PERINGKASAN MULTI-DOKUMEN EKSTRAKTIF MENGGUNAKAN METODE *SENTENCE*-BERT DAN *MAXIMAL MARGINAL RELEVANCE* (MMR) PADA TEKS BERITA *ONLINE* BAHASA INDONESIA

SKRIPSI

Disusun Sebagai Salah Satu Syarat
untuk Memperoleh Gelar Sarjana Komputer
pada Program Studi Informatika

Disusun oleh:
Al Ferro Putra Yusanda
24060122140164

DEPARTEMEN INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
2026

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini :

Nama : Al Ferro Putra Yusanda

NIM : 24060122140164

Judul : Peringkasan Multi-Dokumen Ekstraktif Menggunakan Metode *Sentence*-BERT dan *Maximal Marginal Relevance* (MMR) pada Teks Berita *Online* Bahasa Indonesia

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Semarang, 23 Juni 2026

Al Ferro Putra Yusanda

24060122140164

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Diponegoro, saya yang bertanda tangan di bawah ini:

Nama	: Al Ferro Putra Yusanda
NIM	: 24060122140164
Program Studi	: Informatika
Departemen	: Informatika
Fakultas	: Sains dan Matematika
Jenis Karya	: Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan **Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty Free Right)** kepada Universitas Diponegoro atas karya ilmiah saya yang berjudul:

Peringkasan Multi-Dokumen Ekstraktif Menggunakan Metode Sentence-BERT dan Maximal Marginal Relevance (MMR) pada Teks Berita Online Bahasa Indonesia

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-eksklusif ini Universitas Diponegoro berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan skripsi saya tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Semarang, 23 Juni 2026

Al Ferro Putra Yusanda

24060122140164

HALAMAN PENGESAHAN

Judul : Peringkasan Multi-Dokumen Ekstraktif Menggunakan Metode *Sentence*-BERT dan *Maximal Marginal Relevance* (MMR) pada Teks Berita *Online* Bahasa Indonesia

Nama : Al Ferro Putra Yusanda

NIM : 24060122140164

Telah diujikan pada sidang tugas akhir dan dinyatakan lulus pada tanggal 23 Juni 2026

Semarang, 23 Juni 2026

Pembimbing 1,

Dr. Retno Kusumaningrum, S.Si., M.Kom.

NIP. 198104202005012001

Pembimbing 2,

Priyo Sidik Sasongko, S.Si., M.Kom.

NIP. 197007051997021001

Mengetahui,

Ketua

Departemen Informatika

Dr. Aris Sugiharto, S.Si., M.Kom.

NIP. 197108111997021004

Ketua

Panitia Penguji Tugas Akhir

Dr. Helmie Arif Wibawa, S.Si., M.Cs.

NIP. 197805162003121001

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Peringkasan Multi-Dokumen Ekstraktif Menggunakan Metode *Sentence*-BERT dan *Maximal Marginal Relevance* (MMR) pada Berita *Online* Bahasa Indonesia" sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer di Program Studi Informatika.

Dalam proses penyusunan skripsi ini, penulis telah menerima banyak dukungan, bimbingan, dan motivasi dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Kusworo Adi, S.Si., M.T., selaku Dekan Fakultas Sains dan Matematika Universitas Diponegoro.
2. Dr. Aris Sugiharto, S.Si., M.Kom., selaku Ketua Departemen Informatika Universitas Diponegoro.
3. Nurdin Bahtiar, S.Si., M.T., selaku Koordinator Tugas Akhir.
4. Dr. Retno Kusumaningrum, S.Si., M.Kom., selaku Dosen Pembimbing I yang telah menyediakan waktu, tenaga, dan pikiran untuk membimbing penulis.
5. Priyo Sidik Sasongko, S.Si., M.Kom., selaku Dosen Pembimbing II yang telah menyediakan waktu, tenaga, dan pikiran untuk membimbing penulis.
6. Keluarga dan rekan-rekan penulis, yang mendukung proses eksperimen, diskusi, semangat, dan dukungan moral maupun materi dalam setiap langkah penulis.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan, baik dari segi ruang lingkup maupun pendekatan yang digunakan. Oleh karena itu, kritik dan saran yang membangun dari berbagai pihak sangat penulis harapkan demi penyempurnaan karya ini. Akhir kata, semoga skripsi ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang *Natural Language Processing* dan kecerdasan buatan

Semarang, 23 Juni 2026

Penulis

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Diponegoro, saya yang bertanda tangan di bawah ini:

Nama	:	Al Ferro Putra Yusanda
NIM	:	24060122140164
Program Studi	:	Informatika
Departemen	:	Informatika
Fakultas	:	Sains dan Matematika
Jenis Karya	:	Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan **Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty Free Right)** kepada Universitas Diponegoro atas karya ilmiah saya yang berjudul:

Peringkasan Multi-Dokumen Ekstraktif Menggunakan Metode Sentence-BERT dan Maximal Marginal Relevance (MMR) pada Teks Berita Online Bahasa Indonesia

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-eksklusif ini Universitas Diponegoro berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan skripsi saya tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Semarang, 23 Juni 2026

Materai
Rp. 10.000,-

Al Ferro Putra Yusanda

24060122140164

ABSTRAK

Fenomena *information overload* pada portal berita daring berbahasa Indonesia sering kali memicu tingginya tingkat redundansi informasi lintas artikel, sehingga menyulitkan pembaca dalam memperoleh intisari peristiwa secara efisien. Penelitian ini mengusulkan metode peringkasan multi-dokumen ekstraktif yang mengintegrasikan *Agglomerative Clustering* sebagai algoritma pemilah topik, representasi semantik *Sentence*-BERT (SBERT) menggunakan varian *paraphrase-multilingual-MiniLM-L12-v2*, serta algoritma *Maximal Marginal Relevance* (MMR) sebagai mekanisme penyeleksi kalimat anti-redundan. Pendekatan komputasi berbasis *dense embedding* ini diimplementasikan guna mengatasi keterbatasan representasi statistik konvensional dalam mendeteksi makna parafrasa. Pengujian arsitektur dilakukan terhadap *benchmark dataset* dari portal berita Liputan6 yang terdistribusi ke dalam enam kategori topik utama. Kualitas ringkasan dievaluasi menggunakan metrik ROUGE terhadap *human ground truth* yang telah tervalidasi reliabilitasnya dengan nilai kesepakatan aktual (*Percentage Agreement*) sebesar 63,21%. Hasil eksperimen komputasional pada skenario tingkat kompresi 20%, 30%, 40%, dan 50% memvalidasi bahwa tingkat kompresi 30% merupakan titik keseimbangan (*sweet spot*) paling optimal antara kepadatan informasi utama dan keberagaman frasa. Pada ambang batas kompresi tersebut, metode usulan secara keseluruhan meraih nilai rata-rata performa kompetitif dengan *F1-Score* ROUGE-1 sebesar 0,4750, ROUGE-2 sebesar 0,3034, dan ROUGE-L sebesar 0,3391. Penelitian ini membuktikan secara empiris bahwa integrasi pemahaman semantik mendalam dan metrik diversitas komputasional efektif untuk mereduksi redundansi leksikal sekaligus mempertahankan akurasi faktual teks berita bagi pembaca digital.

Kata Kunci : Peringkasan Multi-Dokumen Ekstraktif, *Sentence*-BERT, *Maximal Marginal Relevance*, ROUGE.

ABSTRACT

The phenomenon of information overload in Indonesian online news portals often triggers a high level of information redundancy across articles, making it difficult for readers to efficiently grasp the core of an event. This study proposed a multi-document extractive text summarization method that integrated Agglomerative Clustering for topic isolation, Sentence-BERT (SBERT) semantic representation utilizing the *paraphrase-multilingual-MiniLM-L12-v2* variant, and the Maximal Marginal Relevance (MMR) algorithm as an anti-redundant sentence selection mechanism. This dense embedding-based computational approach was implemented to overcome the limitations of conventional statistical representations in detecting paraphrased meanings. The architectural testing was conducted on the Liputan6 benchmark dataset distributed into six main news categories. The quality of the summaries was evaluated using ROUGE metrics against human ground truth, which was validated for reliability with a Percentage Agreement of 63.21%. Computational experimental results across 20%, 30%, 40%, and 50% compression scenarios demonstrated that the 30% compression rate was the most optimal sweet spot for balancing main information density and phrase diversity. At this compression threshold, the proposed method achieved a competitive overall average performance with an F1-Score of 0.4750 for ROUGE-1, F1-Score of 0.3034 for ROUGE-2, and F1-Score of 0.3391 for ROUGE-L. This research empirically proves that the integration of deep semantic understanding and computational diversity metrics effectively reduces lexical redundancy while maintaining the factual accuracy of news texts for digital readers.

Keywords : Multi-Document Extractive Summarization, Sentence-BERT, Maximal Marginal Relevance, ROUGE.

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I
PENDAHULUAN

Bab ini menjelaskan kerangka dasar dari penelitian yang terdiri dari latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan penelitian, dan sistematika penulisan. Seluruh bagian tersebut disusun untuk memberikan gambaran umum mengenai arah, fokus, serta ruang lingkup penelitian secara menyeluruh.

1.1. Latar Belakang

Era digital telah mengubah cara masyarakat mengonsumsi informasi. Pertumbuhan pengguna internet di Indonesia mencapai 221,5 juta jiwa pada tahun 2024, dengan penetrasi sebesar 79,5% dari total populasi (Prasetiyo dkk., 2024). Kondisi ini memunculkan fenomena *information overload*, di mana pengguna dibanjiri volume berita yang sangat besar dengan konten yang seringkali redundan pada topik yang sama (Guo dkk., 2024). Masalah ini diperparah oleh praktik jurnalisme *online* di Indonesia yang saat ini mengutamakan kecepatan publikasi dibandingkan akurasi, sehingga satu peristiwa dapat diliput oleh puluhan portal berita dengan sudut pandang yang hampir identik dan repetitif (Mahendra & Sukartik, 2024).

Redundansi pada berita digital di Indonesia semakin parah karena banyak media saling mengutip satu sama lain dari sumber yang sama. Kondisi ini tidak lepas dari besarnya jumlah media online di Indonesia. Terdata oleh Dewan Pers mencatat ada 964 media online yang sudah terverifikasi hingga tahun 2023, sementara secara keseluruhan jumlah outlet media di Indonesia diperkirakan mencapai lebih dari 45.000 (Antara News, 2024; Kompas.id, 2023). Dengan persaingan yang sangat ketat, banyak media akhirnya memilih meliput topik yang sama berulang kali dengan sedikit perbedaan, sehingga informasi yang beredar menjadi sangat repetitif. Pengulangan seperti ini pada akhirnya membebani pembaca dan memperparah fenomena *information overload* (Peng dkk., 2021). Fakta ini juga didukung oleh *Reuters Institute Digital News Report* 2024, yang menemukan bahwa 39% pembaca di berbagai negara mengaku kelelahan akibat terlalu banyaknya berita—angka ini naik 11 poin persentase dibanding lima tahun sebelumnya (Newman dkk., 2024). Di Indonesia sendiri kondisinya bahkan lebih parah; laporan *Digital News Report* 2025 mencatat bahwa 75% konsumen berita Indonesia mengaku sering atau sesekali menghindari berita, lebih tinggi

dari rata-rata global yang sebesar 71% (Steele, 2025). Permasalahan diperparah dengan maraknya praktik *clickbait*, di mana media *online* lebih memprioritaskan judul sensasional demi mendongkrak kuantitas klik daripada menyajikan informasi substansial. Akibatnya, pembaca terpaksa menelusuri keseluruhan isi artikel hanya untuk memverifikasi konteks peristiwa yang sebenarnya (Trustyanda dkk., 2021). Tingginya volume teks yang redundan dan terfragmentasi di ruang digital ini berujung pada inefisiensi waktu serta tenaga pembaca dalam mengekstraksi inti informasi yang mereka butuhkan (Widyassari dkk., 2022).

Peringkasan teks otomatis (*automatic text summarization*) muncul sebagai solusi untuk mengatasi *information overload* (Widyassari dkk., 2022). Terdapat dua pendekatan utama yaitu berupa ekstraktif dan abstraktif. Metode abstraktif rentan terhadap anomali halusinasi, yakni menghasilkan informasi yang gramatikal namun faktanya keliru atau tidak pernah ada di dalam dokumen asli (Ji dkk., 2023; Maynez dkk., 2020). Mengingat berita jurnalistik menuntut akurasi faktual yang tinggi, pendekatan ekstraktif menjadi pilihan yang lebih rasional karena menjamin setiap kalimat berasal langsung dari sumber asli. Metode pendekatan ekstraktif dalam peringkasan masih sangat efektif untuk korpus berita berbahasa Indonesia (Koto dkk., 2020).

Arsitektur peringkasan teks secara umum terbagi menjadi dua kategori berdasarkan jumlah sumber dokumen masukan, yaitu peringkasan dokumen tunggal (*single-document summarization*) dan peringkasan multi-dokumen (*multi-document summarization*). Peringkasan dokumen tunggal menitikberatkan pada proses ekstraksi informasi dari satu sumber teks saja. Peringkasan multi-dokumen memiliki tingkat kompleksitas komputasi yang jauh lebih tinggi karena algoritma harus memproses serta mensintesis informasi dari berbagai dokumen berbeda yang membahas topik serupa (Mawaridi dkk., 2024). Tantangan utama dalam peringkasan multi-dokumen mencakup tingginya tingkat redundansi kalimat, variasi gaya penulisan, serta risiko inkonsistensi fakta antar-artikel berita (Koto dkk., 2020; Widyassari dkk., 2022). Mesin peringkas memerlukan mekanisme pemilihan informasi yang cerdas guna menyaring bagian paling representatif dari seluruh sumber dokumen tersebut.

Proses peringkasan multi-dokumen memerlukan kepastian penunggalan tema dari seluruh artikel masukan sebelum pemrosesan semantik di tingkat kalimat dieksekusi. Aliran berita digital yang bersifat heterogen dan acak menuntut adanya penyaringan di tingkat dokumen (*document-level*) agar fakta dari berbagai sub-topik tidak bercampur secara rancu.

Ketiadaan pengelompokan awal pada pemrosesan multi-dokumen akan memicu penurunan drastis pada koherensi ringkasan, karena mesin berisiko menyatukan konteks dari sub-peristiwa yang sebenarnya berbeda (Lennox dkk., 2023). Oleh sebab itu, tahapan pengelompokan dokumen (*document clustering*) menjadi instrumen krusial untuk mengisolasi teks secara spesifik sebelum proses ekstraksi dimulai. Mengingat jumlah peristiwa atau topik berita yang muncul di internet tidak dapat diprediksi secara pasti, metode *Agglomerative Clustering* dipilih sebagai solusi utama. Metode ini telah terbukti sangat tangguh dalam memilah dan mengelompokkan dokumen-dokumen berita secara hierarkis tanpa mengharuskan sistem untuk menentukan jumlah klaster mutlak K secara subjektif di awal pemrosesan (Wicaksana dkk., 2018; Hanley dkk., 2025).

Berbagai penelitian terdahulu telah mengeksplorasi metode ekstraksi dan penilaian relevansi kalimat secara komputasional untuk korpus berita bahasa Indonesia. Pendekatan awal umumnya menggunakan metode statistik konvensional seperti TF-IDF, namun metode ini memiliki keterbatasan fundamental karena hanya mempertimbangkan frekuensi kemunculan kata tanpa memahami konteks semantik. Akibatnya, dua kalimat bermakna sama yang menggunakan kosakata berbeda dinilai tidak memiliki kemiripan, sehingga memicu tingginya redundansi pada hasil ringkasan (Gunawan dkk., 2019). Berupaya memperbaiki kelemahan tersebut, fokus penelitian selanjutnya berkembang pada dua arah solusi utama berupa perbaikan pemetaan topik dan penanganan redundansi teks. Untuk memperbaiki kerangka topik, diterapkan teknik pengelompokan dokumen seperti *Latent Dirichlet Allocation* (LDA) dan *Significance Sentence* (Widjanarko dkk., 2018), hingga kombinasi *K-Means* dan LDA yang mampu mencatatkan skor evaluasi ROUGE-1 sebesar 0,6199 (Twinandilla dkk., 2018). Sementara itu, untuk menanggulangi masalah tingginya redundansi secara spesifik, algoritma seleksi *Maximal Marginal Relevance* (MMR) banyak dieksplorasi. Integrasi algoritma MMR ini terbukti secara empiris sangat andal dalam menyeimbangkan relevansi dan diversitas kalimat, serta secara signifikan mampu mereduksi duplikasi informasi lintas dokumen dibandingkan metode yang murni berbasis probabilitas (Gunawan dkk., 2019; Mawaridi dkk., 2024).

Kinerja model pengelompokan dan seleksi tersebut masih bertumpu pada probabilitas statistik leksikal (pencocokan kata di permukaan teks) meskipun telah menunjukkan beberapa peningkatan tren akurasi. Mengatasi keterbatasan metode yang tidak mengenali makna semantik tersebut, perkembangan model *transformer* BERT hadir merevolusi

pemrosesan bahasa natural melalui representasi kontekstual yang mendalam (Devlin dkk., 2019). Berdasarkan bukti empiris dari literatur terdahulu, representasi berbasis BERT secara konsisten mengungguli metode statistik konvensional dalam menangkap struktur bahasa Indonesia yang kompleks (Subakti dkk., 2022). Mengingat arsitektur dasar BERT tidak efisien untuk komputasi perbandingan antar-kalimat, dikembangkanlah *Sentence*-BERT (SBERT) yang menggunakan jaringan kembar (*Siamese Network*). SBERT terbukti mampu menghasilkan *sentence embeddings* secara seragam, sehingga kalkulasi kemiripan makna kalimat untuk mendeteksi parafrasa dapat dihitung secara cepat menggunakan *Cosine Similarity* (Reimers & Gurevych, 2019). Keputusan untuk menjadikan SBERT sebagai solusi representasi semantik didukung kuat oleh penelitian terbaru yang memvalidasi bahwa integrasi *embeddings* SBERT pada kerangka peringkasan teks ekstraktif terbukti sangat tangguh dalam menghasilkan ringkasan yang koheren secara naratif dan optimal dalam mengontrol redundansi informasi (Annisa dkk., 2026).

Mekanisme evaluasi kuantitatif memiliki peran krusial dalam mengukur kualitas hasil peringkasan secara objektif guna mendampingi pengembangan arsitektur komputasi. Kajian literatur terbaru secara konsisten menggunakan metrik ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*) sebagai standar validasi utama. Skenario pengujian ini mencakup ROUGE-1 untuk mengukur perolehan informasi dasar tingkat kata (*unigram*), ROUGE-2 untuk kelancaran struktur frasa (*bigram*), dan ROUGE-L untuk mengevaluasi koherensi struktur kalimat secara utuh (*longest common subsequence*) (Hartawan dkk., 2024). Di samping pemilihan metrik, penentuan rasio kompresi turut menjadi variabel penentu dalam evaluasi model. Mesin peringkas perlu diuji melalui variasi persentase pemotongan untuk mengukur stabilitas kinerjanya. Penelitian terkait peringkasan ekstraktif berita berbahasa Indonesia membuktikan bahwa pengujian pada rentang kompresi 20% hingga 50% dari dokumen asli merupakan ambang batas yang paling komprehensif guna menemukan titik keseimbangan optimal antara retensi fakta esensial dengan tingkat keringkasan teks (Girsang & Amadeus, 2023).

Penelitian ini memposisikan metode usulan untuk mengisi celah riset tersebut menggunakan korpus berita Liputan6 melalui pengembangan alur komputasi yang komprehensif. Tahapan awal dilakukan dengan mengintegrasikan algoritma *Agglomerative Clustering* untuk memilah dan mengelompokkan dokumen berita heterogen ke dalam klaster topik peristiwa yang seragam (Koto dkk., 2020). Langkah berikutnya mengeksekusi

pemahaman semantik menggunakan *Sentence*-BERT (SBERT) dan menerapkan algoritma *Maximal Marginal Relevance* (MMR) sebagai mekanisme seleksi akhir. Peran dari algoritma ini secara matematis dirancang untuk mengekstraksi inti informasi yang paling relevan sekaligus mengeliminasi kemunculan kalimat redundan lintas dokumen melalui kalibrasi parameter penyeimbang λ (*lambda*) (Mawaridi dkk., 2024). Kualitas ringkasan yang dihasilkan kemudian dievaluasi secara komprehensif menggunakan metrik ROUGE pada variasi kompresi tersebut terhadap *human ground truth* sebagai standar rujukan. Untuk menjamin objektivitas standar rujukan ini, uji reliabilitas antar-anotator menggunakan *Cohen's Kappa* diimplementasikan guna mengukur tingkat kesepakatan kognitif peringkas manusia, sehingga data standar emas yang dihasilkan bersifat konsisten dan terbebas dari bias subjektivitas (Badshah & Sajjad, 2025; Röttger dkk., 2022). Secara praktis, metode usulan ini diharapkan bermanfaat bagi jurnalis, peneliti, serta masyarakat umum yang membutuhkan intisari informasi, sekaligus menjadi tolok ukur (*benchmark*) bagi pengembangan teknologi pemrosesan bahasa alami (NLP) di masa depan.

Urgensi penelitian ini secara keseluruhan dilatarbelakangi oleh tingginya kebutuhan akan solusi komputasional yang mampu menyajikan intisari fakta berita secara komprehensif, padat, dan minim redundansi. Sinergi antara pengelompokan dokumen (*Agglomerative Clustering*), ketajaman pemahaman semantik (*Sentence-BERT*), serta ketepatan seleksi kalimat (MMR) diproyeksikan mampu menjadi arsitektur cerdas yang tangguh dalam memitigasi fenomena *information overload*. Melalui rancangan evaluasi kuantitatif yang tervalidasi, penelitian ini dilaksanakan untuk membuktikan sejauh mana arsitektur NLP tersebut dapat memberikan efisiensi kognitif yang optimal bagi pembaca berita digital di Indonesia.

1.2. Rumusan Masalah

Penelitian ini merumuskan beberapa pokok permasalahan sebagai tindak lanjut dari uraian latar belakang tersebut:

1. Bagaimana merancang dan mengintegrasikan metode usulan yang terdiri dari *Agglomerative Clustering*, representasi semantik *Sentence-BERT*, serta algoritma seleksi *Maximal Marginal Relevance* (MMR) untuk melakukan tugas peringkasan multi-dokumen pada teks berita berbahasa Indonesia?

2. Bagaimana tingkat performa metode usulan tersebut saat dievaluasi menggunakan metrik ROUGE (ROUGE-1, ROUGE-2, dan ROUGE-L) terhadap *human ground truth* sebagai standar rujukan?

1.3. Tujuan

Tujuan dari penelitian ini adalah untuk membangun dan menguji kinerja mesin peringkasan multi-dokumen ekstraktif pada teks berita berbahasa Indonesia dengan mengintegrasikan algoritma *Agglomerative Clustering* sebagai mekanisme pengelompokan topik, *Sentence*-BERT untuk menghasilkan representasi semantik kalimat dalam bentuk *sentence embeddings*, serta algoritma *Maximal Marginal Relevance* (MMR). Selain itu, penelitian ini juga bertujuan untuk menganalisis pengaruh variasi tingkat kompresi dalam menentukan titik keseimbangan informasi yang optimal, serta mengevaluasi efektivitas algoritma tersebut secara komprehensif menggunakan metrik ROUGE dengan membandingkan luaran mesin terhadap *human ground truth* sebagai standar rujukan hasil ringkasan manusia.

1.4. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat yang langsung dirasakan oleh masyarakat luas, jurnalis, maupun peneliti di tengah tingginya arus informasi digital. Pengguna mendapatkan akses instan terhadap inti sebuah peristiwa secara utuh, tanpa harus menghabiskan waktu menyaring puluhan artikel yang repetitif. Pendekatan semantik ini memberikan terobosan dalam kajian *Natural Language Processing* (NLP) bahasa Indonesia dengan membuktikan bahwa model mampu mengeliminasi redundansi teks secara cerdas tanpa kehilangan konteks. Selain itu, karena metode yang digunakan murni mengambil kalimat-kalimat asli dari sumber berita (tidak mengarang kalimat baru), ringkasan yang disajikan dijamin akurat dan sesuai fakta. Hal ini sangat membantu pengguna terhindar dari informasi yang menyesatkan atau kalimat yang dikarang secara keliru oleh mesin pemroses.

1.5. Batasan Penelitian

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan untuk memperjelas ruang lingkup dan interpretasi hasil, antara lain:

1. Sumber data teks murni dibatasi pada artikel berita berbahasa Indonesia yang diambil dari portal berita Liputan6. Pengujian difokuskan pada enam kategori spesifik, yaitu

Kesehatan, Otomotif, Bisnis, Saham, Bola, dan Tekno. (berita pada tanggal 12 Januari 2026 hingga 18 Januari 2026).

2. Pengukuran kualitas hasil ringkasan dibatasi pada evaluasi komputasional secara kuantitatif menggunakan metrik ROUGE (ROUGE-1, ROUGE-2, dan ROUGE-L) yang berfokus pada seberapa besar irisan informasi antara keluaran mesin dengan *human ground truth*. Penelitian ini tidak mencakup pengujian kualitatif terkait koherensi paragraf, tata bahasa, maupun tingkat keterbacaan (*readability*) oleh pakar bahasa atau panelis manusia.

1.6. Sistematika Penulisan

Sistematika penulisan skripsi ini disusun untuk memberikan gambaran yang terstruktur mengenai isi penelitian. Skripsi terdiri dari lima bab utama, yang masing-masing membahas aspek berbeda dari penelitian ini.

BAB I PENDAHULUAN

Bab ini memberikan gambaran mengenai latar belakang permasalahan, rumusan masalah, tujuan, manfaat, dan batasan penelitian. Sistematika penulisan skripsi terkait pengembangan mesin peringkas multi-dokumen ekstraktif menggunakan metode *Sentence*-BERT dan *Maximal Marginal Relevance* (MMR)

BAB II TINJAUAN PUSTAKA

Bab ini memberikan kajian pustakan yang berhubungan dengan tema skripsi sebagai landasan untuk perumusan dan analisis permasalahan pada skripsi. Kajian pustaka yang digunakan meliputi, metode *Agglomerative Clustering*, teori pra-pemrosesan teks, uji reliabilitas *Cohen's Kappa, Sentence*-BERT, MMR, serta metrik evaluasi ROUGE.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan tahapan sistematis dalam pelaksanaan penelitian yang dimulai dari proses pengumpulan data mentah. Rangkaian pemrosesan selanjutnya meliputi pengelompokan topik berita, penyusunan teks rujukan (ground truth), pengujian *Cohen's Kappa*, pra-pemrosesan teks, ekstraksi representasi semantik SBERT, seleksi kalimat menggunakan algoritma MMR, hingga evaluasi kinerja ringkasan dengan metrik ROUGE.

BAB IV HASIL DAN ANALISA

Bab ini menguraikan lingkungan komputasi, hasil eksekusi skenario eksperimen, dan interpretasi kinerja mesin peringkas. Uraian pembahasan mencakup hasil validasi *Cohen's Kappa*, studi kasus peringkasan dokumen, analisis evaluasi skor ROUGE pada enam kategori berita yang berbeda, peninjauan kelemahan metode usulan, serta demonstrasi implementasi prototipe antarmuka program.

BAB V PENUTUP

Bab ini menjabarkan kesimpulan akhir dari seluruh rangkaian pengujian dan analisis yang telah dilakukan pada bab-bab sebelumnya. Saran pengembangan untuk penelitian lebih lanjut di masa mendatang juga dirumuskan pada bagian ini guna menyempurnakan kualitas peringkasan metode usulan.

BAB II
TINJAUAN PUSTAKA

Bab ini menguraikan kajian *state-of-the-art* serta landasan teoretis yang menjadi fondasi perancangan model peringkasan, mulai dari tahapan *preprocessing*, *Agglomerative Clustering*, uji *Cohen's Kappa*, penjelasan konsep *word* dan *sentence embedding*, arsitektur BERT dan *Sentence*-BERT (SBERT), algoritma seleksi *Maximal Marginal Relevance* (MMR), tingkat kompresi peringkasan, serta metrik evaluasi ROUGE.

2.1. *State-of-the-Art*

Bagian ini memaparkan rangkuman dari berbagai penelitian terdahulu yang relevan dengan topik peringkasan teks otomatis, baik dalam domain dokumen tunggal maupun multi-dokumen. Tinjauan ini mencakup evolusi metode dari pendekatan statistik konvensional hingga pemanfaatan model bahasa berbasis *Transformer* yang menjadi landasan utama dalam penelitian ini. Sebagai ringkasan, Tabel 2.1 di bawah ini menyajikan perbandingan mendalam antara penelitian-penelitian terdahulu yang menjadi rujukan dalam pengembangan mesin peringkasan ini.

Tabel 2.1 Daftar Penelitian Dengan Topik Peringkasan Multi-Dokumen

No.	Penelitian (Tahun)	Masalah/ Topik	Metode	Pendekatan	Temuan
1.	Widjanarko dkk. (2018)	Multi-Dokumen Berita Indonesia	LDA + Significance Sentence dengan hyperparameter alpha 0.01, beta 0.1	Ekstraktif	Membuktikan bahwa pemodelan topik (LDA) mampu mengenali dan menekan redundansi kalimat berita secara efektif, divalidasi dari tingginya skor *cosine similarity* rata-rata (0.931) pada tingkat kompresi 50%.
2.	Gunawan dkk. (2019)	Multi-Dokumen Berita Indonesia	TextRank + MMR	Ekstraktif	Menemukan bahwa integrasi metode seleksi MMR berhasil meningkatkan koherensi ringkasan dan secara signifikan mengurangi duplikasi informasi dibandingkan metode berbasis probabilitas murni (mencapai ROUGE-1: 0.45).
3.	Twinandilla dkk. (2018)	Multi-Dokumen Berita Indonesia	K-Means + LDA	Ekstraktif	Mencapai skor ROUGE-1 sebesar 0.6199 melalui pengelompokan topik terlebih dahulu.

No.	Penelitian (Tahun)	Masalah/ Topik	Metode	Pendekatan	Temuan
4.	Koto dkk. (2020)	Multi-Dokumen Berita Indonesia	IndoBERT Pretrained Model	Ekstraktif & Abstraktif	Menciptakan *benchmark* dataset terbesar; model ekstraktif meraih ROUGE-1 sebesar 41,08.
5.	Mawaridi dkk. (2024)	Multi-Dokumen Berita Indonesia	LDA + MMR	Ekstraktif	Nilai parameter $\lambda = 0.7$ pada MMR menghasilkan skor ROUGE terbaik pada 2.500 artikel berita.
6.	Subakti dkk. (2022)	*Text Clustering* Indonesia	BERT vs TF-IDF	Ekstraktif	Membuktikan arsitektur *Transformer* (BERT) secara konsisten mengungguli metode pembobotan tradisional (TF-IDF) dalam memahami konteks kalimat berbahasa Indonesia.
7.	Hanley dkk. (2025)	Pengelompokan Topik Multi-Dokumen Berita	Hierarchical Clustering	Pra-pemrosesan Ekstraktif	Memvalidasi bahwa pengelompokan hierarkis (*agglomerative*) sangat tangguh untuk mengisolasi sub-topik berita heterogen tanpa perlu menentukan jumlah klaster mutlak di awal pemrosesan.
8.	Annisa dkk. (2026)	Peringkasan Teks Semantik & Kontrol Redundansi	Sentence-BERT (SBERT) + Algoritma Seleksi	Ekstraktif	Membuktikan bahwa integrasi *embeddings* SBERT pada kerangka ekstraktif sangat tangguh dalam menghasilkan ringkasan koheren dan optimal mengontrol redundansi informasi.
9.	Girsang & Amadeus (2023)	Evaluasi Kualitas Kompresi Berita Indonesia	Ant System Algorithm / Evaluasi ROUGE	Ekstraktif	Membuktikan bahwa pengujian pada rentang kompresi 20% hingga 50% dari dokumen asli merupakan ambang batas evaluasi paling komprehensif untuk menjaga keseimbangan retensi fakta esensial.

Tabel 2.1 menyajikan rentetan fase perkembangan metode pada domain peringkasan teks bahasa Indonesia. Penelitian fundamental dalam ranah ini diawali oleh Koto dkk. (2020) yang secara resmi memperkenalkan korpus Liputan6 sebagai *benchmark dataset* berskala besar. Melalui pengujian komprehensifnya, mereka membuktikan bahwa pendekatan ekstraktif masih sangat relevan untuk menjaga akurasi faktual berita, dengan capaian *baseline* ROUGE-1 sebesar 41,08.

Pemodelan ekstraktif awal umumnya mengandalkan fitur statistik probabilistik untuk menyaring informasi. Sebagai contoh, Widjanarko dkk. (2018) dan Twinandilla dkk. (2018)

memanfaatkan metode *Latent Dirichlet Allocation* (LDA) serta *K-Means clustering* untuk memetakan topik dokumen. Walaupun skor evaluasinya menunjukkan tren peningkatan, metode yang murni bergantung pada frekuensi kata kunci ini kesulitan menangani masalah tumpang tindih informasi. Guna mengatasi masalah tersebut, algoritma *Maximal Marginal Relevance* (MMR) kemudian banyak diterapkan. Gunawan dkk. (2019) dan Mawaridi dkk. (2024) memvalidasi bahwa integrasi MMR (terutama dengan penyesuaian parameter usulan $\lambda = 0{,}7$) sangat efektif menekan duplikasi informasi lintas dokumen. Gunawan dkk., 2019 memodifikasi model *TextRank* dengan menambahkan MMR untuk reduksi redundansi, menggunakan pembobotan TF-IDF sebagai fitur ekstraksi awal. Evaluasi empiris pada tingkat kompresi 20% membuktikan bahwa MMR efektif meningkatkan koherensi ringkasan dan menekan duplikasi dengan capaian rata-rata *F-measure* ROUGE-1 sebesar 0,45. Dalam studi yang lebih baru, Mawaridi dkk. (2024) memadukan pemodelan topik LDA dengan algoritma MMR pada 2.500 artikel berita Indonesia. Penelitian tersebut memvalidasi bahwa penggunaan parameter penyeimbang berupa λ (*lambda*) sebesar 0,7 memberikan titik paling optimal antara relevansi dan keberagaman informasi, menghasilkan skor ROUGE-1 sebesar 0,488 pada kompresi 50%. Meskipun algoritma MMR terbukti tangguh dalam seleksi kalimat, representasi fitur yang digunakan pada kedua penelitian tersebut masih terbatas pada pencocokan kata di permukaan teks (*surface-level matching*).

Kehadiran arsitektur *Transformer* menawarkan solusi pemahaman makna teks secara kontekstual untuk mengatasi kelemahan metode konvensional tersebut. Terkait pemrosesan bahasa Indonesia, Subakti dkk. (2022) memberikan bukti empiris bahwa ekstraksi fitur berbasis representasi BERT secara konsisten mengungguli metode pembobotan tradisional (TF-IDF). Karena arsitektur dasar BERT memiliki beban komputasi yang terlalu berat jika digunakan membandingkan kalimat satu per satu, varian *Sentence*-BERT (SBERT) kemudian dikembangkan untuk memetakan kalimat ke dalam ruang vektor padat secara efisien. Ketangguhan metode usulan ini juga baru saja dikonfirmasi oleh penelitian Annisa dkk. (2026), yang membuktikan bahwa *embeddings* SBERT sangat tangguh dalam menghasilkan ringkasan yang koheren sekaligus mengontrol redundansi pada skenario ekstraktif.

Penanganan dokumen berita digital memerlukan tahapan pengelompokan awal sebelum proses pemahaman semantik dieksekusi. Kajian dari Hanley dkk. (2025) memvalidasi bahwa metode klastering hierarkis (*Agglomerative Clustering*) sangat andal

untuk mengisolasi sub-topik berita yang heterogen tanpa perlu repot menebak jumlah klaster secara mutlak di awal pemrosesan. Di sisi lain, dari segi pengujian kualitas, model komputasi juga perlu dievaluasi ketahanannya secara objektif. Girsang & Amadeus (2023) membuktikan bahwa evaluasi ROUGE pada rentang kompresi 20% hingga 50% dari dokumen asli merupakan ambang batas paling ideal untuk mengukur keseimbangan antara retensi fakta esensial dengan tingkat keringkasan teks.

Tinjauan literatur tersebut memperlihatkan celah penelitian (*research gap*) pada ranah peringkasan teks multi-dokumen. Mayoritas penelitian sebelumnya masih memisahkan atau mengkaji keunggulan pemahaman semantik SBERT, ketangguhan algoritma seleksi MMR, dan metode pengelompokan awal secara terpisah. Belum ada satupun kajian yang memadukan elemen-elemen tersebut secara komprehensif dalam satu alur komputasi simultan untuk korpus berita berbahasa Indonesia. Oleh karena itu, penelitian ini hadir memposisikan diri untuk mengisi celah tersebut. Metode usulan pada penelitian ini akan mengintegrasikan *Agglomerative Clustering* untuk pengelompokan topik teks di awal, pengekstraksian makna kalimat menggunakan SBERT untuk mengatasi masalah pendeteksian parafrasa, serta penyaringan kalimat akhir menggunakan algoritma diversitas MMR yang akan diuji secara ketat kemampuannya pada variasi kompresi 20% hingga 50%.

2.2. *Pre-processing*

Text Preprocessing adalah tahapan krusial dalam pemrosesan bahasa alami (NLP) yang bertujuan untuk membersihkan, menstandarisasi, dan mereduksi dimensi data teks mentah sebelum diolah oleh algoritma ekstraksi fitur. Pada konteks peringkasan teks dokumen berita yang bersumber dari portal daring, teks sering kali memuat struktur ireguler yang dapat mendistorsi perhitungan bobot semantik kalimat (Widyassari dkk., 2022).

2.2.1. *Noise Cleaning*

Noise cleaning merupakan proses pembersihan deretan teks dari karakter-karakter yang tidak merepresentasikan makna semantik secara komputasional. Pada dataset berita daring seperti Liputan6 Koto dkk. (2020), *noise* ini umumnya berbentuk tautan (URL), tag pemformatan HTML, tagar (*hashtag*), nomor telepon, maupun karakter simbolik khusus. Penghapusan elemen ini menjamin bahwa matriks fitur yang terbentuk murni berisikan entitas kebahasaan.

2.2.2. Tokenisasi (*Tokenization*)

Segmentasi urutan teks menjadi entitas diskrit yang lebih kecil (*token*) dieksekusi melalui tahapan tokenisasi (Halimah dkk., 2022). Algoritma tokenisasi dalam pemrosesan teks tingkat dokumen beroperasi pada dua level utama, yakni pemecahan paragraf menjadi deretan kalimat (*sentence splitting*) dan pemecahan kalimat menjadi kata tunggal (*word tokenization*). Pemecahan paragraf bertujuan untuk mempersiapkan seluruh kandidat kalimat yang akan diekstraksi oleh mesin. Pemecahan kalimat menjadi kata tunggal berfungsi untuk memfasilitasi komputasi frekuensi kemunculan istilah secara akurat.

2.2.3. *Case Folding*

Tahapan penyeragaman seluruh karakter alfabet dalam dokumen teks menjadi huruf kecil (*lowercase*) dilakukan melalui proses *case folding* (Mawaridi dkk., 2024). *Case folding* adalah tahapan penyeragaman seluruh karakter alfabet dalam dokumen menjadi huruf kecil (*lowercase*). Proses ini mengeliminasi ambiguitas mesin dalam mengidentifikasi kata yang sama namun memiliki kapitalisasi berbeda (misalnya kata "Pelatih" di awal kalimat dan "pelatih" di tengah kalimat), sehingga konsistensi komputasi jarak vektor tetap terjaga.

2.3. Metode *Agglomerative Clustering*

Metode *Agglomerative Clustering* merupakan salah satu *clustering* hierarki dengan pendekatan *bottom-up* yang mengelompokkan dokumen berdasarkan tingkat kemiripannya (Wicaksana dkk., 2018). Pada proses peringkasan multi-dokumen ekstraktif, pengelompokan topik teks sangat penting untuk memastikan algoritma tidak mencampuradukkan informasi dari konteks berita yang berbeda sebelum diringkas (Subakti dkk., 2022). Dengan mengonversi judul ke dalam pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF), model representasi dapat memetakan pentingnya sebuah kata dalam dokumen terhadap seluruh korpus.

Penggabungan fitur ini dengan algoritma *Agglomerative Clustering* memungkinkan dokumen dengan kemiripan semantik tinggi mengelompok secara otomatis hingga mencapai ambang batas (*threshold*) tertentu, yang memastikan setiap klaster hanya memuat peristiwa yang seragam (Dhini dkk., 2023). Penelitian dengan metode *Agglomerative Clustering* ini, menggunakan *library* dari *sklearn.cluster.AgglomerativeClustering*. Tahapan metode *Agglomerative Clustering* yaitu:

1. Ekstraksi fitur judul dokumen menjadi representasi vektor numerik menggunakan metode TF-IDF. Persamaan 2.1 digunakan pada penelitian ini untuk menghitung bobot kata:

$$W_{t,d} = tf_{t,d} \times \log\left(\frac{N}{df_t}\right) \quad (2.1)$$

Keterangan:

T : Total keseluruhan kata unik (*vocabulary* atau leksikon) yang diekstrak dari seluruh dokumen.

$\mathrm{W_{t,d}}$: Bobot akhir kata ke$-t$ $(t = 1,2,\dots,T)$ pada dokumen ke- d $(d = 1,2,\dots,N)$

$\mathrm{tf_{t,d}}$: Jumlah kemunculan kata (frekuensi) ke$-t$ $(t = 1,2,\dots,T)$ pada dokumen ke$-d$ $(d = 1,2,\dots,N)$.

N : Banyaknya keseluruhan dokumen

df_t : Jumlah dokumen yang mengandung kata ke-t $(t = 1,2,\dots,T)$

2. Jarak kemiripan topik antar-dokumen diukur menggunakan *Cosine Similarity*. Metode ini mengevaluasi sudut kosinus dari dua vektor tanpa memedulikan magnitudonya (Han dkk., 2011), menghasilkan koefisien berentang 0 hingga 1 di mana angka 1 merepresentasikan kemiripan semantik mutlak (Lahitani dkk., 2016). Formulasi perhitungan *Cosine Similarity* antara vektor dokumen A dan B dijabarkan pada Persamaan 2.2

$$\mathrm{Sim}(A,B) = \frac{\sum_{i=1}^{V} A_i B_i}{\sqrt{\sum_{i=1}^{V} A_i^2}\sqrt{\sum_{i=1}^{V} B_i^2}} \quad (2.2)$$

Keterangan:

$\mathrm{Sim(A,B)}$: Nilai kemiripan *Cosine* antara dokumen A dan B

A_i : Bobot TF-IDF representasi kata ke-i $(i = 1,2,\dots,V)$ pada dokumen A.

B_i : Bobot TF-IDF representasi kata ke-i $(i = 1,2,\dots,V)$ pada dokumen B.

V : Total keseluruhan kata unik (vocabulary atau leksikon) yang menyusun ruang dimensi vektor pada klaster tersebut.

Persamaan 2.3 digunakan untuk mengonversi nilai kemiripan tersebut menjadi jarak aktual (*Cosine Distance*):

$$D_{A,B} = 1 - \mathrm{Sim}(A,B) \quad (2.3)$$

Keterangan :

$D_{A,B}$: *Cosine Distance* antara dokumen A dengan dokumen B

3. Inisialisasi setiap dokumen judul berita sebagai satu *cluster* tunggal. Jika terdapat M dokumen, maka pada awalnya terdapat M buah *cluster* yang berdiri sendiri.
4. Cari sepasang *cluster* yang memiliki jarak terdekat berdasarkan matriks jarak (nilai $D_{A,B}$ paling minimum). Gabungkan kedua *cluster* terdekat tersebut menjadi satu *cluster* baru.
5. Setelah *cluster* digabungkan, hitung ulang jarak antar *cluster* gabungan tersebut dengan *cluster* lainnya menggunakan kriteria *Average Linkage* (rata-rata jarak seluruh anggota antar klaster). Jika jarak terdekat antar *cluster* yang tersisa masih berada di bawah atau sama dengan nilai *Distance Threshold* yang ditentukan, maka kembali ke langkah 4. Jika jarak terdekat sudah melebihi nilai *Distance Threshold*, maka proses akan dihentikan.

2.4. Uji *Cohen's Kappa*

Objektivitas data rujukan (*Ground Truth*) memegang peranan krusial pada pengembangan mesin peringkasan teks. Metrik statistik *Cohen's Kappa* (κ) digunakan untuk mengukur tingkat kesepakatan antara dua anotator (peringkas manusia) terhadap suatu subjek tertentu dengan memperhitungkan faktor peluang atau kebetulan (McHugh, 2012). Penggunaan metrik ini dalam pemilihan *ground truth* peringkasan teks bertujuan untuk memastikan bahwa informasi yang dianggap penting oleh manusia bersifat konsisten, memiliki reliabilitas yang tinggi, dan tidak bias secara personal. Hal ini merupakan standar baku dalam penyusunan dataset evaluasi pada tugas *Natural Language Processing* (NLP) (Röttger dkk., 2022).

Dua komponen utama harus dihitung terlebih dahulu guna menentukan nilai reliabilitas tersebut, yaitu proporsi kesepakatan aktual (p_o) dan proporsi kesepakatan peluang (p_e) Persamaan 2.4 dan 2.5 mendefinisikan formulasi matematis untuk kedua komponen tersebut secara berurutan:

$$p_o = \frac{\sum f_{ii}}{N} \quad \text{.....(2.4)}$$

Keterangan:

p_o : Proporsi kesepakatan aktual (*observed agreement*) antara dua anotator.

f_{ii} : Total frekuensi kesepakatan aktual di mana kedua anotator memilih kategori yang sama pada indeks kategori ke-i ($i = 1, 2, \dots, n$).

n : Jumlah kategori klasifikasi (pada penelitian ini $n = 2$, mencakup label "pilih" dan "buang").

N : Total keseluruhan unit informasi yang dievaluasi.

$$p_e = \sum_{k=1}^{n} \left(\frac{R_k}{N} \times \frac{C_k}{N} \right) \quad \text{.....(2.5)}$$

Keterangan:

p_e : Proporsi kesepakatan peluang (*expected agreement*).

R_k : Jumlah total unit observasi yang diklasifikasikan ke kategori ke-k ($k = 1,2, \dots, n$) oleh anotator pertama (merepresentasikan total marginal baris).

C_k : Jumlah total unit observasi yang diklasifikasikan ke kategori ke-k($k = 1,2, \dots, n$) oleh anotator kedua (merepresentasikan total marginal kolom).

n : Jumlah kategori klasifikasi (pada penelitian ini $n = 2$, mencakup label "pilih" dan "buang").

N : Total keseluruhan unit informasi yang dievaluasi.

Persamaan matematis dari *Cohen's Kappa* didefinisikan pada Persamaan 2.6 berikut:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad \text{.....(2.6)}$$

Keterangan:

p_o : Proporsi kesepakatan aktual yang diamati antar anotator.

p_e : Proporsi kesepakatan yang diharapkan terjadi secara kebetulan.

Interpretasi nilai κ merujuk pada kriteria baku dari Landis & Koch (1977), di mana nilai $> 0{,}60$ menunjukkan kesepakatan yang kuat (*substantial agreement*) dan $> 0{,}80$ menunjukkan kesepakatan yang hampir sempurna (*almost perfect agreement*). Standar ini secara luas terus diadaptasi dalam riset evaluasi metrik *machine learning* dan NLP modern untuk memastikan konsistensi pemahaman teks pada anotasi evaluasi (Badshah & Sajjad, 2025).

Penelitian ini menyertakan perhitungan akurasi mentah atau *Percentage Agreement* guna menanggulangi kelemahan rumus *Cohen's Kappa* yang sangat sensitif terhadap ketidakseimbangan kelas observasi, sebuah fenomena yang disebut *Kappa Paradox* (Feinstein & Cicchetti, 1990). Perhitungan metrik tersebut bertujuan untuk mengukur proporsi kesepakatan faktual antara dua pihak peringkas tanpa mempertimbangkan reduksi dari faktor peluang kebetulan. Persamaan 2.7 menjabarkan formulasi matematis yang digunakan untuk menghitung persentase kesepakatan aktual tersebut:

$$PA = \frac{A_m + A_b}{N} \times 100 \quad (2.7)$$

Keterangan:

PA : *Percentage Agreement* (Persentase kesepakatan murni antar-anotator).

A_m : Jumlah unit observasi (kalimat) yang disepakati secara bersama-sama oleh anotator untuk dipilih masuk ke dalam ringkasan.

A_b : Jumlah unit observasi (kalimat) yang disepakati secara bersama-sama oleh anotator untuk dibuang (tidak relevan).

N : Total keseluruhan unit observasi (kalimat) yang dievaluasi pada pengujian tersebut.

2.5. Algoritma *Fuzzy Matching*

Algoritma *Fuzzy Matching* atau pencocokan *string* samar adalah teknik komputasional yang digunakan untuk mengidentifikasi tingkat kemiripan antara dua buah teks meskipun tidak identik seratus persen. Dalam pemrosesan bahasa alami (NLP), teknik ini sangat krusial untuk mengatasi variasi penulisan, sinonim, atau parafrasa ringan pada arsitektur ekstraksi teks (Chandrasekaran & Mago, 2021). Salah satu pendekatan *fuzzy matching* yang paling umum pada tingkat kalimat adalah dengan menghitung rasio irisan kata (*lexical overlap*). Metode ini mengevaluasi seberapa banyak kata dari kalimat sumber yang berhasil dipertahankan di dalam kalimat target. Pendekatan ini sering diimplementasikan sebagai dasar evaluasi otomatis untuk metrik persetujuan (*agreement*) karena kemampuannya yang tangguh dalam mentolerir perubahan struktur frasa tanpa menghilangkan makna inti informasi (Fabbri dkk., 2021).

Algoritma *Fuzzy Matching* pada penelitian ini mengeksekusi pencocokan berdasarkan rasio irisan leksikal (*lexical overlap*) antara kalimat pada data induk dengan kalimat hasil

ringkasan anotator. Perhitungan tingkat kemiripan ini diformulasikan secara matematis melalui Persamaan 2.8 berikut:

$$Kemiripan(S_{asli}, S_{ringkasan}) = \frac{|S_{asli} \cap S_{ringkasan}|}{|S_{asli}|} \quad \text{.....(2.8)}$$

Keterangan:

S_{asli} : merepresentasikan himpunan kata penyusun kalimat asli pada data induk.

$S_{ringkasan}$: merepresentasikan himpunan kata penyusun kalimat pada ringkasan anotator.

$|S_{asli} \cap S_{ringkasan}|$: adalah jumlah kata yang beririsan (muncul di kedua kalimat).

$|S_{asli}|$: adalah total keseluruhan kata pada kalimat asli.

Sistem pemetaan dikonfigurasi menggunakan ambang batas kemiripan (*threshold*) sebesar 0,5 atau 50%. Apabila nilai $Kemiripan(S_{asli}, S_{ringkasan}) \geq 0{,}5$, maka algoritma secara otomatis memberikan skor biner (1) yang berarti anotator "sepakat" memilih kalimat tersebut. Sebaliknya, jika nilai kemiripan bernilai $< 0{,}5$, maka akan diberikan skor (0).

2.6. *Word* dan *Sentence Embedding*

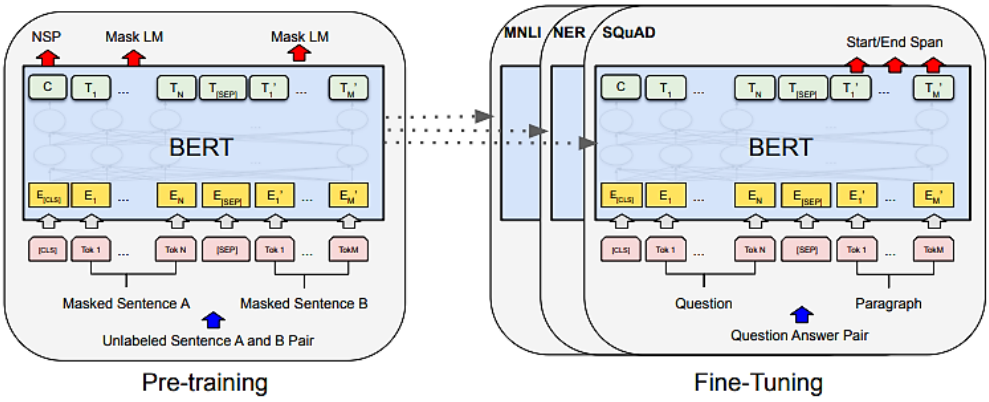
Embedding merupakan paradigma modern dalam *Natural Language Processing* (NLP) yang merepresentasikan teks ke dalam bentuk vektor padat (*dense vector*) berdimensi tetap dan bernilai riil. Berbeda dengan pendekatan leksikal tradisional seperti TF-IDF yang menghasilkan matriks *sparse* (dipenuhi nilai nol) serta mengabaikan urutan kata, *embedding* mampu menangkap kedekatan makna semantik berdasarkan asumsi distribusi linguistik (Minaee dkk., 2021). Representasi kata tingkat awal, seperti *Word2Vec* atau *GloVe*, memetakan satu kata ke dalam satu vektor statis tanpa memedulikan perbedaan makna ganda (polisemi) pada konteks kalimat yang berbeda (Jurafsky & Martin, 2024).

Perkembangan arsitektur pemrosesan bahasa selanjutnya melahirkan representasi berbasis konteks (*context-aware embeddings*) yang mampu mengalokasikan nilai vektor secara dinamis untuk kata yang sama bergantung pada susunan sintaksis yang mengelilinginya (Kalyan dkk., 2021). *Sentence Embedding* secara lebih spesifik menyandikan komposisi makna dari barisan kata kontekstual di dalam satu kalimat utuh menjadi sebuah representasi vektor tunggal. Mekanisme ini memastikan bahwa kalimat-

kalimat yang membawa pesan serupa, meskipun menggunakan variasi kosakata yang sama sekali berbeda (parafrasa), akan menempati titik koordinat geometris yang berdekatan di dalam ruang vektor (Reimers & Gurevych, 2019). Pemetaan semantik pada tingkat kalimat ini menjadi instrumen krusial bagi mesin komputasi untuk mendeteksi kesamaan topik dan mengeliminasi redundansi informasi pada tugas peringkasan multi-dokumen (Wang dkk., 2020).

2.7. BERT (*Bidirectional Encoder Representations from Transformer*)

BERT merupakan model bahasa mutakhir berbasis arsitektur *Transformer* yang dikembangkan oleh *Google* untuk mengatasi keterbatasan model bahasa sekuensial terdahulu. Inovasi utama dari BERT adalah proses *pre-training* yang sepenuhnya dua arah (*bidirectional*) , yang memungkinkannya untuk merepresentasikan makna kata berdasarkan konteks kalimat secara utuh dari sisi kiri dan kanan secara bersamaan (Devlin dkk., 2019). Dalam domain pemrosesan teks, representasi data menggunakan BERT telah terbukti secara empiris mampu meningkatkan akurasi *clustering* karena kemampuannya dalam memahami semantik teks yang lebih dalam dibandingkan metode representasi statistik tradisional (Subakti dkk., 2022).



Gambar 2.1 Prosedur *Pre-Training* dan *Fine-Tuning* BERT (Devlin dkk., 2019)

Mekanisme aliran informasi pada arsitektur model bahasa BERT diilustrasikan Devlin dkk. (2019) seperti pada Gambar 2.1. Karakteristik paling menonjol pada diagram tersebut ditunjukkan oleh arah tanda panah yang saling menyilang dan terhubung ke segala arah pada setiap lapisannya (*bidirectional*). Visualisasi ini merepresentasikan bahwa dalam memproses sebuah teks, BERT tidak membaca urutan kata secara searah (misalnya hanya dari kiri ke kanan), melainkan memproses setiap kata dengan melihat keseluruhan konteks kata di sisi

kiri dan kanannya secara serentak. Mekanisme interaksi antar-kata yang menyeluruh inilah yang memungkinkan model untuk menangkap makna teks secara utuh dan mendalam.

2.7.1. Konstruksi Vektor BERT

Model BERT mengonstruksi representasi awal sebuah teks melalui penjumlahan tiga lapis vektor *embedding* diskrit secara elemen demi elemen (*element-wise addition*). Proses penyusunan ini bertujuan untuk membekali model dengan informasi leksikal, posisi sintaksis, dan identitas segmen kalimat sebelum barisan token memasuki lapisan *Transformer* (Devlin dkk., 2019). Formulasi matematis untuk mengalkulasi vektor masukan awal E dijabarkan pada Persamaan 2.9:

$$E = E_{\text{token}} + E_{\text{position}} + E_{\text{segment}} \quad \text{.....(2.9)}$$

Keterangan:

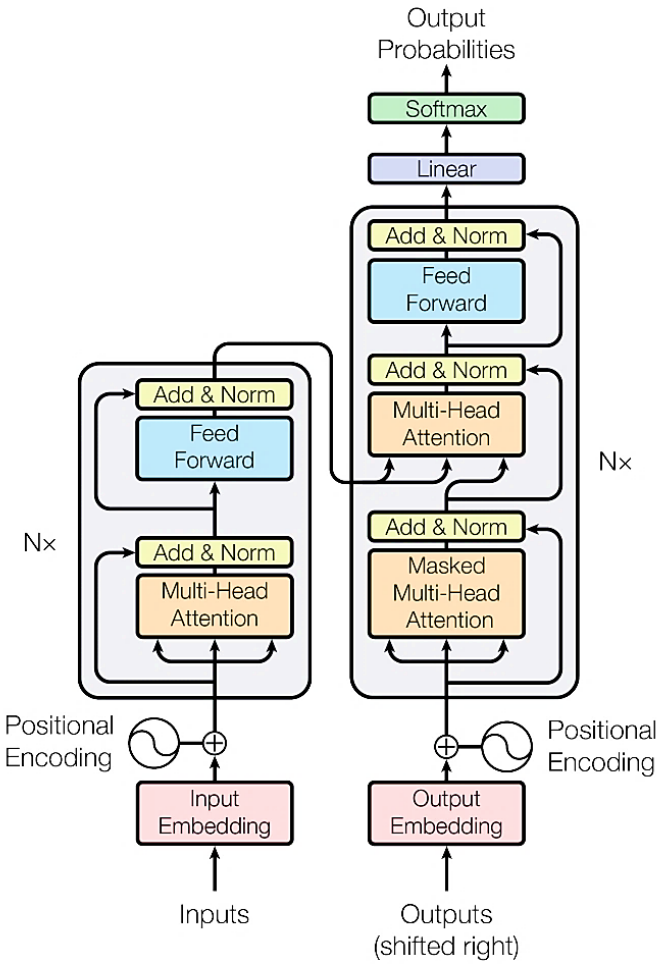
E : Vektor representasi masukan gabungan yang siap dikirim ke lapisan *Transformer*

E_{token} : Vektor representasi numerik dari unit kata atau sub-kata (*subword tokens*) hasil pemrosesan algoritma *WordPiece*.

E_{position} : Vektor penanda posisi urutan token di dalam kalimat untuk memitigasi hilangnya informasi sekuensial akibat hilangnya arsitektur rekursif.

E_{segment} : Vektor pembeda identitas kalimat untuk memisahkan pasangan baris teks masukan pada tugas perbandingan kalimat tunggal.

Vektor representasi masukan gabungan tersebut selanjutnya diteruskan sebagai data umpan ke dalam arsitektur utama *Transformer*. Vaswani dkk. (2017) mengilustrasikan cetak biru arsitektur *Transformer* secara utuh pada Gambar 2.2. Model BERT secara spesifik hanya mengadopsi blok *Encoder* (ditunjukkan pada tumpukan lapisan sebelah kiri gambar) guna mengekstraksi pemahaman dua arah (*bidirectional*), tanpa melibatkan blok *Decoder* di sebelah kanan yang umumnya ditujukan untuk tugas generatif.



Gambar 2.2 Arsitektur Dasar *Transformer* (Vaswani dkk., 2017)

2.7.2. Mekanisme *Self-Attention* dan *Multi-Head Attention*

Mekanisme *Self-Attention* merupakan fondasi utama arsitektur *Transformer* yang bertugas menghitung bobot keterkaitan antar-token kata di dalam satu urutan kalimat secara simultan. Lapisan komputasi ini mentransformasikan token masukan menjadi tiga matriks proyeksi linear yang berbeda, yaitu *Queries* (Q), *Keys* (K), dan *Values* (V) (Vaswani dkk., 2017). Proses evaluasi kedekatan makna semantik dilakukan melalui operasi perkalian titik (*scaled dot-product*) antara matriks Q dan K, yang kemudian dinormalisasi menggunakan fungsi *softmax* sebelum dikalikan dengan matriks V (Devlin dkk., 2019). Formulasi matematis untuk operasi *Self-Attention* tunggal ini dijabarkan pada Persamaan 2.10:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad \text{.....(2.10)}$$

Keterangan:

Q : Matriks *Queries* (representasi token yang mencari informasi).

K : Matriks *Keys* (representasi token yang mengandung indeks informasi)

V : Matriks *Values* (nilai konten semantik dari token).

d_k : Dimensi tersembunyi (*hidden dimension*) dari *attention head*.

Arsitektur *Transformer* selanjutnya memperluas kapasitas pemetaan kontekstual ini melalui penerapan mekanisme *Multi-Head Attention*. Fitur ini membagi proyeksi linier matriks *Queries* (Q), *Keys* (K), dan *Values* (V) ke dalam h buah sub-ruang koordinat geometris (*attention heads*) yang diproses secara paralel (Vaswani dkk., 2017). Fungsi kalkulasi agregat untuk menggabungkan luaran seluruh *attention heads* dirumuskan pada Persamaan 2.11 dan Persamaan 2.12:

$$\text{MultiHead}(Q,K,V) = Concat(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W^O \quad \text{.....(2.11)}$$

$$\text{head}_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) = softmax\left(\frac{\left(QW_i^Q\right)\left(KW_i^K\right)^T}{\sqrt{d_k}}\right)VW_i^V \quad \text{.....(2.12)}$$

Keterangan:

head_i : Hasil luaran matriks kontekstual dari operasi *Self-Attention* spesifik pada tahapan iterasi indeks ke-i

W_i^Q, W_i^K, W_i^V : Matriks proyeksi parameter bobot latih khusus untuk setiap *attention head* ke-i . ($i = 1, 2, \dots, h$)

W^O : Matriks proyeksi linier akhir untuk mengembalikan dimensi vektor gabungan ke ukuran semula.

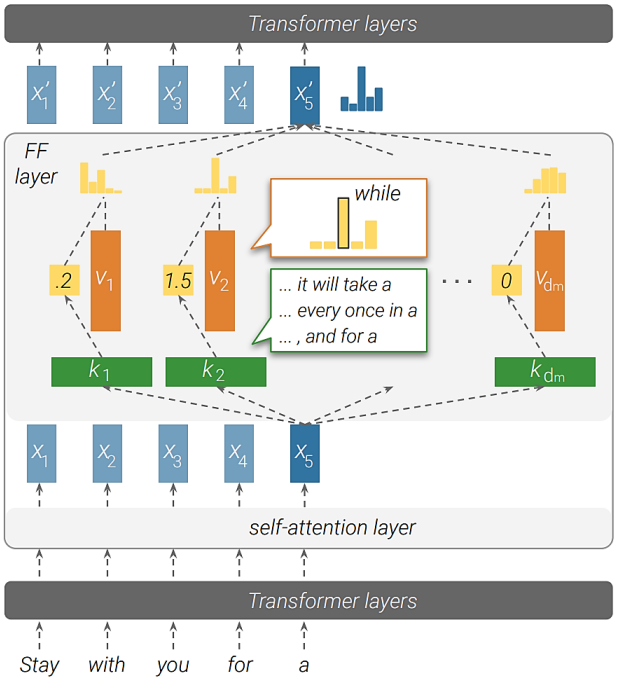
Concat : Operasi penggabungan baris matriks luaran dari seluruh attention heads.

d_k : Nilai penskalaan dimensi dari komponen *attention head* guna mencegah lonjakan nilai *dot-product* yang ekstrem.

H : Jumlah total *attention heads* atau sub-ruang proyeksi paralel yang dijalankan di dalam model (berjumlah 12 heads pada model MiniLM-L12).

2.7.3. *Position-Wise Feed-Forward Network* (FFN)

Setiap blok *Transformer* pada arsitektur BERT dilengkapi dengan jaringan saraf tiruan *Position-Wise Feed-Forward Network* (FFN) yang beroperasi tepat setelah mekanisme *Multi-Head Attention*. Lapisan komputasi ini bertindak sebagai memori penyimpan (*key-value memories*) yang memproses representasi kontekstual teks dan memetakannya ke dalam tingkat abstraksi semantik yang lebih tinggi (Geva dkk., 2021). Mekanisme operasional lapisan FFN yang bertindak selayaknya memori *Key-Value* diilustrasikan secara visual pada Gambar 2.3.



Gambar 2.3 Ilustrasi Lapisan *Feed-Forward* sebagai Memori *Key-Value* (Geva dkk., 2021)

Berdasarkan ilustrasi tersebut, matriks parameter bobot pertama (W_1) bertindak sebagai kunci (*keys*) yang mendeteksi pola linguistik tertentu dari lapisan *attention*, sedangkan matriks parameter bobot kedua (W_2) bertindak sebagai nilai (*values*) yang merespons dan mendistribusikan pola tersebut ke probabilitas kosakata berikutnya (Geva dkk., 2021).

Model BERT dan variannya memodifikasi arsitektur FFN standar dengan mengganti fungsi aktivasi ReLU (*Rectified Linear Unit*) menjadi GELU (*Gaussian Error Linear Unit*). Fungsi GELU mengombinasikan properti non-linier dengan pembobotan stokastik yang membuat proses optimasi gradien pada jaringan saraf dalam (*deep neural network*) berjalan jauh lebih stabil dan mulus (Hendrycks & Gimpel, 2023). Formulasi matematis untuk

operasi linier dan aktivasi non-linier pada FFN dijabarkan melalui Persamaan 2.13 berikut (Devlin dkk., 2019):

$$\mathrm{FFN}(x) = \mathrm{GELU}(xW_1 + b_1)W_2 + b_2 \quad \text{.....(2.13)}$$

Keterangan:

x : Matriks representasi vektor hasil luaran dari lapisan *Multi-Head Attention*.

W_1, W_2 : Matriks parameter bobot linear (weights) yang dipelajari mesin pada lapisan tersembunyi pertama dan kedua.

b_1, b_2 : Vektor bias komputasi untuk menyesuaikan pergeseran distribusi nilai aktivasi.

GELU : Fungsi aktivasi non-linier *Gaussian Error Linear Unit* yang mentransformasikan matriks sebelum diproyeksikan kembali ke dimensi asalnya.

2.8. *Sentence*-BERT

Lapisan *Transformer* paling atas pada arsitektur BERT menghasilkan luaran berupa kumpulan vektor *hidden state* kontekstual tingkat token $(h_1, h_2, \dots, h_N)$ secara spesifik. Struktur output bawaan BERT ini memunculkan kendala komputasi yang masif apabila diaplikasikan langsung untuk mengukur kemiripan pasangan kalimat di dalam korpus besar. Evaluasi kedekatan makna semantik menggunakan arsitektur dasar BERT mengharuskan kedua kalimat digabungkan sebagai masukan tunggal untuk melewati mekanisme *cross-attention* lintas lapisan secara berulang (Reimers & Gurevych, 2019).

Kerangka kerja *Sentence-BERT* (SBERT) hadir sebagai solusi optimasi untuk menanggulangi hambatan komputasi kombinatorial tersebut. Arsitektur jaringan kembar (*Siamese Network*) diterapkan pada model ini untuk memodifikasi struktur pemrosesan BERT konvensional (Reimers & Gurevych, 2019). SBERT memproses setiap baris kalimat secara independen guna menghasilkan representasi *sentence embeddings* berdimensi tetap (*fixed-size*). Langkah pemrosesan mandiri ini memungkinkan perhitungan matriks kemiripan lintas dokumen diselesaikan dalam hitungan detik menggunakan metrik *Cosine Similarity*.

Penelitian ini secara spesifik mengimplementasikan model *pre-trained* paraphrase-*multilingual-MiniLM-L12-v2*. Model bahasa tersebut merupakan versi distilasi efisien yang memiliki 12 lapisan *Transformer* dengan hasil luaran berupa vektor padat (*dense vector*) berdimensi 384 (Wang dkk., 2020). Ukuran dimensi 384 merupakan karakteristik struktural dari arsitektur MiniLM yang dikembangkan melalui teknik *deep self-attention distillation*. Reduksi ukuran dimensi tersembunyi (*hidden size*) dari standar BERT-*Base* sebesar 768 menjadi 384 memfasilitasi model MiniLM untuk mencapai akselerasi kecepatan inferensi $2{,}7x$ hingga $5{,}3x$ lebih cepat dengan tetap mempertahankan akurasi performa yang kompetitif pada berbagai *benchmark* NLP (Wang dkk., 2020).

Ekstraksi representasi semantik kalimat pada arsitektur SBERT melewati tahapan matematis utama yang mengonvergensikan nilai dari level token menuju level kalimat tunggal. Tahapan kontekstualisasi tingkat token (*token-level contextualization*) mengeksploitasi fungsi *Multi-Head Attention* dan *Position-Wise Feed-Forward Network* yang telah dideklarasikan sebelumnya untuk menghasilkan matriks *hidden state* kontekstual (h_t). Tahapan berikutnya menjalankan pemadatan vektor (*sentence pooling*) untuk mengonversi matriks token yang bervariasi menjadi satu vektor berukuran tetap. SBERT menerapkan operasi *Mean Pooling* untuk menghitung nilai rata-rata elemen dari seluruh vektor *hidden state* kontekstual hasil luaran lapisan ke-12 arsitektur MiniLM. Formulasi matematis untuk memperoleh vektor representasi semantik kalimat v dijabarkan pada Persamaan 2.14 berikut (Reimers & Gurevych, 2019):

$$v = \frac{1}{N}\sum_{t=1}^{N} h_t \quad \text{.....(2.14)}$$

Keterangan:

v : Vektor representasi semantik kalimat berdimensi padat 384.

N : Jumlah total token riil di dalam kalimat (mengabaikan token *padding*).

h_t : Vektor *hidden state* kontekstual pada lapisan terakhir untuk token ke-t dengan batasan indeks iterasi $t = 1, 2, \dots, N$.

Model berbasis *paraphrase-multilingual* terbukti memiliki tingkat akurasi semantik yang lebih tinggi dalam memproses data teks berbahasa Indonesia dibandingkan dengan model multibahasa lainnya (Dhini dkk., 2023). Kumpulan vektor SBERT pada metode peringkasan ekstraktif dimanfaatkan untuk mengidentifikasi titik pusat topik dari sebuah

klaster dokumen yang disebut sebagai Vektor *Centroid*. Perhitungan rata-rata (*mean*) dari seluruh dimensi vektor kalimat menghasilkan nilai *Centroid* tersebut, sesuai dengan definisi pada Persamaan 2.15.

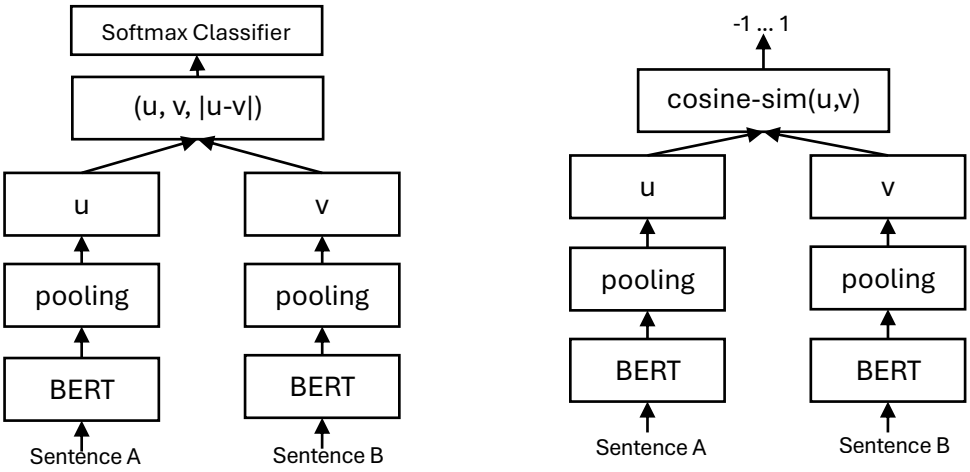
$$C = \frac{1}{n}\sum_{i=1}^{n} v_i \quad \text{.....(2.15)}$$

Keterangan:

C : Vektor *Centroid* (representasi inti topik dari suatu klaster).

n : Jumlah total kalimat di dalam satu klaster topik tersebut.

v_i : Vektor embedding hasil ekstraksi model untuk kalimat ke-i $(i = 1,2, \dots, n)$.



Gambar 2.4 Arsitektur Jaringan *Siamese* pada *Sentence*-BERT untuk Pembanding Kalimat (Reimers & Gurevych, 2019)

Arsitektur jaringan kembar (*Siamese Network*) pada SBERT (Gambar 2.4) merepresentasikan mekanisme aliran data spesifik saat fase pelatihan (*fine-tuning*) (Reimers & Gurevych, 2019). Berbeda dengan BERT konvensional, SBERT memproses setiap kalimat secara independen dan memadatkannya melalui lapisan *Pooling* menjadi vektor berdimensi tetap (*fixed-sized sentence embeddings*). Pada fase implementasi (*inference*) di penelitian ini, komputasi diputus tepat setelah lapisan *Pooling* tanpa mengeksekusi *Softmax Classifier* maupun *Cosine Similarity* bawaan model. Representasi vektor mandiri tersebut ditarik keluar untuk diolah oleh algoritma eksternal, sebuah inovasi yang memungkinkan komputasi kemiripan makna jutaan kalimat selesai dalam hitungan detik, jauh lebih efisien dibandingkan arsitektur BERT standar (Reimers & Gurevych, 2019).

2.9. *Maximal Marginal Relevance* (MMR)

Tahap pemilihan kalimat (*sentence selection*) bertujuan menyeleksi kalimat paling signifikan sekaligus mereduksi redundansi informasi secara logis. Pendekatan tersebut selaras dengan perkembangan teknik peringkasan teks mutakhir. Supriyono dkk., 2024 menegaskan dalam tinjauan komprehensifnya bahwa pengelolaan redundansi informasi yang tumpang tindih lintas dokumen merupakan tantangan fundamental dalam peringkasan multi-dokumen.

Algoritma Maximal Marginal Relevance (MMR) menjadi instrumen utama dalam mengatasi hambatan redundansi pada tahapan seleksi penelitian ini. Carbonell & Goldstein (1998) pertama kali menggagas algoritma ini sebagai metode temu balik informasi (*Information Retrieval*) yang menyusun daftar dokumen berdasarkan urutan relevansi terhadap kueri pengguna. Metode MMR terbukti adaptif untuk optimasi berbasis reinforcement learning (Mao dkk., 2020) serta sangat handal dalam peringkasan teks berbahasa Indonesia (Mawaridi dkk., 2024). Penelitian ini menerapkan pemilihan kalimat menggunakan metrik MMR = λ Rel(s) - (1-λ) max Sim(s, S), di mana λ menyeimbangkan relevansi dan diversitas (keragaman).

Tahap penentuan kalimat yang akan diekstraksi menjadi ringkasan dilakukan menggunakan algoritma *Maximal Marginal Relevance* (MMR). MMR beroperasi dengan menyeimbangkan dua metrik utama yaitu Relevansi (kedekatan makna kandidat kalimat dengan *Centroid*) dan Diversitas (perbedaan makna kandidat kalimat dengan kalimat-kalimat yang sudah terpilih sebelumnya). Tujuannya adalah untuk menghasilkan ringkasan yang padat informasi, namun minim pengulangan (redundansi). Skor MMR secara standar dihitung menggunakan persamaan 2.16.

$$MMR = \arg \max_{S_i \in R \setminus S} \left[\lambda \cdot Sim_1(S_i, C) - (1 - \lambda) \max_{S_j \in S} S\,im_2(S_i, S_j) \right] \quad \text{.....(2.16)}$$

Keterangan:

S_i : Kandidat kalimat yang sedang dievaluasi

C : Vektor *Centroid*

S_j : Kalimat yang sudah terpilih masuk ke dalam ringkasan

Sim : Fungsi *Cosine Similarity* (Kemiripan Kosinus)

λ : Parameter penyeimbang (ditetapkan sebesar 0.7 pada penelitian ini)

Perhitungan magnitudo perkalian kedua vektor dieksekusi secara komputasional menggunakan model representasi embeddings berdasarkan formulasi pada Persamaan 2.17 berikut:

$$Sim_2(S_i, S_j) = \frac{\sum_{k=1}^{384}(S_{i,k} \times S_{j,k})}{\sqrt{\sum_{k=1}^{384}(S_{i,k})^2} \times \sqrt{\sum_{k=1}^{384}(S_{j,k})^2}} \quad \text{.....(2.17)}$$

Keterangan:

k : mewakili indeks dimensi dari vektor SBERT (dari 1 hingga 384).

$S_{i,k}$: Nilai pada dimensi ke-k untuk kalimat kandidat ke-i ($i = 1, 2, \dots, N$ dan $k = 1, 2, \dots, 384$).

$S_{j,k}$: Nilai pada dimensi ke-k untuk kalimat ke-j di dalam himpunan ringkasan ($j = 1,2, \dots, |S|$ dan $k = 1,2, \dots,384$).

2.10. Tingkat Kompresi Peringkasan

Mesin peringkas mengeksekusi proses ekstraksi kalimat secara iteratif hingga mencapai batas panjang maksimal yang didefinisikan sebagai tingkat kompresi (*compression rate*). Parameter kompresi ini merepresentasikan rasio perbandingan antara panjang teks ringkasan buatan mesin dengan total panjang keseluruhan dokumen sumber (Widyassari dkk., 2022). Penentuan batas kompresi memegang peranan krusial dalam metode ekstraktif karena secara langsung memengaruhi kualitas kepadatan informasi. Rasio kompresi yang terlalu rendah (ringkasan terlampau pendek) berisiko menghilangkan konteks esensial berita, sedangkan rasio yang terlalu tinggi (ringkasan terlampau panjang) akan memicu peningkatan kembali tingkat redundansi kalimat. Penelitian ini merancang skenario ekstraksi pada empat variasi batas kompresi, yakni 20%, 30%, 40%, dan 50%, guna mengidentifikasi titik keseimbangan paling optimal antara kepadatan informasi dan keberagaman teks.

2.11. Metrik Evaluasi ROUGE

ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*) adalah metrik baku untuk mengevaluasi kualitas peringkasan teks secara otomatis dengan mengukur tingkat *overlap* n-gram antara ringkasan buatan mesin dan ringkasan buatan manusia/pakar (Sai

dkk., 2022). ROUGE menghitung kata-kata yang tumpang tindih antara ringkasan mesin dan ringkasan referensi serta bobot masing-masing. Unit yang sesuai, seperti n-gram (n-kata), urutan kata, dan pasangan kata antara ringkasan mesin dan ringkasan referensi dihitung untuk melakukan pengukuran nilai ROUGE. ROUGE-N mengukur jumlah n-gram yang cocok antara teks yang dihasilkan oleh mesin peringkasan dan ringkasan manual yang dilakukan manusia (Mawaridi dkk., 2024). Secara matematis, perhitungan metrik ROUGE melibatkan perbandingan antara ringkasan buatan manusia yang didefinisikan sebagai S_{ref} (*Reference Summary*) dan ringkasan buatan mesin yang didefinisikan sebagai S_{sys} (*System Summary*). Evaluasi diukur melalui tiga komponen utama, yaitu *Recall* (tingkat penemuan informasi), *Precision* (tingkat ketepatan mesin), pada Persamaan 2.18 dan 2.19 berikut:

$$Recall_{ROUGE-N} = \frac{\sum_{gram_n \in S_{ref}} Count_{match}(gram_n)}{\sum_{gram_n \in S_{ref}} Count(gram_n)} \quad \text{.....(2.18)}$$

$$Precision_{ROUGE-N} = \frac{\sum_{gram_n \in S_{ref}} Count_{match}(gram_n)}{\sum_{gram_n \in S_{sys}} Count(gram_n)} \quad \text{.....(2.19)}$$

Keterangan:

n	:	Panjang gram yang dievaluasi
$Count_{match}(gram_n)$	:	Jumlah n-gram yang sama (beririsan) antara ringkasan mesin dan ringkasan rujukan.
$\sum Count(gram_n)$	:	Total keseluruhan n-gram pada teks yang bersangkutan (S_{ref} untuk total kata/frasa peringkas manusia, S_{sys} untuk total kata/frasa mesin).

ROUGE-N mengukur tingkat irisan berdasarkan n-gram (potongan kata). Untuk ROUGE-1, nilai $n = 1$ (*unigram*), sedangkan untuk ROUGE-2, nilai $n = 2$ (*bigram*). Varian ROUGE-1 mengukur tingkat kecocokan kata tunggal atau *unigram* antara teks ringkasan dengan referensi. Secara makna, ROUGE-1 bertugas untuk melihat apakah kosa kata dasar yang mendefinisikan inti peristiwa (seperti nama tokoh, lokasi, atau kata kerja utama) berhasil ditangkap oleh algoritma. Peran ROUGE-1 pada penelitian ekstraktif ini sangat krusial sebagai indikator pembuktian awal bahwa algoritma seleksi yang dibangun tidak salah dalam memilih kalimat yang memuat fakta dasar berita.

Varian ROUGE-2 mengevaluasi kecocokan susunan dua kata yang berurutan atau *bigram* secara lebih spesifik. Makna dari pengukuran ini adalah untuk melihat apakah kata-

kata yang diekstrak memiliki konteks frasa yang masuk akal, bukan sekadar kata acak (misalnya, frasa "harga saham" atau "hasil imbang"). Mengingat penelitian ini menggunakan metode usulan berjenis ekstraktif yang langsung mengambil wujud kalimat asli dari dokumen, perolehan nilai ROUGE-2 yang baik akan membuktikan bahwa algoritma mampu memilih kalimat yang susunan frasanya secara tata bahasa paling relevan dengan pemahaman manusia.

Mekanisme ROUGE-L menilai koherensi dan kelancaran struktur kalimat secara utuh menggunakan prinsip *Longest Common Subsequence* (LCS). Berbeda dengan n-gram yang memotong kata secara kaku, LCS mencari urutan deret kata terpanjang yang muncul secara searah di kedua teks, meskipun ada beberapa kata yang menyela di tengahnya. Makna penggunaan ROUGE-L adalah untuk memastikan seberapa lancar dan senada alur cerita hasil ringkasan tersebut saat dibaca. Pada penelitian peringkasan dokumen berita, ROUGE-L berperan sebagai validasi pamungkas bahwa kalimat-kalimat utuh yang dipilih oleh metode usulan tidak merusak kronologi atau narasi berita aslinya. Perhitungan *Recall* dan *Precision* untuk ROUGE-L dirumuskan pada Persamaan 2.20 dan 2.21.

$$Recall_{ROUGE-L} = \frac{LCS(S_{ref},S_{sys})}{|S_{ref}|} \quad (2.20)$$

$$Precision_{ROUGE-L} = \frac{LCS(S_{ref},S_{sys})}{|S_{sys}|} \quad (2.21)$$

Keterangan:

$LCS(S_{ref}, S_{sys})$: Panjang jumlah kata dari urutan deret terpanjang yang beririsan antara ringkasan referensi dan ringkasan mesin.

$|S_{ref}|$: Panjang total kata pada ringkasan rujukan pakar manusia.

$|S_{sys}|$: Panjang total kata pada ringkasan buatan mesin peringkas.

Setelah nilai *Recall* dan *Precision* didapatkan dari masing-masing varian ROUGE, nilai performa akhir mesin peringkas dihitung menggunakan metrik *F1-Score* yang merupakan titik keseimbangan dari kedua komponen tersebut. Rumus *F1-Score* dijabarkan pada Persamaan 2.22:

$$F1\text{-}Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (2.22)$$

Pengujian simultan menggunakan ketiga varian ROUGE tersebut sejalan dengan standar evaluasi riset pemrosesan bahasa alami modern. Kajian literatur terbaru memvalidasi bahwa kombinasi ROUGE-1, ROUGE-2, dan ROUGE-L merupakan tolok ukur yang paling komprehensif untuk menguji kelayakan model peringkasan bahasa Indonesia, karena ketiganya secara hierarkis mampu membedah akurasi mesin mulai dari level kosa kata dasar, keluwesan frasa, hingga kelancaran alur logika narasi (*narrative flow*) (Hartawan dkk., 2024).

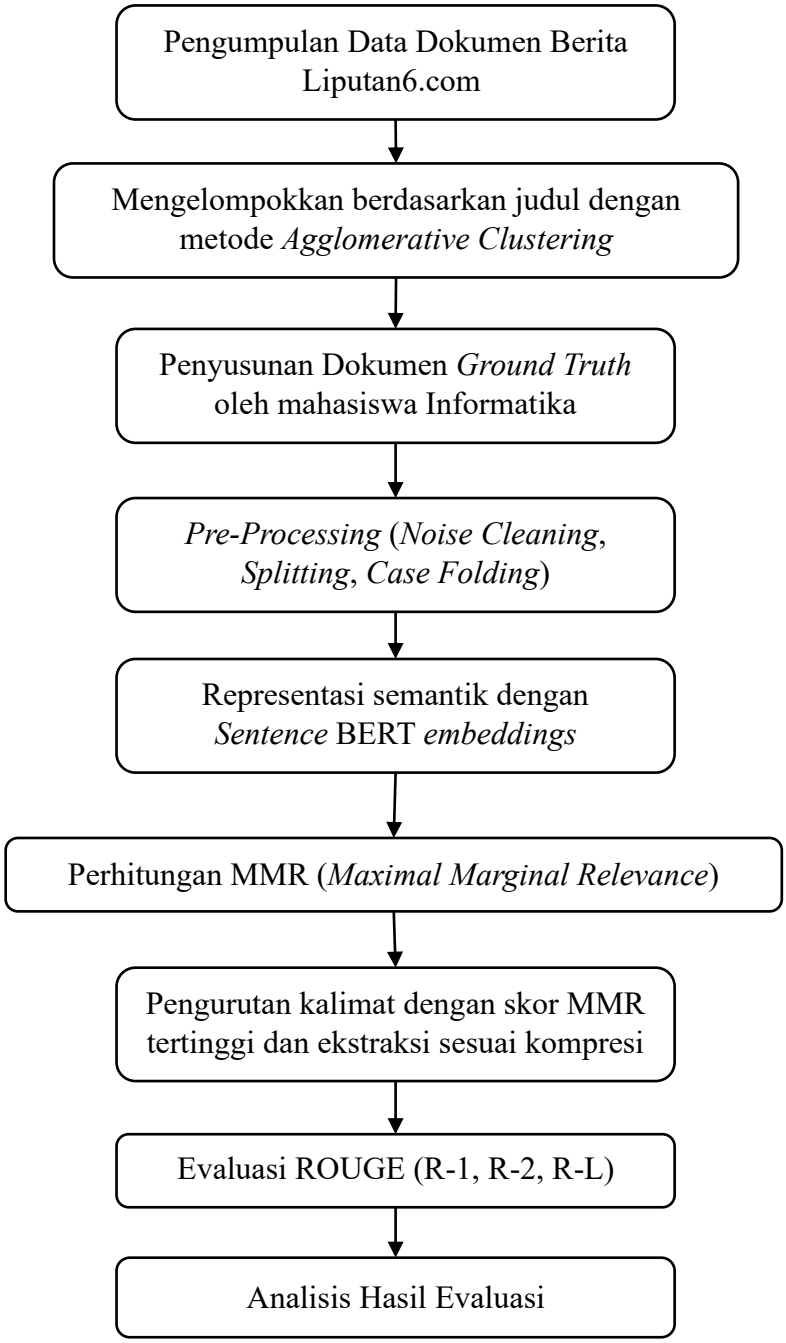
BAB III
METODOLOGI PENELITIAN

Bab ini menjelaskan tahapan-tahapan yang dilakukan dalam penelitian untuk mengimplementasikan mesin peringkasan dengan metode ekstraktif pada berita multi-dokumen berbasis *Sentence*-BERT (SBERT) dan *Maximal Marginal Relevance* (MMR). Evaluasi dilakukan dengan tiga metrik ROUGE yaitu berupa ROUGE-1, ROUGE-2, serta ROUGE-L.

3.1. Alur Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahapan yang sistematis dan berjenjang, mulai dari pengumpulan data mentah hingga evaluasi hasil ringkasan. Tahapan penelitian diawali dengan akuisisi data melalui teknik *web scraping* pada portal berita Liputan6 untuk mengumpulkan artikel dalam berbagai kategori. Data mentah yang diperoleh kemudian dikelompokkan menggunakan metode *Agglomerative Clustering* berbasis pembobotan TF-IDF dan *Cosine Distance* dengan ambang batas 0,8 guna memastikan dokumen berada dalam satu konteks peristiwa yang sama. Secara beriringan, proses penyusunan *human ground truth* dilakukan oleh dua orang anotator terhadap setiap klaster valid, di mana konsistensi rujukan tersebut divalidasi melalui uji reliabilitas *Cohen's Kappa* untuk menjamin standar evaluasi yang objektif dan bebas dari bias subjektivitas.

Teks berita di dalam klaster diproses melalui tahapan pra-pemrosesan yang meliputi *noise cleaning*, *sentence splitting*, dan *case folding* guna memperoleh format data yang seragam. Setiap kalimat hasil pembersihan dikonversi menjadi representasi vektor padat menggunakan model *Sentence*-BERT (*paraphrase-multilingual-MiniLM-L12-v2*) dengan dimensi 384. Perhitungan vektor *centroid* dilakukan setelahnya sebagai representasi inti topik klaster untuk memfasilitasi mekanisme pemilihan. Algoritma *Maximal Marginal Relevance* (MMR) dengan parameter $\lambda = 0{,}7$ diimplementasikan untuk menyeleksi kalimat secara iteratif berdasarkan keseimbangan skor relevansi dan diversitas semantik. Hasil akhir penelitian berupa ringkasan ekstraktif pada variasi tingkat kompresi 20%, 30%, 40%, dan 50% dievaluasi menggunakan metrik ROUGE terhadap human ground truth sebagai standar rujukan. Secara keseluruhan, penggambaran dari arsitektur alur kerja mesin peringkas dapat dilihat pada Gambar 3.1.



Gambar 3.1 Bagan alur Gambaran Umum Penelitian

3.2. Pengumpulan dan Persiapan Data

Pengumpulan data pada penelitian ini dilakukan dengan teknik *web scraping* terhadap portal berita daring Liputan6. Proses ekstraksi data berjalan secara otomatis menggunakan bahasa pemrograman Python. Untuk mengatasi sistem keamanan anti-bot dan pemuatan halaman dinamis (*dynamic rendering*), ekstraksi dilakukan menggunakan *library* Selenium (*Undetected ChromeDriver*) yang dikonfigurasi dengan simulasi gulir perlahan (*smooth scrolling*) dan injeksi parameter *?page=all* untuk memastikan seluruh teks termuat utuh.

Selanjutnya, *BeautifulSoup* digunakan untuk mem-*parsing* elemen HTML dan membuang elemen pengganggu seperti iklan.

Proses *scraping* dijalankan di komputer lokal pada bulan Januari 2026, menargetkan rentang publikasi artikel berita selama satu minggu, yaitu dari 12 Januari 2026 hingga 18 Januari 2026. Sistem diatur sedemikian rupa untuk menelusuri kategori-kategori spesifik seperti Bola, Tekno, Bisnis, Saham, Otomotif, dan Kesehatan. Selama proses penelusuran (*crawling*), sistem ini akan mengakses setiap tautan URL berita yang valid. Dari berbagai metadata yang tersedia pada struktur halaman web (tanggal tayang berita, judul, isi, atau tag kategori), sistem *scrapper* secara spesifik hanya mengekstraksi dua komponen utama, yaitu bagian judul berita (*title*) dan isi berita utama (*content*).

Pemilihan kedua atribut tersebut didasarkan pada fungsi judul sebagai representasi inti topik dan isi berita sebagai narasi komprehensif yang akan diproses oleh metode usulan. Data hasil ekstraksi disimpan ke dalam format JSON dengan distribusi merata, yaitu 50 judul berita untuk masing-masing kategori. Total 300 artikel berita ini merupakan contoh himpunan data mentah (*raw dataset*) utuh yang bersiap memasuki tahapan pengelompokan awal menggunakan metode usulan *Agglomerative Clustering*. Setelah dokumen-dokumen tersebut dipilah secara komputasional ke dalam klaster topik yang koheren, barulah setiap klaster akan dieksekusi proses peringkasannya untuk menghasilkan ringkasan komputasional serta diolah secara manual oleh manusia untuk menghasilkan standar rujukan (*ground truth*). Penetapan kuantitas sebanyak 300 dokumen ini secara empiris tidak akan memengaruhi kualitas dasar kecerdasan model. Hal ini dikarenakan arsitektur SBERT diposisikan secara eksklusif sebagai pengekstraksi fitur semantik pada fase inferensi murni, bukan untuk fase pelatihan ulang (*fine-tuning*). Praktik penggunaan model bahasa *pre-trained* sebagai pengekstraksi fitur tetap (*fixed feature extractor*) ini diakui valid dalam literatur pemrosesan bahasa alami karena performa utamanya bersumber dari basis pengetahuan prapelatihan awal, bukan dari volume data target (Peters dkk., 2019).

Total 300 dokumen berita tersebut dikumpulkan menjadi data mentah (korpus) yang siap diproses pada tahapan pengelompokan topik selanjutnya.[1] Tabel 3.1 menyajikan representasi sebagian dari data dokumen berita yang berhasil diekstrak.

[1] Yusanda, A. F. P. (2026). Korpus data berita Liputan6 untuk peringkasan multi-dokumen ekstraktif (Google Drive, 2026). Tautan: https://bit.ly/dataset_berita_liputan6_12-Januari-2026

Tabel 3.1 Contoh Data Dokumen Berita pada Kategori Tertentu

No.	Judul Berita	Kategori	Isi Berita
1.	Liverpool vs Burnley: The Reds Gagal Jauhi Manchester United	Bola	Liputan6.com, Jakarta -LagaLiverpool vs Burnleypada lanjutanLiga Inggris2025/2026 sudah berakhir. Laga di Anfield, Sabtu (17/1/2026) malam WIB, berakhir 1-1. Terus menekan sejak menit pertama, tuan rumah mendapat peluang emas lewat titik putih. **[…].**
2.	Barcelona Lolos Dramatis di Santander, Joan Garcia Panen Pujian	Bola	Liputan6.com, Jakarta -Barcelonamemastikan tiket ke perempat finalCopa del Reyusai menaklukkanRacing Santanderpada babak 16 besar, Jumat (16/01/2026). Kemenangan 2-0 itu terasa jauh dari kata mudah bagi tim asuhan Hansi Flick. Bertanding di El Sardinero, Barcelona tampil dominan dalam penguasaan bola dan tekanan. **[…]**.
…	…	…	…
51.	Penjualan Mobil 2025 Capai Target Meski di Bawah Angka 2024	Otomotif	Liputan6.com, Jakarta -Penjualan mobilpada 2025 berhasil mencapai target yang telah ditetapkan Gabungan Industri Kendaraan Bermotor Indonesia (Gaikindo). Hal tersebut, dikarenakan pencapaian yang cukup baik saat tutup tahun. **[…]**.
…	…	…	…
101.	Mengenal ADAM, Sindrom Penurunan Hormon yang Diam-Diam Dialami Banyak Pria	Kesehatan	Liputan6.com, Jakarta -Penurunanhormonpadapriabukan sekadar isu medis eksklusif, tapi realitas yang bisa memengaruhi kualitas hidup. Salah satu kondisi yang kini makin banyak dibicarakan adalah ADAM (Androgen Deficiency inAgingMale). Sindrom penurunan hormon testosteronyang kerap terjadi perlahan tapi berdampak nyata. ADAM bukan sekadar istilah baru. **[…]**.
…	…	…	…
300.	Bursa Saham Asia Bervariasi, Bank Sentral Korsel Pertahankan Suku Bunga	Saham	Liputan6.com, Jakarta -Bursa saham AsiaPasifik bergerak bervariasi pada perdagangan Kamis, (15/1/2026). Pergerakanbursa sahamAsia Pasifik terjadi di tengahinvestormenilai keputusan kebijakan terbaruBank Sentral Korea Selatan. Mengutip CNBC, Bank Sentral Korea Selatan mempertahankan suku bunga di 2,5%. **[…].**

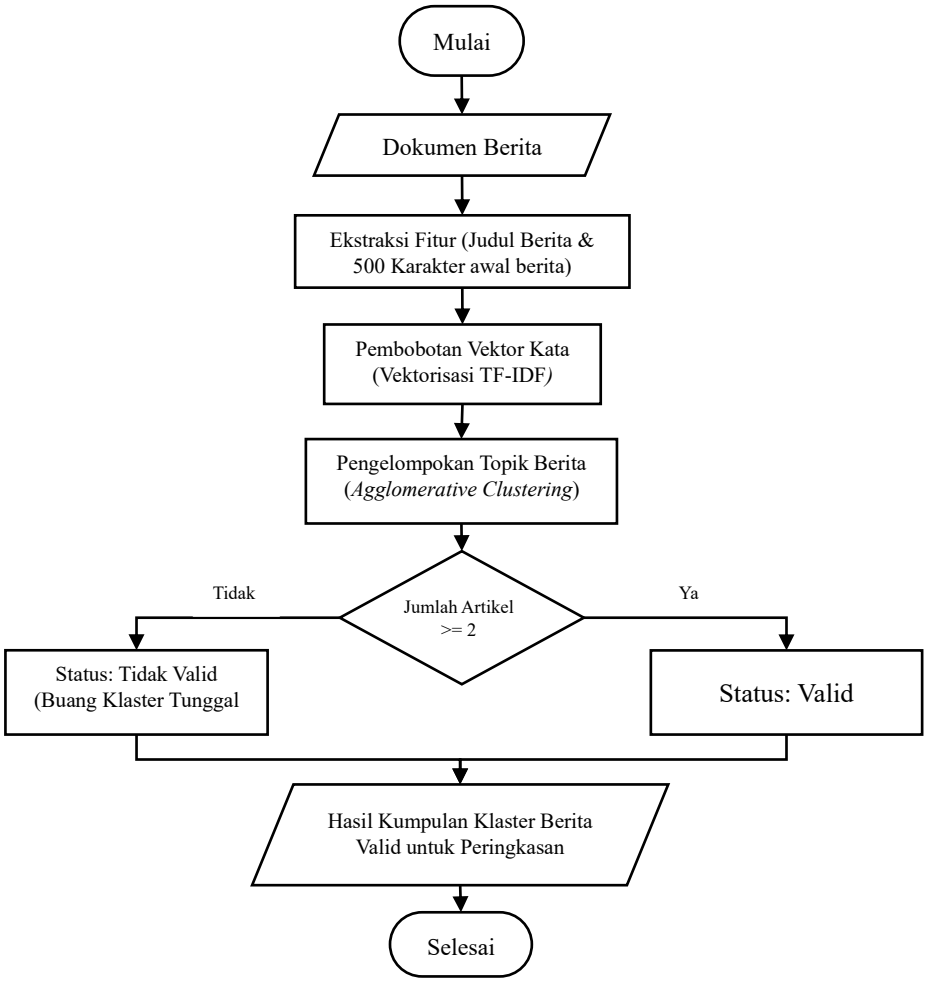
3.3. Pengelompokan Topik Berita

Kumpulan berita harian memiliki ragam peristiwa yang sangat acak dan heterogen. Oleh karena itu, tahapan pra-pemrosesan memilah dokumen-dokumen tersebut ke dalam klaster topik yang sama menggunakan metode usulan *Agglomerative Clustering*. Algoritma

ini menggunakan pendekatan *bottom-up* untuk mengelompokkan berita ke dalam sub-topik yang spesifik, guna memastikan bahwa ringkasan multi-dokumen yang dihasilkan nantinya tetap berfokus pada satu konteks peristiwa yang koheren (Subakti dkk., 2022).

Tahap pengelompokan pada penelitian ini mengekstraksi teks judul yang digabungkan dengan satu paragraf pembuka (teras berita atau *lead*) terlebih dahulu, sebagai representasi utama dokumen untuk proses pengelompokan. Pemilihan atribut ini didasarkan pada prinsip piramida terbalik dalam struktur jurnalistik, di mana elemen esensial sebuah peristiwa (5W1H) selalu dikompresi pada bagian judul dan paragraf pertama. Paragraf pertama atau *lead* berita (yang memuat unsur *Who, What, When, Where, Why, How*) umumnya berukuran sekitar 300 hingga 600 karakter (sekitar 50-70 kata) sebagai pemantik utama pembaca (Latif dkk, 2023). Membatasi ekstraksi hanya pada teras berita terbukti secara konseptual lebih efektif dalam merepresentasikan inti topik dibandingkan menggunakan keseluruhan isi teks yang cenderung memuat banyak *noise* atau informasi sekunder. Gabungan teks terpilih ini kemudian direpresentasikan ke dalam bentuk ruang vektor spasial menggunakan pembobotan TF-IDF, sejalan dengan landasan teori mengenai ekstraksi fitur leksikal pada Subbab 2.3 (Wicaksana dkk., 2018).

Algoritma pengelompokan menghitung tingkat kedekatan antar-dokumen berita menggunakan metrik *Cosine Distance* (jarak kosinus). Proses *bottom-up* ini akan terus menggabungkan artikel-artikel yang memiliki kedekatan topik hierarkis hingga mencapai batas pemotongan (*distance threshold*) sebesar 0,8. Penentuan *threshold* jarak 0,8 ini ditetapkan secara empiris dengan mempertimbangkan karakteristik matriks vektor TF-IDF pada teks pendek (judul dan teras berita) yang memiliki tingkat *sparsity* (kekerangan) sangat tinggi. Karena teks pendek memiliki perpotongan kata yang lebih sedikit, ambang batas 0,8 memberikan titik ekuilibrium yang optimal agar dokumen dengan topik peristiwa yang sama, dapat tetap tergabung ke dalam satu klaster, tanpa mencampurkan peristiwa yang sama sekali berbeda (Subakti dkk., 2022). Pada akhir tahapan ini, klaster yang berhasil terbentuk dengan memuat minimal dua artikel akan diloloskan untuk dieksekusi pada tahap peringkasan semantik selanjutnya. Alur kerja proses *clustering* tersebut disajikan pada Gambar 3.2.



Gambar 3.2 Bagan Metode *Agglomerative Clustering*

3.3.1. Pembobotan Kata (TF-IDF)

Tahap komputasi awal yaitu menghitung frekuensi kemunculan kata (*Term Frecuency*/TF) pada setiap dokumen. Untuk memberikan gambaran yang transparan mengenai asal muasal kata (*term*) pada matriks perhitungan, bagian ini mengilustrasikan simulasi pembobotan menggunakan tiga contoh dokumen judul berita dari kategori "Bola". Ketiga sampel dokumen mentah tersebut (selanjutnya disebut D1, D2, dan D3) adalah sebagai berikut:

1. D1: "Liverpool Gagal Kalahkan Burnley"
2. D2: "Laga Liverpool vs Burnley Gagal"
3. D3: "MU Berhasil Kalahkan Man City"

Langkah selanjutnya mengeksekusi perhitungan nilai *Inverse Document Frequency* (IDF) menggunakan formula $\log_{10}(N/DF)$, di mana $N = 3$ merepresentasikan jumlah

contoh dokumen berita yang digunakan, dan DF (*Document Frequency*) adalah jumlah dokumen yang memuat kata tersebut. Penjabaran asal kata hasil *tokenization* beserta hasil perhitungan nilai frekuensinya dapat dilihat secara lengkap pada Tabel 3.2.

Tabel 3.2 Hasil Perhitungan nilai TF-IDF

***Term* (kata)**	**TF(D1)**	**TF(D2)**	**TF(D3)**	**DF**	**Perhitungan DF**	**Nilai DF**
liverpool	1	1	0	2	$\log_{10}(3/2)$	0,176
gagal	1	1	0	2	$\log_{10}(3/2)$	0,176
kalahkan	1	0	1	2	$\log_{10}(3/2)$	0,176
burnley	1	1	0	2	$\log_{10}(3/2)$	0,176
laga	0	1	0	1	$\log_{10}(3/1)$	0,477
vs	0	1	0	1	$\log_{10}(3/1)$	0,477
mu	0	0	1	1	$\log_{10}(3/1)$	0,477
berhasil	0	0	1	1	$\log_{10}(3/1)$	0,477
man	0	0	1	1	$\log_{10}(3/1)$	0,477
city	0	0	1	1	$\log_{10}(3/1)$	0,477

Perhitungan bobot untuk kata "liverpool" pada dokumen 1 (D1) dijabarkan sebagai berikut:

$$W_{liverpool,D1} = tf_{liverpool,D1} \times \log\left(\frac{N}{df_{liverpool}}\right)$$

$$W_{liverpool,D1} = 1 \times \log\left(\frac{3}{2}\right) \approx 0{,}176$$

Perhitungan bobot akhir TF-IDF (W) mengalikan matriks TF dengan nilai IDF. Representasi hasil akhir membentuk vektor numerik (*vector token*) untuk masing-masing kata di dalam dokumen. Tabel 3.3 mendemonstrasikan matriks vektor bobot tersebut.

Tabel 3.3 Matriks Vektor Bobot TF-IDF Dokumen

No.	**Kata**	**D1**	**D2**	**D3**
1.	liverpool	0,176	0,176	0,000
2.	gagal	0,176	0,176	0,000
3.	kalahkan	0,176	0,000	0,176
4.	burnley	0,176	0,176	0,000
5.	laga	0,000	0,477	0,000
6.	vs	0,000	0,477	0,000
7.	mu	0,000	0,000	0, 477
8.	berhasil	0,000	0,000	0, 477
9.	man	0,000	0,000	0, 477
10.	city	0,000	0,000	0, 477

3.3.2. Perhitungan Jarak (*Cosine Distance*)

Algoritma *clustering* akan menghitung jarak kedekatan antar dokumen atau *cosine distance*. Sebagai contoh, perhitungan jarak antara Dokumen 1 (D1) dan Dokumen 2 (D2) adalah sebagai berikut:

1. Menghitung *Dot Product* ($D1 \cdot D2$):

$$(D1 \cdot D2) = (0{,}176 \times 0{,}176) + (0{,}176 \times 0{,}176) + (0) + (0{,}176 \times 0{,}176) + (0) = 0{,}0929$$

2. Menghitung Panjang Vektor ($\left||D1|\right| dan \left||D2|\right|$):

$$\left||D1|\right| = \sqrt{0{,}176^2 + 0{,}176^2 + 0{,}176^2 + 0{,}176^2} = \sqrt{0{,}1239} \approx 0{,}352$$

$$\left||D2|\right| = \sqrt{0{,}176^2 + 0{,}176^2 + 0{,}176^2 + 0{,}477^2 + 0{,}477^2} = \sqrt{0{,}5479} \approx 0{,}740$$

3. Menghitung *Cosine Similarity*:

$$\text{Sim}(D1, D2) = \frac{0{,}0929}{0{,}352 \times 0{,}740} = \frac{0{,}0929}{0{,}2604} \approx 0{,}356$$

4. Menghitung *Cosine Distance*:

$$Dist(D1, D2) = 1 - 0{,}356 = 0{,}644$$

Prosedur serupa menghitung jarak D1 dengan D3 dan menghasilkan nilai *Cosine Distance* sebesar 0,909 akibat irisan tunggal pada kata "kalahkan". Dokumen D2 dan D3 menghasilkan jarak maksimal sebesar 1,00 karena tidak memiliki irisan kata sama sekali.

3.3.3. Proses Pembentukan Klaster

Algoritma menyusun Matriks Jarak Awal (*Initial Distance Matrix*) berdasarkan akumulasi perhitungan sebelumnya seperti yang tersaji pada Tabel 3.4. Konfigurasi metode *Agglomerative Clustering* pada penelitian ini menetapkan batas henti (*Distance Threshold*) pada nilai 0,8.

Tabel 3.4 Matriks Jarak (*Distance Matrix*) Antar Dokumen Berita

	D1	**D2**	**D3**
D1	0,000	0,644	0,909
D2	0,644	0,000	1,000
D3	0,909	1,000	0,000

Algoritma selanjutnya mengeksekusi penggabungan dokumen secara bertahap menggunakan kriteria kedekatan jarak terkecil melalui rangkaian siklus iterasi komputasi berikut:

1. Iterasi 1: Algoritma mencari nilai jarak terkecil pada matriks, yaitu jarak antara D1 dan D2 (0,644). Karena nilai $0{,}644 < 0{,}8$ (di bawah *threshold*), maka D1 dan D2 digabungkan membentuk Klaster 1.
2. Iterasi 2: Jarak antara Klaster 1 dengan D3 dievaluasi menggunakan *Average Linkage* (rata-rata jarak). Jaraknya adalah $\frac{(0{,}909+1{,}00)}{2} \approx 0{,}954$. Karena nilai $0{,}954 > 0{,}8$ (melebihi *threshold*), maka proses penggabungan dihentikan (stop).

Batas *threshold* 0,8 memfasilitasi algoritma untuk mengelompokkan dokumen mentah menjadi beberapa klaster. Proses validasi mengabaikan klaster dengan jumlah artikel kurang dari dua (berstatus "Tidak Valid") guna menjaga fokus penelitian pada peringkasan multi-dokumen. Klaster dengan populasi lebih dari satu artikel menerima status "Valid" untuk dieksekusi pada tahap peringkasan. Tabel 3.5 menampilkan hasil *clustering* pada kategori "Bola" beserta status validasi jumlah populasinya.

Tabel 3.5 Hasil Clustering Judul Berita dengan Kategori “Bola”

Cluster ID	Jumlah Artikel	Judul Artikel	Status
0	3	Michael Carrick Kembalikan Marwah Old Trafford Sebagai 'Tempat Ajaib' Usai Derbi Manchester	Valid
		Roy Keane Telan Ludah Sendiri: Michael Carrick Bungkam Kritik dengan Kemenangan Derby	
		Michael Carrick Sejak Awal Yakin MU Akan Kalahkan Man City	
1	2	Manchester United Menang Tanpa Dominasi Bola: Rahasia 32 Persen Penguasaan yang Kalahkan Man City Harry Maguire Menjawab Keraguan: Jadi 'Unsung Hero' Manchester United Saat Hantam Man City	Valid
2	3	Mengagumi Kylian Mbappe, Bisa Main dan Cetak Gol Sambil menahan Rasa Sakit	Valid
		Bos Real Madrid Bela Vinicius Junior yang Dihujat Madridista	
		Real Madrid vs Levante: Alvaro Arbeloa Petik Kemenangan Pertama	
3	2	PSG Tikung Chelsea! Wonderkid Barcelona Ini Tinggalkan La Masia Usai Dibujuk Luis Enrique	Valid
		Barcelona Terancam Kehilangan Wonderkid, Dro Fernandez Siap Angkat Kaki Januari Ini	
4	2	Arsenal Sudah Kebanjiran Gelandang, Tapi Tetap Tertarik Datangkan ‘Pedri Serbia’	Valid
		Arsenal Aman dari Konflik Internal, Mikel Arteta Pastikan Masalah Declan Rice Sudah Selesai	

Cluster ID	Jumlah Artikel	Judul Artikel	Status
5	5	Man Utd vs Man City, Carrick Langsung Kasih Paham Amorim Cara Maksimalkan Pemain	Valid
		Manchester United Sikat Man City, Carrick Pamer: Lihat Tuh Haaland Kita Bikin Mati Kutu!	
		Man Utd vs Man City: 3 Gol Dianulir, Setan Merah Sikat Rival Sekota	
		Man Utd vs Man City Mau Mulai, Ini Link Live Streaming Liga Inggris di Vidio	
		Man Utd vs Man City, Dapatkan Link Live Streaming Liga Inggris di Vidio	
6	2	Liverpool Gagal Kalahkan Burnley, Arne Slot: Gini Aja Terus!	Valid
		Liverpool vs Burnley: The Reds Gagal Jauhi Manchester United	
7	2	Manchester United Diharap Jangan Lepas 2 Pemain Ini: Bisa Menghidupkan Kembali Setan Merah! Keputusan Manchester United Angkat Carrick Sebagai Bos Interim Didukung Juara Piala Dunia 2002 Ini	Valid
8	1	Sukses di SEA Games 2025, Atlet Tinju Indonesia Dapat Kado Spesial dari Polri dan TNI	Tidak Valid

Proses clustering mengelompokkan seluruh berita ke dalam topik-topik spesifik berdasarkan kesamaan semantik pada masing-masing kategori. Distribusi jumlah klaster yang terbentuk pada Tabel 3.6 bukanlah hasil penentuan manual oleh peneliti, melainkan murni merupakan luaran (*output*) komputasional dari algoritma pengelompokan terhadap sampel data uji yang digunakan.

Tabel 3.6 Hasil Distribusi Klaster per Kategori Berita

No.	Kategori Berita	Jumlah Klaster Topik
1.	Kesehatan	11
2.	Bola	11
3.	Bisnis	13
4.	Otomotif	7
5.	Tekno	8
6.	Saham	8

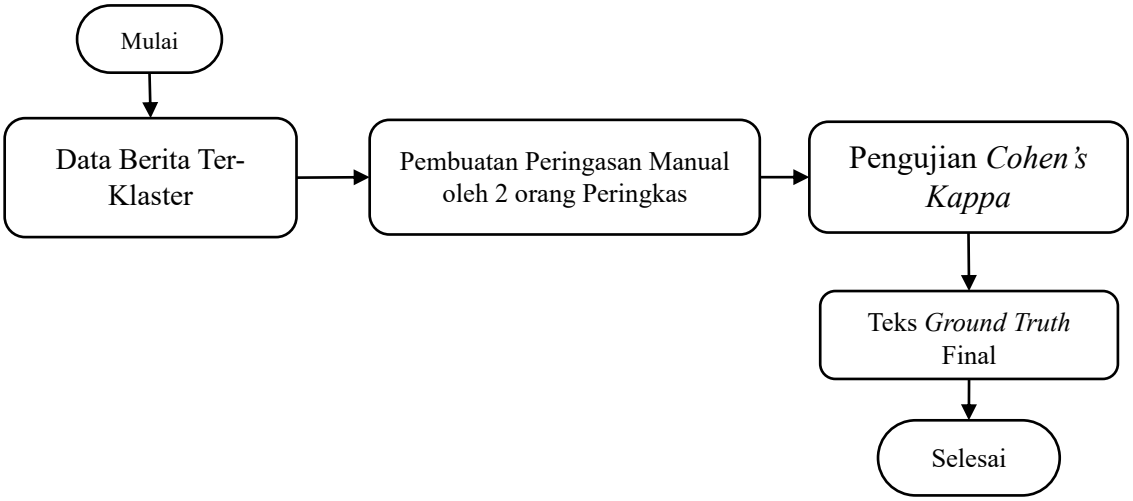
Algoritma *Agglomerative Clustering* secara dinamis memecah data dan membentuk jumlah klaster yang bervariasi karena bergantung penuh pada tingkat kedekatan semantik antar-judul dokumen di setiap kategori. Eksekusi tahap pengelompokan ini pada akhirnya menghasilkan akumulasi total sebanyak 58 klaster valid yang siap dilanjutkan ke tahap peringkasan. Kategori dengan jumlah klaster sedikit mengindikasikan liputan peristiwa

berulang yang terpusat. Kategori dengan jumlah klaster banyak menunjukkan variasi topik berita yang lebih tersebar dan spesifik. Mekanisme validasi tersebut membuktikan bahwa judul berita dengan topik menyimpang otomatis menerima status "tidak valid" dan tereliminasi dari *pipeline* peringkasan.

3.4. Penyusunan *Human Ground Truth* dan Uji *Cohen's Kappa*

Pengukuran kualitas hasil ringkasan dari metode usulan memerlukan sebuah standar emas (*gold standard*) berupa *Human Ground Truth* pada tahap awal evaluasi. Penyusunan *ground truth* ini melibatkan dua orang anotator (Peringkas 1 dan Peringkas 2) yang dipilih dari kalangan mahasiswa tingkat akhir program studi Informatika. Alasan pemilihan ini didasarkan pada kebiasaan mahasiswa Informatika terhadap logika pemrograman yang kaku. Karakteristik tersebut membuat mereka lebih bisa diandalkan untuk memilih kalimat secara eksak (ekstraktif) tanpa tergoda untuk mengarang atau menyusun kalimat baru (abstraktif) layaknya mahasiswa dari rumpun ilmu bahasa.

Agar hasil rujukan benar-benar objektif, peneliti memberikan batasan dan instruksi yang tegas kepada kedua peringkas sebelum mereka mulai bekerja. Mereka diwajibkan untuk membaca seluruh berita di dalam satu klaster topik terlebih dahulu, kemudian menyeleksi kalimat-kalimat utuh yang paling mewakili inti peristiwa. Selama proses penyeleksian ini, peringkas sama sekali tidak diperbolehkan untuk mengubah susunan kata asli atau menyelipkan opini pribadi. Sebagai bukti pertanggungjawaban akademis bahwa proses penyusunan rujukan ini murni hasil pemikiran manual dan independen tanpa bantuan perangkat lunak otomatis, kedua anotator telah menandatangani pakta integritas yang didokumentasikan secara terlampir pada Lampiran 1. Rangkuman alur penyusunan *ground truth* ini dapat dilihat pada Gambar 3.3.



Gambar 3.3 Bagan Alur Penyusunan *Ground Truth*

Pengujian dilakukan diseluruh klaster di setiap kategori berita. Sebagai contoh, pada subbab ini ditampilkan salah satu sampel berita kategori "Bola" dengan klaster topik: “Liverpool vs Burnley”. Berikut hasil peringkasan dari proses penyusunan manual oleh anotator dapat dilihat pada Tabel 3.7.

Tabel 3.7 Contoh Hasil Ringkasan Manual Anotator

Nama Peringkas	**Hasil Peringkasan Manual**
Peringkas 1	Manajer Liverpool, Arne Slot kecewa berat dengan hasil imbang yang didapatkan timnya ketika melawan Burnley. Ia menilai Liverpool seharusnya bisa meraih poin penuh di laga ini. Liverpool awalnya unggul terlebih dahulu berkat gol Florian Wirtz. Namun sayang mereka gagal meraih tiga poin di laga ini setelah Marcus Edwards menyamakan kedudukan di menit ke-65. "Ini bukan kali pertama kami mengalami situasi ini, jadi ini sangat membuat saya frustrasi," buka Slot di laman resmi Liverpool. "Setelah itu, kami menjelma menjadi tim yang bermain terlalu berhati-hati, dan itu membuat kami sulit menciptakan peluang dan hasilnya adalah kami sudah menjalani beberapa pertandingan di mana kami tidak kalah." Ia menilai jika The Reds mampu bermain lebih agresif, mereka bisa mengamankan tiga poin di laga ini. "Di sepak bola, sebuah tim bisa hanya membuat dua peluang namun mereka bisa mencetak gol." Liverpool tidak mau lama-lama meratapi hasil imbang melawan Burnley ini. Karena mereka ditunggu laga lain di tengah pekan ini. The Reds dijadwalkan akan bertandang ke Prancis untuk menghadapi Marseille di lanjutan league phase Liga Champions musim 2025/2026.

Nama Peringkas	Hasil Peringkasan Manual
Peringkas 2	Manajer Liverpool, Arne Slot, merasa sangat frustrasi dengan hasil imbang yang diraih timnya saat melawan Burnley karena mereka seharusnya bisa meraih poin penuh, meskipun Liverpool sempat unggul berkat gol Florian Wirtz namun gagal mempertahankan keunggulan dan membuat banyak peluang. Ia menilai timnya bermain terlalu aman di babak kedua dan mengatakan jika mereka bermain lebih agresif, mereka bisa mengamankan kemenangan. Slot juga menyebutkan, "Ini bukan kali pertama kami mengalami situasi ini, jadi ini sangat membuat saya frustrasi," dan mengkritik ketidakmampuan mereka mengonversi peluang menjadi gol meskipun menguasai bola lebih banyak. Dua tim kemudian menjalani pertandingan berakhir imbang 1-1, dengan Liverpool memanfaatkan penalti yang gagal dieksekusi Dominik Szoboszlai dan kemudian gol dari Wirtz di menit ke-42, sebelum Marcus Edwards menyamakan kedudukan. Mereka memuncaki klasemen sementara dengan 36 poin, unggul satu poin atas Manchester United. Liverpool akan menghadapi Bournemouth setelah ini, sementara Burnley akan meladeni Tottenham Hotspur.

Peneliti menjadikan pemilihan Unit Informasi (UI) sebagai landasan penilaian kesepakatan antar-peringkas. Dalam pendekatan komputasional pada penelitian ini, Unit Informasi didefinisikan secara konkret sebagai kalimat-kalimat tunggal asli penyusun dokumen sumber. Proses ekstraksi Unit Informasi dieksekusi secara otomatis menggunakan modul *Sentence Tokenization* yang memecah teks gabungan klaster berita menjadi barisan kalimat utuh. Seluruh kalimat asli inilah yang kemudian bertindak sebagai data induk (*baseline*) fakta berita.

Pendekatan komputasional ini memberikan jaminan konsistensi tingkat tinggi dalam memetakan fakta berita secara utuh pada skala data yang masif. Penggunaan skrip *Sentence Tokenization* secara terprogram mampu mengekstraksi seluruh dokumen dari 58 klaster ke dalam daftar data induk hanya dalam hitungan detik. Pemecahan teks ini bekerja secara deterministik dengan mendeteksi batas struktural kalimat (seperti tanda titik), sehingga memastikan tidak ada satupun detail peristiwa (5W1H) yang luput akibat faktor kelelahan kognitif manusia (*human error*). Sebagai representasi dari *output* ekstraksi otomatis tersebut, Tabel 3.8 menyajikan sampel penarikan data induk berupa wujud kalimat utuh yang diambil secara acak dari dua klaster berbeda, yakni perwakilan dari kategori Bola dan Saham.

Tabel 3.8 Contoh Data Induk Fakta dari Klaster Topik Tertentu

ID Klaster Berita	Kategori	Informasi Berita Asli	Kode Unit Informasi
Klaster-6	Bola	Arne Slot merasa sangat kecewa dan frustrasi dengan hasil imbang yang diraih timnya.	UI-1
		Liverpool awalnya unggul terlebih dahulu berkat gol yang dicetak oleh Florian Wirtz.	UI-2
		Marcus Edwards berhasil menyamakan kedudukan menjadi 1-1 untuk Burnley.	UI-3
		"Ini bukan kali pertama kami mengalami situasi ini," ujar Slot memberikan evaluasi.	UI-4
		Tim dinilai bermain terlalu aman dan berhati-hati pada babak kedua pertandingan.	UI-5
		Liverpool seharusnya bisa mengamankan kemenangan jika bermain lebih agresif.	UI-6
		Dominik Szoboszlai gagal mengeksekusi tendangan penalti dengan baik.	UI-7
		The Reds dijadwalkan bertandang menghadapi Marseille di Liga Champions.	UI-8
		Statistik menunjukkan Liverpool mendominasi penguasaan bola hingga 60 persen.	UI-9
		Wasit mengeluarkan beberapa kartu kuning akibat tingginya jumlah pelanggaran.	UI-10
		Pertandingan ini dipimpin secara langsung oleh wasit utama di lapangan.	UI-11
		Laga tersebut berlangsung di bawah kondisi cuaca hujan yang cukup deras.	UI-12
Klaster-2	Saham	Rentang pergerakan harga saham BUMI (Rp 404 - Rp 428)	UI-1
		Penguatan IHSG 0,47% ke level 9.075,40 (15 Jan 2026)	UI-2
		Total nilai transaksi harian bursa mencapai Rp 28,3 triliun	UI-3
		Rentang pergerakan harga saham BBRI (Rp 3.730 - Rp 3.850)	UI-4

ID Klaster Berita	Kategori	Informasi Berita Asli	Kode Unit Informasi
		Nilai transaksi harian saham BBRI sebesar Rp 1,3 triliun	UI-5
		Penetapan dividen interim BBRI Rp 137/saham (Tahun Buku 2025)	UI-6
		Rentang pergerakan harga saham BBCA (Rp 7.975 - Rp 8.175)	UI-7
		Nilai transaksi harian saham BBCA sebesar Rp 20,1 triliun	UI-8
		Penutupan sesi pertama BBCA menguat 0,94% di posisi Rp 8.075	UI-9
		Pergerakan IHSG bertahan di zona hijau pada penutupan sesi pertama	UI-10

Guna memetakan keputusan peringkas secara objektif, peneliti menerapkan algoritma *Fuzzy Matching* untuk mencocokkan hasil ringkasan Anotator 1 dan Anotator 2 terhadap daftar kalimat data induk di atas. Algoritma ini mendeteksi tingkat irisan kata (*word intersection*) dengan mengacu pada formulasi irisan leksikal yang telah diuraikan pada Persamaan 2.7. Dalam konteks pengujian ini, kalimat pada data induk bertindak sebagai variabel S_{asli}, sedangkan kalimat hasil ringkasan anotator bertindak sebagai S_{target}. Sistem pemetaan dikonfigurasi menggunakan ambang batas (*threshold*) kemiripan mutlak sebesar 50%. Apabila kalimat ringkasan anotator memiliki irisan leksikal $\geq 50\%$terhadap suatu kalimat di data induk, maka algoritma secara otomatis mencatat skor biner (1) yang berarti anotator "sepakat" memasukkan informasi tersebut. Sebaliknya, jika hasil komputasi bernilai $< 50\%$, akan diberikan skor (0).

Implementasi otomatisasi ini menjadi langkah krusial mengingat pengujian dilakukan pada skala masif. Guna memberikan gambaran matematis mengenai cara kerja logis algoritma pemetaan tersebut, Tabel 3.9 mengilustrasikan simulasi hasil pemetaan biner dari sampel klaster Kategori Bola sebelum nilai akhir *Cohen's Kappa* dihitung oleh sistem.

Tabel 3.9 Matriks Kesepakatan Anotasi pada Hasil Peringkasan Manual

No.	Informasi Berita Asli	Kode Unit Informasi	Peringkas 1	Peringkas 2
1.	Arne Slot merasa sangat kecewa dan frustrasi dengan hasil imbang yang diraih timnya.	UI-1	1	1
2.	Liverpool awalnya unggul terlebih dahulu berkat gol yang dicetak oleh Florian Wirtz.	UI-2	1	1
3.	Marcus Edwards berhasil menyamakan kedudukan menjadi 1-1 untuk Burnley.	UI-3	1	1
4.	"Ini bukan kali pertama kami mengalami situasi ini," ujar Slot memberikan evaluasi.	UI-4	1	1
5.	Tim dinilai bermain terlalu aman dan berhati-hati pada babak kedua pertandingan.	UI-5	1	1
6.	Liverpool seharusnya bisa mengamankan kemenangan jika bermain lebih agresif.	UI-6	1	1
7.	Dominik Szoboszlai gagal mengeksekusi tendangan penalti dengan baik.	UI-7	1	0
8.	The Reds dijadwalkan bertandang menghadapi Marseille di Liga Champions.	UI-8	0	1
9.	Statistik menunjukkan Liverpool mendominasi penguasaan bola hingga 60 persen.	UI-9	0	0
10.	Wasit mengeluarkan beberapa kartu kuning akibat tingginya jumlah pelanggaran.	UI-10	0	0
11.	Pertandingan ini dipimpin secara langsung oleh wasit utama di lapangan.	UI-11	0	0
12.	Laga tersebut berlangsung di bawah kondisi cuaca hujan yang cukup deras.	UI-12	0	0

Perolehan data distribusi berupa ekstraksi terhadap teks berita asli dapat ditinjau pada Tabel 3.9. yaitu total 12 unit informasi utama ($N = 12$) yang dievaluasi oleh kedua anotator peringkasan. Dari keseluruhan unit informasi tersebut, tingkat kesepakatan aktual (p_o) diperoleh distribusi keputusan di mana kedua peringkas sepakat pada 10 unit informasi (6 sepakat memilih dan 4 sepakat membuang). Secara individu, Peringkas 1 dan Peringkas 2 memiliki pola ekstraksi yang identik dengan memilih 7 informasi dan membuang 5

informasi. Selanjutnya dilakukan perhitungan menggunakan rumus *Cohen's kappa*, sebagai berikut:

1. Perhitungan Kesepakatan Aktual

$$(p_o): p_o = \frac{\text{Total Sepakat}}{N} = \frac{10}{12} \approx 0{,}8333$$

2. Perhitungan Peluang Harapan Kebetulan (p_e):

   Peluang harapan (p_e) dihitung dengan menjumlahkan probabilitas Peringkas 1 dan 2 sama-sama menjawab "Ya" dengan probabilitas mereka menjawab "Tidak".

   a. Peluang Harapan "Ya": $P_{ya} = \frac{7}{12} \times \frac{7}{12} \approx 0{,}3403$

   b. Peluang Harapan "Tidak": $P_{tidak} = \frac{5}{12} \times \frac{5}{12} \approx 0{,}1736$

$$p_e = P_{ya} + P_{tidak} = 0{,}3403 + 0{,}1736 = 0{,}5139$$

3. Perhitungan Skor Akhir Cohen's Kappa (κ):

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

$$= \frac{0{,}8333 - 0{,}5139}{1 - 0{,}5139}$$

$$= \frac{0{,}3194}{0{,}4861} \approx 0{,}6570$$

Nilai $\kappa = 0{,}6570$ yang dihasilkan pada perhitungan berada pada rentang 0,61 – 0,80, berdasarkan kriteria interval *Cohen's Kappa* dari Landis & Koch (1977). Angka ini merepresentasikan Kesepakatan Kuat (*Substantial Agreement*). Hal ini membuktikan secara matematis bahwa *ground truth* yang disusun ini memiliki konsistensi pemahaman yang tinggi antar-anotator dan tidak terjadi karena faktor kebetulan semata. Berdasarkan data pada Tabel 3.9, kedua peringkas sepakat memilih 6 informasi dan sepakat membuang 4 informasi dari total 12 Unit observasi. Maka perhitungannya adalah:

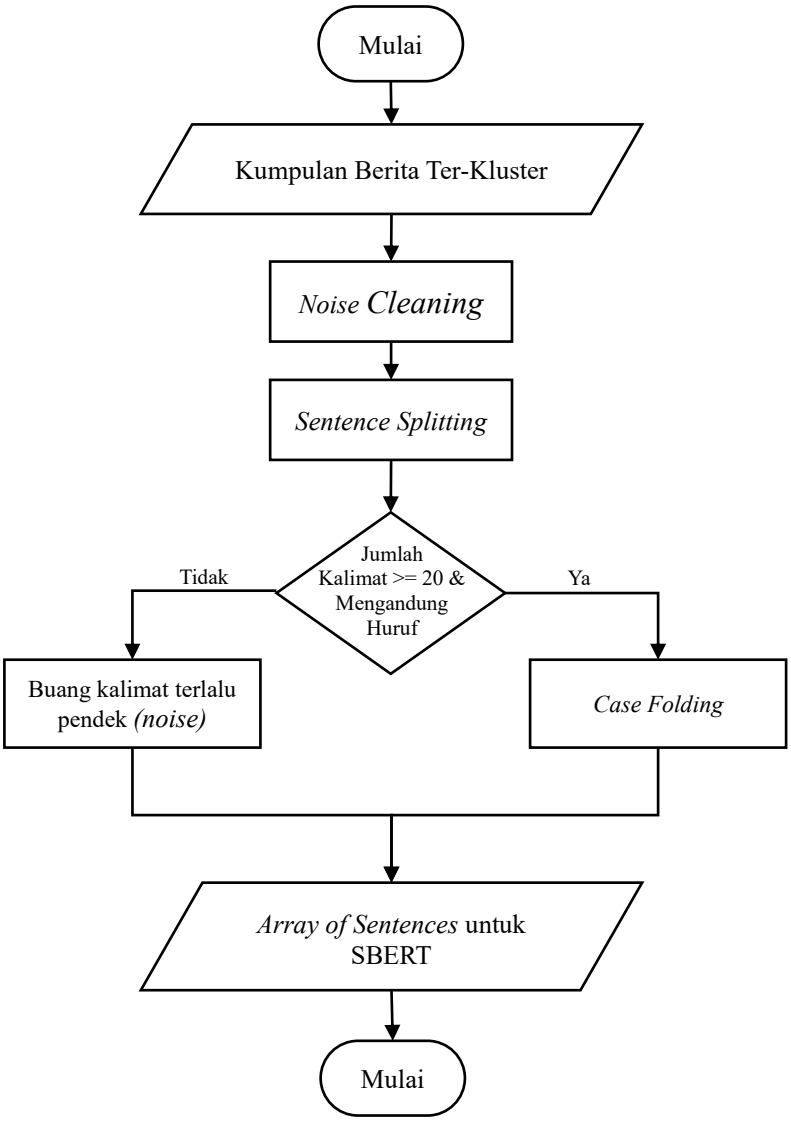
$$Percentage\ Agreement = \frac{6+4}{12} \times 100 = \frac{10}{12} \times 100 \approx 83{,}33$$

Angka 83,33 ini menunjukkan bahwa tingkat persentase kesepakatan aktual antara kedua peringkas sangat tinggi. Penggunaan kedua metrik ini secara berdampingan memberikan gambaran yang lebih komprehensif, di mana *Cohen's Kappa* mengukur reliabilitas yang disesuaikan dengan peluang, sementara *Percentage Agreement* merepresentasikan konsensus faktual di lapangan. Prosedur yang sama kemudian diimplementasikan secara otomatis menggunakan skrip pemrograman untuk menguji seluruh klaster terbentuk dari contoh data uji (total 58 klaster) yang hasilnya akan dipaparkan pada BAB 4.

3.5. Pra-Pemrosesan pada Teks Berita

Teks berita terlebih dahulu dibersihkan dari *noise* sebelum memasuki proses peringkasan. Hal ini agar tidak mengganggu kualitas komputasi mesin peringkas. Teks berita yang diambil secara langsung dari portal daring (Liputan6). Teks berita daring masih mengandung banyak *noise* atau informasi yang tidak relevan serta memuat struktur ireguler yang dapat mendistorsi pembobotan fitur (Widyassari dkk., 2022). Oleh karena itu, tahapan *noise cleaning*, *sentence splitting*, dan *case folding* diimplementasikan secara berurutan untuk menstandarisasi unsur kebahasaan untuk perolehan informasi nantinya.

Proses ini dijalankan pada satu klaster terpilih untuk dilakukan peringkasan. Hasil dari *pre-processing* ini berupa dua versi keluaran untuk setiap kalimat, yang pertama berupa teks asli (untuk ditampilkan pada ringkasan akhir) dan kedua adalah teks bersih (sebagai *input* perhitungan model SBERT). Tahapan *preprocessing* meliputi *noise cleaning*, *sentence splitting*, dan *case folding* ditampilkan dalam bentuk bagan/*flowchart* pada Gambar 3.4.



Gambar 3.4 Bagan Alur *Pre-processing* Dokumen Berita Ter-kluster

Bagian ini menyajikan penerapan algoritma pembersihan (*noise cleaning*, *sentence splitting*, dan *case folding*) pada salah satu sampel paragraf dari dataset berita kategori "Bola" guna memberikan gambaran yang lebih jelas mengenai transformasi teks pra-pemrosesan.

3.5.1. *Noise Cleaning* Kalimat Berita

Proses *noise cleaning* yang digunakan pada penelitian ini adalah *Regular Expression* (*Regex*) yaitu untuk mendeteksi dan menghapus atribut bawaan portal berita, seperti lokasi jurnalis ("Liputan6.com, Jakarta -"). Berikut sebagai contoh prebandingan hasil dari *noise cleaning* yang diterapkan pada salah satu kluster berita pada kategori "Bola", dapat dilihat pada Tabel 3.10. dan Tabel 3.11.

Tabel 3.10 Teks Dokumen Berita Sebelum *Noise Cleaning*

Id Dokumen	Sebelum *Noise Cleaning*
1	Liputan6.com, Jakarta -ManajerLiverpool,Arne Slotkecewa berat dengan hasil imbang yang didapatkan timnya ketika melawan Burnley. Ia menilai Liverpool seharusnya bisa meraih poin penuh di laga ini. Kemarin malam, Liverpool memainkan pertandingan ke-21 EPL 2025/2026 di Anfield**.** **[**….] Liverpool tidak mau lama-lama meratapi hasil imbang melawan Burnley ini. Karena mereka ditunggu laga lain di tengah pekan ini. The Reds dijadwalkan akan bertandang ke Prancis untuk menghadapi Marseille di lanjutan league phase Liga Champions musim 2025/2026. Sumber: Liverpool FC.
2	Liputan6.com, Jakarta -LagaLiverpool vs Burnleypada lanjutanLiga Inggris2025/2026 sudah berakhir. Laga di Anfield, Sabtu (17/1/2026) malam WIB, berakhir 1-1. Terus menekan sejak menit pertama, tuan rumah mendapat peluang emas lewat titik putih. Wasit memberi penalti usai Cody Gakpo dijegal di area terlarang. [….] 23.30 WIB: Aston Villa vs Everton Selasa, 20 Januari 2026 03.00 WIB: Brighton & Hove Albion vs Bournemouth

Tabel 3.11 Teks Dokumen Berita Setelah *Noise Cleaning*

Id Dokumen	Setelah *Noise Cleaning*
1	Manajer Liverpool, Arne Slot kecewa berat dengan hasil imbang yang didapatkan timnya ketika melawan Burnley. Ia menilai Liverpool seharusnya bisa meraih poin penuh di laga ini. Kemarin malam, Liverpool memainkan pertandingan ke-21 EPL 2025/2026 di Anfield. [….] Liverpool tidak mau lama-lama meratapi hasil imbang melawan Burnley ini. Karena mereka ditunggu laga lain di tengah pekan ini. The Reds dijadwalkan akan bertandang ke Prancis untuk menghadapi Marseille di lanjutan league phase Liga Champions musim 2025/2026. Sumber: Liverpool FC.
2	Laga Liverpool vs Burnley pada lanjutan Liga Inggris 2025/2026 sudah berakhir. Laga di Anfield, Sabtu (17/1/2026) malam WIB, berakhir 1-1. Terus menekan sejak menit pertama,tuan rumah mendapat peluang emas lewat titik putih. Wasit memberi penalti usai Cody Gakpo dijegal di area terlarang. [….] 23.30 WIB: Aston Villa vs Everton Selasa, 20 Januari 2026 03.00 WIB: Brighton & Hove Albion vs Bournemouth.

3.5.2. *Splitting* Kalimat Berita

Teks paragraf yang telah bersih dari atribut web kemudian dipecah menjadi array kalimat tunggal berdasarkan deteksi tanda baca akhir (seperti titik), hasilnya dapat dilihat pada Tabel 3.12. Pemecahan ini krusial dilakukan sebelum tanda baca dihapus agar mesin dapat mendeteksi batas akhir kalimat, seperti tanda titik (.) atau tanda tanya (?), secara presisi.

Tabel 3.12 Perbandingan Kalimat Sebelum dan Setelah *Splitting*

Sebelum *Sentence Splitting*	**Setelah *Sentence Splitting***
Manajer Liverpool, Arne Slot kecewa berat dengan hasil imbang yang didapatkan timnya ketika melawan Burnley. Ia menilai Liverpool seharusnya bisa meraih poin penuh di laga ini. Kemarin malam, Liverpool memainkan pertandingan ke-21 EPL 2025/2026 di Anfield. […] Liverpool tidak mau lama-lama meratapi hasil imbang melawan Burnley ini. Karena mereka ditunggu laga lain di tengah pekan ini. The Reds dijadwalkan akan bertandang ke Prancis untuk menghadapi Marseille di lanjutan league phase Liga Champions musim 2025/2026. Sumber: Liverpool FC.	[‘Manajer Liverpool, Arne Slot kecewa berat dengan hasil imbang yang didapatkan timnya ketika melawan Burnley.’,’Ia menilai Liverpool seharusnya bisa meraih poin penuh di laga ini.’,’Kemarin malam, Liverpool memainkan pertandingan ke-21 EPL 2025/2026 di Anfield.’, […] ‘Liverpool tidak mau lama-lama meratapi hasil imbang melawan Burnley ini.’,’Karena mereka ditunggu laga lain di tengah pekan ini.’,’The Reds dijadwalkan akan bertandang ke Prancis untuk menghadapi Marseille di lanjutan league phase Liga Champions musim 2025/2026.’, ‘Sumber: Liverpool FC’]

3.5.3. *Case Folding* Kalimat Berita

Tahapan terakhir pada proses pra-pemrosesan teks adalah penyeragaman bentuk huruf atau *case folding*. Seluruh kalimat berita yang sebelumnya telah bersih dari derau (*noise*) langsung dikonversi secara menyeluruh menjadi huruf kecil (*lowercase*). Standardisasi teks ini diterapkan secara mutlak untuk menghindari ambiguitas komputasi pada pembacaan mesin. Melalui proses ini, komputer tidak akan menganggap kata yang sama namun memiliki kapitalisasi huruf awal yang berbeda sebagai dua variabel teks yang terpisah. Hal ini bertujuan agar akurasi dan stabilitas kalkulasi vektor matematis pada tahap pemodelan selanjutnya tetap terjaga. Sebagai representasi dari tahapan ini, Tabel 3.13 menyajikan hasil pemrosesan *case folding* pada tiga kalimat pertama dari setiap dokumen berita.

Tabel 3.13 Teks Dokumen Berita Setelah *Case Folding*

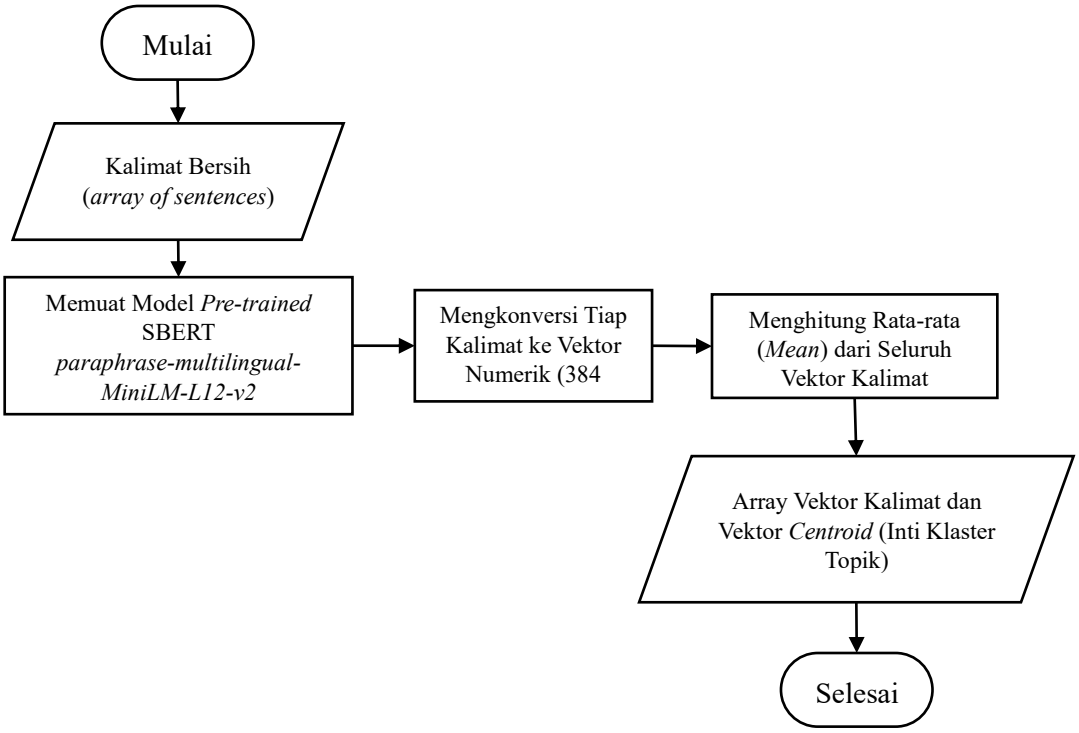
Id Dokumen	**Setelah *Case Folding***
1	[‘manajer liverpool arne slot kecewa berat dengan hasil imbang yang didapatkan timnya ketika melawan burnley’,’ia menilai liverpool seharusnya bisa meraih poin penuh di laga ini’,’kemarin malam liverpool memainkan pertandingan ke epl di anfield’, [**….**]]
2	[‘laga liverpool vs burnley pada lanjutan liga inggris sudah berakhir’, ‘laga di anfield sabtu malam wib berakhir’, ‘terus menekan sejak menit pertama tuan rumah mendapat peluang emas lewat titik putih’, [**….**]]

Hasil pembersihan dokumen berita yang telah membentuk potongan kalimat terstandardisasi, seperti yang ditampilkan pada Tabel 3.13. Kalimat berita disimpan ke dalam

sebuah struktur data *array*. Struktur *array* tersebut diinisialisasi menggunakan variabel bernama *calculate_text* guna membedakannya secara komputasional dari teks artikel mentah (*raw text*) yang belum diproses. Variabel *calculate_text* ini selanjutnya menjadi data masukan (*input*) utama bagi model *Transformer Sentence*-BERT untuk menjalankan tahap pemrosesan representasi semantik.

3.6. Representasi Semantik dengan *Sentence*-BERT

Kalimat-kalimat bersih dari tahap pra-pemrosesan yang tersimpan dalam variabel *calculate_text* dikonversi menjadi representasi numerik agar dapat diproses oleh algoritma seleksi. Penelitian ini mengekstraksi makna semantik kalimat menggunakan model *Sentence-BERT*. Model *paraphrase-multilingual-MiniLM-L12-v2* bertugas mengonversi setiap kalimat secara independen ke dalam ruang vektor berdimensi padat (*dense vector*), sesuai dengan arsitektur *Siamese Network* yang dijabarkan oleh Reimers & Gurevych (2019). Algoritma *Sentence*-BERT kemudian melakukan agregasi nilai dari seluruh kalimat yang telah menjadi vektor untuk menemukan titik pusat informasi klaster, yang direpresentasikan sebagai Vektor *Centroid*. Alur kerja proses ini secara visual dapat dilihat pada Gambar 3.5.



Gambar 3.5 Flowchart Representasi Semantik dengan *Sentence*-BERT

Proses konversi kalimat ke dalam vektor numerik oleh model Sentence-BERT diilustrasikan secara visual pada Gambar 3.5. Model *paraphrase-multilingual-MiniLM-L12-v2* memetakan setiap kalimat bersih menjadi larik numerik desimal kontinu (*floating-point*) sepanjang 384 dimensi. Proses transformasi ini mengekstraksi nilai *hidden state* tingkat token dari lapisan *Transformer* terakhir menggunakan mekanisme *self-attention* sesuai Persamaan 2.10, yang kemudian dipadatkan menggunakan operasi *Mean Pooling* sesuai Persamaan 2.14. Sebagai contoh kalimat yang akan dilakukan representasi semantik menggunakan Untuk penjelasan proses serta perhitungan *Centroid* secara matematis dapat dilihat pada Tabel 3.14 yang menampilkan simulasi komputasi pada 3 kalimat pertama dari salah satu contoh dari data kategori berita tekno, dengan klaster topik "Menangkar Masa Depan AI" yang berjumlah total 3 dokumen artikel berita di dalamnya. Tabel 3.14 menampilkan contoh kalimat beserta jumlah token kata-nya.

Tabel 3.14 Contoh Kalimat dan Jumlah Token Kata

Id	Teks	Token Kata
1.	Di tengah perlombaan global mengadopsi kecerdasan buatan ai talenta indonesia mulai mengambil peran krusial di pusat inovasi dunia	18
2.	product manager di google amerika serikat juan anugraha djuwadi kini menjadi salah satu aktor strategis yang mengarahkan bagaimana teknologi ai dikembangkan agar tetap manusiawi dan berdampak luas bagi pengguna global	30
3.	dalam webinar bertajuk ai streamline your business build internal apps with ai barubaru ini juan membedah peta jalan pengembangan ai yang melampaui sekadar kecanggihan teknis	25

Setiap token t di dalam kalimat tersebut diekstraksi terlebih dahulu oleh mekanisme *self-attention* lapisan *Transformer* ke-12 untuk menghasilkan matriks nilai *hidden state* (h_t) berdimensi 1×384. Operasi *Mean Pooling* kemudian menghitung nilai rata-rata dari seluruh vektor token pada setiap indeks dimensi secara simultan berdasarkan formulasi pada Persamaan 2.9 sebagai berikut:

$$v_1 = \frac{1}{18}\sum_{t=1}^{18} h_t$$

$$v_{1,\text{dim1}} = \frac{0{,}1214 + 0{,}2105 + (-0{,}0841) + \cdots + 0{,}3142}{18} = 0{,}1897$$

$$v_{1,\text{dim2}} = \frac{0{,}0512 + (-0{,}1142) + 0{,}0915 + \cdots + 0{,}1821}{18} = 0{,}0813$$

$$v_{1,\text{dim3}} = \frac{-0{,}3112 + (-0{,}1045) + (-0{,}2118) + \cdots + (-0{,}1892)}{18} = -0{,}2497$$

$$v_{1,\text{dim4}} = \frac{-0{,}0124 + 0{,}0412 + (-0{,}1102) + \cdots + (-0{,}0941)}{18} = -0{,}0704$$

$$v_{1,\text{dim5}} = \frac{-0{,}0812 + (-0{,}1542) + 0{,}0214 + \cdots + (-0{,}1241)}{18} = -0{,}1094$$

Hasil kalkulasi pembarian rata-rata dari tahapan *pooling* tersebut menghasilkan baris vektor representasi semantik final untuk Kalimat 1 (v_1). Potongan matriks nilai untuk 5 dimensi pertama dari total 384 dimensi dimodelkan sebagai berikut:

$$v_1 = [0{,}1897, \quad 0{,}0813, \quad -0{,}2497, \quad -0{,}0704, \quad -0{,}1094, \quad \ldots]_{1\times 384}$$

Langkah perhitungan yang sama diterapkan secara independen oleh sistem terhadap Kalimat 2 (v_2) dan Kalimat 3 (v_3) sehingga menghasilkan nilai luaran ruang vektor sebagai berikut:

$$v_2 = [-0{,}2176, \quad 0{,}0203, \quad -0{,}0594, \quad -0{,}2601, \quad -0{,}1332, \quad \ldots]_{1\times 384}$$

$$v_3 = [-0{,}1454, \quad -0{,}2273, \quad -0{,}0212, \quad -0{,}1799, \quad 0{,}1482, \quad \ldots]_{1\times 384}$$

Nilai desimal kontinu pada ruang geometris inilah yang menjadi landasan bagi model dalam mengenali kedekatan makna kontekstual antar-kalimat berita. Berdasarkan hasil matriks tersebut, algoritma akan menghitung nilai Vektor *Centroid* (C) dengan menjumlahkan nilai pada masing-masing kolom dimensi dari seluruh kalimat, lalu membaginya dengan total kalimat dalam klaster (N).

1. Hitung rata-rata Dim_1:

$$C_{Dim1} = \frac{(0{,}18970) + (-0{,}21763) + (-0{,}14541) + \cdots + (-0{,}28447) + (-0{,}03372)}{115}$$

$$= \frac{-5{,}69641}{115} \approx -0{,}049534$$

2. Hitung rata-rata Dim_2:

$$C_{Dim2} = \frac{0{,}08138 + 0{,}02031 + (-0.22732) + \cdots + (-0{,}13028) + 0{,}334426}{115}$$

$$= \frac{7{,}76606}{115} \approx 0{,}067531$$

3. Hitung rata-rata Dim_3:

$$C_{Dim3} = \frac{-0{,}24977 + (-0{,}05941) + (-0{,}02129) + \cdots + (0{,}06083) + (-0{,}04239)}{55}$$

$$= \frac{-12{,}5887}{115} \approx -0{,}109467$$

Perhitungan ini dieksekusi oleh mesin hingga dimensi ke-384, menghasilkan Vektor *Centroid* utuh (C = [-0,049534;0,067531;-0,109467; -0,124708; [...]; 0,107196; 0,108115]) yang akan menjadi acuan relevansi topik pada tahap algoritma seleksi kalimat (MMR).

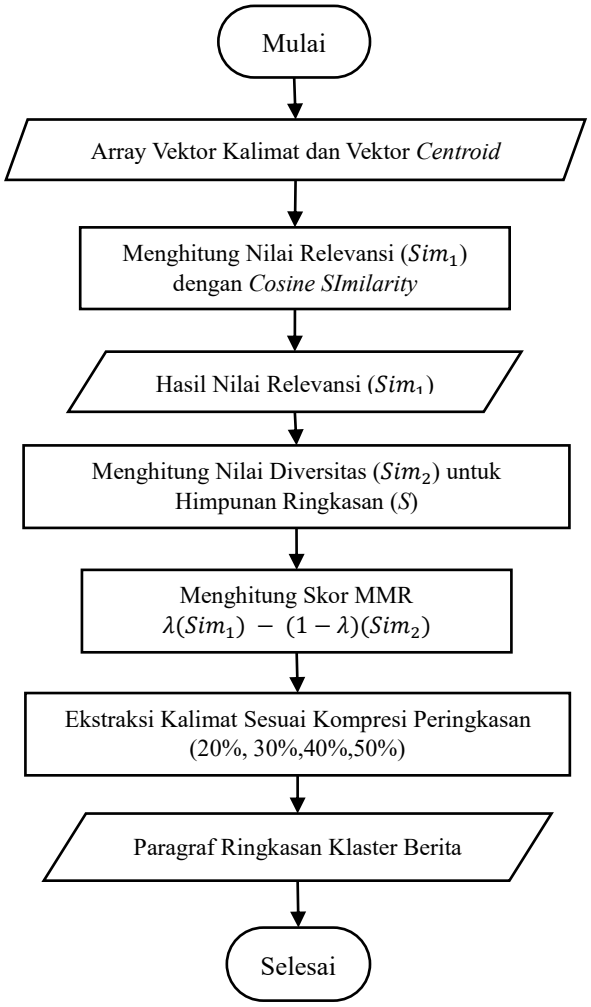
Metode usulan pada penelitian ini mengimplementasikan model *pre-trained* SBERT secara spesifik sebagai mesin pengekstraksi fitur (*feature extractor*). Oleh karena itu, komputasi bawaan di dalam arsitektur SBERT dihentikan secara absolut tepat setelah operasi *Mean Pooling* memproduksi vektor kalimat berdimensi 384. Lapisan klasifikasi lanjutan pada arsitektur asli SBERT tidak dilibatkan, melainkan matriks vektor keluaran tersebut langsung ditarik untuk diproses secara terpisah. Langkah selanjutnya menghitung nilai Vektor *Centroid* secara matematis, yang kemudian diumpankan ke dalam algoritma penyeleksi ekstraktif MMR.

3.7. Perhitungan Skor dengan MMR dan Ekstraksi Kalimat

Tahapan penentuan kalimat yang akan diekstraksi ke dalam draf ringkasan dieksekusi menggunakan algoritma *Maximal Marginal Relevance* (MMR) yang pertama kali digagas oleh Carbonell & Goldstein (1998). Menurut penelitian Tao dkk. (2025), metode yang mengombinasikan algoritma *clustering* dengan MMR secara signifikan mampu mereduksi tingkat redundansi dan meningkatkan kualitas ringkasan dibandingkan metode yang berdiri sendiri.

Algoritma ini menyeimbangkan skor Relevansi (Sim_1) dan Diversitas (Sim_2) dengan parameter *trade-off* ($\lambda/lambda$) yang ditetapkan sebesar 0,7 , ini didasarkan pada karakteristik data berita Indonesia yang menuntut tingkat relevansi tinggi terhadap inti topik guna menjaga keutuhan fakta. Pendekatan ini secara empiris divalidasi oleh penelitian Mawaridi dkk. (2024), yang membuktikan bahwa penggunaan metrik $\lambda = 0{,}7$ pada peringkasan ribuan artikel berita berbahasa Indonesia sukses menghasilkan performa evaluasi ROUGE yang paling optimal dibandingkan nilai parameter lainnya.

Proses seleksi kalimat ini dieksekusi melalui skema komputasi yang beroperasi secara iteratif (berulang). Secara keseluruhan, alur logika dari algoritma MMR termodifikasi yang dibangun pada penelitian ini divisualisasikan pada bagan/*flowchart* Gambar 3.6.



Gambar 3.6 Bagan Perhitungan *Maximum Marginal Relevance*

Iterasi pada proses perhitungan *Maximal Marginal Relevance* (MMR) adalah untuk menentukan jarak antar kalimat dengan *centroid* yang telah ditentukan sebelumnya, pada setiap kalimat teks berita.

1. Iterasi Ke-1 (Pemilihan Kalimat Pertama)

   Pada iterasi pertama, himpunan ringkasan masih kosong ($S = \emptyset$), sehingga skor diversitas max $Sim_2(S_i, S_j)$ bernilai 0,0. Perhitungan MMR difokuskan pada pengalian nilai parameter $\lambda = 0{,}7$ terhadap nilai relevansi termodifikasi:

   a. $Evaluasi\ S_1: MMR(S_1) = 0{,}7 \times 0{,}721913 = 0{,}505339$

   b. $Evaluasi\ S_2: MMR(S_2) = 0{,}7 \times 0{,}700187 = 0{,}490131$

   c. $Evaluasi\ S_3: MMR(S_3) = 0{,}7 \times 0{,}705816 = 0{,}494071$

   Berdasarkan komputasi di atas, kandidat S_1 memperoleh skor tertinggi ($0{,}505339$) dan secara resmi diekstraksi ke dalam himpunan ringkasan. Himpunan ringkasan diperbarui menjadi $S = \{S_1\}$.

2. Iterasi Ke-2 (Perhitungan Diversitas)

   Mesin akan melanjutkan siklus komputasi untuk mencari kalimat kedua dari sisa kandidat kalimat yang belum terpilih, yaitu S_2 dan S_3. Pada tahap ini, himpunan ringkasan sudah tidak lagi kosong karena S_1 telah resmi terpilih pada iterasi pertama ($S = \{S_1\}$). Selanjutnya komponen diversitas/relevansi MMR, yakni $(1 - \lambda) = 0{,}3$, mulai beroperasi. Langkah pertama pada iterasi ini adalah menghitung nilai kemiripan atau diversitas (Sim_2) antara masing-masing kalimat kandidat (S_2) dan (S_8) terhadap kalimat yang sudah ada di dalam himpunan ringkasan (Sim_1)). Nilai diversitas ini dihitung menggunakan metrik *Cosine Similarity* antara representasi vektor kalimat berdimensi 384 yang sebelumnya telah dipetakan oleh model *Sentence-BERT* (SBERT). Secara matematis, tingkat kemiripan antara vektor kalimat kandidat (S_i) terhadap vektor kalimat yang sudah terpilih (S_j) dihitung menggunakan rumus perkalian titik (*dot product*) yang dibagi dengan perkalian magnitudo dari kedua vektor, yang secara komputasional dihitung oleh model SBERT yang memetakan dimensi untuk setiap kalimat. Semakin tinggi nilai komputasi kosinusnya (mendekati angka 1), maka semakin identik makna

semantik dari kedua kalimat tersebut, yang mengindikasikan tingginya tingkat redundansi.

Berdasarkan hasil ekstraksi komputasi mesin peringkas (hasil perhitungan model SBERT), nilai jarak kemiripan untuk kandidat pada iterasi kedua adalah sebagai berikut:

1. Jarak kemiripan makna S_2 terhadap S_1 bernilai tinggi:

$$Sim_2(S_2, S_1) = 0{,}422936$$

2. Jarak kemiripan makna S_3 terhadap S_1 bernilai rendah:

$$Sim_2(S_3, S_1) = 0{,}527220$$

Selanjutnya, algoritma akan memasukkan nilai-nilai tersebut ke dalam persamaan MMR secara utuh. Perhitungan berupa Skor Relevansi dikurangi dengan Skor Redundansi di atas:

1. Evaluasi Kandidat S_2:

   Relevansi: $0{,}7 \times 0{,}700187 = 0{,}490131$

   Redundansi: $0{,}3 \times 0{,}422936 = 0{,}126881$

   Skor Akhir MMR S_2:

$$MMR(S_2) = 0{,}490131 - 0{,}126881 \approx 0{,}363250$$

2. Evaluasi Kandidat S_3:

   Relevansi: $0{,}7 \times 0{,}705816 = 0{,}494071$

   Redundansi: $0{,}3 \times 0{,}527220 = 0{,}158166$

   Skor Akhir MMR S_3:

$$MMR(S_3) = 0{,}494071 - 0{,}158166 \approx 0{,}335905$$

Hasil komputasi Iterasi 2 menunjukkan perbandingan skor akhir antara $S_2(0{,}363250)$ dan $S_3(0{,}335905)$. Fenomena ini membuktikan efektivitas algoritma MMR murni dalam mencegah redundansi informasi. Meskipun kandidat S_3 memiliki modal relevansi awal (Sim_1) yang lebih tinggi dibandingkan $S_2(0{,}705816 > 0{,}700187)$, kalimat S_3 mendapatkan penalti pengurangan yang jauh lebih masif karena isi kalimatnya terlalu mirip secara semantik dengan kalimat S_1 yang sudah terpilih sebelumnya. Oleh karena itu, S_2 memenangkan iterasi kedua, dan himpunan ringkasan diperbarui menjadi $S = \{S_1, S_2\}$.

Siklus komputasi matematika ini dieksekusi oleh mesin secara terus-menerus dan berulang terhadap puluhan kalimat kandidat lainnya hingga batas persentase kompresi dokumen yang ditentukan terpenuhi.

3.8.Evaluasi Kinerja Metode Usulan dengan ROUGE

Pengujian dan evaluasi kualitas paragraf ringkasan yang dihasilkan oleh model peringkasan multi-dokumen secara kuantitatif menjadi tahap terakhir dalam metodologi penelitian ini. Evaluasi model turunan *Natural Language Generation* (NLG) ini dilakukan dengan membandingkan tingkat irisan teks antara keluaran mesin (*System Summary*) dan ringkasan rujukan buatan manusia (*Human Ground Truth*), merujuk pada landasan teoretis evaluasi teks (Sai dkk., 2022).

Metrik ROUGE berfungsi sebagai standar pengukuran utama. Tiga varian metrik ROUGE yang dilaporkan melalui *library rouge-score* menjadi fokus evaluasi, yaitu ROUGE-1 (kata/*unigram*), ROUGE-2 (frasa/*bigram*), dan ROUGE-L (struktur urutan kalimat). Proses evaluasi menghitung tiga komponen utama untuk masing-masing varian ROUGE guna memastikan objektivitas pengujian:

1. *Recall* (R) : Seberapa banyak informasi penting dari *Ground Truth* yang berhasil ditangkap oleh model.
2. *Precision* (P) : Seberapa banyak kata yang dihasilkan mesin yang benar-benar relevan dengan *Ground Truth*.
3. *F1-Score* (F) : Nilai rata-rata harmonik yang menyeimbangkan *Recall* dan *Precision*.

Nilai *F1-Score* ditetapkan sebagai parameter penilaian utama dalam penelitian ini untuk merepresentasikan performa akhir model peringkasan. Karakteristik evaluasi peringkasan ekstraktif dengan variasi alokasi panjang teks keluaran pada setiap tingkat kompresi (20% hingga 50%) mendasari keputusan pemilihan parameter tersebut. Ringkasan dengan tingkat ekstraksi terlalu panjang (misalnya 50%) cenderung menghasilkan nilai *Recall* tinggi akibat besarnya probabilitas irisan kata. Kondisi tersebut memicu penurunan drastis pada nilai *Precision* akibat banyaknya kalimat sekunder (*noise*) lintas dokumen yang ikut terekstraksi. Ringkasan berukuran sangat pendek (misalnya 20%) sebaliknya berpotensi menghasilkan nilai *Precision* tinggi dengan risiko kehilangan banyak informasi esensial yang menyebabkan rendahnya skor *Recall*. *F1-Score* bertindak sebagai rata-rata harmonik

yang menghukum ketimpangan ekstrem antara kedua komponen tersebut. Penggunaan metrik ini menjamin bahwa model peringkasan hanya akan mendapatkan skor tinggi apabila mampu menyajikan informasi esensial secara maksimal sekaligus meminimalisasi ekstraksi kata-kata yang tidak relevan.

Anotator manusia, yaitu 2 orang mahasiswa informatika selaku peringkas manual mengeksekusi proses peringkasan teks secara komprehensif pada setiap klaster berita menggunakan pendekatan ekstraktif guna merangkum inti peristiwa secara murni tanpa intervensi opini pribadi. Teks peringkasan manual hasil standardisasi para anotator ini kemudian berfungsi sebagai acuan evaluasi utama atau *human ground truth*. Perbandingan representatif antara luaran model peringkasan dengan hasil peringkasan manusia pada salah satu sampel klaster topik "Tekno" disajikan secara detail pada Tabel 3.15.

Tabel 3.15 Contoh Hasil Peringkasan Manusia sebagai *Ground Truth*

Hasil Peringkasan Manusia (Human Ground Truth)
Di tengah perlombaan global mengadopsi kecerdasan buatan (AI), talenta Indonesia mulai mengambil peran krusial di pusat inovasi dunia. Product Manager di Google Amerika Serikat, Juan Anugraha Djuwadi, kini menjadi salah satu aktor strategis yang mengarahkan bagaimana teknologi AI dikembangkan agar tetap manusiawi dan berdampak luas bagi pengguna global. Juan, yang merupakan alumnus Columbia University, menegaskan bahwa pengguna akhir seringkali tidak mempedulikan kerumitan algoritma di balik sebuah aplikasi. “Pengguna tidak peduli seberapa canggih teknologi di belakang layar. Mereka peduli apakah solusi itu berguna dan menyelesaikan masalah nyata,” ujar Juan. Kegagalan sistem sebesar satu persen saja dapat berdampak pada jutaan orang. Terkait proses pengambilan keputusan, Juan menawarkan pendekatan moderat antara data dan intuisi. “Data memvalidasi masa kini, namun intuisi mendefinisikan masa depan,” ungkapnya. Juan melihat adanya perbedaan fase kematangan antara pasar Amerika Serikat (AS) dan Indonesia. Jika AS sudah mapan dalam hal monetisasi perangkat lunak, Indonesia masih berada dalam tahap transisi yang dinamis. Ia memprediksi, seiring berkembangnya ekosistem digital di Indonesia, isu mengenai privasi data, etika AI, dan akuntabilitas akan menjadi agenda utama yang tak terelakkan. Adopsi kecerdasan buatan (AI) di kawasan ASEAN terus menunjukkan tren positif. Data terbaru menunjukkan 67 persen organisasi di kawasan ASEAN+ sudah melakukan uji coba AI secara sistematis. “Di lingkungan ASEAN sendiri, strategic AI yang berbasis public money itu sudah nggak praktis,” ujar Budi Janto. Seiring dengan hal itu, fokus perusahaan di ASEAN+ mulai beralih ke Agentic AI. Tercatat peningkatan fokus pada Agentic AI mencapai 91 persen secara tahunan. “Hanya ada 11 persen ini cukup gede dari jumlah company di ASEAN,” ungkap Budi. Terdapat tiga tantangan utama menurut para CIO di ASEAN+, yakni penerapan AI yang dinilai masih belum optimal, keamanan data, serta kualitas data yang belum memadai. Lenovo juga memperkenalkan CIO Playbook 2026. “Tahun lalu kita banyak membahas ekonomi AI. Diskusinya masih berada di tahap uji coba dan pilot. Namun kini, prioritas bisnis telah bergeser ke bagaimana AI mampu mendorong peningkatan pendapatan dan profitabilitas,” ujar Thomas. Data CIO Playbook 2026 menunjukkan faktor peningkatan pendapatan melonjak tujuh peringkat dan kini menempati posisi teratas dalam prioritas bisnis. Saat ini, hampir 90 persen organisasi menyatakan lebih memilih atau tengah beralih ke model penerapan AI hibrida.

Gambaran matematis dari cara kerja evaluasi diilustrasikan melalui Tabel 3.16 dan perhitungan di bawahnya. Ilustrasi tersebut menjabarkan langkah komputasi ROUGE-1, ROUGE-2, dan ROUGE-L dengan membandingkan ringkasan mesin pada tingkat kompresi 30% melawan hasil ringkasan manusia.

Tabel 3.16 Data Hasil Ringkasan Mesin dan Hasil Ringkasan Manusia

Jenis Teks	**Jumlah Kata (Unigram)**	**Jumlah Frasa (Bigram)**	**Barisan Kata Berurutan (*LCS*)**
Ringkasan Mesin (30%)	628 kata	627 frasa	-
Ringkasan Manusia (Ground Truth)	385 kata	384 frasa	-
Jumlah Irisan yang Sama (Overlap)	227 kata	111 frasa	85 kata berurutan

Data pada Tabel 3.16 digunakan untuk menghitung nilai *Recall*, *Precision*, dan *F1-Score* pada metrik ROUGE-1 sebagai berikut:

1. Perhitungan *Recall* ROUGE-1:

$$R_{ROUGE-1} = \frac{227}{385} \approx 0{,}5896$$

2. Perhitungan *Precision* ROUGE-1:

$$P_{ROUGE-1} = \frac{227}{628} \approx 0{,}3615$$

3. Perhitungan *F1-Score* ROUGE-1:

$$F1_{ROUGE-1} = 2 \times \frac{(P_{ROUGE-1} \times R_{ROUGE-1})}{(P_{ROUGE-1} + R_{ROUGE-1})}$$

$$F1_{ROUGE-1} = 2 \times \frac{0{,}3615 \times 0{,}5896}{0{,}3615 + 0{,}5896} = \frac{0{,}426208}{0{,}9511} \approx 0{,}4482$$

Perhitungan tingkat irisan berdasarkan *bigram* untuk metrik ROUGE-2 dijabarkan sebagai berikut:

1. Perhitungan *Recall* ROUGE-2:

$$R_{ROUGE-2} = \frac{111}{384} \approx 0{,}2891$$

2. Perhitungan *Precision* ROUGE-2:

$$P_{ROUGE-2} = \frac{111}{627} \approx 0{,}1770$$

3. Perhitungan *F1-Score* ROUGE-2:

$$F1_{ROUGE-2} = 2 \times \frac{(P_{ROUGE-2} \times R_{ROUGE-2})}{(P_{ROUGE-2} + R_{ROUGE-2})}$$

$$F1_{ROUGE-2} = 2 \times \frac{0{,}1770 \times 0{,}2891}{0{,}1770 + 0{,}2891} = \frac{0{,}102341}{0{,}4661} \approx 0{,}2196$$

Metrik ROUGE-L dievaluasi sebagai perhitungan terakhir. Prinsip evaluasi turunan *n-gram* ini berfokus pada fungsi pencarian *Longest Common Subsequence* (LCS), yakni urutan deret kata terpanjang yang sama secara searah antara ringkasan rujukan (S_{ref}) dan ringkasan mesin (S_{sys}) (Sai dkk., 2022). Panjang LCS sebanyak 85 kata ditemukan pada sampel kalimat tersebut. Rincian perhitungannya adalah sebagai berikut:

1. Perhitungan *Recall* ROUGE-L:

$$R_{ROUGE-L} = \frac{LCS(S_{ref}, S_{sys})}{|S_{ref}|} = \frac{85}{385} \approx 0{,}2208$$

2. Perhitungan *Precision* ROUGE-L:

$$P_{ROUGE-L} = \frac{LCS(S_{ref}, S_{sys})}{|S_{sys}|} = \frac{85}{628} \approx 0{,}1354$$

3. Perhitungan *F1-Score* ROUGE-L:

$$F\text{-}Score_{ROUGE-L} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$= \frac{2 \times 0{,}1354 \times 0{,}2208}{0{,}1354 + 0{,}2208} = \frac{0{,}0598}{0{,}3562} = 0{,}1678$$

Algoritma pada mesin peringkas mengeksekusi komputasi perhitungan serupa secara komprehensif terhadap seluruh dataset yang memuat ratusan bermacam judul berita dari berbagai kategori. Nilai rata-rata performa akhir metode usulan pada sampel tingkat kompresi 30% mencatatkan *F1-Score* dari ROUGE-1 sebesar 0,5500, *F1-Score* dari ROUGE-2 sebesar 0,4162, dan *F1-Score* dari ROUGE-L sebesar 0,2192. Mesin peringkas menghitung skor ROUGE secara keseluruhan di setiap kategori berita melalui skenario tingkat kompresi 20%, 30%, 40%, dan 50% pada implementasi eksperimen sesungguhnya. Rincian hasil peringkasan serta rekapitulasi nilai akhir evaluasi dari seluruh klaster pengujian dipaparkan dan dianalisis secara komprehensif pada Bab IV.

BAB IV
PEMBAHASAN

Bab ini menguraikan hasil dari serangkaian tahapan penelitian yang telah dilakukan, mulai dari distribusi pengelompokan data, analisis proses peringkasan pada studi kasus spesifik, hingga evaluasi akhir performa mesin peringkas. Tujuan dari bab ini adalah untuk membuktikan efektivitas kombinasi model *Sentence-BERT* (SBERT), dengan penerapan algoritma *Maximal Marginal Relevance* (MMR) dalam ekstraksi kalimat.

4.1. Lingkungan Pengembangan Model Peringkasan

Lingkungan proses dibagi menjadi dua bagian utama, yaitu lingkungan eksperimental yang digunakan untuk pengembangan serta pengujian performa algoritma secara mendalam, dan lingkungan implementasi yang berfungsi sebagai wadah fungsional untuk menjalankan model dalam skenario penggunaan di dunia nyata.

4.1.1. Lingkungan Pengembangan Model

Penelitian ini dilakukan menggunakan Google *Colab* dengan bahasa pemrograman Python untuk eksperimen data dan pengujian algoritma dengan spesifikasi sebagai berikut:

CPU	:	Intel(R) Core i5-11400H @ 2.70Hz (12 CPU)
RAM	:	16 GB
GPU	:	NVIDIA Tesla T4
GPU RAM	:	15 GB
DISK	:	112 GB
Library NLP	:	*sentence-transformers* (untuk SBERT), *scikit-learn* (untuk TF-IDF dan *Agglomerative Clustering*), dan *rouge-score* (untuk evaluasi).

4.1.2. Lingkungan Pengembangan Prototipe (*User Interface*)

Penelitian ini diimplementasikan ke dalam sebuah *pipeline* komputasi yang terintegrasi dengan antarmuka web. Unit implementasi ini diberi nama “Berinkin”, yang merupakan akronim dari “Berita Ringkasan Terkini”. Pemilihan nama ini didasari oleh filosofi pohon beringin, di mana ribuan cabang dan daun (merepresentasikan dokumen berita yang redundan) diproses secara cerdas dan dikonvergensikan menjadi satu batang utama

yang kokoh (merepresentasikan satu ringkasan tunggal yang esensial). Adapun spesifikasi lingkungan implementasi yang digunakan adalah sebagai berikut:

Antarmuka (*Client-side)*	:	*framework Next.js* (*TypeScript*) dan *Tailwind CSS*
Logika Komputasi (*Server-side*)	:	*FastAPI* (*Python*)
Database	:	MySQL
Develompent & *Deployment*	:	*Visual Studio Code* , *Vercel*

4.2. Implementasi Model Peringkasan dengan Metode Usulan

Pembangunan arsitektur model peringkasan direalisasikan melalui bahasa pemrograman Python dengan memanfaatkan ragam pustaka pemrosesan bahasa alami (*Natural Language Processing*). Tahapan komputasi ini mengonversi rancangan teoretis pada metodologi penelitian menjadi sebuah *pipeline* program terintegrasi yang berjalan secara sistematis pada Google *Colab*.

Proses eksekusi model diawali dengan memilah dokumen berita mentah menggunakan algoritma *Hierarchical Agglomerative Clustering* berdasarkan kedekatan judul untuk mengisolasi klaster topik yang seragam. Dokumen di dalam klaster valid selanjutnya melewati fungsi pra-pemrosesan teks (*preprocessing*) menggunakan fungsi reguler ekspresi (re) untuk membersihkan unsur *noise*, memotong paragraf menjadi larik kalimat tunggal (*sentence splitting*), serta menyamakan format huruf kecil (*case folding*). Berikut kalimat bersih tersebut kemudian ditransformasikan menjadi representasi numerik (*dense embedding*) berdimensi 384 menggunakan model *pre-trained Sentence*-BERT (SBERT) varian *paraphrase-multilingual-MiniLM-L12-v2*, sekaligus menghitung nilai rata-rata ruang vektor sebagai titik koordinat pusat klaster (*centroid*).

Tahapan akhir berfokus pada pengoperasian algoritma *Maximal Marginal Relevance* (MMR) secara iteratif dengan parameter penyeimbang λ sebesar 0,7. Algoritma mengevaluasi setiap kandidat kalimat dengan menyeimbangkan skor kemiripan kosinus terhadap *centroid* sebagai representasi relevansi, serta memberikan penalti tegas terhadap kalimat yang telah terpilih sebagai representasi diversitas informasi. Proses perulangan komputasi ini menghasilkan urutan bobot nilai terbaik yang secara otomatis menyusun satu kesatuan paragraf ekstraktif yang padat fakta dan bebas dari redundansi leksikal. Tabel 4.1

menyajikan contoh hasil akhir kuantifikasi perankingan kalimat berdasarkan perolehan skor nilai MMR tertinggi (pada tingkat kompresi 20%) dari salah satu klaster dengan judul "Menangkar Masa Depan AI" pada kategori "Tekno" yang siap diumpankan ke dalam metrik evaluasi.

Tabel 4.1 Hasil Akhir Skor MMR Berdasarkan 20% Kompresi Peringkasan

No.	ID	MMR Score	Relevansi	Diversitas	Teks
1.	[doc_2, sent_51]	0,505339	0,721913	0,000000	Dalam skala yang lebih luas, Lenovo mengidentifikasi sejumlah faktor kunci keberhasilan penerapan AI.
2.	[doc_2, sent_5]	0,363250	0,700187	0,422936	Perusahaan mulai mencari cara agar teknologi ini benar-benar berdampak bagi bisnis.
3.	[doc_3, sent_1]	0,335905	0,705816	0,527220	Tren kecerdasan buatan (AI) semakin merambah ke segmensmartphone entry-leveldan menengah.
4.	[doc_3, sent_1]	0,335905	0,705816	0,527220	Tren kecerdasan buatan (AI) semakin merambah ke segmensmartphone entry-leveldan menengah.
5.	[doc_2, sent_59]	0,328440	0,717482	0,579326	Perangkat-perangkat yang kini dibawa Lenovo ke pasar telah memiliki kapabilitas AI.
6.	[doc_1, sent_16]	0,309667	0,650407	0,485395	Jika AS sudah mapan dalam hal monetisasi perangkat lunak, Indonesia masih berada dalam tahap transisi yang dinamis.
7.	[doc_2, sent_48]	0,307778	0,693046	0,591178	Dari 2025 ke 2026, tahap implementasi dan uji coba AI meningkat menjadi 66 persen.
8.	[doc_1, sent_14]	0,296221	0,678077	0,594778	Pendekatan ini menjadi krusial bagi para regulator dan pemimpin perusahaan di Asia Pasifik (APAC) agar tidak hanya bersikap reaktif terhadap tren, tetapi mampu menjadi visioner dalam merancang ekosistem digital.
9.	[doc_2, sent_6]	0,292543	0,629039	0,492613	Bukan lagi soal membangun model AI, tetapi bagaimana mengintegrasikannya dengan perangkat, infrastruktur, dan sistem sudah berjalan.
10.	[doc_2, sent_58]	0,292297	0,660792	0,567525	Sejalan dengan hal ini, prioritas investasi TI pun bergeser ke arah penerapan perangkay AI guna meningkatkan produktivitas dan mendukung proses inferensi AI secara lokal.
11.	[doc_2, sent_60]	0,291337	0,663600	0,577276	Artinya, perangkat tersebut memiliki kemampuan dan kinerja yang diperlukan untuk menjalankan model AI secara lokal di sisi edge .
12.	[doc_2, sent_31]	0,285604	0,678678	0,631570	Melalui riset ini, Lenovo memetakan kesiapan, kapabilitas, dan prioritas perubahan dalam mengadipsi AI di lingkungan bisnis.

No.	ID	MMR Score	Relevansi	Diversitas	Teks
13.	[doc_2, sent_1]	0,277077	0,639850	0,569392	Adopsi kecerdasan buatan (AI) di kawasan ASEAN terus menunjukkan tren positif.
14.	[doc_1, sent_22]	0,275358	0,662709	0,628462	Mereka adalah kontributor strategis yang turut menentukan arah teknologi masa depan yang lebih beretika dan berorientasi pada kepentingan manusia.
15.	[doc_1, sent_4]	0,273084	0,617614	0,530818	Baginya, kunci utama efektivitas AI terletak pada kegunaannya dalam menjawab persoalan nyata di masyarakat.
16.	[doc_1, sent_12]	0,268760	0,617771	0,545599	Ia mengibaratkan data sebagai kompas untuk optimasi efisiensi dan akurasi, namun terobosan besar tetap membutuhkan visi manusia.
17.	[doc_2, sent_13]	0,261964	0,584686	0,491054	Jadi hybrid AI itu bukan pilihan, tapi itu sudah keharusan bagi perusahaan-perusahaan yang di ASEAN, ujarBudi Janto, di Jakarta, baru-baru ini.
18.	[doc_1, sent_3]	0,260887	0,636507	0,615560	Dalam webinar bertajuk AI Streamline Your Business: Build Internal Apps with AI baru-baru ini, Juan membedah peta jalan pengembangan AI yang melampaui sekadar kecanggihan teknis.
19.	[doc_2, sent_43]	0,256634	0,649618	0,660327	Pada 2026 fokus AI didominasi oleh pendekatan berbasis hasil (outcome-based), yang mengisi tiga posisi teratas dalam strategi perusahaan.
20.	[doc_3, sent_19]	0,254324	0,583649	0,514102	Fitur-fitur ini mencakup "Produktivitas Sekali Ketuk" untuk membantu pengguna membuat daftar tugas, catatan, dan menjalankan perintah kerja ringan dengan cepat.
21.	[doc_3, sent_36]	0,254086	0,613007	0,583394	Infinix Smart 10 Plus hadir dengan pengalaman kecerdasan buatan (AI) yang menonjol, salah satu daya tariknya terletak pada kehadiran Folax, asisten virtual berbasis AI.

4.3. Pengujian Kinerja Model Peringkasan

Pengujian kinerja model mesin peringkas dilakukan dengan eksperimen dari data yang diinginkan dan dilakukan pengujian dengan beberapa eksperimen. Data yang digunakan dan skenario eksperimen yang akan dilakukan dijelaskan pada subbab selanjutnya.

4.3.1 Data Uji

Data uji dalam penelitian ini merupakan sekumpulan data yang telah melalui tahap pra-pemrosesan dan siap digunakan dalam seluruh rangkaian pengujian model. Secara fungsional, data uji ini dibagi menjadi dua bagian penting. Pertama, kumpulan judul dan isi

berita mentah yang digunakan sebagai input untuk tahap pengelompokan (*clustering*) dengan jumlah total 50 artikel berita disetiap kategori (Tekno, Bola, Bisnis, Kesehatan, Saham, Otomotif) dengan total keseluruhan data berjumlah 300 judul berita. Judul berita di sini memegang peranan kunci sebagai fitur utama untuk memetakan kesamaan topik antar dokumen.

Kedua, data berupa klaster-klaster berita yang sudah terorganisir dan siap diproses oleh algoritma peringkasan. Pada bagian ini, setiap klaster telah dilengkapi dengan teks *human ground truth* yang disusun secara manual oleh anotator. Teks ringkasan manual tersebut berfungsi sebagai standar emas untuk mengukur tingkat akurasi dan kualitas ringkasan yang dihasilkan oleh mesin melalui perhitungan skor ROUGE, sehingga performa model mesin peringkasan dapat divalidasi secara objektif meggunakan *cohens kappa* dengan persentase kesepakatannya berdasarkan perbandingan langsung antara teks berita asli dan hasil ringkasan manusia.

4.3.2 Skenario Eksperimen

Pengujian performa model pada metode usulan dilakukan terhadap seluruh klaster dokumen valid yang berhasil terbentuk dari *dataset* portal berita daring. Penerapan parameter Lambda (λ) ditetapkan secara statis pada nilai 0,7 guna mengukur konsistensi serta kualitas algoritma MMR dalam mengeksploitasi seleksi kalimat. Nilai konstan tersebut merepresentasikan pembobotan sebesar 70% untuk aspek relevansi kalimat terhadap titik pusat topik (*centroid*) dan 30% untuk aspek diversitas atau keberagaman informasi. Sementara itu, parameter tingkat kompresi peringkasan divariasikan ke dalam empat level, yaitu 20%, 30%, 40%, dan 50%. Skema variasi kompresi ini bertujuan untuk menemukan titik optimal ukuran panjang ringkasan. Melalui batasan tersebut, metode usulan diproyeksikan mampu menyajikan intisari informasi yang padat fakta dengan tingkat redundansi leksikal paling minimal.

4.4. Hasil Uji *Cohen's Kappa* pada Ringkasan Manusia

Uji reliabilitas antar-anotator (*Inter-Annotator Agreement*) dilakukan pada seluruh klaster (58 klaster) di enam kategori berita, guna memastikan objektivitas dan meminimalisir bias personal pada ringkasan rujukan tersebut. Pengujian ini menggunakan dua metrik evaluasi untuk saling melengkapi, yaitu *Cohen's Kappa* dan *Percentage Agreement*

(Persentase Kesepakatan Aktual). Hasil rekapitulasi validasi untuk seluruh kategori disajikan pada Tabel 4.2 berikut.

Tabel 4.2 Hasil Perhitungan *Cohen's Kappa* Kategori Data Uji *Ground Truth*

No.	Kategori	Jumlah Klaster	Rata-rata-*Cohen's Kappa*	Rata-rata *Agreement Rate*
1.	Kesehatan	11	0,2835	67,92%
2.	Otomotif	7	0,2241	60,53%
3.	Bisnis	13	0,2186	56,89%
4.	Saham	8	0,2021	68,13%
5.	Bola	11	0,1047	63,18%
6.	Tekno	8	0,0933	62,59%
	Rata-rata Keseluruhan	58	0,1877	63,21%

Rata-rata skor *Cohen's Kappa* sebesar 0,1877 yang ditunjukkan pada Tabel 4.2 mencerminkan fenomena *Kappa Paradox*. Rendahnya skor ini dipicu oleh ketidakseimbangan kelas (*class imbalance*) yang ekstrem antara jumlah kalimat yang dipilih dan dibuang, serta tingginya redundansi informasi yang memicu subjektivitas leksikal antar-anotator (Röttger dkk., 2022). Mengingat keterbatasan *Kappa* dalam menangani distribusi data yang asimetris, *Percentage Agreement* digunakan sebagai ukuran reliabilitas yang lebih representatif.

Persentase kesepakatan ini didapatkan melalui perhitungan akurasi murni, yaitu rasio antara jumlah kalimat yang disepakati oleh kedua anotator (baik untuk diekstraksi maupun dibuang) terhadap total keseluruhan kalimat dalam suatu klaster. Dominasi jumlah kalimat yang dibuang dalam tugas peringkasan ekstraktif menyebabkan rumus *Cohen's Kappa* memperhitungkannya sebagai peluang kebetulan yang tinggi, sehingga menekan skor akhir secara matematis. Padahal secara faktual, tingkat kesepakatan aktual antar-anotator berada di atas 60%.

Hasil pengujian menunjukkan bahwa rata-rata persentase kesepakatan (*Agreement Rate*) mencapai 63,21%, dengan kategori Kesehatan menyentuh tingkat kesepakatan tertinggi, dan Saham berada di urutan kedua. Tingkat kesepakatan aktual di atas 60% dalam tugas peringkasan ekstraktif manual tergolong wajar, realistis, dan dapat diterima dengan baik. Hal ini membuktikan bahwa kedua anotator memiliki konsensus pemahaman yang kuat dan stabil mengenai intisari topik yang harus dipertahankan. Dengan demikian, himpunan

data *ground truth* ini dinyatakan sah, reliabel, dan sangat layak digunakan sebagai standar referensi utama dalam pengujian performa model peringkas menggunakan metrik ROUGE.

4.5. Hasil Peringkasan dan Evaluasi Model menggunakan ROUGE

Model mesin peringkasan menghasilkan draf ringkasan pada empat variasi tingkat kompresi, yaitu 20%, 30%, 40%, dan 50%, setelah menyelesaikan tahap pengelompokan dan penyeleksian kalimat menggunakan algoritma *Maximal Marginal Relevance* (MMR) murni. Penerapan variasi ini bertujuan untuk mengamati titik keseimbangan (*sweet spot*) antara tingkat kepadatan informasi dan derajat redundansi teks.

Bagian ini membedah satu studi kasus pada berita kategori "Tekno" klaster topik "Menangkar Masa Depan AI" guna memberikan gambaran empiris secara kualitatif (keterbacaan teks) maupun kuantitatif (skor evaluasi). Tabel 4.3 menampilkan perbandingan teks keluaran mesin beserta nilai *F1-Score* ROUGE pada skenario kompresi 20% terhadap *Human Ground Truth*.

Tabel 4.3 Contoh Hasil Peringkasan pada Klaster Kategori "Tekno" dengan Kompresi 20%

Hasil Teks Ringkasan Keluaran Mesin Peringkas (SBERT + MMR Murni) Mesin 20%	**Skor Evaluasi (*F1-Score*)**
Dalam skala yang lebih luas, Lenovo mengidentifikasi sejumlah faktor kunci keberhasilan penerapan AI. Perusahaan mulai mencari cara agar teknologi ini benar-benar berdampak bagi bisnis. Tren kecerdasan buatan (AI) semakin merambah ke segmensmartphone entry-leveldan menengah. Perangkat-perangkat yang kini dibawa Lenovo ke pasar telah memiliki kapabilitas AI. Indonesia masih berada dalam tahap transisi yang dinamis. Dari 2025 ke. 2026, tahap implementasi dan uji coba AI meningkat menjadi 66 persen. Pendekatan ini menjadi krusial bagi para regulator dan pemimpin perusahaan di Asia Pasifik (APAC) agar tidak hanya bersikap reaktif terhadap tren, tetapi mampu menjadi visioner dalam merancang ekosistem digital. Bukan lagi soal membangun model AI, tetapi bagaimana mengintegrasikannya dengan perangkat , infrastruktur, dan sistem sudah berjalan. Sejalan dengan hal ini, prioritas investasi TI pun bergeser ke arah penerapan perangkay AI guna meningkatkan produktivitas dan mendukung proses inferensi AI secara lokal. Artinya, perangkat tersebut memiliki kemampuan dan kinerja yang diperlukan untuk menjalankan model AI secara lokal di sisi edge . Melalui riset ini, Lenovo memetakan kesiapan, kapabilitas, dan prioritas perubahan dalam mengadipsi AI di lingkungan bisnis. Adopsi kecerdasan buatan (AI) di kawasan ASEAN terus menunjukkan tren positif. Mereka adalah kontributor strategis yang turut menentukan arah teknologi masa depan yang lebih beretika dan berorientasi pada kepentingan manusia.	ROUGE-1: 0,4762 ROUGE-2: 0,2889 ROUGE-L: 0,1734

Hasil Teks Ringkasan Keluaran Mesin Peringkas (SBERT + MMR Murni) Mesin 20%	Skor Evaluasi (*F1-Score*)
Baginya, kunci utama efektivitas AI terletak pada kegunaannya dalam menjawab persoalan nyata di masyarakat. Ia mengibaratkan data sebagai kompas untuk optimasi efisiensi dan akurasi, namun terobosan besar tetap membutuhkan visi manusia. Jadi hybrid AI itu bukan pilihan, tapi itu sudah keharusan bagi perusahaan-perusahaan yang di ASEAN, ujarBudi Janto, di Jakarta, baru-baru ini. Dalam webinar bertajuk AI Streamline Your Business: Build Internal Apps with AI baru-baru ini, Juan membedah peta jalan pengembangan AI yang melampaui sekadar kecanggihan teknis. Pada 2026 fokus AI didominasi oleh pendekatan berbasis hasil (outcome-based), yang mengisi tiga posisi teratas dalam strategi perusahaan. Fitur-fitur ini mencakup "Produktivitas Sekali Ketuk" untuk membantu pengguna membuat daftar tugas, catatan, dan menjalankan perintah kerja ringan dengan cepat. Infinix Smart 10 Plus hadir dengan pengalaman kecerdasan buatan (AI) yang menonjol, salah satu daya tariknya terletak pada kehadiran Folax, asisten virtual berbasis AI. Berkat kelincahan dan kematangan tata kelola di kawasan ASEAN menjadikannya sebagai lingkungan uji coba yang ideal untuk Agentic AI. Seiring dengan hal itu, fokus perusahaan di ASEAN+ mulai beralih ke Agentic AI. Penerapan AI Hybrid turut dipengaruhi oleh tingginya biaya pengelolaan data di cloud, serta kekhawatiran terkait keamanan dan privasi data	

Tingkat kompresi 20% pada Tabel 4.2 menunjukkan kecenderungan mesin dalam mengekstrak kalimat-kalimat dengan jarak kosinus terdekat terhadap *centroid* topik secara sangat selektif. Teks yang dihasilkan sangat padat, namun beberapa bagian konteks terasa terpotong karena mesin memprioritaskan kalimat utama dengan bobot relevansi tertinggi tanpa memperluas informasi pendukung. Keterbatasan tersebut tercermin dari *F1-Score* ROUGE-2 dan *F1-Score* ROUGE-L yang relatif lebih rendah dibandingkan tingkat kompresi selanjutnya. Hasil peringkasan pada tingkat kompresi 30% selanjutnya disajikan pada Tabel 4.4.

Tabel 4.4 Contoh Hasil Peringkasan pada Klaster Kategori “Tekno” dengan Kompresi 30%

Hasil Teks Ringkasan Keluaran Mesin Peringkas (SBERT + MMR Murni) Mesin 30%	Skor Evaluasi (*F1-Score*)
Dalam skala yang lebih luas, Lenovo mengidentifikasi sejumlah faktor kunci keberhasilan penerapan AI. Perusahaan mulai mencari cara agar teknologi ini benar-benar berdampak bagi bisnis. Tren kecerdasan buatan (AI) semakin merambah ke segmensmartphone entry-leveldan menengah. Perangkat-perangkat yang kini dibawa Lenovo ke pasar telah memiliki kapabilitas AI. Jika AS sudah mapan dalam hal monetisasi perangkat lunak, Indonesia masih berada dalam tahap transisi yang dinamis.	ROUGE-1: 0,6012 ROUGE-2: 0,4145 ROUGE-L: 0,2645

Hasil Teks Ringkasan Keluaran Mesin Peringkas (SBERT + MMR Murni) Mesin 30%	Skor Evaluasi (*F1-Score*)
Dari 2025 ke 2026, tahap implementasi dan uji coba AI meningkat menjadi 66 persen. Pendekatan ini menjadi krusial bagi para regulator dan pemimpin perusahaan di Asia Pasifik (APAC) agar tidak hanya bersikap reaktif terhadap tren, tetapi mampu menjadi visioner dalam merancang ekosistem digital. Bukan lagi soal membangun model AI, tetapi bagaimana mengintegrasikannya dengan perangkat, infrastruktur, dan sistem sudah berjalan. Sejalan dengan hal ini, prioritas investasi TI pun bergeser ke arah penerapan perangkay AI guna meningkatkan produktivitas dan mendukung proses inferensi AI secara lokal. Artinya, perangkat tersebut memiliki kemampuan dan kinerja yang diperlukan untuk menjalankan model AI secara lokal di sisi edge. Melalui riset ini, Lenovo memetakan kesiapan, kapabilitas, dan prioritas perubahan dalam mengadipsi AI di lingkungan bisnis. Adopsi kecerdasan buatan (AI) di kawasan ASEAN terus menunjukkan tren positif. Mereka adalah kontributor strategis yang turut menentukan arah teknologi masa depan yang lebih beretika dan berorientasi pada kepentingan manusia. Baginya, kunci utama efektivitas AI terletak pada kegunaannya dalam menjawab persoalan nyata di masyarakat. Ia mengibaratkan data sebagai kompas untuk optimasi efisiensi dan akurasi, namun terobosan besar tetap membutuhkan visi manusia. Jadi hybrid AI itu bukan pilihan, tapi itu sudah keharusan bagi yang di ASEAN, ujarBudi Janto, di Jakarta, baru-baru ini. Dalam webinar bertajuk AI Streamline Your Business: Build Internal Apps with AI baru-baru ini, Juan membedah peta jalan pengembangan AI yang melampaui sekadar kecanggihan teknis. Pada 2026 fokus AI didominasi oleh pendekatan berbasis hasil (outcome-based), yang mengisi tiga posisi teratas dalam strategi perusahaan. Fitur-fitur ini mencakup "Produktivitas Sekali Ketuk" untuk membantu pengguna membuat daftar tugas, catatan, dan menjalankan perintah kerja ringan dengan cepat. Infinix Smart 10 Plus hadir dengan pengalaman kecerdasan buatan (AI) yang menonjol, salah satu daya tariknya terletak pada kehadiran Folax, asisten virtual berbasis AI. Berkat kelincahan dan kematangan tata kelola di kawasan ASEAN menjadikannya sebagai lingkungan uji coba yang ideal untuk Agentic AI. Seiring dengan hal itu, fokus perusahaan di ASEAN+ mulai beralih ke Agentic AI. Penerapan AI Hybrid turut dipengaruhi oleh tingginya biaya pengelolaan data di cloud, serta kekhawatiran terkait keamanan dan privasi data. Prioritas kedua adalah reinvensi bisnis berbasis AI. Product Manager di Google Amerika Serikat, Juan Anugraha Djuwadi, kini menjadi salah satu aktor strategis yang mengarahkan bagaimana teknologi AI dikembangkan agar tetap manusiawi dan berdampak luas bagi pengguna global. Kesenjangan kesiapan ini rupanya masih menunjukkan adanya tantangan mendasar dalam penerapan AI. Ia memperkirakan lahirnya erasoftware on-the-fly, di mana perangkat lunak dihasilkan secarareal-timesesuai kebutuhan instan pengguna,	

Hasil Teks Ringkasan Keluaran Mesin Peringkas (SBERT + MMR Murni) Mesin 30%	Skor Evaluasi (*F1-Score*)
baik melalui antarmuka teks maupun suara. Prinsip ini diterjemahkan ke dalam dua filosofi kunci:less is moredanthe details matter. Menurutnya, dalam skala pengguna miliaran seperti di Google, detail sekecil apa pun menjadi isu strategis. Data memvalidasi masa kini, namun intuisi mendefinisikan masa depan, ungkapnya. Beberapa diantaranya yaitu: Jika melihat keseluruhan faktor tersebut, inilah alasan mengapa terjadi pergeseran menuju penerapan AI berbasis hibrida. Adopsi AI tidak lagi terbatas pada fungsi IT seperti analitik data, keamanan siber dan pengembangan perangkat lunak, tetapi juga mengarah ke fungsi non-IT. Inovasi AI yang sukses di wilayah dengan keberagaman sosial tinggi seperti Indonesia dan APAC haruslah bersifat kontekstual dan berbasis pada kepercayaan. com/ Agustin Setyo Wardani) Redmi 14 C tampil dengan peningkatan signifikan pada sektor kamera berkat dukungan kamera utama 50 MP dan teknologi AI. Data terbaru yang terungkap dalam konferensi pers Lenovo Tech Day 2026 menunjukkan 67 persen organisasi di kawasan ASEAN+ sudah melakukan uji coba AI secara sistematis.	

Data pada Tabel 4.3 membuktikan keberhasilan mesin dalam meraih nilai *F1-Score* tertinggi pada tingkat kompresi 30% untuk seluruh metrik, khususnya ROUGE-1 (0,5500) dan ROUGE-2 (0,4162). Kalimat yang ditarik oleh mesin pada ambang batas ini secara kualitatif merupakan inti pembahasan strategi bisnis AI, seperti peralihan ke "*Hybrid AI*" dan "*Agentic AI*". Titik paling optimal pada kepadatan informasi dan koherensi frasa tercapai karena mesin mampu memfilter informasi dengan ketat tanpa kehilangan konteks utama yang krusial. Pengujian selanjutnya dilakukan dengan meningkatkan batas kompresi menjadi 40% dan 50% sebagaimana dirangkum pada analisis Tabel 4.5.

Tabel 4.5 Contoh Hasil Peringkasan pada Klaster Kategori "Tekno" dengan Kompresi 40%

Hasil Teks Ringkasan Keluaran Mesin Peringkas (SBERT + MMR Murni) Mesin 40%	Skor Evaluasi (*F1-Score*)
Dalam skala yang lebih luas, Lenovo mengidentifikasi sejumlah faktor kunci keberhasilan penerapan AI. Perusahaan mulai mencari cara agar teknologi ini benar-benar berdampak bagi bisnis. Tren kecerdasan buatan (AI) semakin merambah ke segmensmartphone entry-leveldan menengah. Perangkat-perangkat yang kini dibawa Lenovo ke pasar telah memiliki kapabilitas AI. Jika AS sudah mapan dalam hal monetisasi perangkat lunak, Indonesia masih berada dalam tahap transisi yang dinamis. Dari 2025 ke 2026, tahap implementasi dan uji coba AI meningkat menjadi 66 persen. Pendekatan ini menjadi krusial bagi para regulator dan pemimpin perusahaan di Asia Pasifik (APAC) agar tidak hanya bersikap reaktif terhadap tren, tetapi mampu menjadi visioner dalam merancang ekosistem digital. Bukan lagi soal membangun model AI, tetapi bagaimana	ROUGE-1: 0,5772 ROUGE-2: 0,4074 ROUGE-L: 0,2919

Hasil Teks Ringkasan Keluaran Mesin Peringkas (SBERT + MMR Murni) Mesin 40%	Skor Evaluasi (*F1-Score*)
mengintegrasikannya dengan perangkat, infrastruktur, dan sistem sudah berjalan. Sejalan dengan hal ini, prioritas investasi TI pun bergeser ke arah penerapan perangkay AI guna meningkatkan produktivitas dan mendukung proses inferensi AI secara lokal. Artinya, perangkat tersebut memiliki kemampuan dan kinerja yang diperlukan untuk menjalankan model AI secara lokal di sisi edge . Melalui riset ini, Lenovo memetakan kesiapan, kapabilitas, dan prioritas perubahan dalam mengadipsi AI di lingkungan bisnis. Adopsi kecerdasan buatan (AI) di kawasan ASEAN terus menunjukkan tren positif. Mereka adalah kontributor strategis yang turut menentukan arah teknologi masa depan yang lebih beretika dan berorientasi pada kepentingan manusia. Baginya, kunci utama efektivitas AI terletak pada kegunaannya dalam menjawab persoalan nyata di masyarakat. Ia mengibaratkan data sebagai kompas untuk optimasi efisiensi dan akurasi, namun terobosan besar tetap membutuhkan visi manusia. Jadi hybrid AI itu bukan pilihan, tapi itu sudah keharusan bagi perusahaan-perusahaan yang di ASEAN, ujar Budi Janto, di Jakarta, baru-baru ini. Dalam webinar bertajuk AI Streamline Your Business: Build Internal Apps with AI baru-baru ini, Juan membedah peta jalan pengembangan AI yang melampaui sekadar kecanggihan teknis. Pada 2026 fokus AI didominasi oleh pendekatan berbasis hasil (outcome-based), yang mengisi tiga posisi teratas dalam strategi perusahaan. Fitur-fitur ini mencakup "Produktivitas Sekali Ketuk" untuk membantu pengguna membuat daftar tugas, catatan, dan menjalankan perintah kerja ringan dengan cepat. Infinix Smart 10 Plus hadir dengan pengalaman kecerdasan buatan (AI) yang menonjol, salah satu daya tariknya terletak pada kehadiran Folax, asisten virtual berbasis AI. Berkat kelincahan dan kematangan tata kelola di kawasan ASEAN menjadikannya sebagai lingkungan uji coba yang ideal untuk Agentic AI. Seiring dengan hal itu, fokus perusahaan di ASEAN+ mulai beralih ke Agentic AI. Penerapan AI Hybrid turut dipengaruhi oleh tingginya biaya pengelolaan data di cloud, serta kekhawatiran terkait keamanan dan privasi data. Prioritas kedua adalah reinvensi bisnis berbasis AI. Product Manager di Google Amerika Serikat, Juan Anugraha Djuwadi, kini menjadi salah satu aktor strategis yang mengarahkan bagaimana teknologi AI dikembangkan agar tetap manusiawi dan berdampak luas bagi pengguna global. Kesenjangan kesiapan ini rupanya masih menunjukkan adanya tantangan mendasar dalam penerapan AI. Ia memperkirakan lahirnya erasoftware on-the-fly, di mana perangkat lunak dihasilkan secarareal-timesesuai kebutuhan instan pengguna, baik melalui antarmuka teks maupun suara. Prinsip ini diterjemahkan ke dalam dua filosofi kunci:less is moredanthe details matter.Menurutnya, dalam skala pengguna miliaran seperti di Google, detail sekecil apa pun menjadi isu strategis. Data memvalidasi masa kini, namun intuisi mendefinisikan masa depan, ungkapnya. Beberapa diantaranya yaitu: Jika melihat keseluruhan faktor tersebut, inilah alasan mengapa terjadi pergeseran menuju penerapan AI berbasis hibrida. Adopsi AI tidak lagi terbatas pada fungsi IT seperti analitik data, keamanan siber dan pengembangan perangkat lunak, tetapi juga mengarah ke fungsi non-IT. Inovasi AI yang sukses di wilayah dengan keberagaman sosial tinggi seperti Indonesia dan APAC haruslah bersifat kontekstual dan berbasis pada kepercayaan. com/ Agustin Setyo Wardani) Redmi 14 C tampil dengan peningkatan signifikan pada sektor kamera berkat dukungan kamera utama 50 MP dan teknologi AI.	

Hasil Teks Ringkasan Keluaran Mesin Peringkas (SBERT + MMR Murni) Mesin 40%	Skor Evaluasi (*F1-Score*)
Data terbaru yang terungkap dalam konferensi pers Lenovo Tech Day 2026 menunjukkan 67 persen organisasi di kawasan ASEAN+ sudah melakukan uji coba AI secara sistematis. Di sisi lain, Lenovoresmi Lenovo Tech Day 2026 dengan tema The Race for Enterprise AI. yang terus berkembang. Di internal Lenovo sendiri, kami telah memanfaatkan AI dalam rantai pasok dan perencanaan logistik, ujar Thomas. Kini, perusahaan tidak hanya melihat pengembalian investasi secara langsung dari AI, tetapi juga peningkatan layanan pelanggan, keterlibatan karyawan yang lebih tinggi, serta pengambilan keputusan yang semakin baik, terang Thomas. Model AI kini dibawa langsung ke lingkungan kerja, sehingga karyawan dituntut mampu memahami dan mengintegrasikan proses berbasis AI ke dalam aktivitas sehari-hari. Pemenangnya di eraagentic A Ibukan yang paling cepat bergerak, tapi mereka bilang yang paling dapat dipercaya, dan paling dapat tata kelolanya baik, tegas Budi. Namun kini, prioritas bisnis telah bergeser ke bagaimana AI mampu mendorong peningkatan pendapatan dan profitabilitas, ujar Thomas. Juan, yang merupakan alumnus Columbia University, menegaskan bahwa pengguna akhir seringkali tidak mempedulikan kerumitan algoritma di balik sebuah aplikasi.	

Nilai *F1-Score* pada seluruh metrik secara konsisten mengalami penurunan signifikan saat batas kompresi diperlebar menjadi 40% dan 50%, di mana skor ROUGE-1 menurun dari 0,5500 menjadi 0,4634. Penurunan ini terjadi akibat anjloknya tingkat ketepatan (*Precision*) secara drastis, meskipun nilai penemuan informasinya (*Recall*) meningkat seiring bertambahnya panjang teks. Penambahan kalimat pada rentang kompresi tinggi terbukti justru memasukkan banyak informasi sekunder atau noise yang tidak terdapat pada rujukan manusia.

Studi kasus ini membuktikan secara empiris bahwa penentuan rasio kompresi sangat memengaruhi efektivitas algoritma seleksi SBERT dan MMR. Tingkat kompresi 30% tervalidasi sebagai batas ekstraksi paling ideal di mana mesin menyajikan fakta paling padat dengan tingkat redundansi minimal. Tahap evaluasi selanjutnya diekspansi ke seluruh klaster dokumen untuk memvalidasi performa metode secara menyeluruh.

4.6. Hasil dan Analisis Evaluasi Model Peringkas dengan ROUGE pada Setiap Kategori Berita

Pengujian ROUGE dilakukan secara spesifik pada enam kategori berita yang berbeda guna memastikan mesin peringkas tidak memiliki bias terhadap gaya bahasa tertentu. Fokus

evaluasi ditekankan pada metrik *F1-Score* dari ROUGE-1 (kata tunggal), ROUGE-2 (frasa berurutan), dan ROUGE-L (struktur urutan).

4.6.1. Hasil Pengujian Kategori Bola

Klaster-klaster dokumen di bawah kategori berita "Bola" menjadi subjek pengujian pertama. Evaluasi ini bertujuan mengamati kemampuan mesin dalam meringkas bahasa jurnalistik olahraga pada berbagai parameter kompresi. Tabel 4.6 menyajikan hasil rata-rata evaluasi metrik ROUGE untuk seluruh klaster pada kategori tersebut.

Tabel 4.6 Nilai Rata-rata *F1-Score* Kategori Bola

Metrik Evaluasi *F1-Score*	**Kompresi**			
	20%	**30%**	**40%**	**50%**
ROUGE-1	0,5048	0,5418	0,5284	0,5304
ROUGE-2	0,3527	0,3721	0,3782	0,4055
ROUGE-L	0,3950	0,4161	0,4165	0,4405

4.6.2. Hasil Pengujian Kategori Tekno

Dokumen berita dengan kategori "Tekno" menjadi fokus pengujian selanjutnya. Kategori ini memiliki karakteristik penggunaan istilah asing dan spesifikasi teknis yang padat. Tabel 4.7 merinci hasil pengujian pada keempat variasi kompresi untuk kategori ini.

Tabel 4.7 Nilai Rata-rata *F1-Score* Kategori Tekno

Metrik Evaluasi *F1-Score*	**Kompresi**			
	20%	**30%**	**40%**	**50%**
ROUGE-1	0,4638	0,4482	0,4247	0,3903
ROUGE-2	0,2515	0,2797	0,2928	0,2804
ROUGE-L	0,2918	0,3079	0,3155	0,3040

4.6.3. Hasil Pengujian Kategori Bisnis

Eksperimen ketiga melibatkan pengolahan klaster dokumen berita "Bisnis" oleh mesin peringkas. Karakteristik utama berita bisnis meliputi banyaknya kutipan tokoh dan pemaparan kebijakan pemerintah maupun perusahaan. Tabel 4.8 merangkum hasil perolehan *F1-Score* untuk kategori Bisnis.

Tabel 4.8 Nilai Rata-rata *F1-Score* Kategori Bisnis

Metrik Evaluasi *F1-Score*	**Kompresi**			
	20%	**30%**	**40%**	**50%**
ROUGE-1	0,4653	0,4344	0,4332	0,3983
ROUGE-2	0,2847	0,2560	0,2829	0,3015
ROUGE-L	0,3150	0,2836	0,2920	0,3081

4.6.4. Hasil Pengujian Kategori Saham

Berita pergerakan "Saham" menjadi pusat perhatian pada eksperimen keempat. Berita kategori ini memiliki tingkat redundansi (*noise*) angka yang sangat tinggi, terutama pada penyebutan kode emiten dan persentase harga. Tabel 4.9 menyajikan hasil perolehan *F1-Score* untuk kategori Saham.

Tabel 4.9 Nilai Rata-rata *F1-Score* Kategori Saham

Metrik Evaluasi *F1-Score*	**Kompresi**			
	20%	**30%**	**40%**	**50%**
ROUGE-1	0,4614	0,4494	0,4107	0,3761
ROUGE-2	0,2803	0,2864	0,2673	0,2562
ROUGE-L	0,3189	0,3228	0,3047	0,2890

4.6.5. Hasil Pengujian Kategori Otomotif

Model pada mesin peringkas diuji menggunakan berita dari kategori "Otomotif" dalam eksperimen kelima. Topik pada kategori ini umumnya berkisar pada peluncuran kendaraan, ulasan spesifikasi, hingga pameran industri. Tabel 4.10 menjabarkan hasil evaluasi ROUGE untuk kategori otomotif.

Tabel 4.10 Nilai Rata-rata *F1-Score* Kategori Otomotif

Metrik Evaluasi *F1-Score*	**Kompresi**			
	20%	**30%**	**40%**	**50%**
ROUGE-1	0,4828	0,4526	0,4473	0,4185
ROUGE-2	0,2969	0,2754	0,2939	0,3062
ROUGE-L	0,3480	0,3139	0,3218	0,3242

4.6.6. Hasil Pengujian Kategori Kesehatan

Kategori "Kesehatan" menjadi sasaran eksperimen terakhir. Gaya bahasa penyuluhan dengan banyak istilah medis dan saran pengobatan menjadi ciri khas utama kategori ini. Tabel 4.11 menampilkan hasil evaluasi kinerja mesin pada kategori kesehatan.

Tabel 4.11 Nilai Rata-rata *F1-Score* Kategori Kesehatan

Metrik Evaluasi *F1-Score*	**Kompresi**			
	20%	**30%**	**40%**	**50%**
ROUGE-1	0,5255	0,5235	0,4981	0,4720
ROUGE-2	0,3425	0,3510	0,3547	0,3499
ROUGE-L	0,3904	0,3905	0,3930	0,3800

4.7. Analisis Rekapitulasi ROUGE Score Keseluruhan *F1-Score*

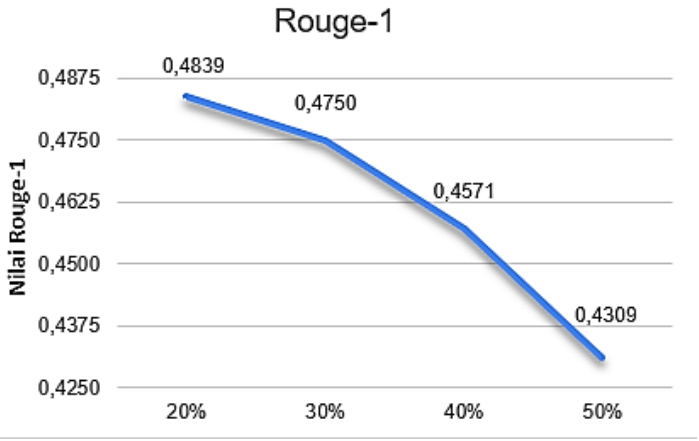
Penyatuan seluruh nilai rata-rata *F1-Score* dari keenam kategori menjadi langkah akhir untuk menemukan nilai rata-rata keseluruhan. Nilai rata-rata tersebut digunakan sebagai kesimpulan final untuk merepresentasikan tingkat keberhasilan dan performa mesin peringkasan multi-dokumen ekstraktif yang telah dibangun. Tabel 4.12 merangkum hasil rekapitulasi nilai rata-rata dari seluruh kategori berita pada keempat level kompresi.

Tabel 4.12 Tabel Metrik Rata-Rata Total *F1-Score* di Semua ROUGE

Metrik Evaluasi *F1-Score*	**Kompresi**			
	20%	**30%**	**40%**	**50%**
ROUGE-1	0,4839	0,4750	0,4571	0,4309
ROUGE-2	0,3014	0,3034	0,3116	0,3166
ROUGE-L	0,3432	0,3391	0,3406	0,3410

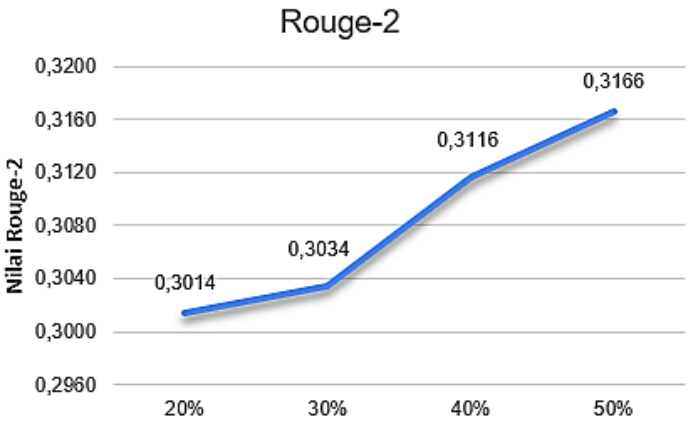
Penilaian kualitatif terhadap perolehan skor ROUGE tersebut menunjukkan bahwa performa metode usulan berada pada kategori terbukti kompetitif dan cukup memuaskan. Angka rata-rata *F1-Score* ROUGE-1 sebesar 0,4839 pada kompresi 20% merepresentasikan tingkat keberhasilan ekstraksi informasi yang tergolong tinggi. Dalam literatur *Natural Language Processing* (NLP), capaian skor ROUGE-1 pada kisaran 0,40 hingga 0,50 untuk tugas peringkasan ekstraktif telah diklasifikasikan sebagai performa *State-of-the-Art* yang sangat tangguh (Koto dkk., 2020; Supriyono dkk., 2024). Pencapaian angka ini membuktikan kemampuan algoritma dalam menangkap esensi faktual secara akurat, mengingat evaluasi n-gram pada metrik ROUGE memiliki batasan teoretis (*upper bound*).

Variasi gaya bahasa (*paraphrasing*) yang tinggi antar peringkas manusia menyebabkan skor kesepakatan (*human agreemen*t) mustahil mencapai angka mutlak 1,00, dengan batas kewajaran maksimal pencocokan leksikal murni umumnya tertahan di sekitar angka 0,50 (Akhmetov dkk., 2022; Sai dkk., 2022). Hal tersebut menegaskan bahwa mesin peringkas terbukti tidak sekadar mengambil kalimat secara acak, melainkan mengandalkan kecerdasan pemetaan semantik yang sejajar dengan nalar ringkasan rujukan pakar manusia. Data komputasi dari Tabel 4.12 diplot ke dalam bentuk grafik kurva garis guna memberikan gambaran visual yang lebih komprehensif terkait fluktuasi kinerja mesin akibat perubahan persentase kompresi. Gambar 4.1, 4.2, dan 4.3 menampilkan visualisasi perbandingan metrik ROUGE-1, ROUGE-2, dan ROUGE-L.



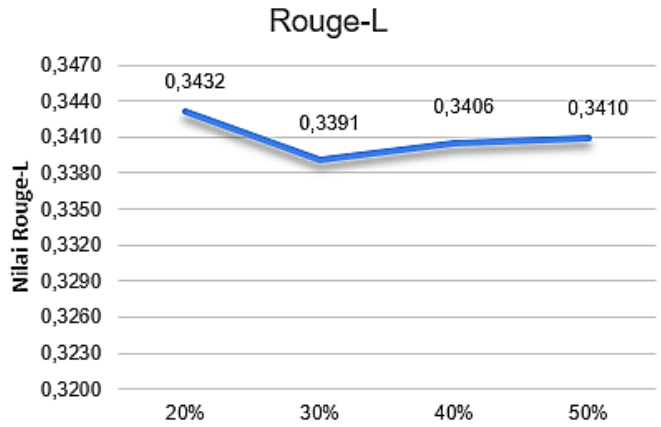
Gambar 4.1 Grafik Pengaruh Tingkat Kompresi Terhadap Kinerja *F1-Score* ROUGE-1 (Kata)

Gambar 4.1 menyajikan tren penurunan pada metrik ROUGE-1. Nilai pencocokan kata tunggal meraih angka tertinggi sebesar 0,4839 pada tingkat kompresi 20% dan terus menurun secara linier hingga 0,4309 pada kompresi 50%. Fenomena ini terjadi karena algoritma MMR dipaksa mengekstraksi kalimat dengan kedekatan absolut terhadap *centroid* pada kompresi ketat, sehingga menghasilkan rasio kepresisian (*Precision*) yang tinggi. Sebaliknya, model mulai memasukkan kalimat sekunder sebagai pelengkap pada kompresi 50% yang memicu masuknya kata-kata *noise*.



Gambar 4.2 Grafik Pengaruh Tingkat Kompresi Terhadap Kinerja *F1-Score* ROUGE-2 (Frasa)

Tren kenaikan ROUGE-2 pada Gambar 4.2 menunjukkan perilaku yang berbeda di mana nilai pencocokan frasa meningkat secara perlahan dari 0,3014 menjadi 0,3166. Probabilitas mesin untuk mengambil urutan kalimat yang memuat struktur bigram utuh (serupa dengan ringkasan manusia) menjadi lebih besar seiring bertambahnya panjang teks ringkasan.



Gambar 4.3 Grafik Pengaruh Tingkat Kompresi Terhadap Kinerja *F1-Score* ROUGE-L (Urutan)

Gambar 4.3 menunjukkan kestabilan pada metrik ROUGE-L. Nilai pencocokan struktur kalimat berada secara stabil pada kisaran angka 0,34 di semua tingkat kompresi. Kondisi ini membuktikan bahwa algoritma MMR yang dibangun mampu mempertahankan struktur gramatikal teks tanpa terpengaruh secara signifikan oleh panjang pendeknya hasil peringkasan.

4.8. Interpretasi Hasil Peringkasan Model

Evaluasi komputasional terhadap dokumen berita dari enam kategori berbeda (Bola, Tekno, Bisnis, Saham, Otomotif, dan Kesehatan) menghasilkan beberapa temuan analitis yang menjelaskan perilaku metode usulan dalam mengeksekusi tugas peringkasan ekstraktif multi-dokumen.

Pertama, tingkat kompresi atau persentase panjang ringkasan secara konsisten memberikan pengaruh yang sangat signifikan terhadap kualitas hasil akhir. Penilaian metrik ROUGE memvalidasi tingkat kompresi 30% sebagai titik keseimbangan paling optimal (*sweet spot*). Pada titik ini, algoritma mampu menangkap kepadatan informasi utama (tercermin dari *F1-Score* ROUGE-1 yang tinggi) sekaligus menjaga koherensi dan kelengkapan frasa antarkalimat (tercermin dari *F1-Score* ROUGE-2 yang solid). Kompresi 20% terbukti terlalu ketat sehingga memotong banyak konteks penting kalimat, sementara melonggarkan batas ekstraksi ke angka 40% hingga 50% justru berisiko memicu masuknya *noise* atau informasi non-esensial.

Kedua, teramati adanya fenomena fluktuasi (*trade-off*) antara komponen *Recall* dan *Precision* seiring dengan penambahan batas kompresi. Ketika kompresi dinaikkan dari 30%

menuju 50%, nilai rata-rata *Precision* secara konsisten mengalami tren penurunan. Hal ini merupakan dampak langsung dari karakteristik matematis algoritma MMR, setelah kalimat-kalimat primer yang paling relevan habis terekstraksi pada iterasi awal, algoritma terpaksa menarik kalimat sekunder demi memenuhi kuota panjang kompresi. Masuknya kalimat sekunder ini menyebabkan nilai kepresisian turun drastis meskipun tingkat penemuan informasinya (*Recall*) secara kumulatif meningkat.

Ketiga, kinerja integrasi SBERT dan MMR menunjukkan performa yang bervariasi bergantung pada kategori teks berita dari sisi karakteristik data. Kategori "Bola" dan "Kesehatan" secara konsisten mencatatkan *F1-Score* ROUGE tertinggi. Temuan ini mengindikasikan ketangguhan model SBERT dalam memetakan vektor representasi pada dokumen yang memiliki struktur penulisan seragam dan kosakata yang terpusat. Kategori "Tekno" sebaliknya mencatatkan performa terendah akibat tingginya penggunaan istilah teknis asing (*anglisisme*) yang variasi penulisannya sangat beragam antar-portal berita. Kondisi ini menyulitkan pencocokan n-gram secara persis pada evaluasi ROUGE, meskipun secara kalkulasi *embedding* kalimat tersebut memiliki kemiripan semantik.

Terakhir, arsitektur metode usulan secara keseluruhan menunjukkan stabilitas komputasi cukup yang tinggi. Integrasi *Agglomerative Clustering* untuk isolasi topik yang dipadukan dengan *Sentence*-BERT dan MMR terbukti mampu menjalankan logika dasarnya tanpa memerlukan aturan modifikasi linguistik tambahan. Algoritma secara stabil mampu memilih kalimat yang paling dekat dengan inti topik (*centroid*) sekaligus memberikan penalti yang tegas terhadap kalimat yang redundan. Ini membuktikan bahwa pemetaan semantik (*dense embedding*) yang dikalibrasi dengan metrik diversitas merupakan pendekatan yang sangat relevan untuk membedah karakteristik teks jurnalistik berbahasa Indonesia.

4.9. Analisis Kelemahan Metode Usulan

Metode usulan menunjukkan efektivitas dalam mengekstraksi inti informasi dan memitigasi redundansi teks, namun beberapa karakteristik bawaan dari arsitektur komputasi ini membatasi ruang lingkup evaluasi sebagai berikut:

1. Karakteristik algoritma yang beroperasi secara murni ekstraktif menjaga keutuhan fakta dengan merangkai ringkasan langsung dari potongan kalimat asli sumber berita. Mekanisme tersebut mengakibatkan transisi antar-kalimat terkadang terasa kurang

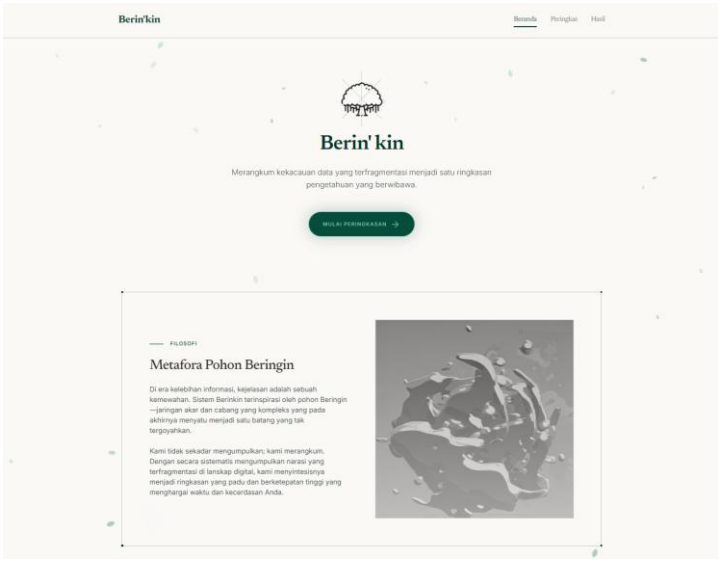
kohesif apabila dibandingkan dengan teks hasil parafrasa manusia sebagai bentuk *trade-off* teknis.

2. Penggunaan parameter penyeimbang pada algoritma MMR ditetapkan secara konstan ($\lambda = 0{,}7$) guna menjaga lingkungan pengujian yang stabil dan terkontrol. Algoritma seleksi ini belum dirancang untuk menyesuaikan nilai parameter secara otomatis terhadap klaster berita dengan tingkat kepadatan informasi ekstrem, meskipun nilai tersebut telah memberikan hasil optimal secara rata-rata.
3. Model representasi SBERT terbukti akurat memetakan semantik bahasa Indonesia baku, namun presisinya berpotensi menurun pada korpus di luar *dataset* Liputan6 yang memuat bahasa kasual, *slang*, atau kiasan jurnalistik akibat keterbatasan distribusi data latih. Pada skenario kebahasaan yang dinamis tersebut, implementasi *Large Language Model* (LLM) secara teoretis akan menghasilkan performa pemahaman konteks yang jauh lebih superior daripada SBERT, karena skala parameternya yang masif memungkinkan adaptasi yang lebih baik terhadap variasi teks tidak baku.

4.10. Implementasi Program

Tahap pengujian eksperimental dilanjutkan dengan implementasi metode usulan SBERT-MMR ke dalam prototipe aplikasi berbasis web bernama Berinkin. Prototipe tersebut mendemonstrasikan fungsionalitas algoritma secara interaktif kepada pengguna akhir. Bagian ini menyajikan representasi visual antarmuka aplikasi guna memberikan gambaran teknis yang komprehensif.

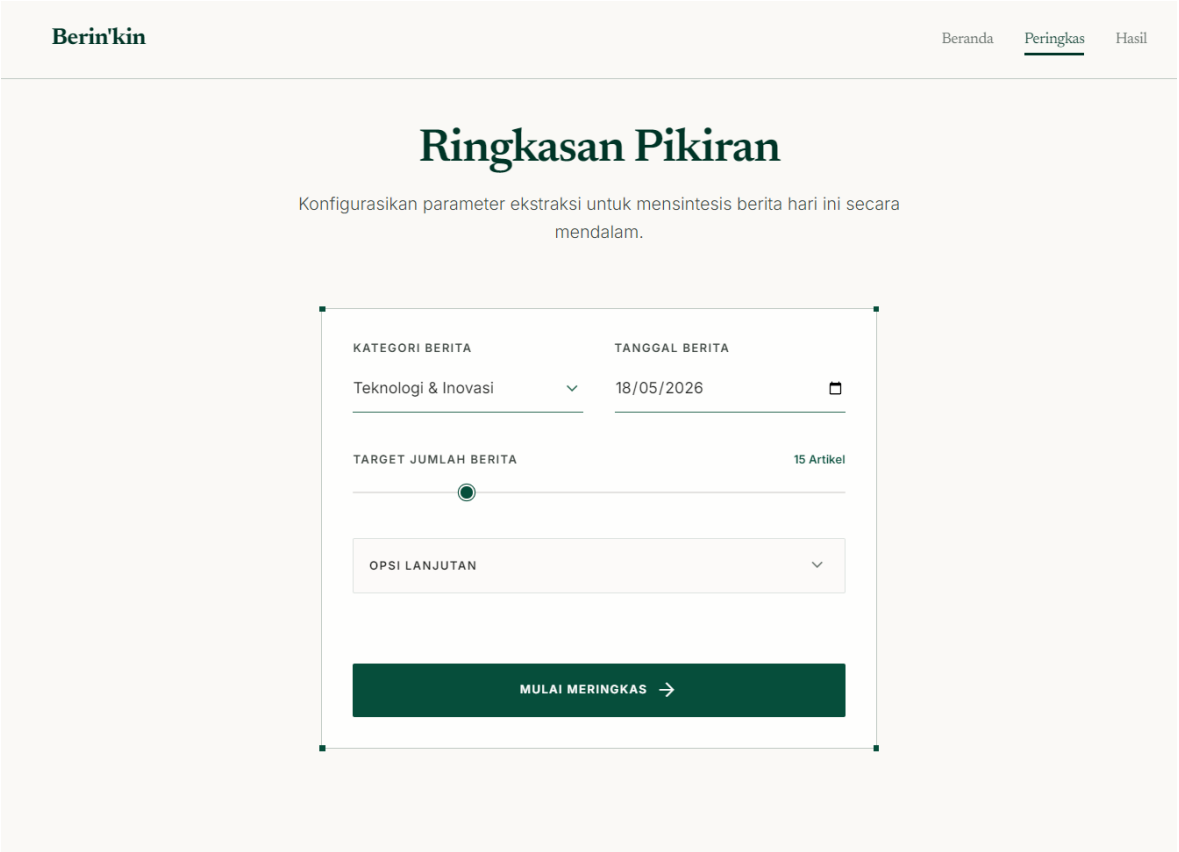
1. Halaman Beranda (Landing Page)



Gambar 4.4 Halaman Beranda Antarmuka Mesin Peringkasan

Gambar 4.4 memperlihatkan halaman yang berfungsi sebagai pintu gerbang utama yang menyambut pengguna dan menjelaskan identitas serta metodologi sistem Berinkin. Menampilkan logo beringin, judul aplikasi, dan *tagline* utama.

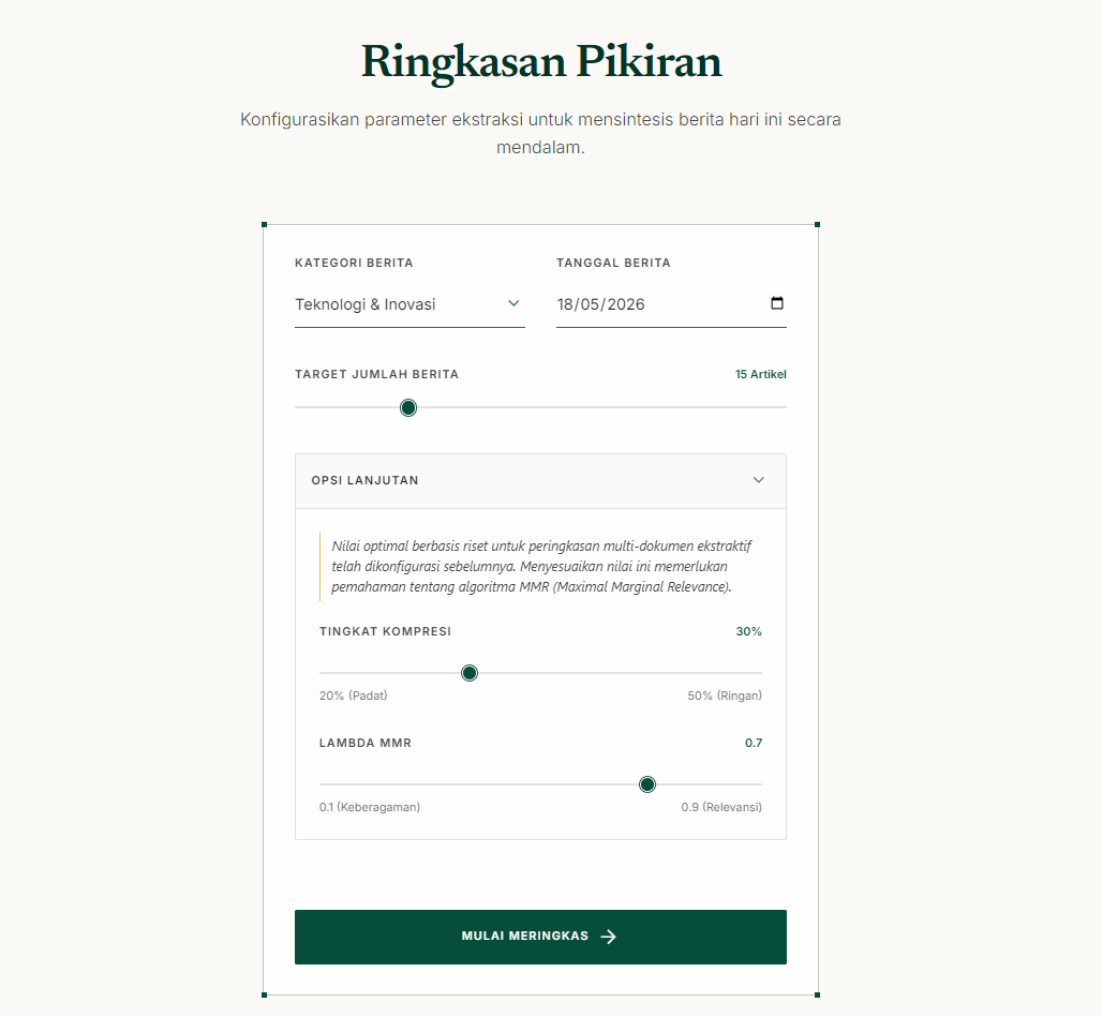
2. Pemilihan Konfigurasi Peringkasan Berita



Gambar 4.5 Menu Pilihan Konfigurasi Peringkasan

Gambar 4.5 menunjukkan halaman sebagai pusat kontrol konfigurasi bagi pengguna untuk menentukan parameter pengambilan dan peringkasan berita. Menampilkan *dropdown* untuk memilih Kategori Berita (Teknologi, Ekonomi, dll), input Tanggal Berita, serta *slider* untuk menentukan Target Jumlah Berita yang akan di-*crawling*.

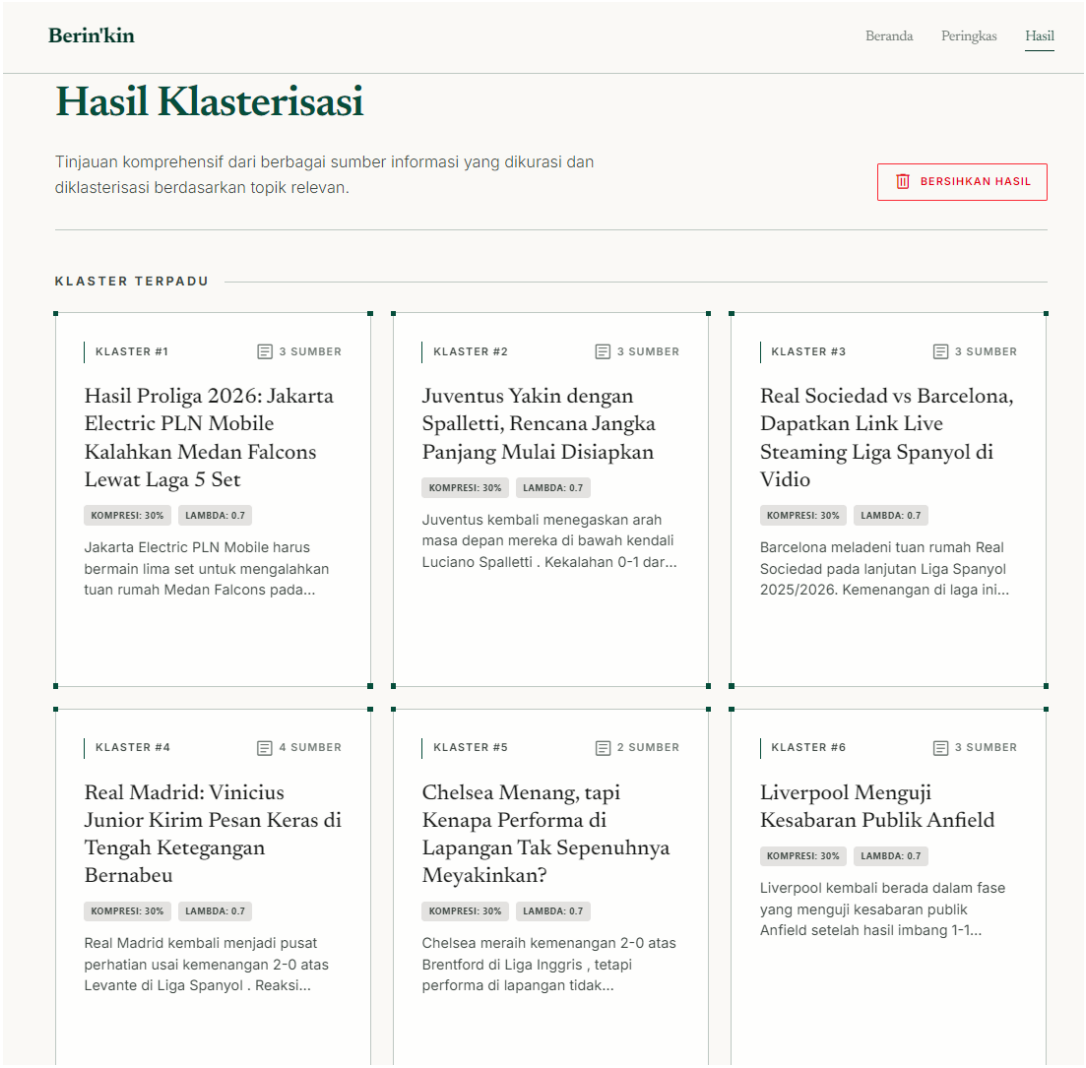
3. Opsi Detail (Tingkat Kompresi Peringkasan dan *Lambda* MMR)



Gambar 4.6 Opsi Lanjutan Konfigurasi Peringkasan

Gambar 4.6 Menampilkan *slider* parameter algoritma. Terdapat pengaturan "Tingkat Kompresi" (20% - 50%) untuk mengatur kepadatan/jumlah kalimat ringkasan, dan "*Lambda* MMR" (0,1 – 0,9) untuk mengatur keseimbangan antara relevansi dan keberagaman kalimat.

4. Daftar Klaster Topik Berita



Gambar 4.7 Menu Daftar Klaster Topik Berita

Gambar 4.7 Menjunjukkan Halaman yang menampilkan seluruh topik yang berhasil diekstraksi dan dikelompokkan oleh algoritma *Agglomerative Clustering* pada hari tersebut. Topik-topik berita yang terbentuk dari gabungan beberapa artikel (*multi-document*) ditampilkan dengan pemetaan *grid* yang rapi.

5. Hasil Peringkasan Berita

Gambar 4.8 Hasil Akhir Peringkasan Ekstraktif Berita

Gambar 4.8 adalah halaman hasil akhir berupa peringkasan ekstraktif dari suatu klaster topik berita yang dipilih pengguna. Menampilkan judul besar dari topik berita yang diringkas, beserta metadata sistem (Label Klaster, Jumlah Sumber, nilai Kompresi, dan nilai *Lambda*), serta paragraf-paragraf hasil ekstraksi kalimat terbaik (Top-N *sentences*) oleh algoritma SBERT & MMR.

BAB V
PENUTUP

5.1. Kesimpulan

Penelitian ini telah berhasil menjawab rumusan masalah terkait integrasi dan evaluasi metode usulan untuk peringkasan multi-dokumen ekstraktif berita berbahasa Indonesia. Kesimpulan dari penelitian ini dijabarkan sebagai berikut:

1. Perancangan mesin peringkasan multi-dokumen ekstraktif berhasil diimplementasikan melalui *pipeline* komputasi yang menggabungkan tiga algoritma secara runut. *Agglomerative Clustering* (dengan *distance threshold* 0,8) terbukti akurat dalam mengisolasi pengelompokan topik berita secara presisi. Model *Sentence*-BERT (varian *paraphrase-multilingual-MiniLM-L12-v2*) selanjutnya secara efektif memetakan representasi semantik kalimat ke dalam ruang vektor berdimensi 384 untuk mengatasi kelemahan deteksi parafrasa pada metode statistik. Tahap penyeleksian akhir dieksekusi secara tangguh menggunakan metrik diversitas algoritma *Maximal Marginal Relevance* (MMR) dengan parameter penyeimbang $\lambda = 0{,}7$ untuk mengekstraksi inti informasi dan memitigasi redundansi teks lintas dokumen.

2. Pengujian menggunakan metrik ROUGE pada berbagai skenario kompresi (20%, 30%, 40%, dan 50%) terhadap *benchmark dataset* Liputan6 membuktikan bahwa tingkat kompresi 30% merupakan konfigurasi paling optimal dalam menentukan titik keseimbangan performa metode usulan. Konfigurasi kompresi 30% tersebut menghasilkan nilai rata-rata keseluruhan *F1-Score* ROUGE-1 sebesar 0,4750, ROUGE-2 sebesar 0,3034, dan ROUGE-L sebesar 0,3391. Keandalan pengujian komputasional ini divalidasi oleh dokumen *human ground truth* yang mencatatkan tingkat kesepakatan aktual (*Percentage Agreement*) sebesar 63,21%, terlepas dari adanya pengaruh fenomena ketidakseimbangan kelas berupa anomali *Kappa Paradox* (skor *Cohen's Kappa* sebesar 0,1877).

Integrasi *dense embedding* berbasis konteks dan metrik diversitas MMR pada metode usulan ini secara keseluruhan menunjukkan efektivitas tinggi dalam memitigasi fenomena *information overload*. Metode usulan ini terbukti mampu menyajikan teks ringkasan berita

yang akurat, minim redundansi leksikal, dan tetap mempertahankan konteks esensial bagi pembaca digital di Indonesia.

5.2. Saran

Analisis performa dan batasan metode usulan yang telah diuraikan menghasilkan beberapa rekomendasi pengembangan untuk penelitian di masa depan.

1. Pengembangan model bahasa dapat diperluas dengan memanfaatkan *Large Language Model* (LLM) yang telah dilatih ulang (*fine-tuned*) khusus untuk domain bahasa Indonesia atau instruksi instruksional guna meningkatkan kedalaman pemahaman konteks dan nuansa bahasa.
2. Penelitian selanjutnya dapat mengembangkan mekanisme MMR yang bersifat dinamis (*dynamic tuning*). Parameter penalti redundansi (λ) dapat dirancang agar menyesuaikan diri secara otomatis berdasarkan perhitungan tingkat penyebaran jarak antar-vektor kalimat di dalam masing-masing klaster topik.
3. Luaran dari metode ekstraktif ini dapat diumpankan ke dalam modul abstraktif guna meningkatkan kelancaran aliran narasi (*readability*). Pendekatan hibrida tersebut memfasilitasi sistem untuk menyusun ulang kalimat secara lebih natural tanpa mengorbankan akurasi faktual yang telah diamankan oleh tahap ekstraksi.
4. Model peringasan dengan metode usulan yang telah teruji ini sangat potensial untuk diintegrasikan secara *end-to-end* ke dalam platform atau aplikasi baca berita harian, sehingga memberikan solusi nyata bagi masyarakat dalam menanggulangi *information overload* secara cepat dan efisien.

DAFTAR PUSTAKA

Akhmetov, I., Mussabayev, R., & Gelbukh, A. (2022). Reaching for upper bound ROUGE score of extractive summarization methods. *PeerJ Computer Science, 8*. https://doi.org/10.7717/peerj-cs.1103

Annisa, R., Sasongko, A., & Maulana, M. S. (2026). MOCS-BERT: Multi-Objective Cuckoo Search with Sentence-BERT Embeddings for Semantically Enhanced Extractive Summarization. *Gazi University Journal of Science*, *39*(2), 957-972. https://doi.org/10.35378/gujs.1827571

Antara News. (2024, 12 Juni). Ketimpangan jumlah media jadi problem sebaran informasi di Indonesia. ANTARA. https://www.antaranews.com/berita/4149099/ ketimpangan-jumlah-media-jadi-problem-sebaran-informasi-di-indonesia

Badshah, S., & Sajjad, H. (2025). Reference-guided verdict: LLMs-as-judges in automatic evaluation of free-form QA. *Proceedings of the 9th Widening NLP Workshop*, 251–267. https://doi.org/10.18653/v1/2025.winlp-main.37

Carbonell, J., & Goldstein, J. (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. *SIGIR '98*, *1*(1), 335. https://doi.org/10.1145/503698.503737

Chandrasekaran, D., & Mago, V. (2021). Evolution of semantic similarity—A survey. *ACM Computing Surveys (CSUR)*, *54*(2), 1–37. https://doi.org/10.1145/3440755

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019a). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4171–4186. https://doi.org/10.18653/v1/N19-1423

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019b). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv*. https://arxiv.org/abs/1810.04805

Dhini, B., Girsang, A., Sufandi, U., & Kurniawati, H. (2023). Automatic essay scoring for discussion forum in online learning based on semantic and keyword similarities. *Asian Association of Open Universities Journal, 18*. https://doi.org/10.1108/AAOUJ-02-2023-0027

Fabbri, A. R., Kryściński, W., McCann, B., Xiong, C., Socher, R., & Radev, D. (2021). SummEval: Re-evaluating summarization evaluation. *Transactions of the Association for Computational Linguistics*, *9*, 391–409. https://doi.org/10.1162/tacl_a_00373

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology, 43*(6), 543–549. https://doi.org/10.1016/0895-4356(90)90158-L

Girsang, A. S., & Amadeus, F. J. (2023). Text summarization using ant system algorithm for Indonesian news article. *Journal of Computer Science*, *19*(3), 321–330. https://doi.org/10.3844/jcssp.2023.321.330

Gunawan, D., Harahap, S. H., & Rahmat, R. F. (2019). Multi-document summarization by using TextRank and maximal marginal relevance for text in Bahasa Indonesia. *2019 International Conference on ICT for Smart Society (ICISS), 7*, 1–5. https://doi.org/10.1109/ICISS48059.2019.8969785

Guo, Z., Wang, X., & Zhang, Y. (2024). Push, overload, and exhaustion: Reactions of college students to information about international news events on social media. *International Journal of Cyber Behavior, Psychology and Learning (IJCBPL)*, *14*(1), 1–17. https://doi.org/10.4018/IJCBPL.362809

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques* (3rd ed.). Waltham, MA: Morgan Kaufmann.

Hanley, H. W. A., Barlas, P., & Durumeric, Z. (2025). Hierarchical level-wise news article clustering via multilingual Matryoshka embeddings. *Proceedings of the Association for Computational Linguistics (ACL)*. Advance online publication. https://aclanthology.org/2025.acl-long.124.pdf

Hartawan, M. S., dkk. (2024). Analyzing PEGASUS model performance with ROUGE on Indonesian news summarization. Sinkron: Jurnal dan Penelitian Teknik Informatika, 8(3), 1542–1550. https://doi.org/10.33395/sinkron.v8i3.13401

Hendrycks, D., & Gimpel, K. (2023). *Gaussian Error Linear Units (GELUs)*. http://arxiv.org/abs/1606.08415

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12). https://doi.org/10.1145/3571730

Jurafsky, D., & Martin, J. H. (2024). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (2nd ed.).

Kalyan, K. S., Rajasekharan, A., & Sangeetha, S. (2021). AMMUS: A survey of transformer-based pretrained models in natural language processing. *arXiv*. http://arxiv.org/abs/2108.05542

Koto, F., Lau, J. H., & Baldwin, T. (2020). Liputan6: A Large-scale Indonesian dataset for text summarization. *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 598–608. https://aclanthology.org/2020.aacl-main.60/

Landis, J. R., & Koch, G. G. (1977). The Measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Lennox, C., Kashyapi, S., & Dietz, L. (2023). Retrieve-cluster-summarize: An alternative to end-to-end training for query-specific article generation. *arXiv preprint arXiv:2310.12361*. https://doi.org/10.48550/arXiv.2310.12361

Latif, A. (2023, 10 Agustus). Berapa Paragraf 'Idealnya' Lead Berita? Medium. https://lalativ.medium.com/berapa-paragraf-idealnya-lead-berita-35dd4bfdcbce

Mahendra, K., & Sukartik, D. (2024). Jurnalisme clickbait di media online Tribunpekanbaru.com dan Riauonline.co.id. *Jurnal Riset Mahasiswa Dakwah dan Komunikasi (JRMDK)*, *6*(1), 118–131.

Mao, Y., Qu, Y., Xie, Y., Ren, X., & Han, J. (2020). Multi-document summarization with maximal marginal relevance-guided reinforcement learning. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, 1737–1751. https://doi.org/10.18653/v1/2020.emnlp-main.136

Mawaridi, B. H., Faisal, M., & Nurhayati, H. (2024). Peringkasan teks berbahasa Indonesia dengan latent Dirichlet allocation dan maximum marginal relevance. *Techno.COM, 23*(3), 623–632. https://doi.org/10.33633/tc.v23i3.9876

Maynez, J., Narayan, S., Bohnet, B., & Mcdonald, R. (2020). On faithfulness and factuality in abstractive summarization. https://github.com/google-research-

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282. https://doi.org/10.11613/BM.2012.031

Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., & Kleis Nielsen, R. (2024). Reuters Institute Digital News Report 2024. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/ digital-news-report/2024

Peng, M., Xu, Z., & Huang, H. (2021). How does information overload affect consumers' online decision process? An event-related potentials study. *Frontiers in Neuroscience*, *15*, 695852. https://doi.org/10.3389/fnins.2021.695852

Peters, M. E., Ruder, S., & Smith, N. A. (2019). To tune or not to tune? Adapting pretrained representations to diverse tasks. Dalam *Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019)* (hlm. 7–14). Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-4302

Prasetiyo, S., Gustiawan, R., & Rizzel Albani, F. (2024). Analisis pertumbuhan pengguna internet di Indonesia. *BIIKMA: Buletin Ilmiah Ilmu Komputer dan Multimedia, 2*(1). https://jurnalmahasiswa.com/index.php/biikma

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3982–3992. https://doi.org/10.18653/v1/D19-1410

Röttger, P., Vidgen, B., Hovy, D., & Pierrehumbert, J. B. (2022). Two contrasting data annotation paradigms for subjective NLP tasks. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 175–190. https://doi.org/10.18653/v1/2022.naacl-main.13

Sai, A. B., Mohankumar, A. K., & Khapra, M. M. (2022). A survey of evaluation metrics used for NLG systems. *ACM Computing Surveys, 55*(2). https://doi.org/10.1145/3485766

Steele, J. (2025). Indonesia. Dalam Digital News Report 2025. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics. ox.ac.uk/digital-news-report/2025/indonesia

Subakti, A., Murf, H., & Hariadi, N. (2022). The performance of BERT as data representation of text clustering. *Journal of Big Data, 9*(1), 15. https://doi.org/10.1186/s40537-022-00564-9

Supriyono, Wibawa, A. P., Suyono, & Kurniawan, F. (2024). A survey of text summarization: Techniques, evaluation and challenges. *Natural Language Processing Journal, 7*. https://doi.org/10.1016/j.nlp.2024.100070

Tao, Q., Yu, B., Sun, Y., Liu, W., & Xu, X. (2025). Automatic text summary method based on optimized K-Means clustering algorithm and Maximal-Marginal-Relevance algorithm. *SSRN*. https://ssrn.com/abstract=4930036

Trustyanda, N., Rizki, C., & Sri, Q. (2021). Budaya clickbait pada judul berita di era digital 4.0. *6*(9). https://doi.org/10.36418/syntax

Twinandilla, S., Adhy, S., Surarso, B., & Kusumaningrum, R. (2018). Multi-document summarization using K-means and latent Dirichlet allocation (LDA) – Significance sentences. *Procedia Computer Science, 135*, 663–670. https://doi.org/10.1016/j.procs.2018.08.220

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008. http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf

Wang, W., Wei, F., Dong, L., Bao, H., Yang, N., & Zhou, M. (2020). MiniLM: Deep Self-attention distillation for task-agnostic compression of pre-trained transformers. *arXiv*. http://arxiv.org/abs/2002.10957

Wicaksana, D., Pandu Adikara, P., & Adinugroho, S. (2018). Clustering dokumen skripsi dengan menggunakan hierarchical agglomerative clustering. *(Vol. 2, Nomor 12)*. http://j-ptiik.ub.ac.id

Widjanarko, A., Kusumaningrum, R., & Surarso, B. (2018). Multi Document Summarization for the Indonesian Language based on Latent Dirichlet Allocation and Significance Sentence. *2018 International Conference on Information and Communications Technology (ICOIACT)*, 520–524. https://doi.org/10.1109/ICOIACT.2018.8350784

Widyassari, A. P., Rustad, S., Shidik, G. F., Noersasongko, E., Syukur, A., Affandy, A., & Setiadi, D. R. I. M. (2022). Review of automatic text summarization techniques & methods. *Journal of King Saud University - Computer and Information Sciences, 34*(4), 1029–1046. https://doi.org/10.1016/j.jksuci.2020.05.006

LAMPIRAN

Lampiran 1:

SURAT PERNYATAAN

Yang bertanda tangan di bawah ini, kami selaku Anotator/Peringkas Independen:

Nama : Ahlis Dinal Bahtiar
NIM : 24060122130088
Status : Mahasiswa Aktif Informatika
Instansi : Universitas Diponegoro

Nama : Muhammad Faiq Assajjad
NIM : 24060122140116
Status : Mahasiswa Aktif Informatika
Instansi : Universitas Diponegoro

Dengan ini menyatakan bahwa dokumen *Human Ground Truth* berupa ringkasan teks berita yang kami kerjakan adalah murni hasil pemikiran manual dan objektif kami sendiri, tanpa campur tangan *software* otomatis maupun pihak lain.

Data *ground truth* ini disusun secara khusus sebagai standar rujukan pengujian kinerja metrik ROUGE untuk skripsi mahasiswa atas nama Al Ferro Putra Yusanda (NIM: 24060122140164) dengan judul penelitian:

"Peringkasan Multi-Dokumen Ekstraktif Menggunakan Metode Sentence-BERT dan Maximal Marginal Relevance (MMR) Pada Teks Berita Online Bahasa Indonesia".

Demikian surat pernyataan ini kami buat dengan sebenar-benarnya untuk dipergunakan semestinya dalam keperluan akademis.

Semarang, 28 April 2026

Anotator 1,

Ahlis Dinal Bahtiar

Anotator 2,

Muhammad Faiq Assajjad