut juga

memberikan pemahaman mengenai pentingnya pemilihan perangkat yang sesuai untuk mendukung proses penelitian.

f. Pentingnya Optimasi dan Evaluasi Model

Selama proses penelitian, saya memahami bahwa keberhasilan suatu model tidak hanya ditentukan oleh arsitektur yang digunakan, tetapi juga oleh proses optimasi yang dilakukan selama pelatihan. Dalam pengembangan model BioBERT, SciBERT, dan PubMedBERT untuk tugas *Named Entity Recognition (NER)* dan *Relation Extraction (RE)*, saya mempelajari bagaimana pengaturan hyperparameter, seperti learning rate, batch size, dan jumlah epoch, dapat memengaruhi performa model. Selain itu, saya juga memperoleh pemahaman mengenai pentingnya evaluasi model menggunakan metrik seperti *precision*, *recall*, dan *F1-score* untuk menilai kualitas hasil yang diperoleh. Pengalaman ini memberikan wawasan bahwa pengembangan model merupakan proses yang memerlukan serangkaian eksperimen dan evaluasi secara berkelanjutan guna memperoleh performa yang optimal.

BAB IV

PENUTUP

4.1 Kesimpulan

Program MBKM Riset/Penelitian di Badan Riset dan Inovasi Nasional (BRIN) dilaksanakan selama enam bulan melalui serangkaian kegiatan penelitian dan pengembangan sistem di bawah bimbingan Ibu Dr. Dipl. Ing. (FH) Anne Parlina, M.P., selaku peneliti pada Pusat Riset Sains Data dan Informasi BRIN. Kegiatan magang yang dilakukan terbagi menjadi dua fokus utama, yaitu pengembangan model ekstraksi informasi pada literatur rumput laut dan pengembangan sistem visualisasi informasi serta hubungan entitas berbasis web. Kedua kegiatan tersebut merupakan bagian dari penelitian yang berjudul “Rancang Bangun Sistem Visualisasi Informasi dan Hubungan Entitas Literatur Rumput Laut Berbasis Web.”

Dalam pelaksanaannya, kegiatan yang dilakukan meliputi studi literatur, perancangan sistem, persiapan dan anotasi dataset, pengembangan serta evaluasi model *Relation Extraction (RE)*, integrasi pipeline ekstraksi informasi, pengembangan dashboard visualisasi berbasis web, hingga penyusunan laporan dan dokumentasi penelitian. Selama periode magang, kemajuan penelitian dipantau melalui kegiatan bimbingan dan presentasi perkembangan yang dilaksanakan secara berkala. Pada fokus pertama, penelitian diarahkan pada pengembangan model ekstraksi informasi menggunakan pendekatan *Named Entity Recognition (NER)* dan *Relation Extraction (RE)* untuk mengidentifikasi entitas serta hubungan antar entitas pada literatur rumput laut.

Dalam pengembangan model RE, dilakukan evaluasi terhadap tiga model berbasis Transformer, yaitu SciBERT, BioBERT, dan PubMedBERT. Berdasarkan hasil pengujian, model PubMedBERT menghasilkan kinerja terbaik dengan nilai *accuracy* sebesar 90,05%, *precision* 90,24%, *recall* 90,05%, dan *F1-score* 90,03%, sehingga dipilih sebagai model yang digunakan dalam sistem. Pada fokus kedua, hasil ekstraksi informasi diintegrasikan ke dalam aplikasi visualisasi berbasis web yang dikembangkan menggunakan *framework Streamlit*. Sistem yang dihasilkan

menyediakan fitur pengelolaan dokumen, penyajian hasil ekstraksi entitas dan relasi, visualisasi hubungan entitas melalui knowledge graph, serta ekspor data hasil ekstraksi untuk mendukung proses analisis lebih lanjut.

Melalui program MBKM Riset/Penelitian di BRIN, mahasiswa memperoleh pengalaman dalam pelaksanaan penelitian, pengelolaan dataset, pengembangan model berbasis Transformer, serta pengembangan aplikasi berbasis web. Selain meningkatkan kompetensi teknis pada bidang *Natural Language Processing (NLP)*, kegiatan magang juga memberikan pemahaman mengenai metodologi penelitian, pentingnya kualitas data, evaluasi model, dan kolaborasi dalam lingkungan riset. Pengalaman tersebut menjadi bekal yang mendukung pengembangan kompetensi akademik dan profesional di bidang sains data dan kecerdasan artifisial.

4.2 Saran

Berdasarkan hasil penelitian dan pengembangan sistem yang telah dilakukan, masih terdapat beberapa aspek yang dapat dikembangkan pada penelitian selanjutnya. Penggunaan dataset literatur rumput laut yang lebih besar dan beragam dapat dilakukan untuk meningkatkan kemampuan model dalam mengidentifikasi entitas dan hubungan antar entitas pada berbagai konteks penulisan ilmiah. Selain itu, cakupan ekstraksi informasi dapat diperluas dengan menambahkan jenis entitas dan relasi yang lebih beragam sehingga representasi pengetahuan yang dihasilkan menjadi lebih lengkap. Dari sisi model, evaluasi terhadap arsitektur dan pendekatan terbaru pada tugas *Named Entity Recognition (NER)* dan *Relation Extraction (RE)* dapat dilakukan untuk memperoleh performa yang lebih baik. Sementara itu, dari sisi sistem, pengembangan dapat diarahkan pada integrasi basis data, mekanisme pembaruan data secara otomatis, serta penambahan fitur analisis dan visualisasi yang lebih interaktif untuk mendukung eksplorasi informasi secara lebih mendalam. Dengan pengembangan tersebut, sistem visualisasi informasi dan hubungan entitas pada literatur rumput laut diharapkan dapat memberikan dukungan yang lebih optimal bagi kegiatan penelitian dan pemanfaatan informasi ilmiah di bidang rumput laut.

DAFTAR PUSTAKA

- Boztepe, E. B., Karakaya, B., Bahadır, K., & Ünlü, İ. (2022). *An Approach for Audio-Visual Content Understanding of Video using Multimodal Deep Learning Methodology*. 5(2). <https://doi.org/10.35377/saucis.05.02.1139765>
- BRIN. *Badan Riset dan Inovasi Nasional*. Diakses 19 Juli 2024. <https://www.brin.go.id/>
- Do, T., Doan, T. T., Le, K., Nguyen, T., Le, D., & Ngo, T. D. (2024). *Key Information Extraction and Recognition from Rich*. 11(4), 569–594. <https://doi.org/10.1142/S2196888824500131>
- Ngo, H. Q., & Nguyen, H. D. (2024). *Ontology Knowledge Map Approach Towards Building Linked Data for Vietnamese Legal Applications* □. 11(2), 323–342. <https://doi.org/10.1142/S2196888824500015>
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Funtowicz, M., Davison, J., Shleifer, S., Platen, P. Von, Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. Le, Gugger, S., Drame, M., ... Rush, A. M. (2020). *Transformers : State-of-the-Art Natural Language Processing*. 38–45.

LAMPIRAN

Biodata Peserta MBKM



Nama : Mursalin
NIM : 231730006
Program Studi : Informatika
Fakultas : Sains dan Teknologi
Universitas : UIN Sultan Maulana Hasanduddin Banten

Biodata Pembimbing BRIN



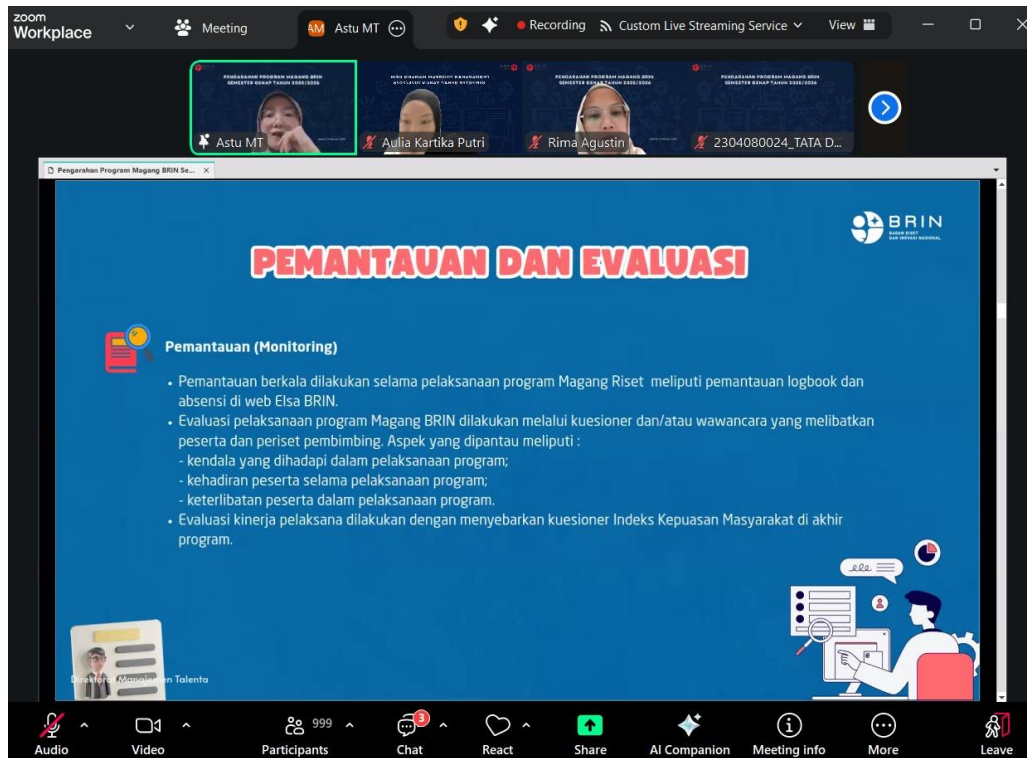
Nama : Dr. Dipl. Ing. (FH). Anne Parlina, MP.
NIP : 197602141994032001
Satuan Kerja : Pusat Riset Sains Data dan Informasi
Jabatan :
Pangkat/Gol :

Biodata Pembimbing Universitas



Nama : Ahmad Tabrani, M.T.I
NIP : 198509172018011002
Jabatan : Ketua Prodi Informatika

Dokumentasi Kegiatan MBKM



Gambar 1. Pertemuan Pengantar MBKM



Gambar 2. Serah Terima Mahasiswa MBKM



Gambar 3. Pertemuan Kelompok Riset Information Retrieval



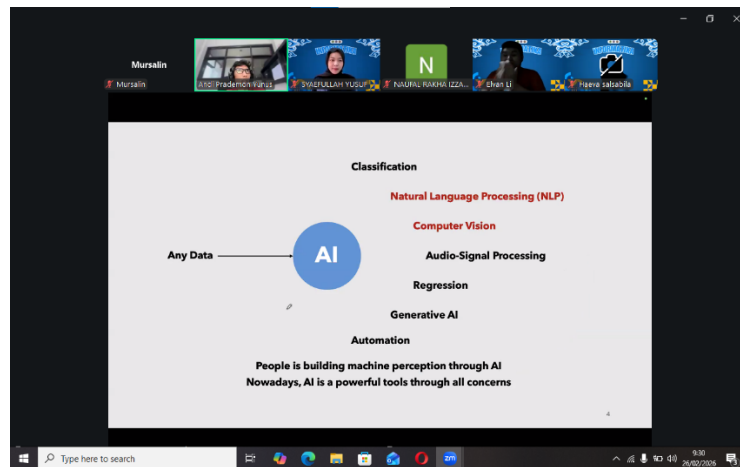
Gambar 4. Pertemuan Mingguan dengan Pembimbing



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

Luaran Kegiatan MBKM

https://drive.google.com/drive/folders/1BbUSOINTsJztqhclcRha77Qx4zNzB7-Y?usp=drive_link