

**PEMBANGKITAN ANOTASI PETA BERDASARKAN EKSTRAKSI DATA
BERITA PENYAKIT ENDEMIK MENGGUNAKAN PENDEKATAN
NATURAL LANGUAGE GENERATION**

SKRIPSI

JEFRY LIANTO

191402027



**PROGRAM STUDI S-1 TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2025**

**PEMBANGKITAN ANOTASI PETA BERDASARKAN EKSTRAKSI DATA
BERITA PENYAKIT ENDEMIK MENGGUNAKAN PENDEKATAN
NATURAL LANGUAGE GENERATION**

**Diajukan untuk melengkapi tugas dan memenuhi syarat memperoleh ijazah
Sarjana Teknologi Informasi**

SKRIPSI

**JEFRY LIANTO
191402027**



**PROGRAM STUDI S-1 TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2025**

PERSETUJUAN

Judul : Pembangkitan Anotasi Peta Berdasarkan Ekstraksi Data
Berita Penyakit Endemik Menggunakan Pendekatan
Natural Language Generation

Kategori : Skripsi

Nama Mahasiswa : Jefry Lianto

Nomor Induk Mahasiswa : 191402027

Program Studi : Sarjana (S-1) Teknologi Informasi

Fakultas : Ilmu Komputer dan Teknologi Informasi

Universitas Sumatera Utara

Medan, 6 Januari 2025

Komisi Pembimbing:

Pembimbing 2,



Ivan Jaya S.Si., M.Kom
NIP. 198407072015041001

Pembimbing 1,



Dedy Arisandi S.T., M.Kom
NIP. 197908312009121002

Diketahui/disetujui oleh
Program Studi S-1 Teknologi Informasi
Ketua



Dedy Arisandi S.T., M.Kom.
NIP. 197908312009121002

PERNYATAAN

PEMBANGKITAN ANOTASI PETA BERDASARKAN EKSTRAKSI DATA BERITA PENYAKIT ENDEMIK MENGGUNAKAN PENDEKATAN NATURAL LANGUAGE GENERATION

SKRIPSI

Saya menyatakan bahwasanya skripsi ini merupakan hasil karya saya sendiri,
terkecuali beberapa ringkasan maupun kutipan yang telah disebutkan sumbernya.

Medan, 6 Januari 2025

Jefry Lianto
191402027

UCAPAN TERIMA KASIH

Segala puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa, atas berkat-Nya yang luar biasa dalam menuntun langkah penulis dalam menyelesaikan skripsi yang berjudul “Pembangkitan Anotasi Peta Berdasarkan Ekstraksi Data Berita Penyakit Endemik Menggunakan Pendekatan Natural Language Generation”, yang di mana menjadi syarat dalam memperoleh gelar sarjana komputer di Program Studi Teknologi Informasi Universitas Sumatera Utara. Dengan penuh syukur, Penulis menyampaikan rasa terima kasih sedalamnya kepada seluruh pihak yang memberikan dukungan, bantuan, dan motivasi dalam proses penyelesaian skripsi ini. Penulis mengucapkan terima kasih kepada:

1. Tuhan Yang Maha Esa yang telah melimpahkan anugrah sehingga penulis dapat menyelesaikan skripsi ini dengan pikiran yang jernih dan sehat.
2. Keluarga penulis, Ibu Fifiana yang selalu memberikan doa, dukungan, saran serta bimbingan tanpa menuntut dan memberatkan penulis dalam menyelesaikan skripsi ini.
3. Bapak Prof. Dr. Muryanto Amin, S.Sos, M.Si selaku Rektor Universitas Sumatera Utara.
4. Ibu Dr. Maya Silvi Lydia B.Sc., M.Sc. selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara.
5. Bapak Dedy Arisandi S.T., M.Kom. selaku Ketua Program Studi S1 Teknologi Informasi Universitas Sumatera Utara dan Dosen Pembimbing I yang telah membimbing, memberikan saran dan masukan kepada penulis hingga akhir penulisan.
6. Bapak Ivan Jaya, S.Si., M.Kom. selaku Sekretaris Program Studi S1 Teknologi Informasi Universitas Sumatera Utara dan Dosen Pembimbing II yang juga telah membimbing, memberikan saran dan masukan kepada penulis hingga akhir penulisan.
7. Ibu Ulfi Andayani S.Kom., M.Kom. selaku Dosen Penguji I yang telah memberikan saran dan masukan kepada penulis sehingga dapat menyelesaikan skripsi ini.
8. Ibu Rossy Nurhasanah S.Kom., M.Kom. selaku Dosen Penguji II yang juga telah memberikan saran dan masukan kepada penulis sehingga dapat menyelesaikan skripsi ini.

9. Seluruh dosen serta staf pegawai di lingkungan Program Studi Teknologi Informasi dan Fasilkom-TI USU, yang telah membantu proses administrasi selama masa perkuliahan
10. Kepada sahabat penulis Hendry Winata, Williams dan Dolok Malau yang selalu mendukung, memberi motivasi dan membantu penulis dalam masa perkuliahan.
11. Seluruh teman-teman Angkatan 2019 yang sudah memberikan pengalaman dan pelajaran serta melewati masa-masa perkuliahan bersama penulis.
12. Seluruh pihak yang tidak dapat penulis sebutkan yang telah memberikan bantuan kepada penulis sehingga mampu menyelesaikan skripsi ini. Penulis mengucapkan terima kasih.

Medan, 6 Januari 2025

Jefry Lianto

ABSTRACT

Dalam era digital, data berita menjadi sumber informasi melimpah, tetapi sering tidak terstruktur, khususnya terkait penyebaran penyakit endemik. Penelitian ini bertujuan untuk mengimplementasikan ekstraksi data otomatis dari berita tentang penyakit endemik untuk menghasilkan anotasi peta menggunakan pendekatan Natural Language Generation (NLG). Sistem yang dikembangkan menggunakan teknik scraping untuk mengumpulkan data dari artikel berita online, lalu diproses dengan peringkasan menggunakan model pre-trained BART dan metode frekuensi. Proses text preprocessing meliputi case folding, tokenizing, dan stopwords removal. Data yang diekstrak digunakan untuk menentukan lokasi geografis dan jenis penyakit, yang kemudian dianotasi ke dalam peta untuk visualisasi dalam Sistem Informasi Geografis (SIG). Evaluasi dilakukan menggunakan berbagai metrik, termasuk BLEU, ROUGE, dan BERTScore, dengan hasil akurasi rata-rata yang menunjukkan performa baik. Nilai rata-rata BERTScore F1 mencapai 0.88, BLEU 0.78, ROUGE-1 0.84, ROUGE-2 0.80, dan ROUGE-L 0.84, yang menandakan bahwa ringkasan yang dihasilkan memiliki kesesuaian tinggi dengan teks asli. Selain itu, evaluasi juga melibatkan penilaian dari tiga kelompok responden (orang umum, tenaga medis, dan ahli bahasa), yang menunjukkan bahwa 80% merasa ringkasan sangat mudah dipahami, 75% jelas, dan 70% mudah dibaca. Temuan ini menunjukkan bahwa pendekatan NLG efektif dalam menghasilkan ringkasan berita yang informatif dan anotasi peta yang akurat untuk memudahkan pemantauan dan pengelolaan penyebaran penyakit endemik di Indonesia.

Kata Kunci : *Natural Language Generation, Anotasi Peta, Ekstraksi Data Berita, Bidirectional and Auto-Regressive Transformers, Metode Frekuensi.*

GENERATION OF MAP ANNOTATIONS BASED ON EXTRACTION OF ENDEMIC DISEASE NEWS DATA USING NATURAL LANGUAGE GENERATION APPROACH

ABSTRACT

In the digital era, news data serves as an abundant information source but is often unstructured, particularly regarding the spread of endemic diseases. This study aims to implement automatic data extraction from news articles about endemic diseases to generate map annotations using a Natural Language Generation (NLG) approach. The developed system utilizes scraping techniques to gather data from online news articles, which are then processed through summarization using the pre-trained BART model and a frequency-based method. The text preprocessing steps include case folding, tokenization, and stopwords removal. The extracted data is used to identify geographic locations and types of diseases mentioned, which are then annotated onto maps for visualization in a Geographic Information System (GIS). The evaluation was conducted using multiple metrics, including BLEU, ROUGE, and BERTScore, showing strong performance with average accuracies. The average BERTScore F1 was 0.88, BLEU was 0.78, ROUGE-1 was 0.84, ROUGE-2 was 0.80, and ROUGE-L was 0.84, indicating high consistency between the summaries and the original texts. Additionally, an evaluation involving three groups of respondents (general public, medical professionals, and linguists) revealed that 80% found the summaries easy to understand, 75% found them clear, and 70% found them easy to read. These findings demonstrate that the NLG approach effectively generates informative news summaries and accurate map annotations to facilitate monitoring and managing the spread of endemic diseases in Indonesia.

Keywords : *Natural Language Generation, Map Annotations, News Data Extraction, Bidirectional and Auto-Regressive Transformers, Frequency-Based Method.*

DAFTAR ISI

HALAMAN JUDUL	i
PERSETUJUAN	ii
PERNYATAAN	iii
UCAPAN TERIMA KASIH	iv
ABSTRACT	vi
<i>ABSTRACT</i>	vii
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Tujuan Penelitian	4
1.4. Batasan Penelitian	4
1.5. Manfaat Penelitian	5
1.6. Metodologi Penelitian	5
1.7. Sistematika Penulisan	6
BAB 2 LANDASAN TEORI	7
2.1. <i>Natural Language Generation</i>	7
2.2. Pemetaan	8
2.3. Berita	8
2.4. Korpus	9
2.5. <i>Machine Learning</i>	10
2.6. <i>Bidirectional and Auto-Regressive Transformers</i>	11
2.7. <i>Term Frequence</i>	14
2.8. <i>Sentence Scoring</i>	15
2.9. Penelitian Terdahulu	16
BAB 3 ANALISIS DAN PERANCANGAN SISTEM	22
3.1. Arsitektur Umum	22
3.2. <i>Data Input</i>	24
3.3. <i>Scrapping Data</i>	25

3.4.	<i>Text Preprocessing</i>	27
3.4.1.	<i>Case Folding</i>	27
3.4.2.	<i>Tokenizing</i>	28
3.5.	<i>Micro Planning A</i>	30
3.5.1.	<i>Bidirectional Encoder dan Autoregressive Decoder</i>	31
3.5.1.	<i>Linear Layer dan Softmax Layer</i>	32
3.6.	<i>Micro Planning B</i>	34
3.6.1.	<i>Chunking Text</i>	34
3.6.2.	<i>Word Frequencies</i>	36
3.6.3.	<i>Sentence Scoring</i>	37
3.7.	<i>Realization</i>	39
3.7.1.	<i>Stopword Removal</i>	39
3.7.2.	<i>Endemic and Location Search</i>	40
3.7.3.	<i>Save Result</i>	42
3.8.	Evaluasi Model	44
3.9.	<i>Map Annotation</i>	46
3.10.	<i>Geographic Information Systems</i>	46
BAB 4 IMPLEMENTASI DAN PENGUJIAN		48
4.1	Implementasi Sistem	48
4.2	Implementasi Bagian Depan	48
4.3	Prosedur Operasional	50
4.4	Pengujian Sistem	50
4.5	Evaluasi Hasil Pengujian	60
BAB 5 KESIMPULAN DAN SARAN		67
5.1	Kesimpulan	67
5.2	Saran	68
DAFTAR PUSTAKA		69
LAMPIRAN		71

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	18
Tabel 3.1 Contoh Sebelum dan Sesudah Proses <i>Case Folding</i>	28
Tabel 3.2 Contoh Sebelum dan Sesudah Proses <i>Tokenizing</i> Satu Kalimat	29
Tabel 3.3 Contoh Sebelum dan Sesudah Proses <i>Tokenizing</i> Satu Kata	30
Tabel 3.4 Contoh Sebelum dan Sesudah Proses <i>Stopword Removal</i>	40
Tabel 4.1. Spesifikasi Hardware dan Software	48
Tabel 4.2. Url Artikel Berita	50
Tabel 4.3. Hasil Pengujian Sistem	51

DAFTAR GAMBAR

Gambar 3.1 Arsitektur Umum	23
Gambar 3.2 Contoh Artikel Berita Endemik DBD	25
Gambar 3.3 Desain Tampilan Main Menu	47
Gambar 3.4 Desain Tampilan Menu Maps	47
Gambar 4.1. Tampilan Main Menu	49
Gambar 4.2. Tampilan Menu Maps	49
Gambar 4.3. Tampilan Peta Untuk Wilayah Provinsi	57
Gambar 4.4. Tampilan Peta Untuk Wilayah Kota	58
Gambar 4.5. Tampilan Peta Untuk Wilayah Kecamatan	58
Gambar 4.6. Tampilan Rekapitulasi Data	59
Gambar 4.7. Tampilan Evaluasi Model	60
Gambar 4.8. Hasil Evaluasi Model BERT, BLEU dan ROUGE	63
Gambar 4.9. Hasil Penilaian Ringkasan Berita Orang Umum	63
Gambar 4.10. Hasil Penilaian Ringkasan Berita Tenaga Medis	64
Gambar 4.11. Hasil Penilaian Ringkasan Berita Ahli Bahasa	65

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Dalam era digital dan perkembangan teknologi informasi, data berita menjadi sumber informasi yang melimpah. Berita menyajikan informasi terkini tentang peristiwa, termasuk penyebaran penyakit endemik di berbagai wilayah. Penyakit endemik adalah penyakit menular yang terjadi secara konsisten di wilayah atau populasi tertentu, disebabkan oleh faktor-faktor seperti iklim, kondisi geografis, dan terbatasnya akses pencegahan serta pengobatan. Penyakit ini bisa menjadi ancaman serius bagi keselamatan manusia jika tidak segera ditangani. Oleh karena itu, negara-negara bahkan organisasi internasional melakukan berbagai upaya untuk mencegah dan mengatasi penyebaran penyakit endemik demi menjaga kesehatan masyarakat (Yuliyanti, 2020).

Namun, berita memiliki keterbatasan dalam menyajikan informasi yang terstruktur dan komprehensif untuk dianalisis secara langsung. Informasi yang tersebar di berbagai media sering kali tidak terstandarisasi, sulit diintegrasikan, dan kurang memadai dalam menggambarkan dimensi geografis yang penting untuk memantau penyebaran penyakit. Data ini juga bersifat terpisah-pisah, sehingga sulit untuk langsung digunakan dalam analisis spasial.

Peta merupakan alat yang penting dalam navigasi dan penyediaan informasi geografis. Peta berperan krusial dalam berbagai aspek kehidupan manusia, dari perencanaan perjalanan hingga penelitian dan analisis di bidang geografi. Dalam konteks penyebaran penyakit, peta sangat berguna untuk memberikan gambaran spasial tentang lokasi terdampak. Namun, keakuratan dan kelengkapan peta sangat penting untuk memastikan bahwa informasi yang disajikan selalu relevan dan terkini. Oleh karena itu, pembaruan dan pengayaan anotasi peta menjadi hal yang sangat diperlukan, terutama untuk menyajikan informasi dinamis seperti penyebaran penyakit endemik.

Sistem Informasi Geografis (SIG) merupakan alat yang sangat berguna dalam konteks ini. SIG memungkinkan integrasi, analisis, dan visualisasi data geografis dari berbagai sumber, sehingga mempermudah pemantauan dan manajemen data spasial. Dengan SIG, informasi mengenai lokasi dan penyebaran penyakit dapat dipetakan

secara visual, memberikan gambaran yang lebih jelas dan terperinci tentang dampak penyakit di berbagai wilayah. Namun, untuk memastikan bahwa peta SIG selalu terkini dan akurat, pembaruan informasi yang efisien dan otomatis sangat penting.

Sebagai solusi, penggunaan peta dengan anotasi berbasis data berita dapat memperkaya informasi yang disajikan, terutama dalam konteks penyebaran penyakit. Dengan mengintegrasikan teknologi *Natural Language Generation* (NLG), data berita terkait endemik dapat diolah dan dianotasi secara otomatis ke dalam peta. NLG adalah cabang kecerdasan buatan yang memungkinkan komputer mengubah data terstruktur menjadi narasi yang dapat dibaca manusia. Dalam pendekatan ini, data berita diolah dan dianalisis menggunakan teknik pemrosesan bahasa alami (*natural language processing*) untuk mengidentifikasi informasi yang relevan. Dalam konteks penelitian ini, data berita tentang penyebaran penyakit dapat diolah dan dianotasi langsung ke dalam peta, memungkinkan visualisasi informasi yang *real-time* dan lebih mudah dipahami. Peta yang dianotasi secara otomatis dengan informasi terkini, seperti lokasi penyebaran dan jumlah kasus, mempermudah pemantauan dan pengelolaan penyebaran penyakit dengan lebih cepat dan akurat.

Penelitian terkait pengembangan alat untuk mendukung visualisasi distribusi penyakit endemik berbasis berita online telah dilakukan menggunakan Google Earth untuk memperbarui data secara berkala dan mengekstraksi data otomatis dari web. Alat ini menampilkan informasi geografis secara real-time atau periodik, memanfaatkan data dari berbagai sumber online yang diintegrasikan langsung ke dalam *Google Earth*. Hasil penelitian menunjukkan bahwa alat ini efektif dalam memperbarui dan menampilkan data geografis, yang bermanfaat untuk visualisasi penyebaran penyakit endemik (Abidin et al., 2015).

Dalam konteks pemantauan penyakit endemik, penelitian mengenai ekstraksi otomatis dari berita online juga dilakukan untuk mengidentifikasi entitas lokasi, tanggal, dan jumlah korban dari insiden penyakit tropis. Menggunakan teknik *Natural Language Processing* (NLP), penelitian ini mengotomatiskan proses identifikasi informasi lokasi, waktu, dan jumlah korban dari berita online. Hasil penelitian ini mendukung pemantauan kesehatan masyarakat secara lebih efisien dengan memanfaatkan data dari berita online (Abidin et al., 2013).

Selain itu, penelitian tentang teknologi *Natural Language Generation* (NLG) telah diterapkan untuk menghasilkan deskripsi otomatis dari data terkait distribusi penyakit endemik. Dalam penelitian ini, data geospasial diubah menjadi kalimat yang menggambarkan lokasi terdampak secara jelas, sehingga membantu pengguna dalam memahami penyebaran penyakit di suatu area. Hasil penelitian menunjukkan bahwa NLG dapat digunakan secara efektif untuk menyampaikan informasi geospasial di dalam sistem visualisasi peta (Zhang, 2008).

Penelitian lainnya terkait visualisasi penyakit endemik di Indonesia menggunakan metode *Named Entity Recognition* (NER) dan pendekatan berbasis aturan untuk mengekstraksi informasi dari berita online. Informasi penting seperti nama penyakit, lokasi, dan waktu diekstraksi dan divisualisasikan pada peta untuk menampilkan distribusi geografis penyakit tropis di Indonesia. Hasil penelitian ini menunjukkan bahwa metode ini sangat mendukung pemantauan distribusi penyakit endemik secara geografis (Atika, 2021).

Penelitian yang berfokus pada identifikasi lokasi dan topik dari berita online juga dilakukan untuk mendukung pemantauan persebaran penyakit endemik di Indonesia. Dengan menggunakan teknik *Natural Language Processing* (NLP) khusus untuk bahasa Indonesia, penelitian ini mengekstrak informasi lokasi dan topik dari berita untuk memperlihatkan distribusi topik dan wilayah yang terkait dengan isu kesehatan lokal. Hasil penelitian ini bermanfaat untuk analisis tren dan pemantauan persebaran isu kesehatan (Aji, 2018).

Dalam mendukung penyajian informasi berita terkait penyakit endemik, penelitian yang menggunakan model transformer, seperti BERT atau BART, untuk peringkasan otomatis berita online juga telah dilakukan. Model transformer ini mampu menghasilkan ringkasan berita yang informatif dan relevan, yang esensinya tetap dipertahankan, sehingga dapat mendukung visualisasi distribusi penyakit endemik dengan ringkasan yang singkat dan efektif (Garg et al., 2021).

Berdasarkan penelitian-penelitian yang telah dilakukan sebelumnya, maka penulis melakukan penelitian berjudul “*Pembangkitan Anotasi Peta Berdasarkan Ekstraksi Data Berita Penyakit Endemik Menggunakan Pendekatan Natural Language Generation*”.

1.2. Rumusan Masalah

Penanganan penyakit endemik seperti Demam Berdarah Dengue (DBD), malaria, dan tuberkulosis (TBC) di Indonesia sering kali menghadapi tantangan dalam hal pelaporan dan penyebaran informasi yang akurat dan mudah dipahami oleh masyarakat. Berdasarkan laporan Kementerian Kesehatan RI (2020), sistem pelaporan saat ini masih menghasilkan data yang kompleks, yang sulit diinterpretasikan oleh masyarakat umum, sehingga menghambat respons cepat dalam penanggulangan wabah. Perbedaan dalam penyajian data dari berbagai sumber berita online juga memperumit upaya memperoleh informasi yang relevan dan akurat terkait persebaran penyakit-penyakit endemik ini di berbagai wilayah. Oleh karena itu, diperlukan suatu sistem yang memanfaatkan pendekatan *Natural Language Generation* (NLG) untuk mengekstraksi dan menyajikan informasi dari berita online tentang penyakit-penyakit endemik secara otomatis. Sistem ini diharapkan dapat menghasilkan anotasi peta dan kalimat yang ringkas serta mudah dipahami, sehingga informasi mengenai DBD, malaria, TBC, dan penyebarannya dapat diakses dengan lebih cepat dan efektif oleh masyarakat serta pemangku kepentingan.

1.3. Tujuan Penelitian

Tujuan penelitian ini, yaitu mengimplementasikan ekstraksi data otomatis dari berita mengenai penyakit endemik untuk menghasilkan anotasi peta menggunakan pendekatan *Natural Language Generation* (NLG).

1.4. Batasan Penelitian

Penelitian ini memiliki batasan sebagai berikut:

- a. Sumber data berita yang digunakan yaitu berita online dari portal berita nasional yang terpercaya (Detik, Kompas, Tempo).
- b. Sumber data berita dibatasi pada teks berita dari sumber-sumber berita *online* pada domain penyakit *endemic* (DBD, malaria, TBC).
- c. Rentang waktu berita dibatasi pada 5 tahun terakhir untuk memastikan relevansi data yang terkini.
- d. Fokus penelitian hanya pada pengembangan pendekatan NLG untuk menghasilkan teks anotasi peta berbahasa Indonesia.

- e. Wilayah dalam penelitian ini akan dibatasi pada tingkat kecamatan dalam wilayah Indonesia untuk memberikan spesifikasi yang lebih jelas dan relevan.

1.5. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

- a. Membantu meningkatkan pemahaman dalam melihat dan memahami hubungan spatial antara berbagai peristiwa atau informasi terkait, sehingga pengguna dapat melihat pola dan distribusi data secara lebih intuitif.
- b. Memudahkan penyajian informasi secara komprehensif, yang berguna dalam pengambilan keputusan atau analisis lebih lanjut oleh lembaga terkait, khususnya dalam konteks pemantauan penyakit endemik.
- c. Mengembangkan pendekatan otomatis untuk pembangkitan anotasi peta, yang dapat menghemat waktu dan usaha manusia dalam memperoleh informasi relevan dengan lebih cepat dan efisien.

1.6. Metodologi Penelitian

Tahapan-tahapan yang akan dilakukan dalam penelitian ini adalah sebagai berikut:

- a. Studi Literatur
Mengumpulkan referensi dari jurnal, buku, artikel dan sumber lainnya terkait Berita, NLG, dan Pemetaan.
- b. Analisis Permasalahan
Menganalisis hasil studi literatur untuk memahami metode NLG dan masalah yang akan diselesaikan, yaitu pembangkitan anotasi peta berdasarkan ekstraksi data berita endemik menggunakan pendekatan NLG.
- c. Perancangan Sistem
Merancang sistem berdasarkan analisis studi literatur, termasuk perancangan arsitektur umum, pengumpulan data, dan perancangan sistem.
- d. Implementasi
Mengimplementasikan rancangan ke dalam kode pemrograman sesuai dengan analisis dan perancangan sebelumnya.

e. Pengujian

Menguji sistem yang telah dibuat untuk mengevaluasi implementasi metode NLG dalam ekstraksi data berita untuk pembangkitan anotasi peta

f. Dokumentasi dan Penyusunan Laporan

Mendokumentasikan dan menyusun laporan hasil evaluasi dari seluruh tahapan penelitian yang telah dilakukan.

1.7. Sistematika Penulisan

Sistematika penulisan skripsi ini terdiri dari lima bagian, yaitu:

Bab 1: Pendahuluan

Bab ini mencakup latar belakang penelitian, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

Bab 2: Landasan Teori

Bab ini menyajikan teori-teori yang diperlukan untuk memahami permasalahan dalam penelitian ini. Termasuk penjelasan tentang teori yang terkait dengan Berita, NLG, Pemetaan, serta beberapa penelitian terdahulu.

Bab 3: Metodologi Penelitian

Bab ini menguraikan arsitektur umum, analisis, dan penerapan metode NLG dalam ekstraksi data berita untuk pembangkitan anotasi peta menggunakan *input link* artikel berita. Selain itu, bab ini juga membahas tentang perancangan sistem.

Bab 4: Implementasi dan Pengujian

Bab ini memaparkan implementasi dari hasil analisis dan perancangan sistem yang telah dijelaskan di Bab 3, serta hasil pengujian terhadap sistem yang telah dibangun.

Bab 5: Kesimpulan dan Saran

Bab ini berisi kesimpulan dari keseluruhan penelitian yang telah dilakukan dan saran-saran untuk pengembangan penelitian selanjutnya.

BAB 2

LANDASAN TEORI

2.1. *Natural Language Generation*

Natural Language Generation (NLG) adalah cabang dari kecerdasan buatan (AI) dan linguistik komputasional yang bertujuan untuk mengembangkan sistem yang mampu menghasilkan teks alami dari data non-linguistik (Gatt & Krahmer, 2018). Sistem NLG biasanya bekerja melalui serangkaian tahapan mulai dari memahami data input, menentukan informasi penting, hingga menyusun kalimat yang sesuai untuk disajikan kepada pengguna dalam bentuk yang mudah dipahami (Reiter & Dale, 2000).

Secara umum, proses NLG dapat dibagi menjadi beberapa langkah utama, yaitu konten determinasi, *microplanning*, dan *realization*:

- a. Konten Determinasi (*Content Determination*): Menentukan informasi atau pesan utama yang akan disampaikan berdasarkan data input.
- b. *Microplanning*: Tahap di mana sistem memutuskan struktur kalimat, gaya bahasa, serta pemilihan kata yang sesuai dengan konteks dan audiens.
- c. *Realization*: Tahap akhir di mana teks dirangkai dalam bahasa alami sehingga dapat dipahami oleh manusia (Gatt & Krahmer, 2018).

NLG juga memiliki beragam aplikasi di berbagai bidang. Salah satu contoh penting adalah dalam jurnalisme otomatis, di mana NLG dapat digunakan untuk menghasilkan berita atau laporan secara otomatis dari data yang ada (Dale, 2016). Selain itu, NLG sering diterapkan dalam sistem layanan pelanggan untuk menghasilkan respons otomatis terhadap pertanyaan umum dan dalam laporan bisnis untuk menyajikan analisis data yang ringkas dan informatif (McKeown, 1992).

Dalam penelitian terbaru, model berbasis deep learning seperti transformer telah banyak digunakan dalam sistem NLG modern, di mana model-model ini mampu menghasilkan teks yang lebih koheren dan relevan dengan konteks. Model transformer seperti BERT dan GPT telah menunjukkan kemampuan yang tinggi dalam berbagai tugas NLG, termasuk peringkasan teks, terjemahan, dan penyusunan kalimat otomatis, sehingga memperluas aplikasi NLG ke ranah yang lebih kompleks dan beragam (Radford et al., 2019).

2.2. Pemetaan

Pemetaan dapat diartikan sebagai proses pengelompokan wilayah berdasarkan faktor-faktor geografis seperti dataran tinggi, pegunungan, serta sumber daya dan potensi penduduk yang memiliki dampak terhadap aspek sosial dan budaya. Proses ini menggunakan skala yang tepat untuk menunjukkan karakteristik wilayah tertentu (Munir, 2012). Sementara itu, peta merupakan representasi dua dimensi yang menggambarkan keseluruhan atau sebagian permukaan bumi pada bidang datar dengan skala yang telah ditentukan (Nasution, 2016). Oleh karena itu, dalam konteks penelitian ini, pemetaan dipahami sebagai pengumpulan data awal yang berfungsi untuk menciptakan peta. Proses ini melibatkan penggambaran distribusi kondisi alamiah suatu wilayah secara geografis dengan memindahkan realitas ke dalam peta dasar yang disesuaikan dengan skala peta.

2.3. Berita

Berita adalah laporan atau informasi mengenai kejadian atau peristiwa yang dianggap penting dan perlu disampaikan kepada publik. Sebagai bagian penting dalam kehidupan manusia, berita menyediakan informasi yang membantu masyarakat untuk memahami peristiwa yang terjadi di sekitar mereka dan di seluruh dunia.

Dalam Kamus Besar Bahasa Indonesia (KBBI), berita diartikan sebagai cerita atau laporan tentang kejadian atau peristiwa yang sedang hangat diperbincangkan di masyarakat (Kementerian Pendidikan dan Kebudayaan, 2018).

Berita dari Kompas, Detik, dan Tempo dipilih sebagai sumber data karena merupakan media massa terkemuka di Indonesia yang memiliki reputasi baik dalam menyajikan informasi yang akurat dan terpercaya. Ketiga media ini mematuhi standar jurnalistik yang ketat dan telah diakui oleh lembaga-lembaga independen seperti Dewan Pers dan Aliansi Jurnalis Independen (AJI).

Kompas dikenal dengan penyajian berita yang mendalam dan objektif, Detik mengedepankan kecepatan dan verifikasi informasi, sementara Tempo terkenal dengan jurnalistik investigatif yang mengedepankan kebenaran. Ketiga media ini memiliki tim redaksi yang berkompeten dan berkomitmen terhadap integritas pemberitaan, menjadikannya pilihan yang valid dan dapat dipercaya dalam penelitian ini.

Penelitian oleh Suryanto & Hidayat (2019) menunjukkan bahwa media-media besar seperti Kompas, Detik, dan Tempo memiliki sistem redaksi yang kuat yang memverifikasi setiap informasi sebelum dipublikasikan, yang menjamin kualitas berita yang disajikan.

2.4. Korpus

Korpus adalah kumpulan teks terstruktur yang berisi kata atau kalimat dalam jumlah besar. Korpus dapat terdiri dari teks dalam satu bahasa (*monolingual*) atau berbagai bahasa (*multilingual*), dan disimpan dalam bentuk file teks. Korpus ini sangat berguna dalam penelitian linguistik, seperti Pengajaran bahasa, Statistik, Dialektologi, serta Linguistik Historis (Bahren, 2021). Korpus memiliki empat karakteristik utama dalam analisis *linguistik*:

a. Sampling dan Representativitas

Dalam pembangunan korpus bahasa, penting untuk memastikan bahwa sampel yang diambil mewakili keragaman bahasa dari sumber yang relevan. Representativitas korpus berarti mencakup berbagai variasi bahasa yang sesuai dengan kebutuhan penelitian, baik dalam konteks, struktur, maupun isi. Untuk memperoleh korpus yang akurat, sampel harus diambil dari sumber dokumen asli yang mencakup variasi penggunaan bahasa yang luas, sehingga informasi yang terkandung di dalamnya benar-benar merefleksikan kecenderungan bahasa yang sedang diteliti.

b. Ukuran Terbatas

Korpus bahasa biasanya memiliki ukuran tertentu, seperti satu juta kata, yang menggambarkan batasan pada jumlah data. Walaupun teks baru dapat ditambahkan, korpus harus tetap terukur dan sesuai rencana riset. Manfaat utama dari penambahan teks adalah memberikan cakupan yang lebih luas terhadap bahasa yang digunakan dalam berbagai konteks. Namun, tantangan dari korpus yang terus bertambah adalah menjaga konsistensi dan representativitas sampel. Oleh karena itu, penting untuk merencanakan secara rinci jumlah sampel dan kata yang akan dikumpulkan pada tahap awal pembangunan korpus, agar hasil akhirnya dapat digunakan secara optimal.

c. Format yang Dapat Dibaca Mesin

Korpus yang disimpan dalam format digital atau dapat dibaca mesin memiliki banyak keuntungan dibandingkan dengan format fisik. Keunggulan utama dari korpus digital adalah kemudahan untuk pencarian dan manipulasi data dengan perangkat lunak secara efisien. Misalnya, pencarian kata atau frasa tertentu dalam korpus digital dapat dilakukan dalam waktu singkat menggunakan perangkat lunak khusus. Selain itu, korpus yang dapat dibaca mesin juga memungkinkan penambahan informasi tambahan secara cepat dan mudah, yang dapat memperkaya analisis data.

d. Referensi Standar

Meskipun tidak termasuk hal yang penting dari definisi suatu korpus, tetapi ada juga pemahaman bahwa korpus merupakan referensi *standard* untuk berbagai Bahasa yang diwakilinya. Hal ini mengandaikan ketersediaan yang luas kepada peneliti lain. Keuntungan dari korpus yang tersedia secara luas adalah bahwa akan memberikan tolok ukur yang dapat digunakan sebagai pembanding dalam studi. Misalnya. Secara langsung dibandingkan dengan hasil yang dipublikasikan (selama metodologi sama) tanpa perlu perhitungan ulang. Korpus *standard* juga berarti penggunaan korpus yang sama digunakan (Dash et al., 2001).

2.5. *Machine Learning*

Machine learning merupakan serangkaian teknik yang dapat membantu dalam menangani dan memprediksi data yang sangat besar dengan cara mempresentasikan data-data tersebut dengan algoritma pembelajaran (Danukusumo, 2017). Istilah *machine learning* pertama kali didefinisikan oleh Arthur Samuel pada tahun 1959. Menurut Arthur Samuel, *machine learning* adalah suatu bidang ilmu komputer yang memberikan kemampuan pembelajaran kepada komputer untuk mengetahui sesuatu tanpa pemrograman yang jelas. Menurut (Mohri et al., 2018) *machine learning* dapat didefinisikan sebagai metode komputasi berdasarkan pengalaman untuk meningkatkan performa atau membuat prediksi yang akurat. Definisi pengalaman disini ialah informasi sebelumnya yang telah tersedia dan bisa dijadikan data pembelajar.

Dalam pembelajaran *machine learning*, terdapat skenario-skenario seperti:

- a. *Supervised Learning* Penggunaan skenario *supervised learning*, pembelajaran menggunakan masukan data pembelajaran yang telah diberi label. Setelah itu membuat prediksi dari data yang telah diberi label.
- b. *Unsupervised Learning* Penggunaan skenario *unsupervised learning*, pembelajaran menggunakan masukan data pembelajaran yang tidak diberi label. Setelah itu mencoba untuk mengelompokkan data berdasarkan karakteristik-karakteristik yang ditemui.
- c. *Reinforcement learning* Pada skenario *reinforcement learning* fase pembelajaran dan tes saling dicampur. Untuk mengumpulkan informasi pembelajar secara aktif dengan berinteraksi ke lingkungan sehingga untuk mendapatkan balasan untuk setiap aksi dari pembelajar.

2.6. *Bidirectional and Auto-Regressive Transformers*

Abstractive summarization adalah salah satu tugas yang paling menantang dalam pemrosesan bahasa alami. Model harus mampu memahami konteks dokumen panjang sambil menghasilkan ringkasan yang memuat informasi penting. Model-model yang berhasil menyelesaikan tugas ini akan mengekstrak pola teks, mencocokkannya dengan pola yang telah dipelajari di *decoder*, dan meregenerasi ringkasan dengan *encoder* bahasa mereka sendiri. Salah satu model atau metode yang dapat dengan baik melakukannya adalah BART (Chen et al., 2023). *Bidirectional and Auto-Regressive Transformers* (BART) adalah jenis model transformator yang memadukan keunggulan pemodelan *bidirectional* (dua arah) dan *auto-regressive* (autoregresif) untuk berbagai tugas pemrosesan bahasa alami (Kanungo et al., 2022).

BART menggunakan arsitektur transformator berbasis NMT (*Neural Machine Translation*) dengan *encoder bidirectional* (mirip BERT) dan *decoder left-to-right* (mirip GPT). *Pre-training* BART dilakukan dalam dua tahap: pertama, teks dirusak dengan fungsi *noising* arbitrer; kedua, model dilatih untuk merekonstruksi teks asli dari teks yang telah dirusak. Teknik-teknik yang digunakan untuk melatih model BART agar lebih efektif dalam berbagai tugas pemrosesan bahasa alami dengan memperbaiki teks yang telah dirusak selama *pre-training* adalah sebagai berikut (Lewis et al., 2019).

a. *Token Masking*

Token-token tertentu dipilih secara acak dan digantikan dengan token [MASK].

b. *Token Deletion*

Beberapa token dihapus secara acak dari *input*, dan model harus mengidentifikasi posisi token yang dihapus tersebut.

c. *Text Infilling*

Sebagian teks dengan panjang tertentu diambil (mengikuti distribusi *Poisson*), kemudian seluruh bagian teks tersebut digantikan dengan *token [MASK]* dan dimasukkan span dengan panjang 0.

d. *Sentence Permutation*

Dokumen dibagi menjadi kalimat-kalimat yang kemudian diacak urutannya.

e. *Document Rotation*

Token tertentu dipilih secara acak, dan dokumen dirotasi berdasarkan token tersebut sehingga dokumen dimulai dari token yang dipilih. Model harus mengidentifikasi bagian mana yang merupakan awal dokumen.

Setelah menerapkan teknik-teknik yang telah dijelaskan sebelumnya, proses BART dalam menghasilkan ringkasan dari teks panjang melibatkan langkah-langkah berikut:

a. *Input data*

Proses dimulai dengan menerima input berupa teks panjang. Sebagai contoh, artikel berita berikut dibaca oleh model:

"Kota Bandung mengalami masalah polusi udara yang semakin parah akibat peningkatan kendaraan bermotor dan aktivitas industri. Dinas Lingkungan Hidup mencatat bahwa kadar partikel PM10 di udara telah melampaui ambang batas yang ditetapkan. Warga diminta untuk mengurangi aktivitas di luar ruangan dan menggunakan masker saat beraktivitas. Pemerintah kota berencana untuk mengimplementasikan kebijakan pembatasan lalu lintas untuk mengurangi emisi kendaraan."

Teks ini memberikan informasi yang cukup panjang dan kompleks mengenai masalah polusi udara di Bandung.

b. *Preprocessing*

Sebelum model BART memproses teks, langkah-langkah preprocessing dilakukan untuk membersihkan dan menyiapkan data. Langkah-langkah ini meliputi:

- *Case Folding*: Mengubah semua huruf dalam teks menjadi huruf kecil untuk konsistensi.
- *Tokenization*: Memecah teks menjadi token, contohnya:
["kota", "bandung", "mengalami", "masalah", "polusi", "udara", "yang", "semakin", "parah", ...]
- *Stopword Removal*: Menghapus kata-kata yang tidak memberikan nilai informasi penting, seperti "yang", "dan", "untuk", sehingga model dapat fokus pada kata-kata yang lebih bermakna.

c. *Encoding*

Setelah preprocessing, teks yang telah dibersihkan diteruskan ke encoder BART. Pada tahap ini, encoder bekerja secara bidirectional, memproses kata-kata dari kedua arah untuk memahami konteks kalimat secara menyeluruh. Misalnya, model memahami bahwa "*polusi udara*" adalah masalah utama, sedangkan "*kendaraan bermotor*" dan "*aktivitas industri*" adalah penyebabnya.

d. *Decoding*

Setelah proses encoding selesai, informasi yang dipelajari diteruskan ke decoder BART. Decoder berfungsi secara autoregressive, di mana model menghasilkan ringkasan dengan memproduksi kata satu per satu.

Dalam hal ini, model mungkin mulai dengan kata "*Polusi*" dan melanjutkan dengan informasi penting:

"Polusi udara di Bandung semakin parah."

"Dinas Lingkungan Hidup mencatat kadar PM10 melampaui batas."

"Warga diminta menggunakan masker."

"Pemerintah berencana membatasi lalu lintas."

e. *Output*

Setelah proses decoding selesai, hasil akhir dari model adalah ringkasan yang dihasilkan. Ringkasan tersebut adalah sebagai berikut:

"Polusi udara di Bandung semakin parah akibat peningkatan kendaraan dan industri. Dinas Lingkungan Hidup mencatat kadar PM10 melampaui batas. Warga diminta untuk menggunakan masker dan pemerintah berencana membatasi lalu lintas."

Dalam ringkasan ini, BART berhasil menyaring informasi kunci dari teks panjang dan menyusunnya dalam format yang lebih singkat, sehingga tetap mempertahankan esensi dari berita tersebut.

BART menunjukkan performa yang sangat baik saat di-*finetune* untuk tugas generasi teks, namun juga bekerja secara efektif untuk tugas pemahaman bahasa. Model ini mencapai hasil *state-of-the-art* dalam tugas dialog abstraktif, tanya-jawab, dan peringkasan abstraktif. Khususnya, BART menghasilkan peningkatan performa sebesar 6 poin ROUGE pada *dataset* XSum, yang berisi ringkasan berita yang sangat abstraktif (Lewis et al., 2019).

2.7. *Term Frequency*

Term Frequency (TF) adalah salah satu metode dalam analisis teks yang digunakan untuk mengukur frekuensi kemunculan suatu kata dalam sebuah dokumen. (Jurafsky et al., 2021). Semakin sering kata tersebut muncul, semakin tinggi nilainya dalam konteks dokumen tersebut. TF membantu menentukan pentingnya suatu istilah dalam dokumen yang sedang dianalisis, terutama jika dibandingkan dengan dokumen lainnya.

Persamaan 2.1 menjelaskan persamaan dari *Term Frequency*.

$$TF(t) = \frac{f(t,d)}{N} \quad (2.1)$$

Keterangan:

t : Istilah yang dianalisis.

$f(t, d)$: Frekuensi kemunculan istilah t dalam dokumen d .

N : Jumlah total kata dalam dokumen d .

Dalam konteks penelitian ini, TF digunakan untuk menilai relevansi kata dalam berita yang berkaitan dengan penyakit endemik. Kata-kata dengan frekuensi tinggi dalam artikel berita sering kali menggambarkan topik utama dan dapat digunakan untuk menentukan kalimat yang penting.

2.8. *Sentence Scoring*

Sentence Scoring adalah teknik yang digunakan untuk memberikan nilai atau skor pada setiap kalimat dalam dokumen berdasarkan beberapa kriteria, seperti frekuensi kata, panjang kalimat, posisi kalimat, atau kemunculan kata kunci tertentu (Lin et al., 2004).

Persamaan 2.2 menjelaskan persamaan dari *Sentence Scoring*.

$$Score(S) = \sum_{t \in S} TF(t) + W_p(S) + W_l(S) \quad (2.2)$$

Keterangan:

$Score(S)$: Skor kalimat S .

$TF(t)$: Frekuensi kata t dalam kalimat S .

$W_p(S)$: Bobot berdasarkan posisi kalimat S dalam dokumen.

$W_l(S)$: Bobot berdasarkan Panjang kalimat S .

Proses *Sentence Scoring*:

- **Frekuensi Kata:** Menggunakan TF untuk menentukan relevansi kata dalam sebuah kalimat. Kalimat dengan banyak kata penting mendapatkan skor lebih tinggi.
- **Posisi Kalimat:** Kalimat awal dan akhir dalam dokumen sering kali mengandung informasi penting.

- Panjang Kalimat: Kalimat yang terlalu panjang atau terlalu pendek dapat memengaruhi skor karena cenderung kurang informatif.

Sentence Scoring membantu memilih kalimat yang relevan dari berita tentang penyakit endemik untuk dianotasi ke dalam peta. Kalimat dengan skor tinggi dianggap mengandung informasi penting dan signifikan.

2.9. Penelitian Terdahulu

Sampai saat ini beberapa penelitian mengenai *Natural Language Generation* sudah banyak dilakukan oleh beberapa peneliti untuk membangkitkan sebuah kalimat diantaranya seperti :

Penelitian T. F Abidin, M. Subianto, T. A. Gani & R. Ferdhiana (2015) *Periodic update and automatic extraction of web data for creating a Google Earth based tool* yaitu membuat sebuah *system Google Earth* berdasarkan ekstraksi berita *online* yang *ter-update*. Dalam penelitian ini, menggunakan *Support Vector Machine* (SVM) untuk mengidentifikasi berita *online* yang terkait yang akan diekstrak dan digunakan dalam membuat *system google earth*.

Penelitian T. F Abidin, R . Ferdhiana & Hajjul Kamil (2013) *Automatic Extraction of Place Entities and Sentences Containing the Date and Number of Victims of Tropical Disease Incidence from the Web* yaitu mengidentifikasi dan mengekstrak lokasi kejadian dan menggunakan *Support Vector Machine* (SVM) untuk menentukan kalimat yang berisi tanggal kejadian dan jumlah korban. Dalam penelitian ini, algoritma berbasis aturan secara otomatis mengidentifikasi lokasi kejadian penyakit tropis dari web yang diberikan.

Penelitian XueYing Zhang (2008) *Route description using Natural Language Generation Technology* yaitu menghasilkan deskripsi rute jalan berdasarkan *dataset* sistem informasi geografis dan teknologi NLG. Dalam penelitian ini, menggunakan pendekatan yang mengambil *dataset* masukan dari GIS dan menggunakan teknik NLG dalam membangun keluaran tekstual dari rencana rute.

Penelitian Atika Dessy (2021) Ekstraksi Informasi Berita *Online* dengan *Named Entity Recognition* (NER) dan Rule-Based untuk Visualisasi Penyakit Tropis di Indonesia yaitu pada penelitian ini penulis memanfaatkan *Named Entity Recognition* (NER) dengan *Library Spacy* untuk ekstraksi informasi penyakit tropis di Indonesia berupa entitas terkait nama penyakit tropis, lokasi kejadian, waktu kejadian dan jumlah korban yang akan di visualisasikan dengan memanfaatkan *Rules-Bases*. *Library SpaCy* membangun ekstraksi informasi atau sistem *Natural Language Processing* memproses teks untuk pembelajaran mendalam pada pengenalan entitas.

Penelitian Aziz Mustika Aji (2018) Ekstraksi Lokasi Dan Topik Pada Berita *Online* Berbahasa Indonesia yaitu pada penelitian ini dibangun sistem ekstraksi informasi khususnya lokasi dan topik dari suatu teks berita *online*. Ekstraksi lokasi dilakukan dengan menggunakan pengenalan entitas bernama (*named entity recognition*) dengan metode *Hidden Markov Model* (HMM). Sedangkan metode ekstraksi topik yang digunakan adalah algoritma term *frequency and proportional document frequency* (TF*PDF).

Penelitian Garg, Apar, *et al.*, (2020) *NEWS Article Summarization with Pre-trained Transformer* yaitu penggunaan model *pre-trained transformer* untuk peningkatan aplikasi NLP, khususnya dalam tugas *text sumarization*. Dua model *pre-trained* yang digunakan, yaitu *Text-to-Text Transfer Transformer* (menggunakan model *pre-trained* T5 dan model yang dihasilkan dari *transfer learning* pada T5) dan *Bidirectional Encoder Representations from Transformers* (menggunakan model BART dan *scratch* model *encoder-decoder*).

Tabel 2.1 Penelitian Terdahulu

NO	Peneliti	Judul Penelitian	Tahun	Yang Dilakukan
1	T. F Abidin, M. Subianto, T. A. Gani & R. Ferdhiana	<i>Periodic update and automatic extraction of web data for creating a Google Earth based tool</i>	2015	Membuat sebuah <i>system Google Earth</i> berdasarkan ekstraksi berita <i>online</i> yang terbaru. Dalam penelitian ini, menggunakan <i>Support Vector Machine</i> (SVM) untuk mengidentifikasi berita <i>online</i> yang terkait yang akan diekstrak dan digunakan dalam membuat <i>system google earth</i> .
2	T. F Abidin, R. Ferdhiana & Hajjul Kamil	<i>Automatic Extraction of Place Entities and Sentences Containing the Date and Number of Victims of Tropical Disease Incidence from the Web</i>	2013	Mengidentifikasi dan mengekstrak lokasi kejadian dan menggunakan <i>Support Vector Machine</i> (SVM) untuk menentukan kalimat yang berisi tanggal kejadian dan jumlah korban. Dalam penelitian ini, algoritma berbasis aturan secara otomatis mengidentifikasi lokasi kejadian penyakit tropis dari web yang diberikan.
3	XueYing Zhang	<i>Route description using Natural Language Generation Technology</i>	2008	Menghasilkan deskripsi rute jalan berdasarkan <i>dataset</i> sistem informasi geografis dan teknologi NLG. Dalam penelitian ini, menggunakan pendekatan yang mengambil <i>dataset</i> masukan dari GIS dan menggunakan teknik NLG dalam

				membangun keluaran tekstual dari rencana rute.
4	Atika Dessy	Ekstraksi Informasi Berita <i>Online</i> dengan <i>Named Entity Recognition</i> (NER) dan <i>Rule-Based</i> untuk Visualisasi Penyakit Tropis di Indonesia	2021	Pada penelitian ini penulis memanfaatkan <i>Named Entity Recognition</i> (NER) dengan <i>Library Spacy</i> untuk ekstraksi informasi penyakit tropis di Indonesia berupa entitas terkait nama penyakit tropis, lokasi kejadian, waktu kejadian dan jumlah korban yang akan di visualisasikan dengan memanfaatkan <i>Rules-Bases</i> . <i>Library SpaCy</i> membangun ekstraksi informasi atau sistem <i>Natural Language Processing</i> memproses teks untuk pembelajaran mendalam pada pengenalan entitas.
5	Aziz Mustika Aji	Ekstraksi Lokasi Dan Topik Pada Berita <i>Online</i> Berbahasa Indonesia	2018	Pada penelitian ini dibangun sistem ekstraksi informasi khususnya lokasi dan topik dari suatu teks berita <i>online</i> . Ekstraksi lokasi dilakukan dengan menggunakan pengenalan entitas bernama (<i>named entity recognition</i>) dengan metode <i>Hidden Markov Model</i> (HMM). Sedangkan metode ekstraksi topik yang digunakan adalah algoritma term <i>frequency and proportional document frequency</i> (TF*PDF).

6	Garg, Apar, <i>et al.</i> ,	<i>NEWS Article Summarization with Pre-trained Transformer</i>	2020	Pada penelitian ini penulis menggunakan model <i>pre-trained transformer</i> untuk peningkatan aplikasi NLP, khususnya dalam tugas <i>text sumarization</i> . Dua model <i>pre-trained</i> yang digunakan, yaitu <i>Text-to-Text Transfer Transformer</i> (menggunakan model <i>pre-trained T5</i> dan model yang dihasilkan dari <i>transfer learning</i> pada T5) dan <i>Bidirectional Encoder Representations from Transformers</i> (menggunakan model BART dan <i>scratch</i> model <i>encoder-decoder</i>).
---	-----------------------------	--	------	---

Penelitian mengenai ekstraksi informasi, peringkasan teks dan pembangkitan kalimat telah banyak dilakukan, khususnya dalam konteks visualisasi data geografis dan penyebaran penyakit. Beberapa penelitian terdahulu telah mengimplementasikan berbagai metode seperti *Support Vector Machine* (SVM), *Named Entity Recognition* (NER), dan teknologi *Natural Language Generation* (NLG) untuk mengekstraksi informasi dari berita *online* dan memvisualisasikannya dalam bentuk peta atau deskripsi teks. Berikut merupakan perbedaan penelitian yang akan dilakukan dengan penelitian sebelum – sebelumnya.

a. Integrasi NLG untuk Anotasi Peta Otomatis

Penelitian ini akan mengimplementasikan pendekatan NLG untuk membangkitkan anotasi peta secara otomatis berdasarkan ekstraksi data berita endemik. Hal ini berbeda dengan penelitian Atika (2021) yang menggunakan NER dan *rule-based* untuk visualisasi, namun tidak memanfaatkan NLG untuk menghasilkan anotasi teks secara otomatis pada peta.

b. Fokus pada Penyakit Endemik dan Anotasi Dinamis

Sementara penelitian sebelumnya seperti Abidin et al. (2013, 2015) dan Atika (2021) membahas penyakit tropis secara umum, penelitian ini lebih spesifik pada penyakit endemik. Ini akan memberikan informasi yang lebih akurat kepada pengguna.

c. Penggunaan Teknologi NLG dalam Bahasa Indonesia

Penelitian Zhang (2008) menunjukkan penggunaan NLG dalam menghasilkan deskripsi rute, namun dalam konteks bahasa dan wilayah yang berbeda. Penelitian ini akan mengembangkan pendekatan NLG untuk membangkitkan teks anotasi peta dalam bahasa Indonesia, yang belum banyak dieksplorasi dalam penelitian sebelumnya.

d. Implementasi pada Sistem Informasi Geografis Wilayah Indonesia

Penelitian ini akan menerapkan hasil ekstraksi dan pembangkitan kalimat langsung pada peta wilayah Indonesia, sehingga memberikan kontribusi praktis dalam pemantauan penyebaran penyakit endemik di Indonesia.

e. Mengatasi Keterbatasan Visualisasi Statis

Penelitian sebelumnya cenderung menghasilkan visualisasi statis atau deskripsi teks terpisah. Penelitian ini akan mengintegrasikan informasi langsung ke dalam peta melalui anotasi yang dibangkitkan, memberikan konteks spasial dan tekstual secara simultan.

BAB 3

ANALISIS DAN PERANCANGAN SISTEM

3.1. Arsitektur Umum

Dalam penelitian ini, metodologi yang digunakan bertujuan untuk membangkitkan anotasi peta berdasarkan ekstraksi data dari berita mengenai penyakit endemik. Arsitektur sistem ini dirancang untuk mengintegrasikan berbagai komponen yang saling terhubung, sehingga informasi yang dihasilkan akurat dan relevan bagi pengguna.

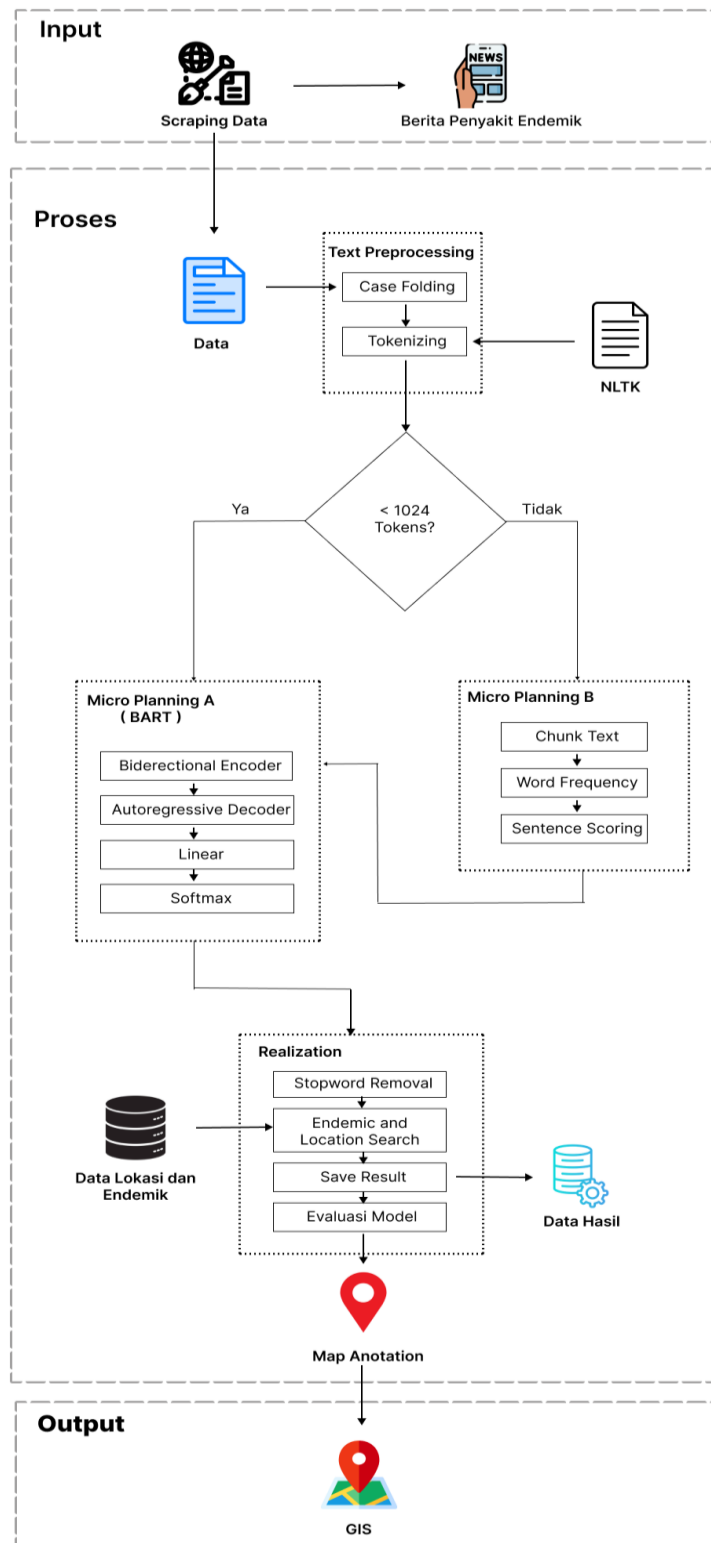
Proses dimulai dengan pengumpulan data dari berbagai portal berita yang melaporkan kasus penyakit endemik. Data yang diperoleh kemudian akan diproses menggunakan web scraping untuk mengumpulkan konten berita secara otomatis. Setelah itu, langkah preprocessing dilakukan untuk membersihkan dan menyiapkan data, meliputi *case folding*, *tokenization*, dan *stopword removal*.

Dalam penelitian ini, dua model digunakan untuk analisis dan pemrosesan data. Model pertama adalah *BART (Bidirectional and Auto-Regressive Transformers)*, yang dipilih karena kemampuannya dalam memahami konteks teks panjang dan menghasilkan ringkasan yang koheren. *BART* menggunakan pendekatan abstractive summarization, yang memungkinkan model untuk merumuskan ulang informasi dengan cara yang lebih natural dan informatif. Pendekatan ini sangat berguna dalam konteks berita, di mana penyampaian informasi yang ringkas dan padat sangat penting.

Model kedua menggunakan pendekatan *word frequency*, yang berfungsi untuk memberikan analisis berbasis statistik dari kata-kata yang paling sering muncul dalam artikel berita. Metode ini memungkinkan peneliti untuk menilai kalimat-kalimat yang relevan berdasarkan frekuensi kata, yang dapat membantu dalam mengidentifikasi informasi kunci dan membuat keputusan yang lebih baik mengenai konten yang akan diringkaskan.

Hasil dari proses peringkasan menggunakan kedua model ini akan dianotasi ke dalam peta menggunakan Sistem Informasi Geografis (GIS). Pemetaan ini bertujuan untuk memberikan visualisasi yang jelas mengenai lokasi penyebaran penyakit,

sehingga memudahkan masyarakat dan pemangku kepentingan dalam memahami situasi kesehatan yang ada.



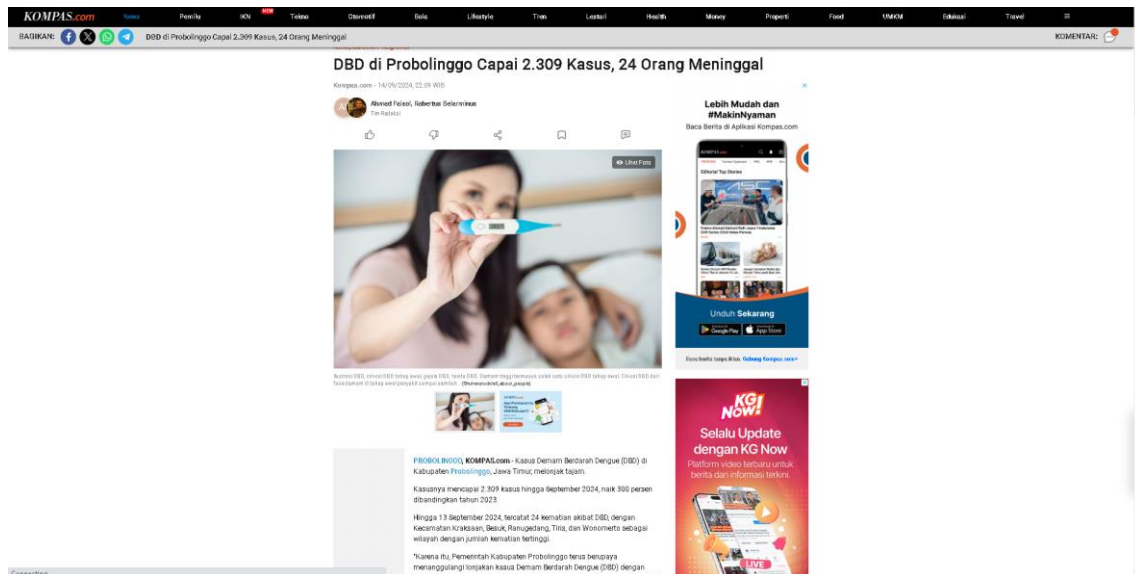
Gambar 3.1 Arsitektur Umum

Gambar 3.1 di bawah ini menunjukkan arsitektur umum dari metodologi penelitian ini, menggambarkan langkah-langkah dari pengumpulan data hingga menghasilkan anotasi peta.

3.2. *Data Input*

Input data untuk penelitian ini diambil dari berbagai portal berita *online* terpercaya yang secara khusus melaporkan dan mendiskusikan penyakit endemik. Penyakit endemik adalah penyakit yang secara konsisten hadir dalam populasi atau wilayah tertentu, seperti demam berdarah, tuberkulosis, dan malaria. Portal berita yang dipilih mencakup media berita nasional seperti Kompas, Detik, dan Tempo, yang sering melaporkan situasi kesehatan di Indonesia. Contoh kasus berita yang diambil meliputi laporan tentang lonjakan kasus demam berdarah di suatu daerah, langkah-langkah pemerintah untuk mengendalikan penyebaran, tips pencegahan bagi masyarakat, artikel tentang upaya penanganan tuberkulosis di daerah tertentu, program pengobatan yang disediakan, tantangan dalam penanggulangan penyakit, serta kisah pasien yang berhasil sembuh. Juga termasuk laporan mengenai penyebaran malaria di wilayah endemik, statistik kasus baru, upaya pemberantasan nyamuk penyebab malaria, dan penelitian terbaru mengenai vaksin malaria.

Menggunakan teknik NLG untuk informasi tersebut dibuat ringkasan berita yang dapat digunakan dalam anotasi peta dan analisis lebih lanjut, serta visualisasi data dengan membuat anotasi peta yang menunjukkan penyebaran penyakit endemik berdasarkan informasi yang dikumpulkan dari berita. Dengan proses pengumpulan dan pengolahan data yang terstruktur, penelitian ini dapat memastikan bahwa informasi yang digunakan akurat, relevan, dan berguna untuk analisis lebih lanjut serta pembuatan anotasi peta yang informatif. Contoh berita yang dapat digunakan berasal dari tautan <https://regional.kompas.com/read/2024/09/14/220942678/dbd-di-probolinggo-capai-2309-kasus-24-orang-meninggal/> yang membahas tentang “DBD di Probolinggo Capai 2.309 Kasus, 24 Orang Meninggal” seperti yang ditunjukkan pada Gambar 3.2 sebagai berikut.



Gambar 3.2 Contoh Artikel Berita Endemik DBD

3.3. *Scraping Data*

Pengambilan data dari artikel berita *online* dilakukan dengan menggunakan teknik *web scraping*, yaitu proses otomatis untuk mengumpulkan informasi dari situs web menggunakan *input link* artikel berita yang diberikan. Dalam konteks ini, penulis menggunakan *BeautifulSoup*, sebuah *library Python* yang kuat untuk mengurai dokumen HTML dan XML. Proses dimulai dengan meng-*input link* artikel berita sebagai sumber data. Langkah pertama adalah mengirim permintaan HTTP ke URL artikel menggunakan *library requests* untuk mendapatkan konten halaman web. Setelah konten halaman web berhasil diambil, *BeautifulSoup* digunakan untuk *parsing* HTML tersebut. *BeautifulSoup* mengubah struktur HTML halaman web menjadi objek yang dapat dengan mudah dinavigasi dan dicari. Dengan objek *BeautifulSoup*, kami dapat mencari elemen-elemen HTML *tag* <p> untuk paragraf mendapatkan teks berita. Penulis menggunakan metode *find_all* untuk menemukan semua *tag* paragraf dan menggabungkan teks dari setiap paragraf tersebut menjadi satu teks lengkap. Proses ini memungkinkan pengumpulan teks berita yang bersih dan terstruktur, bebas dari elemen yang tidak relevan seperti iklan dan tautan. Teks berita yang diperoleh kemudian dapat disimpan untuk keperluan analisis lebih lanjut. Proses *scraping* dilakukan menggunakan fungsi *scrape_text* seperti yang ditunjukkan pada *Pseudocode* berikut.

```

FUNCTION scrape_data(url: STRING) RETURNS STRING
  DECLARE response: HTTPResponse
  DECLARE soup: BeautifulSoup
  DECLARE paragraphs: LIST OF STRING
  DECLARE text: STRING

  response ← kirim permintaan HTTP GET ke url

  IF response berhasil THEN
    soup ← parsing konten response menggunakan
    BeautifulSoup
    paragraphs ← cari semua tag <p> dalam soup
    text ← ""

    FOR EACH paragraph IN paragraphs DO
      text ← text + paragraph.get_text() + " "

    RETURN text
  ELSE
    LOG "Kesalahan saat mengakses URL"
    RETURN ""
END FUNCTION

```

Pseudocode Tahap Scraping

Pseudocode di atas mendefinisikan fungsi *scrape_text*, yang dirancang untuk mengambil teks dari halaman web berdasarkan URL yang diberikan. Fungsi ini menerima parameter url bertipe string dan mengembalikan hasil bertipe string. Berikut adalah langkah-langkah yang dilakukan dalam fungsi ini:

- Mengirim Permintaan: Mengirim permintaan HTTP GET ke URL yang diberikan untuk mendapatkan konten halaman web.
- Parsing HTML: Jika permintaan berhasil, konten HTML akan diparsing menggunakan *BeautifulSoup*, menjadikannya lebih mudah untuk dinavigasi.
- Mencari Tag Paragraf: Menggunakan metode *find_all* untuk menemukan semua tag paragraf (<p>) dalam konten yang diparse.
- Menggabungkan Teks: Teks dari setiap paragraf dikumpulkan menjadi satu string yang lebih panjang.
- Pengembalian Hasil: Mengembalikan teks yang telah digabungkan. Jika terdapat kesalahan saat mengakses URL, mencatat pesan kesalahan dan mengembalikan string kosong.

Contoh 1. Data Hasil Scraping

{“kalimat_asli”: “DBD di Probolinggo Capai 2.309 Kasus, 24 Orang Meninggal Tim Redaksi PROBOLINGGO, KOMPAS.com - Kasus Demam Berdarah Dengue (DBD) di Kabupaten Probolinggo, Jawa Timur, melonjak tajam. Kasusnya mencapai 2.309 kasus hingga September 2024, naik 300 persen dibandingkan tahun 2023. Hingga 13 September 2024, tercatat 24 kematian akibat DBD, dengan Kecamatan Kraksaan, Besuk, Ranugedang, Tiris, dan Wonomerto sebagai wilayah dengan jumlah kematian tertinggi.”}

3.4. Text Preprocessing

Text Preprocessing adalah langkah penting dalam pemrosesan bahasa alami (NLP) untuk mempersiapkan teks mentah menjadi bentuk yang lebih terstruktur dan siap untuk analisis lebih lanjut. Pada penelitian ini pada tahap *Text Preprocessing* dilakukan menggunakan metode *Case Folding* dan *Tokenizing*.

3.4.1. Case Folding

Case folding adalah langkah dalam *preprocessing* teks yang bertujuan untuk mengubah semua huruf dalam teks menjadi huruf kecil. Proses ini penting untuk memastikan konsistensi dalam teks dan menghindari perbedaan yang disebabkan oleh kapitalisasi, sehingga kata-kata seperti “Indonesia” dan “indonesia” diperlakukan sebagai entitas yang sama. *Case folding* biasanya dilakukan pada tahap awal *preprocessing* untuk menyederhanakan teks dan mengurangi variasi yang tidak perlu, sehingga analisis lebih lanjut menjadi lebih efektif dan akurat. Metode *case folding* dilakukan menggunakan fungsi *lower* untuk *string* pada *Python* seperti yang ditunjukkan pada *Pseudocode* berikut.

```
FUNCTION case_folding(teks: STRING) RETURNS STRING
    lower_text ← ubah teks menjadi huruf kecil
    RETURN lower_text
END FUNCTION
```

Pseudocode Proses Case Folding

Fungsi *case_folding* mengubah semua huruf dalam teks menjadi huruf kecil. Langkah ini memastikan konsistensi data dalam proses analisis teks. Setelah konversi, teks yang telah diubah akan dikembalikan sebagai hasil.

Tabel 3.1 Contoh Sebelum dan Sesudah Proses Case Folding

Sebelum Case Folding	Sesudah Case Folding
DBD di Probolinggo Capai 2.309 Kasus, 24 Orang Meninggal Tim Redaksi PROBOLINGGO, KOMPAS.com - Kasus Demam Berdarah Dengue (DBD) di Kabupaten Probolinggo, Jawa Timur, melonjak tajam. Kasusnya mencapai 2.309 kasus hingga September 2024, naik 300 persen dibandingkan tahun 2023. Hingga 13 September 2024, tercatat 24 kematian akibat DBD, dengan Kecamatan Kraksaan, Besuk, Ranugedang, Tiris, dan Wonomerto sebagai wilayah dengan jumlah kematian tertinggi.	dbd di probolinggo capai 2.309 kasus, 24 orang meninggal tim redaksi probolinggo, kompas.com - kasus demam berdarah dengue (dbd) di kabupaten probolinggo, jawa timur, melonjak tajam. kasusnya mencapai 2.309 kasus hingga september 2024, naik 300 persen dibandingkan tahun 2023. hingga 13 september 2024, tercatat 24 kematian akibat dbd, dengan kecamatan kraksaan, besuk, ranugedang, tiris, dan wonomerto sebagai wilayah dengan jumlah kematian tertinggi.

3.4.2. Tokenizing

Tokenizing adalah proses memecah teks menjadi unit-unit kecil yang disebut *token*, seperti kata atau kalimat. Dengan menggunakan *Natural Language Toolkit* (NLTK), sebuah *library Python* yang kuat untuk pemrosesan bahasa alami, teks dapat dipecah dengan cepat dan efisien. NLTK menyediakan fungsi *word_tokenize* untuk memecah teks menjadi kata-kata dan *sent_tokenize* untuk memecah teks menjadi kalimat. Misalnya, *word_tokenize* (“ini adalah contoh teks.”) akan menghasilkan token [“ini”, “adalah”, “contoh”, “teks”, “.”]. Proses *tokenizing* ini penting untuk analisis teks karena memungkinkan algoritma untuk bekerja dengan unit teks yang lebih kecil dan lebih terkelola. Tokenizing dilakukan dengan memecah menjadi dua jenis data, yaitu dipecah dalam satuan kalimat dan dalam satuan kata. Proses ini dilakukan menggunakan teks yang telah diubah ke dalam bentuk *lowercase* atau huruf kecil pada proses *Case*

Folding sebelumnya untuk pemecahan dalam satuan kata sementara untuk satuan kalimat menggunakan data asli. Proses *Tokenizing* menggunakan NLTK dilakukan menggunakan *pseudocode* seperti yang ditunjukkan pada *Pseudocode* sebagai berikut.

```

FUNCTION tokenization(teks: STRING) RETURNS LIST OF STRING
  DECLARE token: LIST OF STRING

  token ← pisahkan teks menjadi token berdasarkan spasi dan
  tanda baca

  RETURN token
END FUNCTION

```

Pseudocode Proses Tokenizing

Fungsi *tokenization* memisahkan teks menjadi daftar token (kata-kata) dengan cara memisahkan berdasarkan spasi dan tanda baca. Langkah ini penting dalam analisis teks karena memungkinkan pemrosesan lebih lanjut pada setiap kata yang terpisah. Hasilnya adalah daftar token yang siap untuk dianalisis.

Tabel 3.2 Contoh Sebelum dan Sesudah Proses *Tokenizing* Satu Kalimat

Sebelum <i>Tokenizing</i>	Sesudah <i>Tokenizing</i>
dbd di probolinggo capai 2.309 kasus, 24 orang meninggal tim redaksi probolinggo, kompas.com - kasus demam berdarah dengue (dbd) di kabupaten probolinggo, jawa timur, melonjak tajam. kasusnya mencapai 2.309 kasus hingga september 2024, naik 300 persen dibandingkan tahun 2023. hingga 13 september 2024, tercatat 24 kematian akibat dbd, dengan kecamatan kraksaan, besuk, ranugedang, tiris, dan wonomerto sebagai wilayah dengan jumlah kematian tertinggi.	['dbd di probolinggo capai 2.309 kasus, 24 orang meninggal tim redaksi probolinggo, kompas.com - kasus demam berdarah dengue (dbd) di kabupaten probolinggo, jawa timur, melonjak tajam.', 'kasusnya mencapai 2.309 kasus hingga september 2024, naik 300 persen dibandingkan tahun 2023.', 'hingga 13 september 2024, tercatat 24 kematian akibat dbd, dengan kecamatan kraksaan, besuk, ranugedang, tiris, dan wonomerto sebagai wilayah dengan jumlah kematian tertinggi.']

Tabel 3.3 Contoh Sebelum dan Sesudah Proses *Tokenizing* Satu Kata

Sebelum <i>Tokenizing</i>	Sesudah <i>Tokenizng</i>
dbd di probolinggo capai 2.309 kasus, 24 orang meninggal tim redaksi probolinggo, kompas.com - kasus demam berdarah dengue (dbd) di kabupaten probolinggo, jawa timur, melonjak tajam. kasusnya mencapai 2.309 kasus hingga september 2024, naik 300 persen dibandingkan tahun 2023. hingga 13 september 2024, tercatat 24 kematian akibat dbd, dengan kecamatan kraksaan, besuk, ranugedang, tiris, dan wonomerto sebagai wilayah dengan jumlah kematian tertinggi.	['dbd', 'di', 'probolinggo', 'capai', '2.309', 'orang', 'meninggal', 'tim', 'kompas.com', '-', 'kasus', 'demam', 'berdarah', 'dengue', '(', 'dbd', ')', 'di', 'kabupaten', 'probolinggo', ',', 'jawa', 'timur', ',', 'melonjak', 'tajam', ',', 'kasusnya', 'mencapai', '2.309', 'kasus', 'hingga', 'september', '2024', ',', 'naik', '300', 'persen', 'dibandingkan', 'tahun', '2023', ',', 'hingga', '13', 'september', '2024', ',', 'tercatat', '24', 'kematian', 'akibat', 'dbd', ',', 'dengan', 'kecamatan', 'kraksaan', ',', 'besuk', ',', 'ranugedang', ',', 'tiris', ',', 'dan', 'wonomerto', 'sebagai', 'wilayah', 'dengan', 'jumlah', 'kematian', 'tertinggi', '.']

3.5. *Micro Planning A*

Micro Planning A menggunakan model BART yang telah dilatih sebelumnya untuk menghasilkan ringkasan dari artikel berita secara efisien, dan proses ini diterapkan jika jumlah token dalam artikel tidak melebihi 1024. Tahapan dalam *Micro Planning A* meliputi *Bidirectional Encoder* yang menganalisis teks dalam dua arah untuk memahami konteks penuh, *Autoregressive Decoder* yang menghasilkan teks secara berurutan dengan mempertimbangkan kata-kata sebelumnya, *Linear Layer* yang mengonversi hasil decoding menjadi vektor yang sesuai dengan vocabulary, serta *Softmax Layer* yang mengubah vektor ini menjadi probabilitas untuk memilih kata yang paling relevan. Hasil ringkasan dari *Micro Planning A* kemudian digunakan dalam tahap realisasi untuk penghapusan kata tidak penting dan pencarian lokasi endemik, sehingga ringkasan lebih akurat dan informatif.

Penggunaan dua metode terpisah, yaitu *Micro Planning A* dengan BART dan *Micro Planning B* dengan metode frekuensi, didasarkan pada keterbatasan model BART yang hanya dapat memproses hingga 1024 token dalam satu waktu (Lewis et al., 2020). Oleh karena itu, *Micro Planning B* diperlukan untuk artikel yang melebihi batas token ini, di mana teks dipecah menjadi beberapa chunk, kemudian dilakukan analisis frekuensi kata untuk mengidentifikasi bagian yang paling relevan sebelum diproses lebih lanjut. Pendekatan ini memastikan bahwa ringkasan tetap fokus pada informasi penting tanpa mengorbankan kelengkapan konten yang ingin disampaikan, terutama untuk teks yang lebih panjang atau kompleks.

3.5.1. *Bidirectional Encoder dan Autoregressive Decoder*

Tahap pertama dalam proses *Micro Planning A* adalah menggunakan *Bidirectional Encoder* untuk memahami konteks penuh dari teks input. Encoder ini menganalisis teks dalam dua arah, memungkinkan setiap kata dipahami dalam konteks sekelilingnya, baik kata sebelumnya maupun kata sesudahnya. Setelah proses encoding selesai, *Autoregressive Decoder* digunakan untuk menghasilkan teks ringkasan secara bertahap, satu token pada satu waktu, berdasarkan kata-kata yang telah dihasilkan sebelumnya. Jika teks melebihi batas panjang 1024 token, teks dipecah menjadi beberapa bagian (chunk) yang diproses secara individu, menjaga agar model tetap dapat memproses teks panjang. Proses pembuatan *Bidirectional Encoder dan Autoregressive Decoder* dilakukan menggunakan *pseudocode* seperti yang ditunjukkan pada *Pseudocode* berikut.

```
FUNCTION bidirectional_autoregressive_summarization(input_text:
STRING) RETURNS ENCODED_TEXT
  INITIALIZE encoder, decoder FROM model
  "sshleifer/distilbart-cnn-12-6"

  # Encoding dengan Bidirectional Encoder
  encoded_text ← encoder(input_text) # Memproses konteks
  teks secara dua arah

  # Decoding dengan Autoregressive Decoder
  decoded_summary ← ""
  FOR token IN encoded_text DO
    next_token ← decoder(token, context=decoded_summary)
    decoded_summary += next_token
  END FOR
```

```

RETURN decoded_summary
END FUNCTION

```

Pseudocode Proses Bidirectional Encoder dan Autoregressive Decoder

Berikut penjelasan fungsi dari proses *Bidirectional Encoder dan Autoregressive Decoder* :

- Encoding dengan *Bidirectional Encoder*: Teks input diproses oleh encoder, yang menghasilkan representasi vektor untuk setiap kata dengan mempertimbangkan konteks sekelilingnya. Ini memastikan bahwa setiap kata dalam teks dipahami dalam konteks penuh, baik sebelum maupun sesudah kata tersebut.
- Decoding dengan *Autoregressive Decoder*: Vektor yang dihasilkan oleh encoder kemudian diolah oleh decoder secara autoregresif, yaitu menghasilkan kata-kata secara berurutan. Pada setiap langkah, token berikutnya dipilih berdasarkan konteks yang telah dihasilkan sebelumnya, menjaga koherensi dan konsistensi ringkasan.
- Pengembalian Ringkasan: Setelah seluruh token dihasilkan, fungsi mengembalikan teks ringkasan yang utuh sebagai output dari tahap ini.

3.5.1. Linear Layer dan Softmax Layer

Setelah teks melalui proses *encoding* dan *decoding*, tahap berikutnya adalah *Linear Layer* dan *Softmax Layer*. *Linear Layer* bertugas mengubah *output* dari *decoder* menjadi representasi yang mencakup seluruh kosakata model, memastikan bahwa setiap kata dalam ringkasan terhubung dengan kata-kata di *vocabulary*. *Softmax Layer* kemudian mengonversi representasi ini menjadi probabilitas, yang memungkinkan model untuk memilih kata dengan probabilitas tertinggi untuk melanjutkan proses peringkasan. Proses ini diulang untuk setiap token hingga ringkasan lengkap terbentuk. Proses pembuatan *Linear Layer* dan *Softmax Layer* dilakukan menggunakan *pseudocode* seperti yang ditunjukkan pada *Pseudocode* berikut

```

FUNCTION linear_softmax_decoding(encoded_text: ENCODED_TEXT)
RETURNS SUMMARY
    INITIALIZE linear_layer, softmax_layer FROM model
    "sshleifer/distilbart-cnn-12-6"
    DECLARE summary: STRING

    FOR each_token IN encoded_text DO
        # Linear layer mengubah token hasil decoding menjadi

```

```

vektor vocabulary
    vocab_vector ← linear_layer(each_token)

    # Softmax layer menghitung distribusi probabilitas pada
vocabulary
    probabilities ← softmax_layer(vocab_vector)

    # Memilih kata dengan probabilitas tertinggi
    next_word ←
SELECT_WORD_WITH_MAX_PROBABILITY(probabilities)

    summary += next_word
END FOR

RETURN summary
END FUNCTION

```

Pseudocode Proses Linear Layer dan Softmax Layer

Berikut penjelasan fungsi dari proses *Linear Layer* dan *Softmax Layer*:

- Transformasi dengan *Linear Layer*: Setiap token hasil *decoding* diubah oleh *Linear Layer* menjadi vektor yang berhubungan dengan semua kata dalam vocabulary model, sehingga memungkinkan pemetaan ke seluruh kosakata yang tersedia.
- Perhitungan dengan *Softmax Layer*: *Softmax Layer* mengonversi vektor tersebut menjadi probabilitas untuk setiap kata dalam *vocabulary*. Kata-kata dengan relevansi konteks tinggi mendapatkan probabilitas yang lebih besar.
- Pemilihan Kata dengan Probabilitas Tertinggi: Kata dengan probabilitas tertinggi dipilih sebagai bagian dari ringkasan. Proses ini diulang untuk setiap token hingga ringkasan terbentuk secara lengkap dan sesuai dengan konteks yang diinginkan.
- Pengembalian Ringkasan Final: Setelah seluruh kata dalam ringkasan dipilih, fungsi mengembalikan teks ringkasan final yang telah disusun.

Contoh 2. Data Hasil Micro Planning A

“ Hingga 13 September 2024, tercatat 24 kematian akibat DBD, dengan Kecamatan Kraksaan, Besuk, Ranugedang, Tiris, dan Wonomerto sebagai wilayah dengan jumlah kematian tertinggi. DBD di Probolinggo Capai 2.309 Kasus, 24 Orang Meninggal Tim Redaksi PROBOLINGGO, KOMPAS.com - Kasus Demam Berdarah Dengue (DBD) di

Kabupaten Probolinggo, Jawa Timur, melonjak tajam. Kasusnya mencapai 2.309 kasus hingga September 2024, naik 300 persen dibandingkan tahun 2023.”

3.6. *Micro Planning B*

Micro Planning B dirancang untuk memproses teks panjang yang melebihi batas maksimum 1024 token pada model BART. Dalam kasus ini, teks dipecah menjadi beberapa bagian yang lebih kecil (*chunk*) agar setiap bagian dapat diproses secara individu tanpa melebihi kapasitas token model. Setelah teks dipecah, setiap *chunk* diproses menggunakan teknik analisis frekuensi kata untuk memastikan bahwa hanya bagian teks yang paling relevan dan informatif yang diproses lebih lanjut. Proses ini memastikan bahwa ringkasan yang dihasilkan tetap berfokus pada informasi penting dari seluruh teks panjang. Proses ini terdiri dari 3 tahapan yaitu *Chunking Text*, *Word Frequencies* dan *Sentence Scoring*.

3.6.1. *Chunking Text*

Chunking Text adalah proses pemecahan teks yang panjang menjadi beberapa bagian atau potongan (*chunk*) yang lebih kecil agar lebih mudah diproses oleh model. Proses ini sangat berguna dalam Natural Language Processing (NLP), terutama ketika bekerja dengan model yang memiliki batasan token seperti BART, yang hanya bisa memproses teks dengan panjang maksimal tertentu (misalnya, 1024 token). Dengan chunking, teks yang panjang dapat dibagi menjadi beberapa bagian yang lebih pendek sehingga setiap bagian dapat dianalisis dan diringkas secara individual. Proses *Chunking Text* dilakukan menggunakan *pseudocode* seperti yang ditunjukkan pada *Pseudocode* berikut

```

FUNCTION chunk_text(text: STRING, max_chunk_size: INTEGER)
RETURNS LIST
    DECLARE chunks: ARRAY
    DECLARE words: ARRAY ← split_text_into_words(text)  #
    Memecah teks menjadi kata-kata

    DECLARE current_chunk: STRING ← ""
    DECLARE current_length: INTEGER ← 0

    FOR EACH word IN words DO
        IF current_length + LENGTH(word) + 1 <= max_chunk_size
        THEN
            current_chunk += word + " "

```

```

        current_length += LENGTH(word) + 1
    ELSE
        ADD current_chunk TO chunks # Menambahkan chunk
yang sudah penuh ke dalam daftar chunk
        current_chunk ← word + " " # Memulai chunk baru
dengan kata saat ini
        current_length ← LENGTH(word) + 1
    END IF
END FOR

# Tambahkan chunk terakhir jika ada kata-kata yang tersisa
IF current_chunk ≠ "" THEN
    ADD current_chunk TO chunks
END IF

RETURN chunks
END FUNCTION

```

Pseudocode Proses Chunking Text

Berikut penjelasan fungsi dari proses *Chunking Text*:

- Inisialisasi Variabel: *Pseudocode* memulai dengan mendeklarasikan variabel *chunks* sebagai daftar kosong untuk menyimpan setiap *chunk*, serta *current_chunk* untuk menyimpan teks sementara dari *chunk* saat ini. Variabel *current_length* mencatat panjang dari *current_chunk* saat ini.
- Pemecahan Teks Menjadi Kata-kata: Teks dipecah menjadi kata-kata individual agar setiap kata dapat diproses satu per satu.
- Penyusunan *Chunk*: Setiap kata diperiksa apakah dapat ditambahkan ke *current_chunk* tanpa melebihi panjang maksimum (*max_chunk_size*). Jika masih muat, kata tersebut ditambahkan ke *current_chunk*, dan *current_length* diperbarui.
- Pembuatan *Chunk* Baru Jika Batas Tercapai: Jika penambahan kata menyebabkan *current_chunk* melebihi *max_chunk_size*, *chunk* saat ini ditambahkan ke daftar *chunks*, dan kata tersebut menjadi awal dari *chunk* baru.
- Penambahan *Chunk* Terakhir: Setelah iterasi selesai, *current_chunk* terakhir yang belum ditambahkan disimpan ke dalam *chunks*.
- Mengembalikan Daftar *Chunk*: Fungsi mengembalikan daftar semua *chunk* yang telah dibuat, siap untuk diproses lebih lanjut.

3.6.2. Word Frequencies

Proses perhitungan *Word Frequencies* dilakukan menggunakan variabel *word_frequencies* yang digunakan sebagai *dictionary* untuk menghitung frekuensi kemunculan setiap kata dalam teks, dengan mengabaikan kata-kata umum yang tidak informatif (*stop words*). Proses dilakukan dengan menggunakan data hasil *Tokenizing* yang telah diubah menjadi kata-kata, kemudian setiap kata yang bukan *stop word* ditambahkan ke *dictionary word_frequencies* dengan kata sebagai kunci dan frekuensi kemunculannya sebagai nilai. Jika kata tersebut sudah ada dalam *dictionary*, nilai frekuensinya akan diberikan nilai tambahan sebesar satu. Tujuan dari *word_frequencies* adalah untuk memberikan dasar penilaian (*scoring*) pada kalimat-kalimat dalam teks berdasarkan pentingnya kata-kata yang mereka miliki, yang kemudian digunakan dalam proses rangkuman teks. Proses *Word Frequencies* dilakukan menggunakan *pseudocode* seperti yang ditunjukkan pada *Pseudocode* berikut.

```

FUNCTION word_frequencies(tokens: LIST OF STRING) RETURNS
DICTIONARY
    DECLARE frequency_dict: DICTIONARY

    FOR EACH token IN tokens DO
        IF token NOT IN frequency_dict THEN
            frequency_dict[token] ← 1
        ELSE
            frequency_dict[token] ← frequency_dict[token] + 1

    RETURN frequency_dict
END FUNCTION

```

Pseudocode Proses Word Frequencies

Fungsi *word_frequencies* digunakan untuk menghitung frekuensi kemunculan setiap kata dalam daftar token yang diberikan. Fungsi ini menerima parameter *tokens*, yang merupakan daftar string (kata-kata), dan mengembalikan hasil dalam bentuk *dictionary* (kamus) yang berisi kata sebagai kunci dan jumlah kemunculannya sebagai nilai. Berikut penjelasan proses *word_frequencies* :

- Inisialisasi Dictionary: Variabel *frequency_dict* digunakan untuk menyimpan hasil frekuensi.
- Menghitung Frekuensi:
 - Untuk setiap token dalam daftar *tokens*, program memeriksa apakah token sudah ada dalam *frequency_dict*.

- Jika token belum ada, token ditambahkan ke dalam dictionary dengan nilai awal 1.
- Jika token sudah ada, nilai frekuensinya akan ditambah satu.
- Pengembalian Hasil: Setelah semua token diperiksa, fungsi mengembalikan *frequency_dict* yang berisi frekuensi kemunculan setiap kata.

Contoh 3. Data Hasil *Word Frequencies*

{'dbd': 3, 'di': 3, 'probolinggo': 3, 'capai': 1, '2309': 2, 'kasus': 3, '24': 2, 'orang': 1, 'meninggal': 1, 'tim': 1, 'redaksi': 1, 'kompas.com': 1, 'demam': 1, 'berdarah': 1, 'dengue': 1, 'kabupaten': 1, 'jawa': 1, 'timur': 1, 'melonjak': 1, 'tajam': 1, 'hingga': 2, 'september': 2, '2024': 2, 'naik': 1, '300': 1, 'persen': 1, 'dibandingkan': 1, 'tahun': 1, '2023': 1, 'tercatat': 1, 'kematian': 2, 'akibat': 1, 'kecamatan': 1, 'kraksaan': 1, 'besuk': 1, 'ranugedang': 1, 'tiris': 1, 'wonomerto': 1, 'sebagai': 1, 'wilayah': 1, 'dengan': 1, 'jumlah': 1, 'tertinggi': 1}

3.6.3. *Sentence Scoring*

Proses *Sentence Scoring* dilakukan menggunakan variabel *sentence_scores*, sebuah *dictionary* yang digunakan untuk menilai setiap kalimat dalam teks berdasarkan frekuensi kata-kata penting yang terkandung di dalamnya. Prosesnya melibatkan iterasi melalui setiap kalimat, tokenisasi kalimat menjadi kata-kata, dan kemudian mengecek setiap kata terhadap *word_frequencies*. Jika kata tersebut ada dalam *word_frequencies*, nilai frekuensi kata tersebut ditambahkan ke skor kalimat dalam *sentence_scores*. Jika kalimat belum ada dalam *dictionary*, kalimat tersebut dimasukkan dengan nilai awal sama dengan frekuensi kata. Hasil akhir *sentence_scores* berisi kalimat-kalimat sebagai kunci dan skor kumulatif mereka berdasarkan kata-kata penting sebagai nilai, yang kemudian digunakan untuk mengurutkan dan memilih kalimat-kalimat yang paling penting untuk rangkuman teks. Proses *Sentence Scoring* dilakukan menggunakan *pseudocode* seperti yang ditunjukkan pada *Pseudocode* berikut

```
FUNCTION sentence_scoring(sentences: LIST OF STRING,
frequency_dict: DICTIONARY) RETURNS LIST OF FLOAT
  DECLARE scores: LIST OF FLOAT
  DECLARE max_score: FLOAT

  max_score ← 0
```

```

FOR EACH sentence IN sentences DO
    score ← 0
    words ← tokenize(sentence)

    FOR EACH word IN words DO
        IF word IN frequency_dict THEN
            score ← score + frequency_dict[word]

    scores.append(score)
    IF score > max_score THEN
        max_score ← score

RETURN scores
END FUNCTION

```

Pseudocode Proses Sentence Scoring

Fungsi *sentence_scoring* digunakan untuk memberikan skor pada setiap kalimat dalam daftar berdasarkan frekuensi kata yang ada dalam dictionary. Fungsi ini menerima dua parameter: *sentences*, yaitu daftar kalimat, dan *frequency_dict*, yaitu dictionary yang berisi frekuensi kemunculan setiap kata.

- Inisialisasi Skor: Variabel *scores* menyimpan hasil skor untuk setiap kalimat, sedangkan *max_score* digunakan untuk melacak skor maksimum yang ditemukan.
- Menghitung Skor:
 - Untuk setiap kalimat, program melakukan tokenisasi untuk mendapatkan daftar kata.
 - Skor untuk kalimat tersebut dihitung dengan menjumlahkan frekuensi setiap kata dalam *frequency_dict*. Jika sebuah kata terdapat dalam dictionary, nilai frekuensinya akan ditambahkan ke skor kalimat.

Contoh 4. Data Hasil Sentence Scoring

{'DBD di Probolinggo Capai 2.309 Kasus, 24 Orang Meninggal Tim Redaksi PROBOLINGGO, KOMPAS.com - Kasus Demam Berdarah Dengue (DBD) di Kabupaten Probolinggo, Jawa Timur, melonjak tajam.': 31, 'Kasusnya mencapai 2.309 kasus hingga September 2024, naik 300 persen dibandingkan tahun 2023.': 17, 'Hingga 13 September 2024, tercatat 24 kematian akibat DBD, dengan Kecamatan Kraksaan,

Besuk, Ranagedang, Tiris, dan Wonomerto sebagai wilayah dengan jumlah kematian tertinggi.'. 25}

3.7. *Realization*

3.7.1. *Stopword Removal*

Stopword Removal adalah langkah penting yang bertujuan untuk menghapus kata-kata umum yang sering muncul tetapi memiliki nilai informasi yang rendah, seperti “dan”, “di”, atau “yang”. Sastrawi menyediakan daftar *stopword* Bahasa Indonesia dan alat untuk menghapus kata-kata ini dari teks. Proses ini dilakukan dengan terlebih dahulu memuat daftar *stopword* dari Sastrawi, kemudian menerapkannya pada teks yang telah dilakukan *tokenizing* dalam bentuk kata-kata untuk menghilangkan kata-kata yang tidak bermakna. Misalnya, setelah *stopword removal*, kalimat “ini adalah contoh teks yang sederhana” akan menjadi “contoh teks sederhana.” Dengan menghilangkan *stopwords*, teks yang diolah menjadi lebih bersih dan fokus pada kata-kata yang lebih bermakna, sehingga meningkatkan efisiensi dan akurasi dalam analisis dan pemodelan teks. Proses *Stopword Removal* dilakukan menggunakan *pseudocode* seperti yang ditunjukkan pada *Pseudocode* berikut.

```
FUNCTION stopwords_removal(teks: STRING) RETURNS STRING
    factory ← StopWordRemoverFactory()
    stopwords_remover ← factory.create_stop_word_remover()
    filtered_text ← stopwords_remover.remove(teks)
    RETURN filtered_text
END FUNCTION
```

Pseudocode Proses Stopword Removal

Fungsi *stopwords_removal* digunakan untuk menghapus kata-kata umum (*stopwords*) dari teks yang diberikan. Fungsi ini menerima parameter teks, yang merupakan string, dan mengembalikan teks yang telah difilter.

Di dalam fungsi, objek *factory* dibuat menggunakan *StopWordRemoverFactory()*, yang bertugas untuk membuat penghapus *stopword*. Kemudian, *stopwords_remover* diinisialisasi dengan memanggil metode *create_stop_word_remover()* dari objek *factory*. Setelah itu, metode *remove(teks)* dari objek *stopwords_remover* digunakan untuk

menghapus stopwords dari teks yang diberikan. Hasil teks yang telah difilter dikembalikan oleh fungsi.

Tabel 3.4 Contoh Sebelum dan Sesudah Proses *Stopword Removal*

Sebelum <i>Stopword Removal</i>	Sesudah <i>Stopword Removal</i>
['dbd', 'di', 'probolinggo', 'capai', '2.309', 'kasus', ',', '24', 'orang', 'meninggal', 'tim', 'redaksi', 'probolinggo', ',', 'kompas.com', '-', 'kasus', 'demam', 'berdarah', 'dengue', '(', 'dbd', ')', 'di', 'kabupaten', 'probolinggo', ',', 'jawa', 'timur', ',', 'melonjak', 'tajam', ',', 'kasusnya', 'mencapai', '2.309', 'kasus', 'hingga', 'september', '2024', ',', 'naik', '300', 'persen', 'dibandingkan', 'tahun', '2023', 'hingga', '13', 'september', '2024', ',', 'tercatat', '24', 'kematian', 'akibat', 'dbd', ',', 'dengan', 'kecamatan', 'kraksaan', ',', 'besuk', ',', 'ranugedang', ',', 'tiris', ',', 'dan', 'wonomerto', 'sebagai', 'wilayah', 'dengan', 'jumlah', 'kematian', 'tertinggi', '.']	dbd probolinggo capai 2.309 kasus, 24 orang meninggal tim redaksi probolinggo, kompas.com - kasus demam berdarah dengue (dbd) kabupaten probolinggo, jawa timur, melonjak tajam. kasusnya mencapai 2.309 kasus hingga september 2024, naik 300 persen dibandingkan tahun 2023. hingga 13 september 2024, tercatat 24 kematian akibat dbd, kecamatan kraksaan, besuk, ranugedang, tiris, wonomerto wilayah jumlah kematian tertinggi.

3.7.2. *Endemic and Location Search*

Proses *Endemic and Location Search* melibatkan beberapa langkah dan data endemik dan lokasi daerah yang sudah tersimpan pada *database*. Pertama, semua entri endemik diambil dari *database* menggunakan *Endemic.query.all()*, dan setiap nama endemik diubah menjadi huruf kecil dan dihapus spasinya dengan *lower().strip()*. Kemudian, teks yang telah diproses diperiksa apakah mengandung nama endemik tersebut. Jika ditemukan, pesan log akan mencatat bahwa endemik tersebut ditemukan dalam teks. Setelah itu, semua entri daerah diambil dari *database* menggunakan *Daerah.query.all()*, dan setiap nama kota diubah menjadi huruf kecil dan dihapus spasinya dengan *lower().strip()*. Teks tersebut kemudian diperiksa untuk keberadaan nama kota tersebut. Jika nama kota ditemukan dalam teks, pesan log akan mencatat bahwa kota tersebut ditemukan dalam teks. Langkah-langkah ini membantu

mengidentifikasi dan mencatat keterkaitan antara teks yang dianalisis dengan endemik dan lokasi geografis dari daerah tertentu. Proses *Endemic and Location Search* dilakukan menggunakan *pseudocode* seperti yang ditunjukkan pada *Pseudocode* berikut.

```

FUNCTION search_location_endemic(location: STRING, endemic: STRING)
RETURNS LIST OF LOCATION
  DECLARE results: LIST OF LOCATION
  DECLARE comparison_data: DICTIONARY

  comparison_data ← retrieve_comparison_data(location, endemic)

  FOR EACH data IN comparison_data DO
    IF data['location'] MATCHES location AND data['endemic'] MATCHES
endemic THEN
      results.append(data)

  RETURN results
END FUNCTION

```

Pseudocode Proses Endemic and Location Search

Fungsi *search_location_endemic* digunakan untuk mencari lokasi berdasarkan penyakit endemik tertentu. Fungsi ini menerima dua parameter: *location* yang merupakan string nama lokasi dan *endemic* yang merupakan string nama penyakit endemik.

- Inisialisasi: Variabel *results* digunakan untuk menyimpan hasil pencarian lokasi, dan *comparison_data* untuk menyimpan data yang diperlukan untuk perbandingan.
- Pengambilan Data: Data perbandingan diambil menggunakan fungsi *retrieve_comparison_data(location, endemic)*, yang akan mengembalikan informasi relevan mengenai lokasi dan penyakit endemik.
- Pencocokan Data: Melalui loop, fungsi memeriksa setiap data dalam *comparison_data*. Jika data lokasi dan penyakit endemik cocok dengan parameter yang diberikan, data tersebut ditambahkan ke dalam daftar hasil.
- Pengembalian Hasil: Fungsi mengembalikan daftar *results* yang berisi lokasi-lokasi yang sesuai dengan kriteria pencarian.

Contoh 5. Data Hasil *Endemic and Location Search*

```
{"kota": "kota banjarmasin", "latitude": "-3.328499", "longitude": "114.589203"}
```

3.7.3. *Save Result*

Tahapan *Save Result* melibatkan beberapa langkah untuk menyimpan hasil yang telah diproses ke dalam *database*. Pertama, objek Hasil dibuat dengan mengisi atribut-atributnya menggunakan data yang telah diproses, seperti ringkasan kalimat (*kalimat_summerization*), teks tanpa stop words (*kalimat_stopword*), teks asli (*kalimat_asli*), nama kota (kota), koordinat geografis (*latitude* dan *longitude*), nama endemik (nama), dan URL asal (url). Kemudian, informasi ini dicatat menggunakan `logging.info` untuk memastikan bahwa data yang akan disimpan telah dihasilkan dengan benar. Setelah itu, objek *new_hasil* ditambahkan ke sesi *database* menggunakan `db.session.add(new_hasil)`, dan perubahan disimpan secara permanen ke *database* dengan `db.session.commit()`. Terakhir, pesan log dicatat untuk mengonfirmasi bahwa data telah berhasil disimpan. Proses *Save Result* dilakukan menggunakan *pseudocode* seperti yang ditunjukkan pada *Pseudocode* berikut.

```

FUNCTION save_result(kalimat_summerization: STRING, filtered_text:
STRING, kalimat_asli: STRING, daerahKota: STRING, latitude: FLOAT,
longitude: FLOAT, endemic_name: STRING, url: STRING) RETURNS VOID
  DECLARE new_hasil: Hasil

  new_hasil ← Hasil(
    kalimat_summerization=kalimat_summerization,
    kalimat_stopword=filtered_text,
    kalimat_asli=kalimat_asli,
    kota=daerahKota,
    latitude=latitude,
    longitude=longitude,
    nama=endemic_name,
    url=url
  )

  logging.info("Data to be saved: {new_hasil}")
  db.session.add(new_hasil)
  db.session.commit()
  logging.info("Data saved successfully")
END FUNCTION

```

Pseudocode Proses Save Result

Fungsi *save_result* digunakan untuk menyimpan hasil ringkasan dan informasi terkait ke dalam basis data. Fungsi ini menerima beberapa parameter yang mencakup kalimat ringkasan, teks yang telah difilter, kalimat asli, nama kota, latitude, longitude, nama penyakit endemik, dan URL.

- Inisialisasi Hasil: Variabel `new_hasil` diinisialisasi dengan membuat objek dari kelas Hasil menggunakan semua parameter yang diberikan.
- Logging Data: Sebelum menyimpan, informasi tentang data yang akan disimpan dicatat menggunakan `logging.info`.
- Penyimpanan ke Basis Data: Data baru ditambahkan ke sesi database dan kemudian dikomit untuk menyimpan perubahan ke dalam basis data.
- Konfirmasi Penyimpanan: Setelah penyimpanan berhasil, pesan konfirmasi dicatat menggunakan `logging.info`.

Contoh 6. Data Hasil *Save Result*

```
{
  "endemic": "dbd",
  "id": 1033,
  "kalimat_asli": "DBD di Probolinggo Capai 2.309 Kasus, 24 Orang Meninggal Tim Redaksi PROBOLINGGO, KOMPAS.com - Kasus Demam Berdarah Dengue (DBD) di Kabupaten Probolinggo, Jawa Timur, melonjak tajam. Kasusnya mencapai 2.309 kasus hingga September 2024, naik 300 persen dibandingkan tahun 2023. Hingga 13 September 2024, tercatat 24 kematian akibat DBD, dengan Kecamatan Kraksaan, Besuk, Ranugedang, Tiris, dan Wonomerto sebagai wilayah dengan jumlah kematian tertinggi.",
  "kalimat_stopword": "dbd probolinggo capai 2.309 kasus, 24 orang meninggal tim redaksi probolinggo, kompas.com - kasus demam berdarah dengue (dbd) kabupaten probolinggo, jawa timur, melonjak tajam. kasusnya mencapai 2.309 kasus hingga september 2024, naik 300 persen dibandingkan tahun 2023. hingga 13 september 2024, tercatat 24 kematian akibat dbd, kecamatan kraksaan, besuk, ranugedang, tiris, wonomerto wilayah jumlah kematian tertinggi.",
  "kalimat_summerization": "Hingga 13 September 2024, tercatat 24 kematian akibat DBD, dengan Kecamatan Kraksaan, Besuk, Ranugedang, Tiris, dan Wonomerto sebagai wilayah dengan jumlah kematian tertinggi. DBD di Probolinggo Capai 2.309 Kasus, 24 Orang Meninggal Tim Redaksi PROBOLINGGO, KOMPAS.com - Kasus Demam Berdarah Dengue (DBD) di Kabupaten Probolinggo, Jawa Timur, melonjak tajam. Kasusnya mencapai 2.309 kasus hingga September 2024, naik 300 persen dibandingkan tahun 2023.",
  "latitude": -7.7649,
  "longitude": 113.2132,
  "nama": "Probolinggo",
  "url": "https://regional.kompas.com/read/2024/09/14/220942678/dbd-di-probolinggo-capai-2309-kasus-24-orang-meninggal"
}
```

3.8. Evaluasi Model

Proses evaluasi model ini bertujuan untuk mengukur kualitas ringkasan yang dihasilkan oleh model dengan membandingkannya dengan kalimat asli menggunakan beberapa metrik evaluasi: *ROUGE*, *BLEU*, dan *BERTScore*. Pertama, fungsi mengambil data kalimat asli dan ringkasan dari database, lalu menghitung *ROUGE* untuk menilai kesamaan n-gram antara kalimat asli dan ringkasan, mencakup skor *ROUGE-1*, *ROUGE-2*, dan *ROUGE-L*. Setelah itu, *BLEU* dihitung dengan mengukur kesamaan urutan kata dalam kalimat yang telah ditokenisasi, dan *BERTScore* digunakan untuk mengevaluasi kesamaan semantik antara kalimat asli dan ringkasan berdasarkan representasi embedding.

Setelah menghitung skor untuk setiap pasangan kalimat, rata-rata nilai dari setiap metrik dievaluasi untuk mendapatkan satu nilai akhir yang mewakili performa model. Hasil rata-rata skor dari metrik *BLEU*, *ROUGE*, dan *BERTScore* kemudian disatukan dalam respons *JSON* untuk memudahkan interpretasi, memberikan pandangan keseluruhan tentang seberapa baik model menghasilkan ringkasan yang akurat dan relevan dari segi struktur dan makna. Proses Evaluasi Model dilakukan menggunakan *pseudocode* seperti yang ditunjukkan pada *Pseudocode* berikut.

```
FUNCTION evaluate_all() RETURNS JSON
  # Ambil kalimat asli dan ringkasan dari database
  DECLARE original_sentences, summarized_sentences

  # Inisialisasi objek evaluasi ROUGE dan BERTScore
  INITIALIZE rouge AS Rouge()

  # Hitung ROUGE untuk semua pasangan
  rouge_scores ← rouge.get_scores(summarized_sentences,
original_sentences, avg=True)

  # Tokenisasi kalimat untuk BLEU
  tokenized_references ← tokenize_sentences(original_sentences)
  tokenized_candidates ← tokenize_sentences(summarized_sentences)

  # Hitung BLEU untuk setiap pasangan
  bleu_scores ← []
  FOR EACH ref, cand IN zip(tokenized_references,
tokenized_candidates) DO
    bleu_score ← sentence_bleu([ref], cand)
    ADD bleu_score TO bleu_scores
```



```

END FOR

# Hitung BERTScore untuk semua pasangan
P, R, F1 ← bertscore(summarized_sentences, original_sentences)

# Hitung rata-rata BLEU dan BERTScore
avg_bleu ← mean(bleu_scores)
avg_bertscore_f1 ← mean(F1)

# Gabungkan hasil ke dalam JSON response
response ← {
  "avg_bleu": avg_bleu,
  "avg_rouge_1": rouge_scores["rouge-1"]['f'],
  "avg_rouge_2": rouge_scores["rouge-2"]['f'],
  "avg_rouge_l": rouge_scores["rouge-l"]['f'],
  "avg_bertscore_f1": avg_bertscore_f1
}

RETURN response
END FUNCTION

```

Pseudocode Proses Evaluasi Model

Berikut penjelasan fungsi dari proses Evaluasi Model:

- Pengambilan Data dari Database: Fungsi *evaluate_all()* dimulai dengan mengambil data kalimat asli (*original_sentences*) dan ringkasan (*summarized_sentences*) dari database, yang masing-masing mewakili teks referensi dan teks hasil ringkasan untuk evaluasi.
- Inisialisasi Evaluasi *ROUGE*: Objek *rouge* diinisialisasi untuk menghitung skor *ROUGE*, yang mengukur kesamaan antara kalimat asli dan ringkasan dari segi *n-gram*. *ROUGE* menghitung metrik seperti *ROUGE-1*, *ROUGE-2*, dan *ROUGE-L*, yang kemudian dirata-ratakan untuk seluruh data.
- Tokenisasi untuk *BLEU*: Teks asli dan ringkasan ditokenisasi menjadi kata-kata individual, karena *BLEU* memerlukan tokenisasi untuk menghitung kesamaan urutan kata dalam kalimat. Hasil tokenisasi ini disimpan sebagai daftar yang terpisah untuk kalimat referensi (*tokenized_references*) dan kandidat (*tokenized_candidates*).
- Perhitungan Skor *BLEU*: Skor *BLEU* dihitung untuk setiap pasangan kalimat antara teks asli dan ringkasan. Skor ini dihitung berdasarkan kemiripan urutan kata, yang dinilai melalui tokenisasi yang telah dilakukan.

- Perhitungan *BERTScore*: Selanjutnya, fungsi menghitung *BERTScore*, yang menggunakan model berbasis *embedding* untuk menilai kesamaan semantik antara kalimat asli dan ringkasan. *BERTScore* menghasilkan nilai *Precision*, *Recall*, dan *F1*, tetapi pada proses ini hanya nilai *F1* yang diambil sebagai indikator performa model.
- Menghitung Rata-Rata Skor: Setelah semua pasangan kalimat dibandingkan, fungsi menghitung rata-rata dari skor *BLEU* dan *BERTScore F1*, serta nilai rata-rata dari skor *ROUGE* (*ROUGE-1*, *ROUGE-2*, *ROUGE-L*) untuk mendapatkan nilai akhir dari keseluruhan data.
- Menyusun *JSON* Hasil Evaluasi: Rata-rata dari setiap metrik kemudian disusun dalam bentuk *JSON* yang mencakup *avg_bleu*, *avg_rouge_1*, *avg_rouge_2*, *avg_rouge_l*, dan *avg_bertscore_f1*. *JSON* ini kemudian dikembalikan sebagai respons, memberikan gambaran tentang seberapa baik model menghasilkan ringkasan yang akurat dan informatif.

3.9. *Map Annotation*

Tahapan *map annotation* dilakukan setelah mendapatkan data lengkap mulai yang dihasilkan dari tahapan sebelumnya. Tahapan ini melibatkan penandaan peta dengan informasi yang telah diproses. Pertama, data yang telah disimpan ke dalam *database*, seperti lokasi geografis (*latitude* dan *longitude*), nama kota (kota), dan nama endemik (nama), diekstraksi. Selanjutnya, menggunakan *OpenStreetMap* (OSM), setiap data lokasi digunakan untuk membuat penanda (*marker*) pada peta. Penanda ini dilengkapi dengan informasi tambahan dalam bentuk *pop-up* atau *tooltip* yang berisi ringkasan teks (*kalimat_summerization*) dan informasi lainnya. Proses ini membantu dalam visualisasi data secara geografis, memungkinkan pengguna untuk melihat distribusi endemik dan informasi terkait lainnya secara intuitif di peta interaktif.

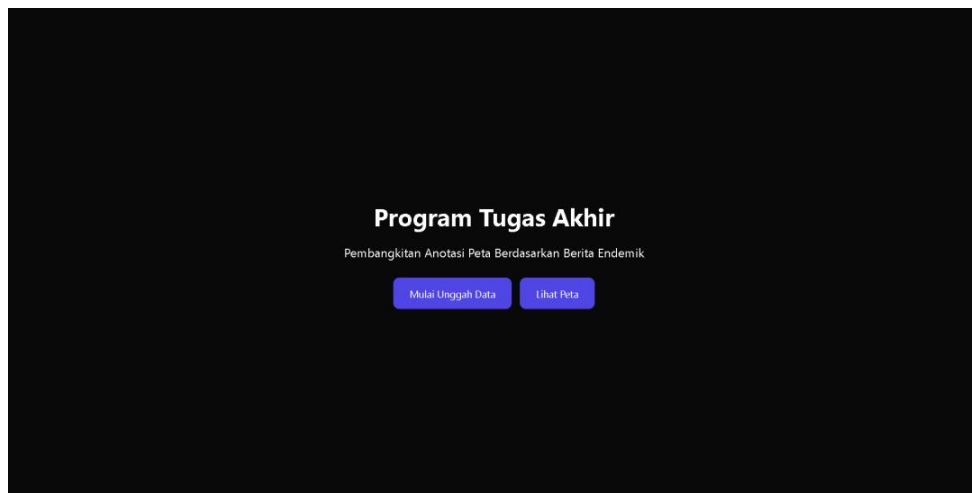
3.10. *Geographic Information Systems*

Geographic Information Systems (GIS) dibuat menggunakan OSM. Pada sistem yang diusulkan, pengguna dapat menggunakan *link* artikel berita sebagai *input* untuk menambahkan *marker* dan informasi sesuai berita tersebut. Sistem dibuat menjadi dua

halaman, yaitu halaman main menu dan *maps*. Penjelasan setiap halaman pada sistem ini adalah sebagai berikut.

a. *Main Menu*

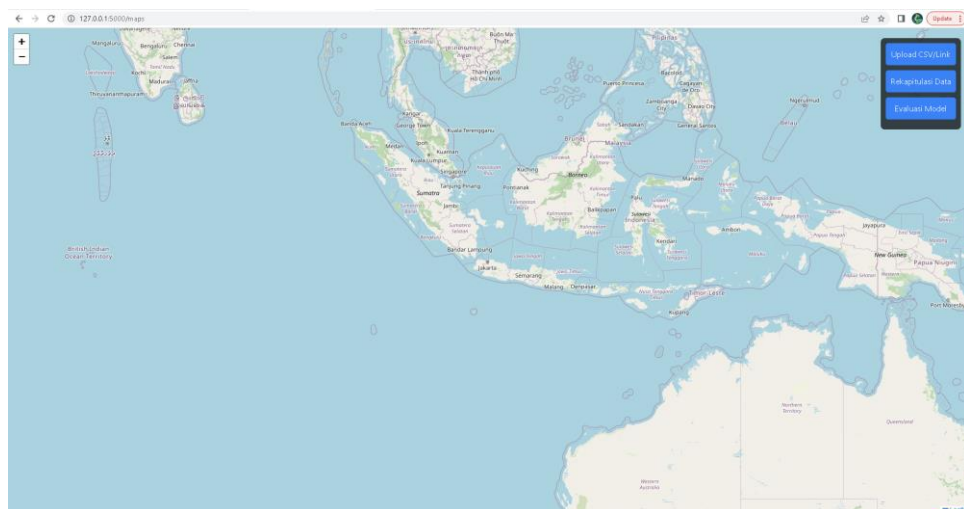
Main menu merupakan halaman utama yang pada sistem yang dirancang. Halaman ini berisi *form* yang dapat digunakan pengguna untuk membuat anotasi peta menggunakan *link* ataupun daftar *link* artikel berita dalam bentuk *file* csv. Desain tampilan main menu ditunjukkan pada Gambar 3.14 sebagai berikut.



Gambar 3.3 Desain Tampilan *Main Menu*

b. *Menu Maps*

Menu *maps* merupakan yang menampilkan titik poin (*marker*) berdasarkan data dari lokasi endemik yang disebutkan pada artikel berita. Setiap *marker* akan memberikan informasi ringkasan dari artikel berita tersebut. Desain tampilan menu *maps* ditunjukkan pada Gambar 3.15 sebagai berikut.



Gambar 3.4 Desain Tampilan Menu *Maps*

BAB 4

IMPLEMENTASI DAN PENGUJIAN

4.1 Implementasi Sistem

Pada tahap implementasi sistem, proses pembangkitan anotasi peta berdasarkan ekstraksi data berita endemik menggunakan pendekatan NLG melibatkan beberapa tahapan seperti yang telah dijelaskan pada metodologi sebelumnya pada Bab 3. Dalam penelitian ini, sistem dikembangkan menggunakan bahasa pemrograman Python. Penelitian dilakukan menggunakan perangkat keras (*hardware*) dan perangkat lunak (*software*) dengan spesifikasi atau versi seperti yang ditunjukkan pada Tabel 4.1 sebagai berikut.

Tabel 4.1. Spesifikasi *Hardware* dan *Software*

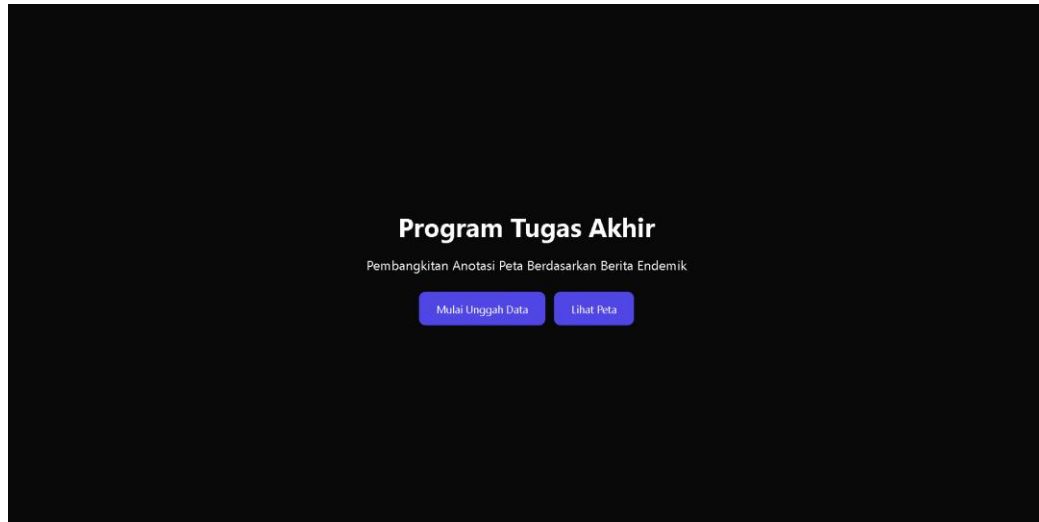
	No	Nama	Spesifikasi atau Versi
<i>Hardware</i>	1	<i>Processor</i>	Intel® Core™ i5-1135G7
	2	<i>Storage</i>	512 GB SSD
	3	<i>RAM</i>	8GB DDR4
	4	<i>VGA</i>	GeForce® MX350
<i>Software</i>	1	<i>Windows 11</i>	64 bit
	2	<i>Visual Studio Code</i>	1.91.1
	3	<i>Flask</i>	3.0.3
	4	<i>BeautifulSoup4</i>	4.12.3
	5	<i>Transformers</i>	4.42.4
	6	<i>NLTK</i>	3.8.1
	7	<i>Sastrawi</i>	1.0.1
	8	<i>MySQL</i>	9.0.0
	9	<i>OpenStreetMap</i>	0.7.61

4.2 Implementasi Bagian Depan

Implementasi bagian depan sistem dilakukan sesuai dengan perencanaan yang sudah dijabarkan pada Bab 3 sebelumnya. Berikut adalah implementasi bagian depan sistem.

a. *Main Menu*

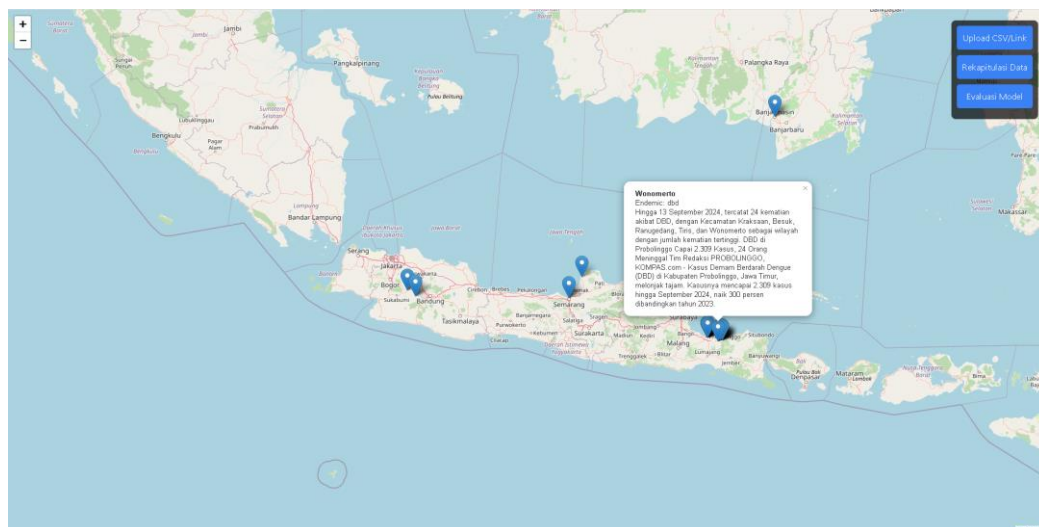
Main menu digunakan untuk *input link* artikel berita dan mengakses menu *maps*. Tampilan *main menu* ditunjukkan pada Gambar 4.1 sebagai berikut.



Gambar 4.1. Tampilan Main Menu

b. *Menu Maps*

Pada menu *maps* pengguna dapat melakukan pencarian lokasi *marker* yang tersedia serta mendapatkan informasi dari artikel berita yang telah di ringkas dalam bentuk *pop up*. Pada menu ini pengguna juga dapat melakukan penghapusan informasi yang diberikan jika sudah tidak relevan dengan kondisi saat ini. Tampilan main *maps* ditunjukkan pada Gambar 4.2 sebagai berikut.



Gambar 4.2. Tampilan Menu Maps

4.3 Prosedur Operasional

Prosedur operasional diawali dengan pengguna memasukkan *link* atau daftar *link* dalam bentuk *file csv* pada *main menu* seperti yang ditunjukkan pada Gambar 4.1. Kemudian sistem akan memproses *link* tersebut dengan mengambil data artikel berita menggunakan metode *scraping* kemudian dilakukan perangkuman dan penentuan jenis endemik dan daerah untuk menentukan *marker* yang akan dibuat pada menu *maps*. Setelah proses tersebut selesai pengguna dapat mengakses menu *maps* untuk mendapatkan informasi ringkasan artikel pada *pop up* setiap *marker* daerah pada artikel berita yang ada di artikel berita dari *link* yang telah di *input*-kan sebelumnya seperti yang ditunjukkan pada Gambar 4.2.

4.4 Pengujian Sistem

Pengujian sistem dilakukan dengan skenario pengecekan *marker* daerah dan hasil peringkasan. Pengujian dilakukan menggunakan lima url artikel berita dari lima portal berita yang berbeda dengan rentang waktu 5 tahun terakhir seperti yang ditunjukkan pada Tabel 4.2. Hasil dari pengujian sistem ditunjukkan pada Tabel 4.3.

Tabel 4.2. Url Artikel Berita

No	URL
1	https://www.detik.com/jateng/berita/d-7262625/dinkes-jateng-catat-90-orang-meninggal-akibat-dbd-terbanyak-di-jepara
2	https://satu.tempo.co/sains/nias-selatan-darurat-bencana-dbd-dan-malaria-ratusan-orang-dirawat-8-meninggal-24565
3	https://www.detik.com/jabar/berita/d-6774812/kasus-dbd-di-karawang-ugal-ugalan-4-orang-meninggal-dunia
4	https://regional.kompas.com/read/2024/09/14/220942678/dbd-di-probolinggo-capai-2309-kasus-24-orang-meninggal
5	https://www.detik.com/jabar/berita/d-7604301/2-473-warga-ciamis-tertular-tbc

Tabel 4.3. Hasil Pengujian Sistem

Data Asli	Data Summary	Lokasi	Endemik	Penilaian
Dinas Kesehatan Provinsi Jawa Tengah (Dinkes Jateng) mencatat ada 90 orang meninggal akibat demam berdarah dengue (DBD). Kasus kematian akibat DBD terbanyak terjadi di Jepara dengan 20 kasus. 90 kasus meninggal tersebar di 35 kabupaten/kota ya memang ada angka kenaikan tapi data yang dilaporkan kebanyakan di e-report jadi kebanyakan puncak kasusnya terjadi di Januari-Februari, kata Kabid Pencegahan dan Pengendalian Penyakit (P2P) Dinkes Jateng, Irma Makiah melalui sambungan telepon, Selasa (26/3/2024). Irma menyebut mayoritas pasien meninggal ialah anak berusia 5-14 tahun. Dia mengingatkan agar orang tua lebih waspada terhadap gejala DBD terutama cara-cara pencegahan dengan memberi makanan bergizi dan istirahat yang cukup.	“90 kasus meninggal tersebar di 35 kabupaten/kota ya memang ada angka kenaikan tapi data yang dilaporkan kebanyakan di e-report jadi kebanyakan puncak kasusnya terjadi di Januari-Februari, kata Kabid Pencegahan dan Pengendalian Penyakit (P2P) Dinkes Jateng, Irma Makiah melalui sambungan telepon, Selasa (26/3/2024). Dinas Kesehatan Provinsi Jawa Tengah (Dinkes Jateng)	Jepara, Jawa Tengah	DBD	Sesuai

Data Asli	Data Summary	Lokasi	Endemik	Penilaian
	mencatat ada 90 orang meninggal akibat demam berdarah dengue (DBD). Dia mengingatkan agar orang tua lebih waspada terhadap gejala DBD terutama cara-cara pencegahan dengan memberi makanan bergizi dan istirahat yang cukup.			
TEMPO.CO, Jakarta - Kasus demam berdarah dengue (DBD) dan malaria merebak di wilayah Kabupaten Nias Selatan, Sumatera Utara, sepanjang Januari hingga Juli 2024. Setidaknya ada tujuh kecamatan yang terdampak meliputi Pulau-Pulau Batu, Pulau-Pulau Batu Timur, Pulau-Pulau Batu Barat, Pulau-Pulau Batu Utara, Simauk, Tanah Masa dan Hibala, dengan jumlah 562 warga terjangkit. "Sebanyak delapan orang meninggal dan	Kasus DBD dan malaria di Kabupaten Nias Selatan, Sumatera Utara, merebak sejak Januari hingga Juli 2024, menginfeksi 562 warga di tujuh	Nias Selatan, Sumate ra Utara	DBD, Malaria	Sesuai

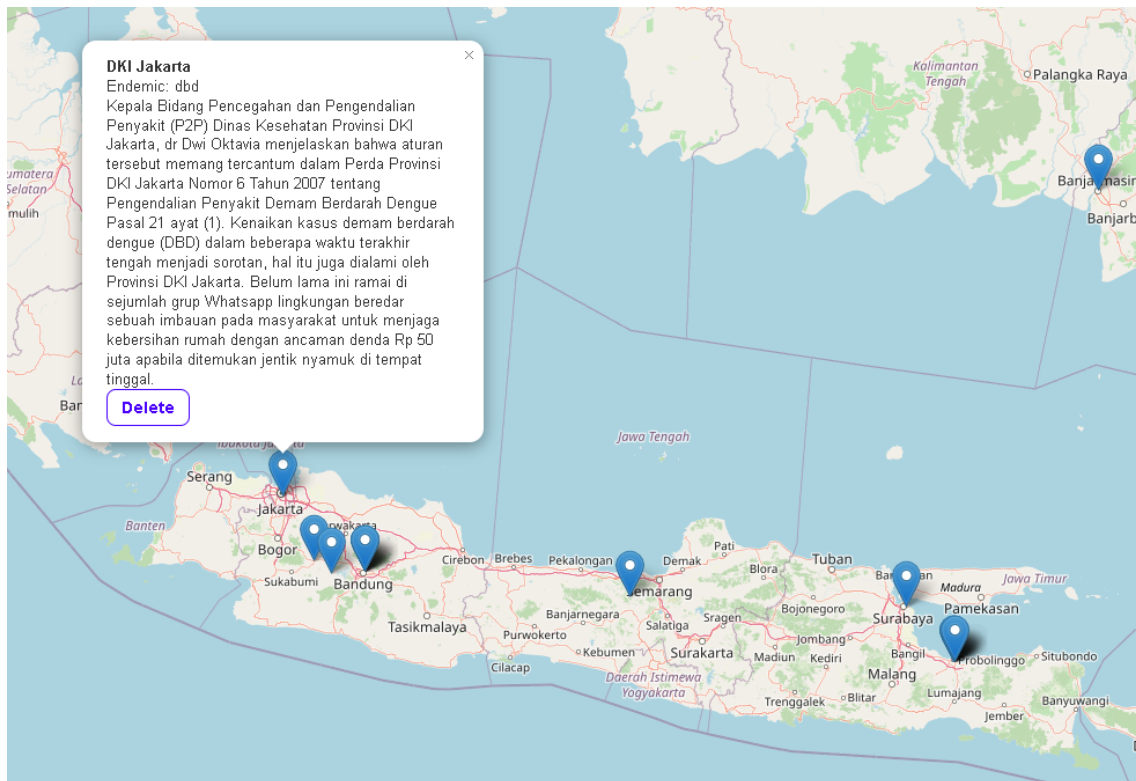
Data Asli	Data Summary	Lokasi	Endemik	Penilaian
554 warga lainnya dirawat dan telah dinyatakan sembuh," kata Kepala Pusat Data, Informasi dan Komunikasi Kebencanaan Badan Nasional Penanggulangan Bencana (BNPB), Abdul Muhari, melalui keterangan tertulis, Kamis 15 Agustus 2024.	kecamatan. Delapan orang meninggal, sementara 554 lainnya telah dirawat dan dinyatakan sembuh, menurut BNPB			
Kasus demam berdarah dengue (DBD) di Karawang menggila. Sejak tahun ini saja, total ada 507 kasus terkonfirmasi dan empat orang meninggal dunia. \"Dari awal tahun sampai pertengahan Juni ini tercatat ada 507 kasus pasien terjangkit DBD di Karawang, empat di antaranya bahkan dilaporkan telah meninggal dunia,\" ujar Kepala Bidang Pencegahan dan Pengendalian Penyakit Dinas Kesehatan (Dinkes) Karawang Yayuk Sri Rahayu saat dihubungi, Kamis (15/6/2023). Peningkatan kasus DBD tersebut, Kata Yayuk, justru bertolak belakang dengan pengaruh cuaca, yang cenderung panas, dan curah hujan yang rendah di Karawang.	“Dari awal tahun sampai pertengahan Juni ini tercatat ada 507 kasus pasien terjangkit DBD di Karawang, empat di antaranya bahkan dilaporkan telah meninggal dunia,\" ujar Kepala Bidang Pencegahan dan Pengendalian Penyakit Dinas Kesehatan (Dinkes) Karawang Yayuk Sri	Karawang	DBD	Sesuai

Data Asli	Data Summary	Lokasi	Endemik	Penilaian
	<p>Rahayu saat dihubungi, Kamis (15/6/2023). Peningkatan kasus DBD tersebut, Kata Yayuk, justru bertolak belakang dengan pengaruh cuaca, yang cenderung panas, dan curah hujan yang rendah di Karawang. Kasus demam berdarah dengue (DBD) di Karawang menggila.</p>			
<p>DBD di Probolinggo Capai 2.309 Kasus, 24 Orang Meninggal Tim Redaksi PROBOLINGGO, KOMPAS.com - Kasus Demam Berdarah Dengue (DBD) di Kabupaten Probolinggo, Jawa Timur, melonjak tajam. Kasusnya mencapai 2.309 kasus hingga September 2024, naik 300 persen dibandingkan tahun 2023. Hingga 13 September 2024, tercatat 24 kematian</p>	<p>Hingga 13 September 2024, tercatat 24 kematian akibat DBD, dengan Kecamatan Kraksaan, Besuk, dan</p>	<p>Probolinggo, Jawa Timur, Kraksaan, Besuk, Ranugedang,</p>	DBD	Sesuai

Data Asli	Data Summary	Lokasi	Endemik	Penilaian
akibat DBD, dengan Kecamatan Kraksaan, Besuk, Ranugedang, Tiris, dan Wonomerto sebagai wilayah dengan jumlah kematian tertinggi.	Ranugedang, Tiris, dan Wonomerto sebagai wilayah dengan jumlah kematian tertinggi. DBD di Probolinggo Capai 2.309 Kasus, 24 Orang Meninggal Tim Redaksi PROBOLINGGO, O, KOMPAS.com - Kasus Demam Berdarah Dengue (DBD) di Kabupaten Probolinggo, Jawa Timur, melonjak tajam. Kasusnya mencapai 2.309 kasus hingga September 2024, naik 300 persen dibandingkan tahun 2023.	Tiris, Wonomerto		

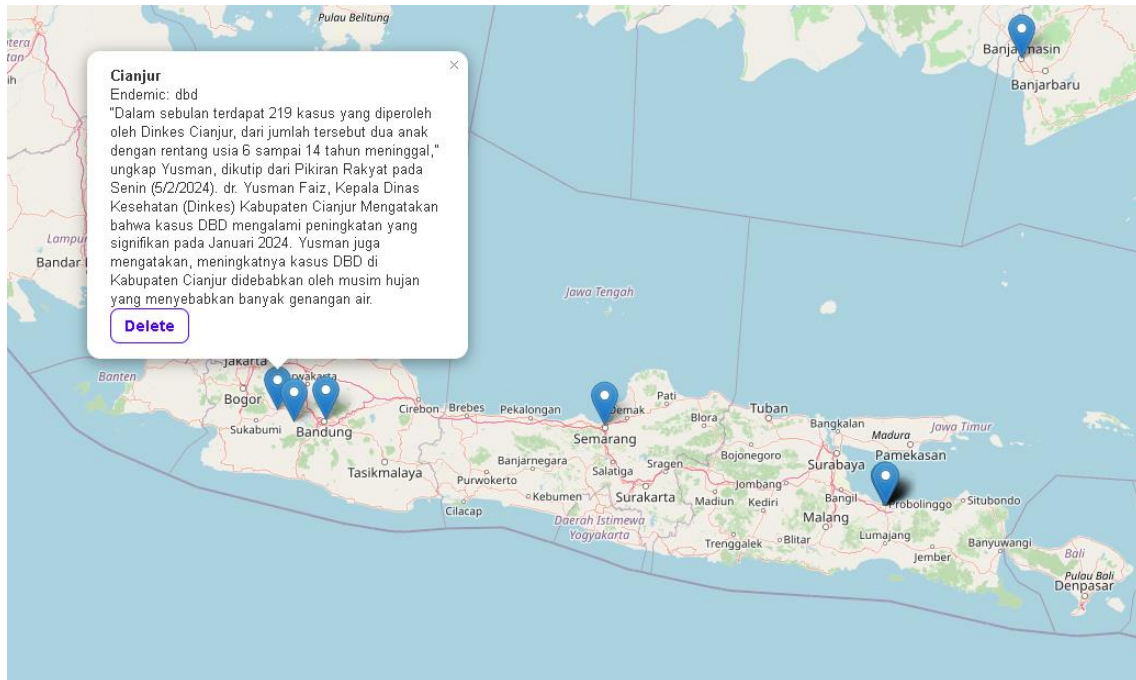
Data Asli	Data Summary	Lokasi	Endemik	Penilaian
<p>Ciamis - Dinas Kesehatan (Dinkes) Kabupaten Ciamis menemukan 2.473 kasus tuberkulosis (TBC) hingga September 2024. Penderita terindikasi dari berbagai kalangan dan tersebar di 37 puskesmas yang ada di Kabupaten Ciamis. Puskesmas tersebut juga menangani pengobatan penderita TBC. "Jumlahnya sampai September sebanyak 2.473 kasus. Untuk kasus TBC ini semakin banyak ditemukan semakin bagus. Harus menemukan sebanyak-banyaknya. Supaya dapat melakukan penanganan pengobatan bagi penderita dan bisa meminimalisir penularan," ujar Kabid P2P Dinkes Ciamis Edis Herdis, Kamis (24/10/2024).</p>	<p>Dinas Kesehatan Kabupaten Ciamis mencatat 2.473 kasus TBC hingga September 2024, tersebar di 37 puskesmas. Menurut Kabid P2P Dinkes Ciamis, Edis Herdis, semakin banyak kasus yang ditemukan, semakin baik untuk mendukung pengobatan dan pencegahan penularan.</p>	Ciamis	TBC	Sesuai

Dari hasil pengujian, dapat disimpulkan bahwa sistem mampu menentukan lokasi dan menghasilkan ringkasan berita yang relevan dengan nama lokasi serta konteks artikel. Penentuan jenis endemik sesuai dengan nama penyakit yang disebutkan. Namun, jika terdapat beberapa lokasi dalam berita, sistem akan menambahkan marker sesuai jumlah lokasi teridentifikasi, meskipun beberapa mungkin tidak terkait dengan kasus endemik. Sistem dioptimalkan untuk memilih lokasi yang paling spesifik, seperti menampilkan titik kota jika dalam berita terdapat provinsi dan kota.



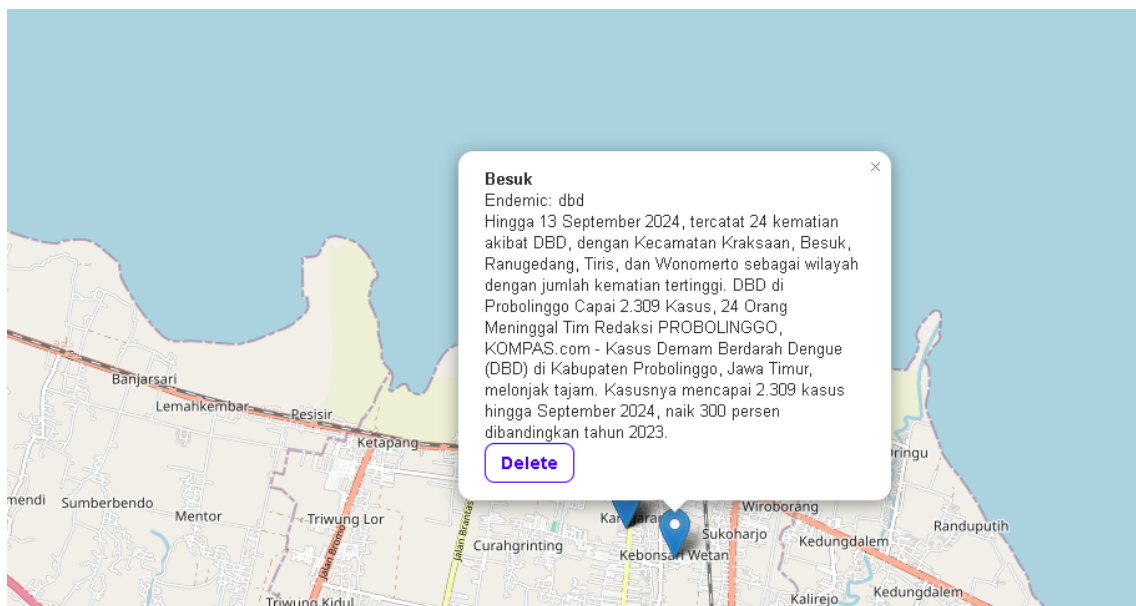
Gambar 4.3. Tampilan Peta Untuk Provinsi

Pada Gambar 4.3, sistem berhasil mengidentifikasi dan menampilkan lokasi endemik di tingkat provinsi, di mana informasi terkait penyakit yang terdeteksi dalam berita, seperti DBD, ditampilkan. Setiap provinsi yang teridentifikasi sebagai lokasi endemik diberi marker dengan penjelasan lebih lanjut mengenai jumlah kasus dan kondisi di provinsi tersebut. Informasi ini didapatkan dari proses ekstraksi berita yang menyebutkan lokasi provinsi secara spesifik.



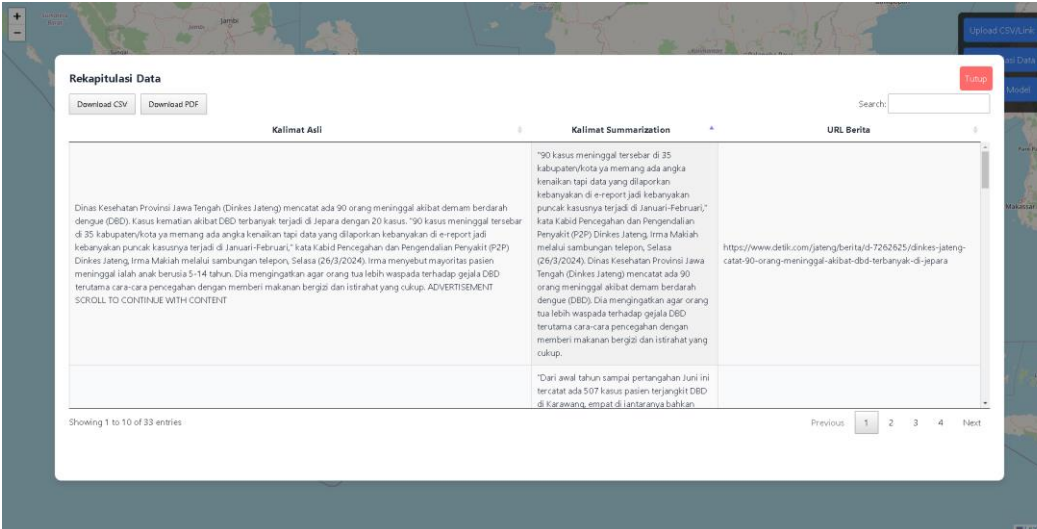
Gambar 4.4. Tampilan Peta Untuk Kota

Gambar 4.4 memperlihatkan hasil pemetaan pada tingkat kota. Sistem mampu mengekstrak lokasi kota dari teks berita dan menampilkannya di peta, disertai informasi tambahan seperti jumlah kasus dan kondisi penyebaran penyakit di kota tersebut. Marker pada peta memberikan ringkasan berita terkait kondisi endemik di kota yang disebutkan.



Gambar 4.5. Tampilan Kota Untuk Kecamatan

Pada Gambar 4.5, hasil ekstraksi berita memungkinkan sistem menampilkan peta hingga ke tingkat kecamatan. Setiap kecamatan yang disebutkan dalam berita diekstraksi secara otomatis dan dipetakan dengan akurat. Marker di setiap kecamatan memberikan rincian lebih lanjut tentang jumlah kasus dan dampak dari penyakit endemik yang dilaporkan di lokasi tersebut, membantu pengguna memahami distribusi kasus endemik secara lebih rinci di tingkat lokal.



Rekapitulasi Data

Download CSV Download PDF

Search:

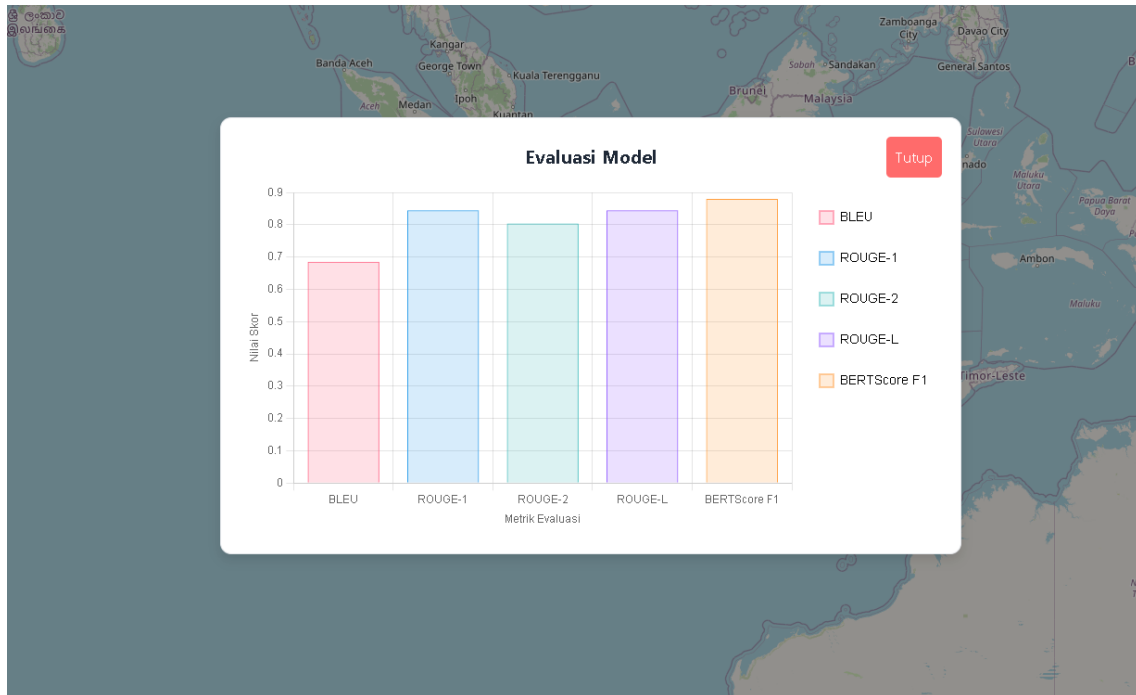
Kalimat Asli	Kalimat Summarization	URL Berita
Dinas Kesehatan Provinsi Jawa Tengah (Dinkes Jateng) mencatat ada 90 orang meninggal akibat demam berdarah dengue (DBD). Kasus kematian akibat DBD terbanyak terjadi di Jepara dengan 20 kasus. "90 kasus meninggal tersebar di 35 kabupaten/kota ya memang ada angka kematian tapi data yang dilaporkan kebanyakan di e-report jadi kebanyakan puncak kasusnya terjadi di Januari-Februari," kata Kabid Pencegahan dan Pengendalian Penyakit (P2P) Dinkes Jateng, Irma Maliah melalui sambungan telepon, Selasa (26/3/2024). Dinkes Jateng mencatat ada 90 orang meninggal akibat demam berdarah dengue (DBD). Dia mengingatkan agar orang tua lebih waspada terhadap gejala DBD terutama cara-cara pencegahan dengan memberi makanan bergizi dan istirahat yang cukup. ADVERTISEMENT SCROLL TO CONTINUE WITH CONTENT	"90 kasus meninggal tersebar di 35 kabupaten/kota ya memang ada angka kematian tapi data yang dilaporkan kebanyakan di e-report jadi kebanyakan puncak kasusnya terjadi di Januari-Februari," kata Kabid Pencegahan dan Pengendalian Penyakit (P2P) Dinkes Jateng, Irma Maliah melalui sambungan telepon, Selasa (26/3/2024). Dinkes Jateng mencatat ada 90 orang meninggal akibat demam berdarah dengue (DBD). Dia mengingatkan agar orang tua lebih waspada terhadap gejala DBD terutama cara-cara pencegahan dengan memberi makanan bergizi dan istirahat yang cukup.	https://www.detik.com/jateng/berita/d-7262625/dinkes-jateng-catat-90-orang-meninggal-akibat-dbd-terbanyak-di-jepara
	"Dari awal tahun sampai pertengahan Juni ini tercatat ada 507 kasus pasien terinfeksi DBD di Karawang, empat di antaranya bahkan"	

Showing 1 to 10 of 33 entries

Previous 1 2 3 4 Next

Gambar 4.6. Tampilan Rekapitulasi Data

Gambar 4.6 menunjukkan tampilan Rekapitulasi Data, di mana hasil ekstraksi berita ditampilkan dalam bentuk tabel. Setiap baris tabel memuat Kalimat Asli, Kalimat Summarization, dan URL Berita terkait. Tampilan ini memudahkan pengguna untuk melihat ringkasan dari berita yang diekstraksi dan melakukan analisis lebih lanjut terkait distribusi kasus endemik di berbagai lokasi yang disebutkan dalam berita. Pada tampilan ini, pengguna bisa mencari berita terkait serta dapat mengunduh tabel dalam format csv ataupun dalam bentuk pdf.



Gambar 4.7. Tampilan Evaluasi Model

Gambar 4.7 menampilkan Evaluasi Model, yang memperlihatkan skor hasil evaluasi model dengan metrik seperti *BLEU*, *ROUGE-1*, *ROUGE-2*, *ROUGE-L*, dan *BERTScore F1*. Visualisasi ini memberikan gambaran mengenai performa model dalam merangkum berita berdasarkan beberapa metrik akurasi yang dihasilkan dari proses evaluasi, sehingga pengguna dapat melihat seberapa baik ringkasan yang dihasilkan oleh sistem.

4.5 Evaluasi Hasil Pengujian

Pada bagian ini dilakukan dua evaluasi yaitu evaluasi menggunakan berbagai metrik, termasuk *ROUGE*, *BLEU*, dan *BERTScore* untuk mengukur kesesuaian antara ringkasan yang dihasilkan oleh sistem dan kalimat asli dan evaluasi tingkat *naturalness* dari ringkasan berita yang dihasilkan oleh sistem. Evaluasi ini dilakukan dengan memilih lima contoh ringkasan berita yang dihasilkan oleh sistem.

Evaluasi pertama dilakukan dengan menggunakan kombinasi metrik *ROUGE*, *BLEU*, dan *BERTScore* untuk mendapatkan informasi mengenai tingkat kesamaan, keterbacaan, dan relevansi antara ringkasan dan kalimat asli. Metrik yang digunakan masing-masing memberikan informasi terkait aspek yang berbeda, seperti kesesuaian kata (*ROUGE*), kualitas bahasa (*BLEU*), dan kesamaan semantik (*BERTScore*).

Metrics	Average Score
BERTScore F1	0.880598783493042
BLEU	0.786191967321536
ROUGE-1	0.8452235765257344
ROUGE-2	0.8043759947149519
ROUGE-L	0.8452235765257344

Gambar 4.8. Hasil Evaluasi Model *BERT*, *BLEU* dan *ROUGE*

Pada Gambar 4.8, dapat dilihat bahwa nilai evaluasi yang diperoleh dari metrik BERTScore F1, BLEU, dan berbagai varian ROUGE (ROUGE-1, ROUGE-2, dan ROUGE-L) memberikan hasil yang cukup baik. BERTScore F1 menunjukkan nilai tertinggi sebesar 0.880, yang menandakan bahwa ringkasan yang dihasilkan oleh sistem memiliki kesamaan semantik yang tinggi dengan teks aslinya. Hal ini menunjukkan bahwa model berhasil menangkap makna utama dari artikel yang diringkas.

Nilai BLEU sebesar 0.786 juga memberikan hasil yang baik, meskipun metrik ini lebih menitikberatkan pada kesamaan n-gram antara ringkasan dan teks asli. Sedangkan ROUGE-1 dan ROUGE-L menunjukkan nilai F-Score yang konsisten di atas 0.84, menandakan bahwa ringkasan yang dihasilkan cukup akurat dalam hal kesamaan token dengan kalimat asli, baik dalam n-gram individu maupun urutan kalimat secara keseluruhan.

Namun, nilai ROUGE-2 yang lebih rendah (0.804) mengindikasikan bahwa kesamaan berdasarkan dua kata berurutan dalam ringkasan dan kalimat asli lebih sedikit. Meskipun demikian, hasil ini tetap menunjukkan bahwa ringkasan yang dihasilkan secara umum cukup baik dan relevan, dengan peluang untuk meningkatkan performa melalui optimasi lebih lanjut." Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa pendekatan Natural Language Generation (NLG) yang digunakan dalam penelitian ini efektif dalam menghasilkan ringkasan yang informatif dan relevan, meskipun masih terdapat ruang untuk perbaikan di aspek tertentu.

Evaluasi system kedua dilakukan dengan menilai tingkat *naturalness* dari ringkasan berita yang dihasilkan oleh sisteem. Penilaian *naturalness* dilakukan oleh orang umum usia mulai dari 20-30, tenaga medis sebagai pengguna sistem, dan ahli bahasa. Responden diberi contoh ringkasan berita yang telah disiapkan, kemudian

diminta untuk menilai ringkasan tersebut dengan skala 1 hingga 5 pada tiga aspek *naturalness*, yaitu *readability*, *clarity*, dan *understanding*.

a. Readability

Readability adalah aspek yang mengukur kemudahan membaca ringkasan berita. Skala penilaian meliputi:

- 1: menyatakan ringkasan berita sangat sulit dibaca
- 2: menyatakan ringkasan berita sulit dibaca
- 3: menyatakan ringkasan berita cukup dapat dibaca
- 4: menyatakan ringkasan berita mudah dibaca
- 5: menyatakan ringkasan berita sangat mudah dibaca

b. Clarity

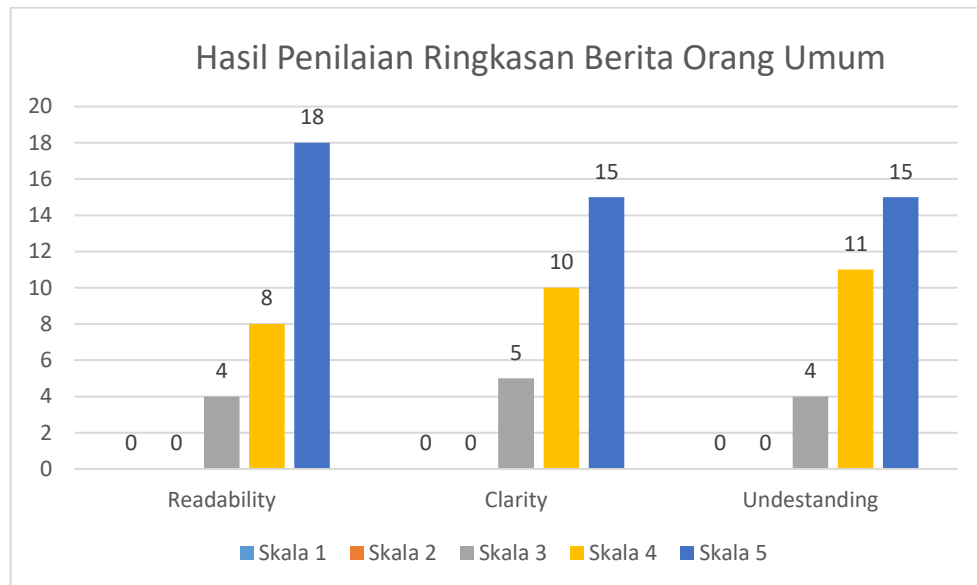
Clarity adalah aspek yang mengukur kejelasan penyampaian informasi dalam ringkasan berita. Skala penilaian meliputi:

- 1: menyatakan ringkasan berita sangat tidak jelas
- 2: menyatakan ringkasan berita tidak jelas
- 3: menyatakan ringkasan berita cukup jelas
- 4: menyatakan ringkasan berita jelas
- 5: menyatakan ringkasan berita sangat jelas

c. Understanding

Understanding adalah aspek yang mengukur pemahaman pembaca terhadap informasi yang disampaikan dalam ringkasan berita. Skala penilaian meliputi:

- 1: menyatakan ringkasan berita sangat sulit dipahami
- 2: menyatakan ringkasan berita sulit dipahami
- 3: menyatakan ringkasan berita cukup dapat dipahami
- 4: menyatakan ringkasan berita mudah dipahami
- 5: menyatakan ringkasan berita sangat mudah dipahami



Gambar 4.5. Hasil Penilaian Ringkasan Berita Orang Umum

Pada Gambar 4.5, tingkat *naturalness* dari setiap aspek penilaian untuk lima contoh ringkasan berita yang dihasilkan oleh sistem dinilai oleh 30 responden berusia 20-30 tahun. Hasil evaluasi menunjukkan bahwa:

a. Readability

Hasil ini menunjukkan bahwa mayoritas responden (18 dari 30) menilai ringkasan berita sangat mudah dibaca (Skala 5), sementara 8 responden menilai mudah dibaca (Skala 4), dan 4 responden menilai cukup dapat dibaca (Skala 3). Tidak ada responden yang menilai ringkasan berita sulit dibaca (Skala 2) atau sangat sulit dibaca (Skala 1).

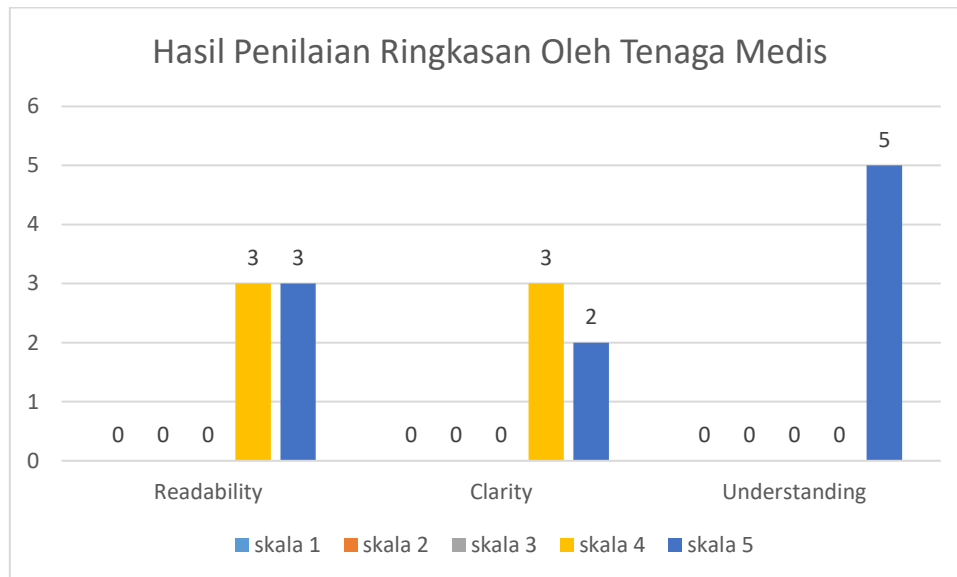
b. Clarity

Penilaian kejelasan menunjukkan bahwa separuh responden (15 dari 30) menilai ringkasan berita jelas (Skala 5), sementara 10 responden menilai mudah sekali dibaca (Skala 4), dan 5 responden menilai cukup dapat dibaca (Skala 3). Tidak ada responden yang menilai ringkasan berita tidak jelas (Skala 2), atau sangat tidak jelas (Skala 1).

c. Understanding

Penilaian pemahaman menunjukkan bahwa mayoritas responden (15 dari 30) merasa ringkasan berita sangat mudah dipahami (Skala 5), sementara 11

responden menilai berita mudah dipahami (Skala 4) dan 4 responden menilai beerita cukup dapat dipahami (Skala 3). Tidak ada responden yang menilai ringkasan berita sulit dipahami (Skala 2), atau sangat sulit dipahami (Skala 1).



Gambar 4.6. Hasil Penilaian Ringkasan Berita Tenaga Medis

Pada Gambar 4.6, tingkat *naturalness* dari setiap aspek penilaian untuk lima contoh ringkasan berita yang dihasilkan oleh sistem dinilai oleh 5 responden tenaga medis. Hasil evaluasi menunjukkan bahwa:

a. Readability

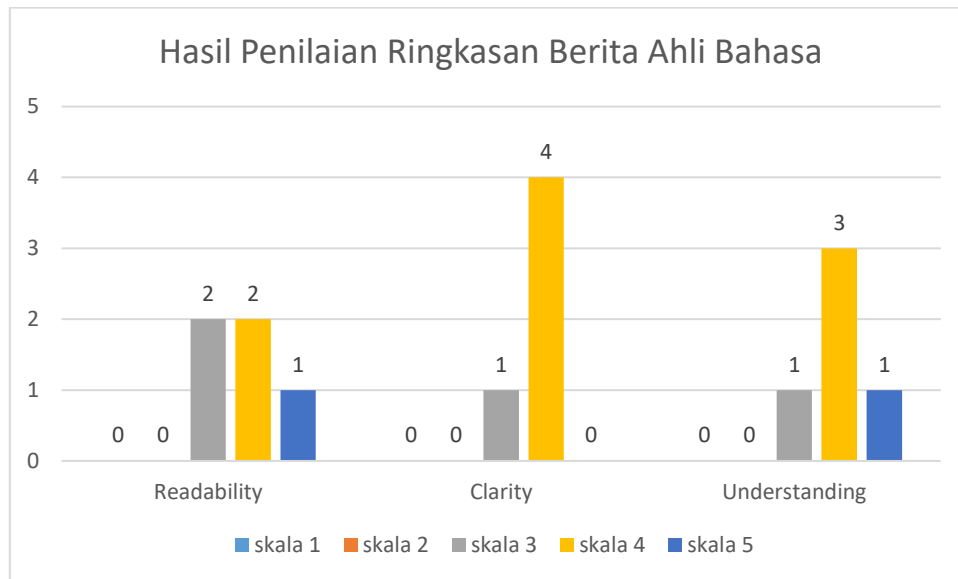
Mayoritas tenaga medis menilai ringkasan cukup mudah dibaca. Sebanyak dua responden menilai pada skala 4, dua responden menilai pada skala 3, dan satu responden memberikan penilaian terbaik di skala 5.

b. Clarity

Dalam aspek kejelasan, dua responden memberikan penilaian pada skala 4, sementara tiga responden memberikan penilaian di skala 3. Tidak ada yang menilai di bawah skala tersebut.

c. Understanding

Penilaian pemahaman menunjukkan mayoritas tenaga medis memberikan penilaian tertinggi di skala 5, dengan empat dari lima responden memberikan penilaian tersebut. Hanya satu responden yang menilai pada skala 4, menunjukkan bahwa mayoritas merasa ringkasan sangat mudah dipahami.



Gambar 4.7. Hasil Penilaian Ringkasan Berita Ahli Bahasa

Pada Gambar 4.7, tingkat *naturalness* dari setiap aspek penilaian untuk lima contoh ringkasan berita yang dihasilkan oleh sistem dinilai oleh 5 responden ahli bahasa. Hasil evaluasi menunjukkan bahwa:

a. Readability

Penilaian terhadap kemudahan membaca menunjukkan hasil yang beragam. Dua responden menilai pada skala 3, sementara dua responden menilai pada skala 2, dan satu responden memberikan penilaian terbaik di skala 5.

b. Clarity

Sebanyak empat ahli bahasa memberikan penilaian pada skala 4 untuk kejelasan ringkasan, sementara satu ahli bahasa memberikan penilaian pada skala 3. Tidak ada yang memberikan penilaian di bawah skala tersebut.

c. Understanding

Dalam aspek pemahaman, tiga ahli bahasa memberikan penilaian pada skala 4, sementara satu responden masing-masing memberikan penilaian pada skala 3 dan skala 5.

Berdasarkan hasil penilaian dari ketiga kelompok responden (orang umum, tenaga medis, dan ahli bahasa), dapat disimpulkan bahwa:

a. Readability

Orang umum memberikan penilaian yang lebih tinggi terhadap kemudahan membaca dibandingkan tenaga medis dan ahli bahasa. Ahli bahasa memberikan penilaian yang lebih kritis terhadap aspek ini, menunjukkan bahwa ada kebutuhan untuk menyesuaikan ringkasan agar lebih mudah dibaca oleh kalangan yang lebih terdidik dalam linguistik.

b. Clarity

Semua kelompok memberikan penilaian yang baik untuk kejelasan, meskipun tenaga medis dan ahli bahasa memberikan penilaian sedikit lebih rendah dibandingkan orang umum. Ini menunjukkan bahwa aspek kejelasan cukup baik tetapi masih bisa ditingkatkan.

c. Understanding

Semua kelompok menilai pemahaman terhadap ringkasan secara umum baik, dengan orang umum dan tenaga medis memberikan penilaian tertinggi. Ahli bahasa, meskipun memberikan penilaian yang baik, masih memiliki perbedaan dalam interpretasi mereka terhadap pemahaman teks.

Hasil evaluasi menunjukkan bahwa ringkasan berita yang dihasilkan sistem memiliki tingkat keterbacaan dan kejelasan yang baik di mata orang umum dan tenaga medis. Namun, untuk ahli bahasa, terdapat kebutuhan untuk meningkatkan keterbacaan, mengingat penilaian yang lebih kritis dari mereka. Secara keseluruhan, semua kelompok responden menilai pemahaman terhadap ringkasan secara positif, namun masih ada ruang untuk peningkatan di beberapa aspek, terutama dalam hal keterbacaan bagi kalangan yang lebih terdidik dalam linguistik.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dari penelitian pembangkitan anotasi peta berdasarkan ekstraksi data berita menggunakan pendekatan *Natural Language Generation* (NLG), dapat diambil beberapa kesimpulan sebagai berikut.

- a. Pembangkitan anotasi peta dilakukan dengan ekstraksi data artikel berita endemik yang didapatkan dari metode *scraping*, kemudian dilakukan peringkasan menggunakan model *pre-trained* atau menggunakan metode frekuensi dengan melewati proses *text preprocessing* menggunakan metode *case folding*, *tokenizing* dan *stopword removal*. Hasil data yang telah diekstrak kemudian dilakukan pencarian pada *database* yang berisi data jenis endemik dan daerah (*latitude* dan *longitude*) yang terdapat pada data artikel berita tersebut kemudian disimpan ke dalam *database* dengan hasil yang sudah sesuai dengan struktur untuk anotasi peta. Menggunakan data tersebut dilakukan pembuatan *Sistem Informasi Geografis* (SIG) yang menarik dan informatif.
- b. Penerapan metode *Natural Language Generation* (NLG) dengan model *pre-trained* dan metode frekuensi berhasil menghasilkan ringkasan berita yang baik dalam aspek *readability*, *clarity*, dan *understanding*. Hasil evaluasi dari 40 responden menunjukkan tingkat kepuasan yang tinggi, dengan rata-rata penilaian 4.5 dari 5 untuk kegunaan dan relevansi informasi. Selain itu, evaluasi juga dilakukan menggunakan berbagai metrik, termasuk BLEU, ROUGE, dan BERTScore, dengan hasil akurasi rata-rata yang menunjukkan performa baik. Nilai rata-rata BERTScore F1 mencapai 0.88, BLEU 0.78, ROUGE-1 0.84, ROUGE-2 0.80, dan ROUGE-L 0.84, yang menandakan bahwa ringkasan yang dihasilkan memiliki kesesuaian tinggi dengan teks asli. Dengan demikian, sistem ini terbukti efektif dalam menyajikan informasi yang relevan dan berkualitas tinggi, sekaligus memberikan landasan yang kuat untuk pengembangan lebih lanjut dalam meningkatkan pengalaman pengguna.

5.2 Saran

Berdasarkan hasil dari penelitian pembangkitan anotasi peta berdasarkan ekstraksi data berita endemik menggunakan pendekatan *Natural Language Generation* (NLG), dapat diambil beberapa saran untuk penelitian berikutnya sebagai berikut.

- a. Menggunakan model *pre-trained* lainnya dalam pembuatan ringkasan artikel berita untuk pembangkitan anotasi peta.
- b. Menggunakan metode lainnya seperti *Named-Entity Recognition* (NER) dalam menentukan jenis endemik dan daerah endemik.
- c. Menambahkan algoritma yang mampu memahami konteks lokasi dalam berita, sehingga dapat membedakan lokasi utama kejadian wabah dan lokasi lain yang terdapat pada berita.
- d. Menggunakan metode *transfer learning* untuk menyesuaikan model *pre-trained* dengan keperluan spesifik seperti peringkasan artikel berita endemik dalam bahasa Indonesia.

DAFTAR PUSTAKA

- Abidin, T. F., Ferdhiana, R., & Kamil, H. (2013). Automatic extraction of place entities and sentences containing the date and number of victims of tropical disease incidence from the web. *Journal of Emerging Technologies in Web Intelligence*, 5(3), 302–309.
- Abidin, T. F., Subianto, M., Gani, T. A., & Ferdhiana, R. (2015). Periodic update and automatic extraction of web data for creating a Google Earth based tool. *2015 International Conference on Advanced Computer Science and Information Systems (ICACISIS)*, 293–296.
- Aji, A. M. (2018). *Ekstraksi Lokasi Dan Topik Pada Berita Online Berbahasa Indonesia*. Universitas Gadjah Mada.
- Atika, D. (2021). *Ekstraksi Informasi Berita Online dengan Named Entity Recognition (NER) dan Rule-Based untuk Visualisasi Penyakit Tropis di Indonesia*. Universitas Sumatera Utara.
- Bahren, P. (2021, November 7). *Prof. Bahren: Korpus Sangat Bermanfaat bagi peneliti Bahasa*. Fakultas Humaniora Universitas Islam Negeri Maulana Malik Ibrahim Malang.<https://humaniora.uinmalang.ac.id/index.php/component/content/article/73-berita/373-prof-bahren-korpus-sangat-bermanfaat-bagi-peneliti-bahasa>
- Chen, T., Wang, X., Yue, T., Bai, X., Le, C. X., & Wang, W. (2023). Enhancing abstractive summarization with extracted knowledge graphs and multi-source transformers. *Applied Sciences*, 13(13), 7753.
- Danukusumo, K. P. (2017). *Implementasi deep learning menggunakan convolutional neural network untuk klasifikasi citra candi berbasis GPU*. UAJY.
- Dash, N. S., Chaudhuri, B. B., Rayson, P., Wilson, A., McEnery, T., Hardie, A., & Khoja, S. (2001). Corpus-based empirical analysis of form, function and frequency of characters used in Bangla. *Published in Rayson, P., Wilson, A., McEnery, T., Hardie, A., and Khoja, S.,(Eds.) Special Issue of the Proceedings of the Corpus Linguistics 2001 Conference, Lancaster: Lancaster University Press. UK, 13, 144–157.*

- Garg, A., Adusumilli, S., Yenneti, S., Badal, T., Garg, D., Pandey, V., Nigam, A., Gupta, Y. K., Mittal, G., & Agarwal, R. (2021). NEWS article summarization with pretrained transformer. *Advanced Computing: 10th International Conference, IACC 2020, Panaji, Goa, India, December 5–6, 2020, Revised Selected Papers, Part I 10*, 203–211.
- Gatt, A., & Krahmer, E. (2018). Survey of the State of the Art in Natural Language Generation: Core Tasks, Applications and Evaluation. *Journal of Artificial Intelligence Research*, 61, 65–170.
- Kanungo, Y. S., Das, G., & Negi, S. (2022). Cobart: Controlled, optimized, bidirectional and auto-regressive transformer for ad headline generation. *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 3127–3136.
- KN, A. M. (2018). *Panduan Produksi Acara TV Non-Drama*. Gramedia Widiasarana Indonesia. <https://books.google.co.id/books?id=AJNuDwAAQBAJ>
- Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *ArXiv Preprint ArXiv:1910.13461*.
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.
- Munir, A. (2012). Ilmu ukur wilayah dan sistem informasi geografis. *Jakarta: Kencana Prenada Media Group*.
- Nasution, K. (2016, March 5). *PETA : Definisi, Fungsi, Komponen, Jenis, Tahap-Tahap Pemetaan, dan Jenis-jenis Alat dalam Pemetaan*. KURNIA BLOG. <https://kurnia-12.blogspot.com/2016/03/peta-definisi-fungsi-komponen.html>
- Yuliyanti, W. D. (2020). Upaya World Health Organization (WHO) melalui Global Malaria Programme (GMP) dalam Mengatasi Penyakit Endemik Malaria di Indonesia Tahun 2016-2019. *Universitas Islam Negeri Sunan Ampel Surabaya*.
- Zhang, X. (2008). Route description using natural language generation technology. *Information Retrieval Technology: 4th Asia Information Retrieval Symposium, AIRS 2008, Harbin, China, January 15-18, 2008 Revised Selected Papers 4*, 454–459.

LAMPIRAN

Lampiran I.

Tabel Hasil Pengujian Sistem

No	URL	Data Asli	Data Summary	Lokasi	Endemik	Penilaian
1.	https://surabaya.kompas.com/read/2024/06/19/211712878/kasus-demam-berdarah-di-gresik-meningkatkan-korban-jawa-timur-ada-250-pasien-atau-kasus-demam-berdarah-pada-periode-januari-sampai-minggu-kedua-juni-2024-kemudian-120-kasus-pada-rentang-januari-hingga-juli-2023- sementara-januari-hingga-juli-2022-sebanyak-172-kasus-atau-pasien	Kasus Demam Berdarah di Gresik Meningkat, Ada Korban Jiwa Tim Redaksi KOMPAS.com - Penyakit demam berdarah kembali melanda Kabupaten Gresik, Jawa Timur. Korban yang terkena tidak hanya meningkat dibanding tahun lalu tetapi juga ada yang meninggal dunia. Menurut data Dinas Kesehatan (Dinkes) Gresik, ada 250 pasien atau kasus demam berdarah pada periode Januari sampai minggu kedua Juni 2024. Kemudian, 120 kasus pada rentang Januari hingga Juli 2023, sementara Januari hingga Juli 2022 sebanyak 172 kasus atau pasien.	Kasus Demam Berdarah di Gresik Meningkat, Ada Korban Jiwa Tim Redaksi KOMPAS.com - Penyakit demam berdarah kembali melanda Kabupaten Gresik, Jawa Timur. Menurut data Dinas Kesehatan (Dinkes) Gresik, ada 250 pasien atau kasus demam berdarah pada periode Januari sampai minggu kedua Juni	Gresik, Jawa Timur	DBD	Sesuai

			2024. Kemudian, 120 kasus pada rentang Januari hingga Juli 2023, sementara Januari hingga Juli 2022 sebanyak 172 kasus atau pasien.			
2.	https://www.detik.com/jabar/berita/d-7464590/dbd-di-kota-tasikmalaya-tembus-1-230-kasus-4-orang-meninggal	Kasus warga terjangkit demam berdarah dengue (DBD) di Kota Tasikmalaya sampai medio 2024 ini mencapai 1.230 kasus lebih. Empat orang diantaranya meninggal dunia. Salah satu penyebab lonjakan kasus penyakit menular ini, diakibatkan nyamuk DBD di Kota Tasikmalaya dianggap kebal pestisida. Nyamuk DBD di Kota Tasikmalaya sudah tak mempan lagi difogging. "Jadi beberapa waktu lalu kami Dinas Kesehatan Kota Tasikmalaya bersama Bapelkes Kemenkes melakukan penelitian,	"Jadi beberapa waktu lalu kami Dinas Kesehatan Kota Tasikmalaya bersama Bapelkes Kemenkes melakukan penelitian,	Tasikmalaya	DBD	Sesuai

		diketahui dari beberapa sampel nyamuk DBD ternyata sudah resisten terhadap pestisida," kata Kepala Dinas Kesehatan Kota Tasikmalaya, Uus Supangat, Selasa (30/7/2024).					Selasa (30/7/2024). Salah satu penyebab lonjakan kasus penyakit menular ini, diakibatkan nyamuk DBD di Kota Tasikmalaya dianggap kebal pestisida. Kasus warga terjangkit demam berdarah dengue (DBD) di Kota Tasikmalaya sampai medio 2024 ini mencapai 1.230 kasus lebih.
3.	https://diskominfo.kaltimprov.go.id/kesehatan/jumlah-positif-malaria-di-	Samarinda - Penyakit malaria di Provinsi Kalimantan Timur mengalami peningkatan, berdasarkan data ditemukan 61 kasus positif tahun 2023 Peta sebaran kasus positif malaria di Kaltim ditemukan di	Samarinda - Penyakit malaria di Provinsi Kalimantan Timur mengalami peningkatan, berdasarkan	Provinsi	Malaria	Sesuai	

kaltim- capai-61- kasus	<p>Penajam Paser Utara 2 kasus, Kutai Timur 38 kasus, Berau 2 kasus, Kutai Barat 3 kasus dan Balikpapan 16 kasus. Dilihat dari potensi penyebaran malaria di Kaltim, Penajam Paser Utara masih menjadi daerah tertinggi potensi malaria dilihat dari Annual Parasite Incidence (API). API merupakan indikator yang digunakan untuk menentukan tren morbiditas malaria dan menentukan endemik suatu daerah. Kepala Dinkes Kaltim dr Jaya Mualimin mengatakan Penajam Paser Utara masih jadi yang tertinggi untuk angka kesakitan malaria (API) diangka 6,44 poin.</p>	<p>data ditemukan 61 kasus positif tahun 2023 Peta sebaran kasus positif malaria di Kaltim ditemukan di Penajam Paser Utara 2 kasus, Kutai Timur 38 kasus, Berau 2 kasus, Kutai Barat 3 kasus dan Balikpapan 16 kasus. Dilihat dari potensi penyebaran malaria di Kaltim, Penajam Paser Utara masih menjadi daerah tertinggi potensi malaria dilihat dari Annual Parasite Incidence (API). Kepala Dinkes Kaltim dr Jaya</p>	<p>Barat, Kutai Timur, Berau, Balikpapan, Samari nda</p>
-------------------------------	--	---	--

			Mualimin mengatakan Penajam Paser Utara masih jadi yang tertinggi untuk angka kesakitan malaria (API) diangka 6,44 poin.			
4.	https://www.detik.com/jatim/berita/d-6651881/81-753-warga-jatim-alami-tbc-penyumbang-tertinggi-kota-surabaya	Kasus Tuberkulosis (TBC) yang ditemukan sepanjang tahun 2022 ada 81.753 atau 74% dari estimasi 107.547 yang ditemukan di Jawa Timur. Kasus terbanyak di Jawa Timur, ada di Kota Surabaya dengan jumlah kasus sebanyak 10.741. "Untuk estimasi kasus TBC 107.547, beban dari pusat ke Jatim harus menemukan ini. Yang ketemu 81.753 atau 74%. Sisanya PR," kata Kadinkes Jatim Erwin Ashta Triyono kepada wartawan, Minggu (2/4/2023). Erwin mengatakan, dari 81.753 kasus TBC, paling banyak TBC sensitif obat.	Kasus Tuberkulosis (TBC) yang ditemukan sepanjang tahun 2022 ada 81.753 atau 74% dari estimasi 107.547 yang ditemukan di Jawa Timur. Sisanya PR," kata Kadinkes Jatim Erwin Ashta Triyono kepada wartawan, Minggu (2/4/2023). Erwin mengatakan,	Surabaya, Jawa Timur	TBC	Sesuai

Sedangkan daerah yang paling tinggi ditemukan kasus TBC adalah Kota Surabaya.

ADVERTISEMENT

dari 81.753 kasus TBC, paling banyak TBC sensitif obat.

SCROLL TO CONTINUE
WITH CONTENT

5.	https://www.kompas.id/baca/utusan/2023/12/08/dbd-renggut-77-jiwa-di-kalbar-sepanjang-g-2023	Hingga pekan ke-48 tahun ini, terdapat 6.939 kasus demam berdarah dengue di Kalimantan Barat, dengan 77 orang meninggal. Masyarakat diminta terus waspada di tengah curah hujan yang meningkat. Kepala Dinas Kesehatan Provinsi Kalimantan Barat Erna Yulianti PONTIANAK, KOMPAS — Sebanyak 77 orang meninggal akibat demam berdarah dengue atau DBD di Kalimantan Barat sepanjang tahun 2023. Penyebaran DBD semakin meningkat saat musim hujan. Berdasarkan data Dinas Kesehatan (Dinkes) Kalbar, tercatat 6.939 kasus DBD sepanjang tahun ini. Kasusnya	Kepala Dinas Kesehatan Provinsi Kalimantan Barat Erna Yulianti PONTIANAK, KOMPAS — Sebanyak 77 orang meninggal akibat demam berdarah dengue atau DBD di Kalimantan Barat sepanjang tahun 2023. Hingga pekan ke-48 tahun ini, terdapat 6.939 kasus demam berdarah dengue di	Kalima ntan Barat, Landak , Mempa wah, Sangga u, Kubu Raya, Pontian ak	DBD	Sesuai
----	---	--	--	---	-----	--------

tersebar di 14	Kalimantan
kabupaten/kota. Beberapa	Barat, dengan
daerah rawan adalah Kubu	77 orang
Raya, Sanggau, Landak,	meninggal.
dan Mempawah. Sejauh	Beberapa
ini, baru Sanggau yang	daerah rawan
menyatakan kejadian luar	adalah Kubu
biasa DBD.	Raya, Sanggau,
	Landak, dan
	Mempawah.

Lampiran II. Kuesioner Evaluasi Kalimat Hasil Anotasi Berita Endemik

Evaluasi Kalimat Hasil Anotasi Berita Endemik berdasarkan Readability, Clarity, dan Understanding

Instruksi:

Berikut ini adalah data berita asli dan kalimat summarization yang dihasilkan oleh sistem BART. Bacalah keduanya dengan cermat, kemudian berikan penilaian pada aspek **Readability (Keterbacaan)**, **Clarity (Kejelasan)**, dan **Understanding (Pemahaman)**. Skala penilaian adalah dari 1 (sangat buruk) hingga 5 (sangat baik).

Data Berita 1 (Link

: https://semarangkota.go.id/p/5842/dinkes_kota_semarang_ingatkan_waspadadbd)

Dinas Kesehatan (Dinkes) Kota Semarang mengingatkan warga tetap waspada penyakit demam berdarah dengue (DBD). Apalagi, hujan masih kerap melanda ibu kota Jawa Tengah.

Kepala Dinkes Kota Semarang, Moh Abdul Hakam mengatakan, kasus DBD mencapai 90 kasus hingga akhir Maret atau pekan ke-12 pada tahun ini. Sedangkan kasus demam dengue (DD) mencapai 1.400 kasus.

"Temuan kami cukup banyak karena kami menyebar rapid test. Kalau dulu kena DB tshu saat di RS. Sekarang rapid test sudah bisa dilakukan secara langsung di puskesmas," jelas Hakam, Senin (15/4/2024).

Pasca banjir beberapa waktu lalu, Dinkes melalui puskesmas gencar melakukan rapid test untuk deteksi dini terhadap masyarakat yang memeriksakan diri ke puskesmas dengan gejala mengarah ke DBD maupun DD.

Rentang usia paling dominasi kasus DBD adalah usia remaja dengan persentase 41 persen. Disusul, dewasa 22 persen. Kemudian, balita memiliki persentase kasus sembilan persen dan pralansia delapan persen. Sisnya, kasus DBD pada neonatal hanya satu persen dan lansia juga satu persen.

Kasus DBD didominasi laki-laki dengan persentase 52 persen. Sejauh ini, daerah paling tinggi kasus DBD justru di Banyumanik.

Dia meminta wilayah dengan tinggi kasus DBD atau masuk peta kerawanan merah untuk melakukan pemberantasan jentik nyamuk (PJN) dua kali dalam satu minggu.

"Normalnya seminggu sekali, tapi di peta kerawanan merah kami melakukan dua kali seminggu. Ini strategi paling jitu. Kalau fogging hanya membuat nyamuk tidak sadar selama 10 jam," ujarnya.

Jika suatu wilayah memiliki angka bebas jentik (ABJ) di bawah 90 persen, dia memastikan, wilayah tersebut terdapat kasus DBD atau DD. Oleh karena itu, dia meminta masyarakat untuk melakukan PJN secara rutin.

Kalimat yang dihasilkan :

Jika suatu wilayah memiliki angka bebas jentik (ABJ) di bawah 90 persen, dia memastikan, wilayah tersebut terdapat kasus DBD atau DD. Kepala Dinkes Kota Semarang, Moh Abdul Hakam mengatakan, kasus DBD mencapai 90 kasus hingga akhir Maret atau pekan ke-12 pada tahun ini. Sekarang rapid test sudah bisa dilakukan secara langsung di puskesmas," jelas Hakam, Senin (15/4/2024).

Penilaian Aspek Untuk Data Berita 1

1. Readability (Keterbacaan)

Seberapa mudah kalimat yang dihasilkan ini dibaca dan dipahami dari segi struktur dan tata bahasa?

2. Clarity (Kejelasan Informasi)

Apakah kalimat yang dihasilkan ini menyampaikan informasi dengan jelas dan tidak membingungkan?

3. Understanding (Pemahaman Konten)

Apakah kalimat yang dihasilkan mudah dipahami oleh pembaca (baik profesional medis maupun masyarakat umum) dan sesuai dengan informasi yang diberikan dalam berita asli?

Keterangan : Skala penilaian adalah dari 1 (sangat buruk) hingga 5 (sangat baik).

	1	2	3	4	5
Readability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Clarity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Understanding	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Data Berita 2 (Link : <https://dutatv.com/dinkes-catat-ada-57-kasus-dbd-di-banjarmasin-3-meninggal-dunia/>)

Banjarmasin, DUTA TV — Kasus Demam Berdarah Dengue (DBD) di Kota Banjarmasin, sudah tercatat sekitar 57 kasus hingga bulan Mei ini. Dari data terakhir Dinas Kesehatan Kota Banjarmasin, pada bulan April lalu ada sekitar lima orang yang menderita DBD. Sedangkan kasus yang cukup tinggi terjadi pada bulan Januari lalu, dengan 23 kasus dan Februari 17 kasus. Peningkatan kasus tersebut dikarenakan peralihan musim atau pancaroba. Untuk kasus kematian tercatat tiga orang. Ketiganya merupakan anak umur dibawah 10 tahun. Pihak Dinas Kesehatan Kota Banjarmasin juga terus menginstruksikan jajaran dibawahnya untuk melaksanakan pengamanan dengan cepat dan tepat. Mengantisipasi kembali melonjaknya kasus DBD, kepala Dinas Kesehatan Kota Banjarmasin, M. Ramadhan mengimbau kepada seluruh warga terus melaksanakan 3M Plus, dan menjaga lingkungan mereka, serta menerapkan perilaku hidup sehat. Ramadhan juga menambahkan, jika ada warga yang terlihat tanda seperti DBD, agar segera membawa ke Fasyenkes terdekat agar bisa segera ditangani.

Kalimat yang dihasilkan :

Mengantisipasi kembali melonjaknya kasus DBD, kepala Dinas Kesehatan Kota Banjarmasin, M. Ramadhan mengimbau kepada seluruh warga terus melaksanakan 3M Plus, dan menjaga lingkungan mereka, serta menerapkan perilaku hidup sehat. Banjarmasin, DUTA TV — Kasus Demam Berdarah Dengue (DBD) di Kota Banjarmasin, sudah tercatat sekitar 57 kasus hingga bulan Mei ini. Dari data terakhir Dinas Kesehatan Kota Banjarmasin, pada bulan April lalu ada sekitar lima orang yang menderita DBD.

Penilaian Aspek Untuk Data Berita 2

1. Readability (Keterbacaan)

Seberapa mudah kalimat yang dihasilkan ini dibaca dan dipahami dari segi struktur dan tata bahasa?

2. Clarity (Kejelasan Informasi)

Apakah kalimat yang dihasilkan ini menyampaikan informasi dengan jelas dan tidak membingungkan?

3. Understanding (Pemahaman Konten)

Apakah kalimat yang dihasilkan mudah dipahami oleh pembaca (baik profesional medis maupun masyarakat umum) dan sesuai dengan informasi yang diberikan dalam berita asli?

Keterangan : Skala penilaian adalah dari 1 (sangat buruk) hingga 5 (sangat baik).

	1	2	3	4	5
Readability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Clarity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Understanding	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Data Berita 3 (Link: <https://beritacikarang.com/januari-mei-2023-ada-366-kasus-dbd-di-kabupaten-bekasi-turun-dibanding-tahun-lalu/>)

BERITACIKARANG.COM, CIKARANG PUSAT – Dinas Kesehatan Kabupaten Bekasi mencatat terdapat 366 kasus Demam Berdarah Dengue (DBD) yang terjadi di wilayahnya pada periode Januari – Mei 2023. Jumlah kasus ini menurun dibandingkan 2022.

Kepala Seksi Pencegahan dan Penanggulangan Penyakit Menular Dinas Kesehatan Kabupaten Bekasi Ahmad Nurfaal merinci kasus DBD yang terjadi selama 2023, yakni 65 kasus pada Januari 2023, 53 kasus pada Februari 2023, 66 kasus pada Maret 2023, 76 kasus pada April 2023 dan 106 kasus pada Mei 2023.

“Jumlah kasus tahun ini memang banyak. Meski demikian, jumlahnya masih jauh di bawah kasus yang terjadi pada periode yang sama tahun lalu,” ungkapnya, Kamis (09/06).

Adapun jumlah kasus DBD pada tahun 2022 lalu total berjumlah 537 kasus. Rincian 73 kasus pada Januari 2022, 46 kasus pada Februari 2022, 155 kasus pada Maret 2022, 140 kasus pada April 2022 dan 123 kasus pada Mei 2022.

“Meskipun angkanya masih jauh dari tahun lalu, kami tetap waspada. Kami terus berupaya mengendalikan kasus DBD dengan mengajak masyarakat menerapkan Perilaku Hidup Bersih dan Sehat (PHBS) dan melakukan Gerakan 1 Rumah 1 Jumantik (Juru Pemantau Jentik),” ungkapnya.

Ahmad Nurfaal menekankan agar masyarakat Kabupaten Bekasi dapat melakukan PSN (Pemberantasan Sarang Nyamuk) dengan cara 3M plus yakni menguras tempat penampungan air, menutup rapat semua tempat penampungan air agar nyamuk tidak bisa masuk serta menimbun atau mendaur ulang limbah barang bekas yang sudah tidak terpakai supaya tidak dijadikan tempat berkembangbiak nyamuk.

“Masyarakat harus memastikan setiap lingkungan rumah baik di dalam rumah dan di luar rumah bebas jentik. Lakukan pemantauan rumah sendiri seminggu sekali. Lakukan Gerakan bersama di lingkungan masing masing tingkat RT. Misalnya Jumat bersih atau Minggu bersih,” pungkasnya. (dim)

Kalimat yang dihasilkan :

Rincian 73 kasus pada Januari 2022, 46 kasus pada Februari 2022, 155 kasus pada Maret 2022, 140 kasus pada April 2022 dan 123 kasus pada Mei 2022. Ahmad Nurfalah menekankan agar masyarakat Kabupaten Bekasi dapat melakukan PSN (Pemberantasan Sarang Nyamuk) dengan cara 3M plus yakni menguras tempat penampungan air, menutup rapat semua tempat penampungan air agar nyamuk tidak bisa masuk serta menimbun atau mendaur ulang limbah barang bekas yang sudah tidak terpakai supaya tidak dijadikan tempat berkembangbiak nyamuk. Kami terus berupaya mengendalikan kasus DBD dengan mengajak masyarakat menerapkan Perilaku Hidup Bersih dan Sehat (PHBS) dan melakukan Gerakan 1 Rumah 1 Jumantik (Juru Pemantau Jentik),” ungkapnya.

Penilaian Aspek Untuk Data Berita 3

1. Readability (Keterbacaan)

Seberapa mudah kalimat yang dihasilkan ini dibaca dan dipahami dari segi struktur dan tata bahasa?

2. Clarity (Kejelasan Informasi)

Apakah kalimat yang dihasilkan ini menyampaikan informasi dengan jelas dan tidak membingungkan?

3. Understanding (Pemahaman Konten)

Apakah kalimat yang dihasilkan mudah dipahami oleh pembaca (baik profesional medis maupun masyarakat umum) dan sesuai dengan informasi yang diberikan dalam berita asli?

Keterangan : Skala penilaian adalah dari 1 (sangat buruk) hingga 5 (sangat baik).

	1	2	3	4	5
Readability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Clarity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Understanding	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Data Berita 4 (Link : <https://sukabumiberita.com/januari-april-2024-sebanyak-639-warga-kota-sukabumi-kena-dbd/>)

[SukabumiBerita.com](https://sukabumiberita.com)—Angka kasus Demam Berdarah Dengue (DBD) di Kota Sukabumi, ternyata masih tinggi. Terbukti, dari data Dinas Kesehatan (Dinkes) sepanjang Januari hingga April 2024 sebanyak 639 warga menderita penyakit tersebut.

Kabid Pencegahan dan Pengendalian Penyakit (P2P) Dinkes Kota Sukabumi Wita Darmawanti mengungkapkan dari sebanyak 639 kasus satu di antaranya meninggal dunia pada April 2024 lalu,” kata Wita kepada wartawan, Selasa (7/5/2024).

Apabila melihat dari data yang ada, kasus DBD terjadi paling banyak pada April. Sebab itu, saat musim hujan masyarakat perlu waspada dari berbagai penyakit salah satunya penularan DBD. “Chikungunya dapat ditularkan melalui gigitan nyamuk aedes aegypti. Kondisi ini umumnya ditandai dengan demam tinggi dan nyeri sendi secara mendadak,” katanya.

Menurutnya, genangan air hujan bisa menjadi tempat berkembang biaknya nyamuk aedes aegypti yang merupakan penyebab DBD. “Sebab itu, kami terus berupa melakukan upaya penanggulangan dan pencegahan, terkait penyakit yang harus diwaspadai saat musim penghujan ini. Upaya yang kita lakukan yaitu Promosi Kesehatan (Promkes) dan edukasi. Pelayanan kesehatan tetap siap,” katanya.

Dinkes terus berupaya menurunkan angka kasus DBD ini dengan berbagai cara. Misalnya, mensosialisasikan pencegahan ke masyarakat tentang 3 M Plus, yakni menguras, menutup, mendaur dan mencegah, serta mengedukasi masyarakat agar menerapkan Perilaku Hidup Bersih dan Sehat (PHBS). “Terpenting itu, masyarakat harus bisa melakukan gerakan PSN (Pemberantasan Sarang Nyamuk),” katanya.

Masyarakat agar menerapkan Gerakan 1 Rumah 1 Jumantik (G1R1J). Program itu bertujuan untuk mengoptimalkan segenap anggota keluarga menjadi Juru Pemantau Jentik (Jumantik) di rumahnya masing-masing.

“Kami juga mengaktifkan kelompok operasional penanggulangan DBD atau Pokjanal di berbagai tingkatan RT, RW, kelurahan, kecamatan, hingga tingkat kota, sehingga dengan berbagai upaya itu diharapkan mampu menekan jumlah kasus DBD di Kota Sukabumi,” katanya.

Kalimat yang dihasilkan :

Misalnya, mensosialisasikan pencegahan ke masyarakat tentang 3 M Plus, yakni menguras, menutup, mendaur dan mencegah, serta mengedukasi masyarakat agar menerapkan Perilaku Hidup Bersih dan Sehat (PHBS). Kabid Pencegahan dan Pengendalian Penyakit (P2P) Dinkes Kota Sukabumi Wita Darmawanti mengungkapkan dari sebanyak 639 kasus satu di antaranya meninggal dunia pada April 2024 lalu,” kata Wita kepada wartawan, Selasa (7/5/2024). “Terpenting itu, masyarakat harus bisa melakukan gerakan PSN (Pemberantasan Sarang Nyamuk),” katanya.

Penilaian Aspek Untuk Data Berita 4

1. Readability (Keterbacaan)

Seberapa mudah kalimat yang dihasilkan ini dibaca dan dipahami dari segi struktur dan tata bahasa?

2. Clarity (Kejelasan Informasi)

Apakah kalimat yang dihasilkan ini menyampaikan informasi dengan jelas dan tidak membingungkan?

3. Understanding (Pemahaman Konten)

Apakah kalimat yang dihasilkan mudah dipahami oleh pembaca (baik profesional medis maupun masyarakat umum) dan sesuai dengan informasi yang diberikan dalam berita asli?

Keterangan : Skala penilaian adalah dari 1 (sangat buruk) hingga 5 (sangat baik).

	1	2	3	4	5
Readability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Clarity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Understanding	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Data Berita 5 (Link : <https://inspira.tv/kasus-dbd-meningkat-dua-anak-di-cianjur-meninggal-dunia/>)

BANDUNG INSPIRA – Kasus Demam Berdarah Dengue (DBD) di Kabupaten Cianjur meningkat secara signifikan. Menurut laporan beberapa rumah sakit di Kabupaten Cianjur, pada awal 2024 terdapat ratusan warga yang terjangkit DBD.

dr. Yusman Faiz, Kepala Dinas Kesehatan (Dinkes) Kabupaten Cianjur Mengatakan bahwa kasus DBD mengalami peningkatan yang signifikan pada Januari 2024.

“Dalam sebulan terdapat 219 kasus yang diperoleh oleh Dinkes Cianjur, dari jumlah tersebut dua anak dengan rentang usia 6 sampai 14 tahun meninggal,” ungkap Yusman, dikutip dari Pikiran Rakyat pada Senin (5/2/2024).

Yusman juga mengatakan, meningkatnya kasus DBD di Kabupaten Cianjur didebabkan oleh musim hujan yang menyebabkan banyak genangan air. Sebagian besar pasien yang terjangkit DBD merupakan pasien dalam usia produktif.

“Jadi kan sekarang musim hujan, banyak genangan-genangan air. Maka dari itu pencegahan terus dilakukan dengan fogging.

Beberapa daerah di Kabupaten Cianjur yang terjangkit DBD merupakan daerah yang memiliki kepadatan penduduk yang tinggi seperti Kecamatan Cianjur, Cilaku, dan Karangtengah.

Berdasarkan data yang dikumpulkan oleh beberapa rumah sakit di Kabupaten Cianjur, rumah sakit yang memiliki pasien DBD paling banyak adalah RSUD Cianjur dengan jumlah pasien DBD kurang lebih 30 orang.

Dinas Kesehatan Kabupaten Cianjur mengimbau agar masyarakat waspada, serta melakukan pemberantasan nyamuk dengan cara menguras genangan air dan menutup saluran air agar menghambat pertumbuhan jentik nyamuk.

“Apabila warga memang membutuhkan fogging, secara bertahap akan kita penuhi tapi kita akan buat skala prioritas dulu di lokasi dengan kasus DBD terbanyak,” tutur Yusman.

Sementara itu, Dinas Kesehatan mengalami kendala dalam fasilitas fogging karena keterbatasan dalam biaya, alat dan SDM sehingga fogging belum dapat dilakukan.

“Saat ini fogging dilakukan hanya sesuai dengan permintaan dari warga. Sebelum dilakukan fogging pun kita harus menerjunkan tim surveilans dulu untuk memastikan adanya korban juga pencarian jentik nyamuk *Aedes aegypti* dan *Aedes albopictus*. Jika sudah dipastikan ada, baru bisa dilakukan fogging,” ujar Yusman. (Tina)**

Kalimat yang dihasilkan :

Dalam sebulan terdapat 219 kasus yang diperoleh oleh Dinkes Cianjur, dari jumlah tersebut dua anak dengan rentang usia 6 sampai 14 tahun meninggal,” ungkap Yusman, dikutip dari Pikiran Rakyat pada Senin (5/2/2024). dr. Yusman Faiz, Kepala Dinas Kesehatan (Dinkes) Kabupaten Cianjur Mengatakan bahwa kasus DBD mengalami peningkatan yang signifikan pada Januari 2024. Yusman juga mengatakan, meningkatnya kasus DBD di Kabupaten Cianjur didebabkan oleh musim hujan yang menyebabkan banyak genangan air.

Penilaian Aspek Untuk Data Berita 5

1. Readability (Keterbacaan)

Seberapa mudah kalimat yang dihasilkan ini dibaca dan dipahami dari segi struktur dan tata bahasa?

2. Clarity (Kejelasan Informasi)

Apakah kalimat yang dihasilkan ini menyampaikan informasi dengan jelas dan tidak membingungkan?

3. Understanding (Pemahaman Konten)

Apakah kalimat yang dihasilkan mudah dipahami oleh pembaca (baik profesional medis maupun masyarakat umum) dan sesuai dengan informasi yang diberikan dalam berita asli?

Keterangan : Skala penilaian adalah dari 1 (sangat buruk) hingga 5 (sangat baik).

	1	2	3	4	5
Readability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Clarity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Understanding	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Penutup

Terima kasih atas partisipasi Anda dalam survei ini. Jawaban Anda akan sangat membantu dalam meningkatkan kualitas kalimat hasil sistem kami.

(Silakan tuliskan komentar tambahan Anda mengenai kalimat yang dihasilkan dan saran perbaikan, jika ada)

Your answer
