

**SISTEM RANGKUMAN OTOMATIS BERBASIS AI
UNTUK REDUKSI INFORMATION OVERLOAD
DALAM KOMUNITAS FINANSIAL TELEGRAM
BERJENJANG**

Proposal Tugas Akhir

Oleh

**Jonathan Wiguna
18222019**



**PROGRAM STUDI SISTEM DAN TEKNOLOGI INFORMASI
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG
Oktober 2025**

LEMBAR PENGESAHAN

SISTEM RANGKUMAN OTOMATIS BERBASIS AI UNTUK REDUKSI INFORMATION OVERLOAD DALAM KOMUNITAS FINANSIAL TELEGRAM BERJENJANG

Proposal Tugas Akhir

Oleh

Jonathan Wiguna
18222019

Program Studi Sistem dan Teknologi Informasi
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung

Proposal Tugas Akhir ini telah disetujui dan disahkan
di Bandung, pada tanggal 27 Oktober 2025

Pembimbing

Dr. Ir. John Doe, M.T.
NIP. 123456789

DAFTAR ISI

DAFTAR GAMBAR	v
DAFTAR TABEL	vi
DAFTAR KODE	vii
I PENDAHULUAN	1
I.1 Latar Belakang	1
I.2 Rumusan Masalah	2
I.3 Tujuan	3
I.4 Batasan Masalah	3
I.5 Metodologi	4
I.5.1 Metodologi Penelitian	4
I.5.2 Metodologi Pengembangan	5
II STUDI LITERATUR	7
II.1 <i>Information Overload</i> dan <i>Cognitive Load Theory</i>	7
II.1.1 Konsep dan Dampak <i>Information Overload</i>	7
II.1.2 Landasan Teori: <i>Cognitive Load Theory</i>	8
II.1.3 Ekonomi Perhatian di Media Sosial	9
II.2 <i>Text Summarization</i> dengan <i>Large Language Models</i>	10
II.2.1 Peringkasan Berbasis <i>Transformer</i>	10
II.2.2 Peringkasan Multi-Dokumen dan Sintesis Informasi	11
II.2.3 Peringkasan Percakapan dan Dialog	12
II.2.4 Memori Jangka Panjang dalam Sistem Dialog	12
II.2.5 Metrik Evaluasi untuk Peringkasan	13
II.3 Analisis Sentimen untuk Media Sosial Finansial	13
II.3.1 Model Bahasa untuk Analisis Sentimen Indonesia	13
II.3.2 Analisis Sentimen Domain Finansial	14
II.3.3 Pemanfaatan Sinyal Komunitas dalam Peringkasan	15
II.3.4 Klasifikasi Emosi versus Sentimen untuk Investor Ritel	15
II.4 <i>Named Entity Recognition</i> untuk Teks Finansial Indonesia	16
II.4.1 NER untuk Bahasa Indonesia	17
II.4.2 NER Domain Finansial	17
II.5 <i>Topic Modeling</i> dengan BERTopic	18
II.5.1 BERTopic: Metodologi dan Keunggulan	18

II.5.2	Perkembangan Terkini dalam <i>Topic Modeling</i> Neural	19
II.6	Telegram sebagai Platform Komunitas	19
II.6.1	Karakteristik dan Dinamika Platform Telegram	19
II.6.2	Analisis Skala Besar Komunitas Telegram	20
II.7	Arsitektur Pemrosesan Data <i>Real-Time</i>	21
II.7.1	<i>Pipeline</i> NLP untuk Pemrosesan <i>Stream</i> Teks	21
II.7.2	Arsitektur Berbasis <i>Event</i> dan Performa	22
II.7.3	Pertimbangan <i>Micro-batching</i> dan Frekuensi Data	22
II.7.4	Tantangan Inferensi LLM dalam Sistem <i>Real-Time</i>	23
II.8	Peringkasan untuk Media Sosial dan Percakapan Online	24
II.8.1	Pemanfaatan Sinyal Komunitas dalam Peringkasan	24
II.8.2	Peringkasan Percakapan <i>Multi-speaker</i>	25
II.8.3	Peringkasan Ekstrim untuk Konten Media Sosial	25
II.9	Kesenjangan Penelitian dan Kontribusi	25
III	ANALISIS MASALAH	28
III.1	Analisis Kondisi Saat Ini	28
III.1.1	Model Konseptual Komunitas Telegram Eksisting	28
III.1.2	Identifikasi Masalah Sistem Saat Ini	31
III.1.2.1	<i>Information Overload</i> Ekstrim di Grup PIRANHA	31
III.1.2.2	<i>Information Overload</i> Tinggi di Grup Advanced dengan <i>Coverage Gap</i>	32
III.1.2.3	Fragmentasi Informasi Lintas Struktur Hierarkis	33
III.1.2.4	Keterbatasan Kapasitas Kognitif dan Dampaknya terhadap Kualitas Keputusan	34
III.2	Analisis Kebutuhan	35
III.2.1	Identifikasi <i>Stakeholder</i> dan <i>Pain Points</i>	35
III.2.1.1	Anggota Grup PIRANHA (Target Primer)	35
III.2.1.2	Anggota Grup Advanced (Target Sekunder)	36
III.2.1.3	Anggota Cuap Cuap dan Swing Plan	36
III.2.1.4	Michael Yeoh dan Tim Administrator	36
III.2.2	Kebutuhan Fungsional	37
III.2.3	Kebutuhan Non-Fungsional	38
III.3	Analisis Alternatif Solusi	40
III.3.1	Alternatif Pendekatan untuk Mereduksi <i>Information Overload</i>	41
III.3.1.1	Alternatif 1: Ekspansi <i>Manual Curation</i> (Status Quo yang Ditingkatkan)	41
III.3.1.2	Alternatif 2: Bot Filter Berbasis Aturan (<i>Rule-Based Keyword Bot</i>)	41

III.3.1.3	Alternatif 3: Bot Rangkuman Berbasis <i>AI</i> dengan NLP Khusus Do- main (DIUSULKAN)	42
III.3.1.4	Alternatif 4: Solusi SaaS <i>Generic</i> dengan LLM <i>Third-Party</i>	43
III.3.2	Analisis Pemilihan Solusi	44
IV	DESAIN KONSEP SOLUSI	47
V	RENCANA SELANJUTNYA	48

DAFTAR GAMBAR

I.1	Siklus metodologi CRISP-DM (Chapman dkk. 2000)	5
III.1	Model konseptual sistem komunikasi komunitas Telegram Michael Yeoh saat ini	29

DAFTAR TABEL

III.1	Karakteristik grup dalam komunitas Telegram Michael Yeoh	30
III.2	Kebutuhan fungsional sistem rangkuman otomatis	37
III.3	Kebutuhan non-fungsional sistem rangkuman otomatis	39
III.4	<i>Decision matrix</i> untuk pemilihan alternatif solusi	45

DAFTAR KODE

BAB I

PENDAHULUAN

Bab I adalah pendahuluan yang menceritakan keseluruhan isi dari tugas akhir. Bab ini diawali dengan penjelasan mengenai topik dan latar belakang tugas akhir, dilanjutkan dengan rumusan masalah, tujuan tugas akhir, serta batasan masalah yang dibahas di tugas akhir. Selain itu, Bab I akan menjelaskan tentang pendekatan dan metode yang digunakan di metodologi.

I.1 Latar Belakang

Revolusi digital telah memicu ledakan produksi data; volume data global diproyeksikan akan mencapai 181 zettabyte pada tahun 2025, meningkat drastis dari hanya 2 zettabyte pada tahun 2010 (International Data Corporation March 2021). Pertumbuhan ini melahirkan tantangan *information overload*, di mana volume informasi melampaui kapasitas kognitif individu untuk memprosesnya (Eppler dan Mengis 2004). Menurut *Cognitive Load Theory*, kapasitas memori kerja manusia yang terbatas membuat kualitas pengambilan keputusan dapat menurun hingga 50% saat dihadapkan pada beban informasi berlebih (Sweller 1988; Arnold, Goldschmitt, dan Rigotti June 2023).

Masalah ini menjadi akut dalam komunitas online di media sosial, di mana laju pesan yang tinggi dapat menenggelamkan informasi krusial (Nematzadeh dkk. 2019). Konsekuensinya adalah atensi selektif hingga kelumpuhan dalam pengambilan keputusan (Edmunds dan Morris 2000). Fenomena ini relevan di Indonesia, di mana platform seperti Telegram telah menjadi ekosistem dominan untuk komunitas finansial berskala besar. Didukung oleh fitur seperti kapasitas grup hingga 200.000 anggota, Telegram menjadi medium pilihan dengan lebih dari 27 juta pengguna di Indonesia untuk diskusi investasi *real-time* (AppMagic 2024; Perlo dkk. 2025).

Dominasi Telegram ini berjalan paralel dengan lonjakan jumlah investor ritel di In-

donesia, yang mencapai lebih dari 6 juta investor pada September 2024, dengan pertumbuhan 744.000 investor baru di tahun tersebut (Bursa Efek Indonesia September 2024). Mayoritas investor baru ini adalah generasi muda yang sangat bergantung pada komunitas media sosial untuk pengambilan keputusan investasi (Otoritas Jasa Keuangan 2022; Junaidi dan Nurhidayah 2023). Ketergantungan ini menciptakan masalah dimana akses informasi justru berujung pada *information overload* yang parah, sehingga investor kesulitan memilah sinyal berkualitas dari kebisingan informasi. Tantangan ini terefleksi secara nyata dalam komunitas finansial Michael Yeoh, yang memiliki lebih dari 22.000 anggota dalam struktur hierarkis empat tingkat (MY Cuap Cuap, MY Swing Plan, MY Advanced Group, dan MY PIRANHA). *Information overload* paling akut terjadi di grup PIRANHA yang terdiri dari 2.000 investor ultra-premium dengan volume diskusi melebihi 200 pesan per jam, di mana fragmentasi informasi antar grup semakin memperburuk masalah.

Penelitian akademis sebelumnya cenderung berfokus pada analisis sentimen dari forum publik seperti X (Twitter) dan Reddit untuk tujuan prediksi pergerakan harga saham (misalnya, Bollen, Mao, dan Zeng (2011); Pawar dkk. (2022); Cheng, Ge, dan Li (2023)). Sementara itu, solusi praktis seperti bot Telegram yang ada saat ini bersifat generik, tidak mampu melakukan sintesis lintas-grup, dan tidak dioptimalkan untuk domain finansial Indonesia yang spesifik (misalnya, Xu dkk. (2021); Beebom (2023)). Kedua pendekatan tersebut tidak dirancang untuk menyelesaikan masalah operasional internal sebuah komunitas.

Kondisi *information overload* parah dan fragmentasi informasi akibat struktur komunitas berjenjang ini menyoroti permasalahan signifikan yang perlu diteliti. Fenomena ini tidak hanya berisiko mengurangi efektivitas anggota dalam mengambil keputusan investasi, tetapi juga mengikis nilai dari komunitas itu sendiri.

I.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan pada subbab I.1, rumusan masalah yang akan dibahas pada tugas akhir ini adalah:

1. Bagaimana merancang dan mengembangkan sistem rangkuman otomatis berbasis *AI* yang dapat mereduksi *information overload* melalui sintesis intelijen lintas-grup dalam komunitas finansial Telegram berjenjang?
2. Bagaimana meningkatkan efektivitas sistem rangkuman otomatis tersebut berdasarkan *feedback* dan evaluasi dari pengguna komunitas?

I.3 Tujuan

Berdasarkan latar belakang dan rumusan masalah yang telah dijelaskan, tujuan dari tugas akhir ini adalah:

1. Merancang, membangun, dan mengevaluasi sistem *prototype* rangkuman otomatis berbasis *AI* yang mampu melakukan sintesis intelijen lintas-grup untuk mereduksi *information overload* dalam komunitas finansial Telegram.
2. Melakukan iterasi perbaikan sistem berdasarkan *feedback* pengguna untuk meningkatkan efektivitas dan *usability* sistem dalam konteks komunitas yang diteliti.

I.4 Batasan Masalah

Batasan masalah digunakan untuk menjaga pelebaran dari pokok masalah dan mengarahkan tugas akhir menjadi lebih spesifik. Batasan masalah dari tugas akhir ini adalah:

1. Konteks penelitian dan implementasi sistem terbatas secara eksklusif pada ekosistem komunitas investasi Michael Yeoh di Telegram, yang terdiri dari empat grup berjenjang: MY Cuap Cuap (18.000 anggota, *read-only* untuk informasi umum), MY Swing Plan (grup sinyal *trading* dengan format terstruktur), MY Advanced Group (4.000 anggota premium dengan komunikasi interaktif dua arah), dan MY PIRANHA (2.000 anggota ultra-premium dengan volume diskusi tertinggi melebihi 200 pesan per jam).
2. Arsitektur sistem menggunakan tumpukan teknologi (*technology stack*) yang telah ditentukan, meliputi model NLP *cahya/bert-base-indonesian-NER* untuk ekstraksi entitas finansial, *indonesian-roberta-base-emotion-classifier* untuk analisis emosi, BERTopic untuk pemodelan topik, serta pemanfaatan Groq API dengan model bahasa Llama 3.1 70B untuk proses generasi rangkuman.
3. Keluaran (*output*) utama dari sistem adalah rangkuman terstruktur yang dihasilkan secara periodik setiap jam (11:00, 12:00, 14:00, 15:00, 16:00) sebagai suplemen rangkuman manual resmi yang ada. Penelitian ini tidak mencakup pengembangan sistem peringatan (*alert*) *real-time* atau fitur percakapan interaktif.
4. Sumber data untuk analisis terbatas pada pesan berbasis teks yang ada di dalam grup-grup Telegram yang ditentukan. Sistem tidak mengintegrasikan sumber data eksternal seperti API harga saham *real-time* atau umpan berita finansial dari portal media.
5. Sistem menerapkan prinsip privasi dengan TIDAK memproses pesan yang

telah di-*unsend* oleh pengirim, mengakui bahwa tindakan *unsend* merupakan indikasi eksplisit penarikan kembali informasi dari domain publik komunitas. Implementasi ini memastikan bahwa hanya konten yang secara sadar dipublikasikan secara permanen oleh anggota yang akan dianalisis oleh sistem.

I.5 Metodologi

Pengerjaan tugas akhir ini menggunakan dua metodologi, yaitu metodologi untuk penelitian dan metodologi pengembangan.

I.5.1 Metodologi Penelitian

Metodologi penelitian yang digunakan adalah *Design Science Research Methodology* (DSRM). Pendekatan DSRM dipilih karena relevansinya yang tinggi dengan penelitian di bidang sistem informasi yang bertujuan untuk merancang dan mengevaluasi artefak teknologi inovatif guna memecahkan masalah di dunia nyata. Proses DSRM terdiri dari enam tahapan yang akan diikuti dalam penelitian ini:

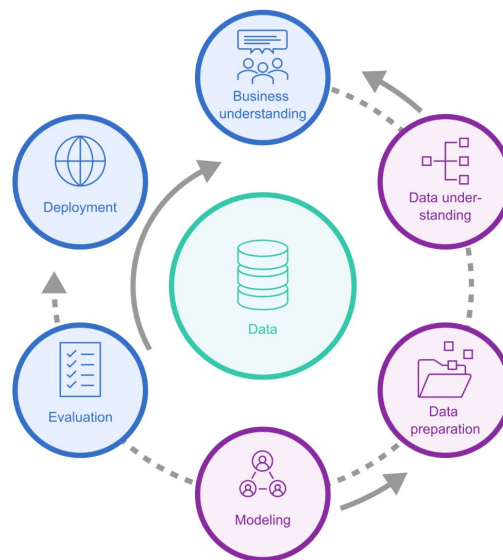
1. **Identifikasi Masalah & Motivasi:** Tahap ini telah diuraikan dalam latar belakang, di mana masalah spesifik yang diidentifikasi adalah *information overload* dan fragmentasi informasi dalam komunitas finansial Telegram yang besar dan berjenjang. Motivasi utamanya adalah kebutuhan untuk meningkatkan efisiensi akses informasi bagi anggota komunitas.
2. **Perumusan Tujuan Solusi:** Berdasarkan masalah yang ada, tujuan solusi didefinisikan secara kuantitatif dan kualitatif. Tujuan utamanya adalah menghasilkan sebuah artefak (sistem) yang terbukti mampu mereduksi beban informasi, yang akan diukur melalui akurasi identifikasi topik dan efisiensi akses informasi yang dirasakan pengguna.
3. **Desain & Pengembangan:** Tahap ini merupakan aktivitas inti dalam menciptakan artefak penelitian. Pada tugas akhir ini, artefak yang dirancang dan dikembangkan adalah sebuah sistem purwarupa untuk rangkuman otomatis. Proses ini mencakup perancangan arsitektur, implementasi *pipeline* pemrosesan data, dan integrasi model *NLP*.
4. **Demonstrasi:** Pada tahap ini, artefak yang telah dibangun akan didemonstrasikan untuk menyelesaikan masalah yang telah diidentifikasi. Sistem purwarupa akan dioperasikan menggunakan data percakapan nyata dari komunitas Michael Yeoh untuk menunjukkan kemampuannya dalam mengubah data mentah menjadi rangkuman yang terstruktur.
5. **Evaluasi:** Artefak akan dievaluasi secara sistematis untuk mengukur sejauh

mana tujuan solusi tercapai. Evaluasi akan membandingkan kondisi sebelum dan sesudah penggunaan sistem, menggunakan metrik yang telah ditetapkan seperti akurasi, tingkat kompresi pesan, dan umpan balik kualitatif dari pengguna melalui survei.

6. **Komunikasi:** Tahap terakhir adalah mengkomunikasikan masalah, artefak yang dibangun, serta hasil evaluasinya kepada audiens akademis. Komunikasi ini diwujudkan melalui penulisan laporan tugas akhir ini serta presentasi dalam forum seminar dan sidang.

I.5.2 Metodologi Pengembangan

Metodologi pengembangan yang digunakan dalam penelitian ini adalah *Cross Industry Standard Process for Data Mining* (CRISP-DM). Metodologi CRISP-DM dipilih karena relevansinya yang tinggi untuk proyek berbasis *machine learning* dan *natural language processing*, serta kemampuannya dalam memberikan kerangka kerja terstruktur untuk proyek *data mining* di berbagai industri (Chapman dkk. 2000). CRISP-DM terdiri dari enam fase utama yang bersifat iteratif dan siklik, memungkinkan pengulangan dan penyempurnaan di setiap tahap pengembangan. Metodologi CRISP-DM dapat dilihat pada Gambar I.1.



Gambar I.1 Siklus metodologi CRISP-DM (Chapman dkk. 2000)

Metodologi CRISP-DM yang diadaptasi untuk pengembangan sistem rangkuman otomatis ini terdiri dari enam fase berikut:

1. **Business Understanding:** Memahami kebutuhan komunitas finansial Telegram dan mendefinisikan tujuan proyek dari perspektif bisnis, yaitu mere-

duksi *information overload* melalui rangkuman otomatis.

2. **Data Understanding:** Mengumpulkan dan menganalisis data pesan dari tiga grup Telegram target untuk memahami karakteristik, volume, dan pola komunikasi dalam komunitas.
3. **Data Preparation:** Membersihkan, mentransformasi, dan mempersiapkan data pesan untuk pemrosesan NLP, termasuk normalisasi teks dan ekstraksi fitur linguistik.
4. **Modeling:** Mengintegrasikan model NLP untuk analisis sentimen, ekstraksi entitas, identifikasi topik, dan generasi rangkuman menggunakan *large language model*.
5. **Evaluation:** Mengevaluasi kualitas rangkuman yang dihasilkan sistem melalui metrik objektif dan *feedback* subjektif dari pengguna komunitas.
6. **Deployment:** Mengimplementasikan sistem secara operasional dalam komunitas, termasuk penjadwalan otomatis dan penyampaian rangkuman periodik ke kanal Telegram.

Penjelasan detail mengenai aktivitas dan keluaran dari setiap fase metodologi CRISP-DM dalam konteks penelitian ini akan diuraikan lebih lanjut pada Bab III (Analisis dan Perancangan Sistem).

BAB II

STUDI LITERATUR

Bab ini menyajikan tinjauan komprehensif terhadap literatur yang relevan dengan pengembangan sistem rangkuman otomatis berbasis *AI* untuk mereduksi *information overload* dalam komunitas finansial Telegram. Studi literatur ini mencakup tujuh domain utama: (1) *Information Overload* dan *Cognitive Load Theory*, (2) *Text Summarization* dengan *Large Language Models*, (3) Analisis Sentimen untuk Media Sosial Finansial, (4) *Named Entity Recognition* untuk Teks Finansial Indonesia, (5) *Topic Modeling* dengan BERTopic, (6) Telegram sebagai Platform Komunitas, dan (7) Arsitektur Pemrosesan Data *Real-Time*. Setiap domain memberikan landasan teoretis dan metodologis yang mendukung desain, implementasi, dan evaluasi sistem yang dikembangkan dalam penelitian ini.

II.1 *Information Overload* dan *Cognitive Load Theory*

Information overload merupakan fenomena di mana individu dihadapkan pada volume informasi yang melampaui kapasitas kognitif mereka untuk memproses secara efektif (Eppler dan Mengis 2004). Dalam konteks komunitas online, khususnya platform media sosial seperti Telegram, masalah ini menjadi semakin akut karena tingginya laju produksi dan distribusi informasi. Bagian ini mengulas teori-teori fundamental yang menjelaskan bagaimana *information overload* memengaruhi pengambilan keputusan, serta mengidentifikasi strategi mitigasi yang dapat diterapkan melalui sistem otomatis berbasis *AI*.

II.1.1 Konsep dan Dampak *Information Overload*

Penelitian seminal oleh Eppler dan Mengis (2004) menyediakan landasan teoretis yang komprehensif tentang konsep *information overload*. Berdasarkan tinjauan literatur dari berbagai disiplin ilmu termasuk *Organization Science*, Akuntansi, Pemasaran, dan Sistem Informasi Manajemen, mereka mendefinisikan *information*

overload sebagai kondisi di mana kualitas pengambilan keputusan menurun ketika kuantitas informasi yang tersedia melampaui kapasitas pemrosesan individu. Studi tersebut mengidentifikasi hubungan inverted U-curve antara kuantitas informasi dan kualitas keputusan, yang menunjukkan bahwa setelah mencapai titik optimal, penambahan informasi justru menurunkan performa kognitif. Temuan ini sangat relevan dengan konteks komunitas Telegram finansial yang menghasilkan ratusan hingga ribuan pesan per hari, yang berpotensi menempatkan anggota komunitas pada sisi kanan kurva tersebut di mana *overload* terjadi.

Shahrzadi dkk. (2024) memperkuat pemahaman ini melalui *scoping review* terbaru yang menganalisis penyebab, konsekuensi, dan strategi penanganan *information overload*. Penelitian mereka mengidentifikasi tiga kategori penyebab utama: (1) faktor personal seperti keterbatasan kapasitas kognitif dan kurangnya keterampilan manajemen informasi, (2) karakteristik informasi seperti ambiguitas, kompleksitas, dan ketidakpastian, dan (3) parameter teknologi informasi termasuk intensitas notifikasi dan desain antarmuka. Dampak yang diidentifikasi mencakup penurunan produktivitas, kualitas keputusan yang buruk, stres, dan kelelahan kognitif. Yang paling signifikan bagi penelitian ini adalah identifikasi mereka terhadap strategi mitigasi berbasis teknologi, khususnya penggunaan *filtering tools* dan sistem otomatis untuk mereduksi volume informasi, yang memberikan justifikasi langsung untuk pengembangan bot rangkuman otomatis.

II.1.2 Landasan Teori: *Cognitive Load Theory*

Cognitive Load Theory (CLT), yang dikembangkan oleh Sweller (1988), memberikan kerangka teoretis untuk memahami keterbatasan kapasitas memori kerja manusia dalam pemrosesan informasi. Teori ini mengidentifikasi tiga jenis beban kognitif: (1) *intrinsic load*, yang inheren dengan kompleksitas materi yang dipelajari, (2) *extraneous load*, yang dihasilkan oleh cara informasi disajikan, dan (3) *germane load*, yang berkontribusi pada pembentukan skema dan otomatisasi. Dalam konteks komunitas online, anggota menghadapi *intrinsic load* yang tinggi karena kompleksitas informasi finansial, sementara desain antarmuka platform media sosial sering kali meningkatkan *extraneous load* melalui notifikasi, pesan yang tidak terstruktur, dan kurangnya organisasi informasi.

Skulmowski dan Rey (2022) mengembangkan pemahaman CLT dengan mengadaptasinya untuk lingkungan digital dan pembelajaran daring. Penelitian mereka mengusulkan perspektif baru tentang *extraneous cognitive load* yang mempertimbangkan karakteristik unik media digital, seperti konten yang kaya secara perseptual, distrak-

si visual, dan desain antarmuka yang kompleks. Mereka berargumen bahwa media digital menciptakan bentuk *extraneous load* yang berbeda dari media tradisional, terutama melalui elemen-elemen yang tidak relevan dengan tujuan pembelajaran atau pemrosesan informasi inti. Temuan ini sangat relevan untuk memahami bagaimana antarmuka Telegram yang penuh dengan pesan, emoji, dan media visual dapat meningkatkan beban kognitif ekstrinsik bagi anggota yang mencoba mengekstrak informasi investasi yang relevan.

Penelitian empiris oleh Sidnam-Mauch dan Monge (2024) menerapkan CLT secara khusus pada konteks distraksi media sosial melalui studi dengan 1.026 partisipan. Mereka mengintegrasikan *load theory of attention* dengan kerangka *uses and gratifications* untuk menjelaskan bagaimana media sosial mobile menciptakan distraksi melalui sumber daya kognitif yang terbatas. Studi mereka mendemonstrasikan bahwa individu dengan kapasitas memori kerja yang lebih rendah lebih rentan terhadap distraksi media sosial, dan bahwa *affordances* teknologi seperti notifikasi *push* dan aksesibilitas konstan memperburuk masalah ini. Temuan ini memperkuat argumen bahwa komunitas Telegram finansial dengan lebih dari 22.000 anggota dan aliran pesan yang konstan menciptakan lingkungan yang sangat rentan terhadap *cognitive overload* dan distraksi, yang pada akhirnya dapat menurunkan kualitas pengambilan keputusan investasi.

II.1.3 Ekonomi Perhatian di Media Sosial

Heitmayer (2025) memberikan perspektif teoretis terbaru dengan mengusulkan bahwa perhatian (*attention*) berfungsi sebagai mata uang simbolis universal di media sosial. Penelitian ini memperkenalkan model *dual-stream* yang membedakan antara perhatian yang "mengalir" (*flow attention*) dan perhatian yang "mengeras" (*calcified attention*). *Flow attention* mengacu pada alokasi perhatian yang dinamis dan sementara terhadap konten yang terus berubah, sementara *calcified attention* merujuk pada investasi perhatian jangka panjang yang termanifestasi dalam bentuk *followers*, *likes*, atau keanggotaan komunitas. Dalam konteks komunitas finansial Telegram, kedua bentuk perhatian ini bersaing: anggota harus mengalokasikan *flow attention* mereka untuk memproses aliran pesan yang konstan, sementara *calcified attention* mereka termanifestasi dalam komitmen berkelanjutan terhadap komunitas dan kepercayaan terhadap sumber informasi tertentu seperti Michael Yeoh. Kelangkaan perhatian (*attention scarcity*) yang diidentifikasi dalam penelitian ini menjadi masalah sentral yang harus diatasi oleh sistem rangkuman otomatis dengan memaksimalkan efisiensi alokasi perhatian anggota.

Analisis empiris oleh Nematzadeh dkk. (2019) tentang *information overload* dalam komunikasi grup menggunakan data dari platform Twitch menunjukkan bagaimana peningkatan volume pesan dalam grup dapat menyebabkan transisi dari percakapan yang koheren menjadi "kakofoni" di mana informasi bermakna tenggelam dalam kebisingan. Mereka mengidentifikasi bahwa ketika laju pesan melebihi ambang batas tertentu, kualitas percakapan menurun drastis, partisipasi menjadi tidak merata, dan informasi penting hilang. Fenomena ini sangat relevan dengan komunitas Telegram finansial yang diteliti, di mana pesan dapat muncul dengan laju ratusan per jam selama jam perdagangan aktif. Sistem rangkuman otomatis yang dirancang dalam penelitian ini bertujuan untuk mengatasi transisi menuju kakofoni ini dengan menyaring dan mensintesis informasi kunci secara periodik.

II.2 Text Summarization dengan Large Language Models

Peringkasan teks otomatis telah mengalami evolusi signifikan dengan munculnya model bahasa berbasis *transformer* dan *large language models* (LLMs). Bagian ini mengulas perkembangan metodologi peringkasan teks dari pendekatan berbasis ekstraksi hingga generasi abstraktif dengan LLMs, dengan fokus khusus pada tantangan peringkasan multi-dokumen dan percakapan yang relevan dengan konteks komunitas Telegram.

II.2.1 Peringkasan Berbasis Transformer

Liu dan Lapata (2019) memperkenalkan pendekatan *BERTSum* yang memanfaatkan representasi BERT untuk peringkasan ekstraktif dan abstraktif. Penelitian mereka mendemonstrasikan bahwa model *encoder* tingkat dokumen yang dibangun di atas BERT dengan lapisan *Transformer* antar-kalimat dapat menghasilkan rangkuman berkualitas tinggi. Pendekatan dua tahap *fine-tuning* yang mereka usulkan, di mana model pertama-tama di-*fine-tune* untuk tugas ekstraksi kalimat sebelum dilatih untuk generasi abstraktif, menetapkan metodologi dasar yang diadopsi oleh banyak penelitian berikutnya. Kontribusi kunci dari penelitian ini adalah demonstrasi bahwa model bahasa pra-latih dapat secara efektif menangkap struktur dokumen dan hubungan semantik antar-kalimat, yang esensial untuk menghasilkan rangkuman yang koheren dan informatif.

Pemahaman tentang *trade-off* antara peringkasan ekstraktif dan abstraktif diperluas oleh penelitian Pilault dkk. (2020), yang membandingkan kedua pendekatan menggunakan model *transformer* pada dokumen panjang. Evaluasi manusia mereka mengungkapkan temuan penting: sementara pendekatan abstraktif berbasis *tran-*

sformer unggul dalam koherensi dan kelancaran, metode ekstraktif cenderung mencetak lebih tinggi dalam hal informativeness. Temuan ini menginformasikan desain sistem dalam penelitian ini, di mana pendekatan hibrida digunakan: BERTopic untuk ekstraksi topik dan entitas kunci (pendekatan ekstraktif), dikombinasikan dengan Llama 3.1 70B untuk generasi rangkuman naratif yang koheren (pendekatan abstraktif).

II.2.2 Peringkasan Multi-Dokumen dan Sintesis Informasi

Tantangan peringkasan multi-dokumen, yang sangat relevan dengan konteks sintesis lintas tiga grup Telegram, dibahas secara komprehensif oleh DeYoung dkk. (2024). Penelitian mereka mengajukan pertanyaan fundamental: apakah model peringkasan multi-dokumen modern benar-benar dapat mensintesis informasi, atau hanya melakukan agregasi sederhana? Melalui evaluasi terhadap berbagai model dari *fine-tuned transformers* hingga GPT-4, mereka menemukan bahwa sebagian besar model mengalami kesulitan dalam mensintesis informasi yang bertentangan atau berbeda perspektif, dan sangat sensitif terhadap urutan input dokumen. Temuan kritis ini memiliki implikasi langsung bagi desain sistem dalam penelitian ini: sistem harus dirancang dengan mekanisme eksplisit untuk menangani informasi yang mungkin bertentangan antara grup "Cuap Cuap", "Swing Plan", dan "Advanced Group", dan tidak boleh bergantung semata-mata pada kemampuan sintesis implisit LLM.

Ravaut dkk. (2024) menyediakan wawasan penting tentang bagaimana LLMs memanfaatkan konteks panjang dalam peringkasan. Penelitian mereka mengungkapkan masalah *position bias*, di mana LLMs cenderung memberikan bobot yang tidak merata terhadap informasi berdasarkan posisinya dalam konteks. Mereka mengevaluasi enam LLMs berbeda pada sepuluh dataset dengan lima metrik evaluasi, menemukan bahwa model cenderung memberikan perhatian berlebihan pada informasi di awal dan akhir konteks, sementara mengabaikan informasi di tengah. Untuk mengatasi masalah ini, mereka memperkenalkan benchmark *MiddleSum* dan mengevaluasi metode hierarkis dan inkremental untuk peringkasan. Temuan ini sangat relevan untuk implementasi Llama 3.1 70B dalam penelitian ini, yang harus memproses konteks panjang (potensi ratusan pesan dalam satu jam). Desain sistem harus mempertimbangkan strategi mitigasi *position bias*, seperti pengelompokan pesan berdasarkan topik sebelum peringkasan atau penggunaan rangkuman inkremental.

II.2.3 Peringkasan Percakapan dan Dialog

Gliwa dkk. (2019) memperkenalkan *SAMSum Corpus*, dataset dialog yang dianotasi secara manual untuk peringkasan abstraktif, yang menetapkan bahwa peringkasan percakapan memiliki tantangan unik dibandingkan dengan peringkasan artikel berita. Dialog dan percakapan grup memiliki struktur yang lebih kompleks dengan pergantian pembicara, konteks implisit, referensi anaforik, dan topik yang bercabang. Dataset mereka mendemonstrasikan pentingnya memahami struktur dialog dan dinamika percakapan untuk menghasilkan rangkuman yang akurat dan informatif. Karakteristik ini sangat relevan dengan konteks Telegram, di mana percakapan dapat melibatkan puluhan pembicara dengan topik yang berubah cepat dan sering kali tumpang tindih.

Tian, Xia, dan Song (2024) mengembangkan metodologi peringkasan dialog lebih lanjut dengan mengusulkan pendekatan *Mixture of Experts* (MoE) berbasis LLM. Penelitian mereka memperkenalkan mekanisme *role-oriented routing* yang memilih model atau strategi peringkasan yang sesuai berdasarkan karakteristik informasi percakapan yang berbeda. Pendekatan ini sangat relevan dengan konteks komunitas finansial Telegram yang berjenjang, di mana pesan dari berbagai peran (misalnya, Michael Yeoh sebagai pemimpin komunitas versus anggota biasa) mungkin memerlukan perlakuan berbeda dalam proses peringkasan. Sistem yang dirancang dalam penelitian ini mengadopsi konsep serupa melalui *weighted sentiment analysis* yang memberikan bobot berbeda berdasarkan peran pengirim pesan.

II.2.4 Memori Jangka Panjang dalam Sistem Dialog

Konsep memori jangka panjang dalam sistem percakapan dieksplorasi oleh Wang dkk. (2023), yang mengusulkan metode peringkasan rekursif untuk memungkinkan LLMs mempertahankan konteks percakapan yang panjang. Pendekatan mereka melibatkan pembuatan rangkuman hierarkis dari konteks dialog yang kemudian digunakan untuk menjaga konsistensi dalam percakapan yang berkepanjangan. Konsep ini paralel dengan desain sistem rangkuman per jam dalam penelitian ini, di mana setiap rangkuman per jam dapat dipandang sebagai pembentukan "memori jangka panjang" bagi komunitas. Akumulasi rangkuman per jam ini dapat membentuk representasi terkompresi dari diskusi komunitas yang memungkinkan anggota untuk memahami tren dan perkembangan informasi tanpa harus membaca setiap pesan individual.

II.2.5 Metrik Evaluasi untuk Peringkasan

Evaluasi kualitas rangkuman otomatis merupakan aspek kritis yang dibahas oleh Zhang dkk. (2020). Mereka memperkenalkan *BERTScore*, metrik evaluasi yang menggunakan *contextual embeddings* dari BERT untuk menghitung kesamaan antara rangkuman yang dihasilkan dan rangkuman referensi, mengatasi keterbatasan metrik berbasis *n-gram* seperti ROUGE yang hanya menghitung kecocokan kata eksak. *BERTScore* mendemonstrasikan korelasi yang lebih baik dengan penilaian manusia karena dapat mengenali parafrase dan kesamaan semantik yang tidak terangkap oleh metrik tradisional. Penelitian ini mengadopsi *BERTScore* bersama dengan metrik lain untuk evaluasi komprehensif kualitas rangkuman yang dihasilkan sistem.

Namun, Shen dkk. (2023) memberikan peringatan penting tentang penggunaan LLMs sebagai evaluator otomatis untuk peringkasan abstraktif. Melalui analisis ekstensif terhadap ChatGPT dan GPT-4 sebagai evaluator, mereka menemukan masalah signifikan dalam stabilitas, reliabilitas, dan ketergantungan pada dimensi evaluasi tertentu. Temuan mereka menunjukkan bahwa evaluasi otomatis berbasis LLM tidak dapat sepenuhnya menggantikan evaluasi manusia, terutama untuk aspek kualitatif seperti koherensi naratif dan relevansi kontekstual. Implikasi bagi penelitian ini adalah perlunya kombinasi antara metrik otomatis (*BERTScore*, ROUGE) dan evaluasi subjektif melalui *feedback* pengguna komunitas untuk mendapatkan penilaian holistik terhadap kualitas rangkuman.

II.3 Analisis Sentimen untuk Media Sosial Finansial

Analisis sentimen merupakan komponen esensial dalam memahami dinamika diskusi finansial di media sosial. Bagian ini mengulas perkembangan metodologi analisis sentimen, dengan fokus khusus pada domain finansial dan pemrosesan bahasa Indonesia, serta mengeksplorasi bagaimana sinyal sentimen dari komunitas dapat dimanfaatkan untuk meningkatkan kualitas rangkuman.

II.3.1 Model Bahasa untuk Analisis Sentimen Indonesia

Wilie dkk. (2020) memperkenalkan *IndoNLU*, benchmark komprehensif untuk evaluasi pemahaman bahasa Indonesia yang mencakup dua belas tugas termasuk analisis sentimen. Dataset mereka dilatih pada *Indo4B*, korpus bahasa Indonesia dengan empat miliar kata, dan menyediakan model IndoBERT yang mengungguli *multilingual BERT* (mBERT) pada berbagai tugas. Penelitian ini mendemonstra-

sikan bahwa model bahasa monolingual yang dilatih khusus pada korpus bahasa Indonesia menghasilkan performa superior dibandingkan model multilingual untuk tugas-tugas NLP Indonesia. Kontribusi mereka menetapkan IndoBERT sebagai baseline untuk berbagai aplikasi NLP Indonesia, termasuk implementasi analisis sentimen dalam penelitian ini yang menggunakan *indonesian-roberta-base-emotion-classifier*.

Penelitian oleh Koto dkk. (2020) memperkuat temuan ini dengan memperkenalkan *IndoLEM*, benchmark dataset dan model IndoBERT untuk NLP Indonesia. Mereka mendemonstrasikan bahwa *pre-training* pada korpus bahasa Indonesia yang besar (220 juta kata) menghasilkan representasi bahasa yang lebih baik untuk tugas-tugas *downstream* termasuk analisis sentimen, yang mencapai F1-score 84,13%. Penelitian mereka juga mengeksplorasi *transfer learning* dari model multibahasa, menunjukkan bahwa meskipun model multibahasa dapat memberikan titik awal yang baik, *fine-tuning* pada data Indonesia sangat meningkatkan performa.

Nugroho dkk. (2021) menerapkan *fine-tuning* BERT untuk analisis sentimen pada ulasan aplikasi mobile berbahasa Indonesia, konteks yang memiliki kesamaan dengan teks media sosial karena sifatnya yang informal dan sering mengandung bahasa gaul atau singkatan. Penelitian mereka membandingkan model multibahasa dengan model khusus Indonesia, mendemonstrasikan keunggulan *transfer learning* untuk bahasa Indonesia dengan data berlabel terbatas. Temuan mereka sangat relevan dengan penelitian ini, karena pesan Telegram finansial sering kali menggunakan bahasa informal, singkatan, dan istilah khusus yang memerlukan model yang robust terhadap variasi linguistik.

II.3.2 Analisis Sentimen Domain Finansial

Araci (2019) memperkenalkan *FinBERT*, model BERT pertama yang dilatih khusus pada korpus finansial untuk analisis sentimen. Model ini mencapai peningkatan akurasi 15% dibandingkan BERT umum pada dataset *Financial PhraseBank* dan *FiQA*. Penelitian ini mendemonstrasikan pentingnya *domain-specific pre-training* untuk teks finansial, yang memiliki karakteristik linguistik unik termasuk terminologi khusus, struktur kalimat formal, dan nuansa sentimen yang berbeda dari domain umum. Konsep sentimen dalam konteks finansial lebih kompleks daripada klasifikasi positif-negatif sederhana; istilah seperti "volatilitas tinggi" dapat memiliki konotasi positif atau negatif tergantung konteks strategi investasi. Temuan ini memotivasi penggunaan model analisis sentimen yang disesuaikan untuk konteks finansial dalam penelitian ini, meskipun disesuaikan untuk bahasa Indonesia.

Si dkk. (2014) mengeksplorasi pemanfaatan sentimen media sosial untuk prediksi pergerakan saham dengan mengintegrasikan analisis sentimen Twitter dengan struktur relasi sosial. Penelitian mereka mendemonstrasikan bahwa menggabungkan sentimen dengan informasi jaringan sosial (siapa yang mempengaruhi siapa) secara signifikan meningkatkan akurasi prediksi pergerakan harga saham. Temuan kunci mereka adalah bahwa tidak semua opini memiliki bobot yang sama; pendapat dari individu yang lebih berpengaruh atau kredibel dalam jaringan sosial harus diberi bobot lebih tinggi. Konsep ini memiliki paralel langsung dengan desain sistem dalam penelitian ini, di mana implementasi *weighted sentiment analysis* memberikan bobot berbeda berdasarkan peran pengirim (Michael Yeoh versus anggota biasa), mengakui bahwa tidak semua kontributor dalam komunitas memiliki tingkat pengaruh atau keahlian yang sama.

II.3.3 Pemanfaatan Sinyal Komunitas dalam Peringkasan

Kano dkk. (2018) mengusulkan metodologi inovatif untuk memanfaatkan sinyal popularitas dalam media sosial (seperti *votes*, *shares*, *bookmarks*) sebagai label jarak jauh (*distant labels*) untuk peringkasan ekstraktif percakapan online. Penelitian mereka mendemonstrasikan bahwa metrik keterlibatan komunitas dapat berfungsi sebagai indikator kualitas dan relevansi konten, memisahkan kontribusi konten dari faktor kontekstual. Pendekatan ini sangat relevan dengan konteks komunitas Telegram finansial berjenjang dalam penelitian ini, di mana struktur hierarkis (grup "Cuap Cuap", "Swing Plan", "Advanced Group") dan peran pengguna (pemimpin komunitas versus anggota biasa) dapat dimanfaatkan sebagai sinyal implisit tentang pentingnya informasi. Prinsip pemanfaatan sinyal komunitas ini diadopsi secara langsung dalam penelitian ini, di mana "peran pengguna" (misalnya, admin Michael Yeoh versus anggota biasa) digunakan sebagai proksi untuk "popularitas" atau "kredibilitas" informasi, yang kemudian menjadi dasar bagi fitur *weighted sentiment*. Sistem yang dikembangkan mengimplementasikan *weighted sentiment* yang memberikan bobot lebih tinggi pada pesan dari sumber yang lebih kredibel atau berpengaruh dalam komunitas.

II.3.4 Klasifikasi Emosi versus Sentimen untuk Investor Ritel

Pemilihan *indonesian-roberta-base-emotion-classifier* dalam penelitian ini memerlukan justifikasi teoretis, karena berbeda dari pendekatan analisis sentimen finansial konvensional yang fokus pada klasifikasi bullish/bearish atau positif/negatif. Perbedaan fundamental ini didasarkan pada karakteristik unik investor ritel dibandingkan dengan investor institusional atau analis profesional.

Penelitian perilaku keuangan menunjukkan bahwa keputusan investasi ritel sangat dipengaruhi oleh emosi, terutama ketakutan (*fear*) dan keserakahan (*greed*). Berbeda dengan analis profesional yang dilatih untuk membuat keputusan berdasarkan analisis rasional, investor ritel sering kali membuat keputusan yang didorong oleh reaksi emosional terhadap pergerakan pasar atau diskusi komunitas. Dalam konteks komunitas Telegram dengan lebih dari 22.000 anggota yang sebagian besar adalah investor ritel muda, emosi kolektif yang terekspresikan dalam diskusi dapat berfungsi sebagai indikator sentimen pasar yang lebih akurat daripada sekadar klasifikasi positif/negatif.

Model klasifikasi emosi yang digunakan dalam penelitian ini dapat mendeteksi spektrum emosi yang lebih nuansa, termasuk ketakutan, kegembiraan, kemarahan, dan kesedihan. Emosi-emosi ini memberikan sinyal yang lebih kaya tentang psikologi pasar dibandingkan dengan sentimen biner. Misalnya, ekspresi ketakutan tinggi dalam diskusi komunitas dapat mengindikasikan potensi *panic selling*, sementara kegembiraan berlebihan dapat menandakan *euphoria* yang sering kali mendahului koreksi pasar. Pendekatan berbasis emosi ini selaras dengan teori *behavioral finance* yang mengidentifikasi bias psikologis sebagai faktor kunci dalam pengambilan keputusan investasi ritel.

Lebih lanjut, konteks diskusi informal di Telegram lebih cenderung mengekspresikan emosi autentik dibandingkan dengan laporan finansial formal atau analisis profesional. Anggota komunitas sering kali berbagi reaksi emosional mereka terhadap pergerakan harga, keputusan investasi, atau berita pasar dengan cara yang langsung dan tidak tersaring. Klasifikasi emosi dapat menangkap sinyal-sinyal psikologis ini yang hilang dalam analisis sentimen tradisional, memberikan wawasan yang lebih mendalam tentang suasana hati (*mood*) komunitas dan potensi implikasinya terhadap perilaku investasi kolektif.

II.4 *Named Entity Recognition* untuk Teks Finansial Indonesia

Named Entity Recognition (NER) merupakan tugas fundamental dalam NLP yang bertujuan mengidentifikasi dan mengklasifikasikan entitas bernama dalam teks, seperti nama orang, organisasi, lokasi, dan dalam konteks finansial, nama perusahaan, kode saham (ticker), dan istilah finansial spesifik. Bagian ini mengulas perkembangan metodologi NER untuk bahasa Indonesia dan domain finansial, yang menjadi dasar implementasi ekstraksi entitas dalam sistem yang dikembangkan.

II.4.1 NER untuk Bahasa Indonesia

Koto dkk. (2020) menetapkan benchmark untuk NER bahasa Indonesia melalui dataset *IndoLEM* dan model IndoBERT. Penelitian mereka mendemonstrasikan bahwa IndoBERT mencapai performa *state-of-the-art* pada tugas NER Indonesia, mengungguli mBERT dengan margin yang signifikan. Keberhasilan ini dikaitkan dengan kemampuan model untuk menangkap nuansa linguistik bahasa Indonesia, termasuk struktur morfologi, pola afiksasi, dan konteks penggunaan yang spesifik pada korpus Indonesia. Temuan ini memberikan justifikasi kuat untuk penggunaan *bert-base-indonesian-NER* dalam sistem yang dikembangkan, yang dibangun di atas arsitektur IndoBERT.

Penelitian Khairunnisa, Imankulova, dan Komachi (2020) mengidentifikasi masalah kritis dalam dataset NER Indonesia yang ada: inkonsistensi anotasi yang dapat menurunkan performa model. Mereka menyediakan dataset yang dianotasi ulang dengan standar yang lebih konsisten dan mengevaluasi berbagai pendekatan *deep learning* termasuk BiLSTM-CRF dengan berbagai jenis *embeddings*. Penelitian mereka menekankan pentingnya kualitas data pelatihan dan konsistensi skema anotasi untuk mencapai performa NER yang optimal. Wawasan ini relevan dengan penelitian ini, terutama dalam konteks *fine-tuning* model NER untuk mengenali entitas finansial spesifik yang mungkin tidak terwakili dengan baik dalam dataset NER umum Indonesia.

II.4.2 NER Domain Finansial

Zhang dan Zhang (2022) mengembangkan *FinBERT-MRC*, pendekatan NER finansial yang merumuskan tugas ekstraksi entitas sebagai masalah *machine reading comprehension*. Model mereka mencapai F1-score 92,78% dan 96,80% pada dataset finansial Tiongkok, mengungguli pendekatan BiLSTM-CRF dan BERT-CRF tradisional. Penelitian ini mendemonstrasikan bahwa NER domain finansial memiliki tantangan unik karena variasi semantik dan leksikal yang spesifik, seperti perbedaan antara nama perusahaan dalam konteks berita versus dokumen formal, atau ambiguitas antara nama organisasi dan produk finansial. Meskipun penelitian mereka fokus pada bahasa Tiongkok, prinsip-prinsip metodologis dapat diadaptasi untuk konteks Indonesia, terutama dalam menangani ambiguitas nama perusahaan versus kode saham.

Shah dkk. (2023) menyediakan sumber daya penting untuk NER finansial melalui *FiNER-ORD*, dataset NER finansial *open research* pertama dengan kualitas tinggi.

Dataset ini mengatasi variasi semantik dan leksikal yang unik dalam domain finansial dan menyediakan benchmark untuk berbagai model bahasa pra-latih termasuk varian BERT dan LLMs. Penelitian mereka menekankan bahwa entitas finansial sering kali memiliki representasi multipel (misalnya, "Apple Inc.", "Apple", "AAPL") yang harus dikenali sebagai entitas yang sama. Wawasan ini sangat relevan dengan konteks pasar saham Indonesia, di mana perusahaan mungkin dirujuk dengan nama lengkap, nama singkat, atau kode ticker (misalnya, "Bank Central Asia", "BCA", "BBCA").

II.5 *Topic Modeling* dengan BERTopic

Topic modeling merupakan teknik untuk mengidentifikasi tema atau topik abstrak dalam kumpulan dokumen. Bagian ini mengulas evolusi metodologi *topic modeling* dari pendekatan statistik tradisional ke pendekatan berbasis *neural embeddings*, dengan fokus khusus pada BERTopic yang digunakan dalam sistem yang dikembangkan.

II.5.1 BERTopic: Metodologi dan Keunggulan

Grootendorst (2022) memperkenalkan BERTopic, metodologi *topic modeling* neural yang menggabungkan model bahasa berbasis *transformer* (BERT), *document embeddings*, *clustering* dengan HDBSCAN, dan prosedur *class-based TF-IDF* (c-TF-IDF) untuk mengekstrak representasi topik yang koheren. Arsitektur BERTopic mengatasi keterbatasan pendekatan tradisional seperti *Latent Dirichlet Allocation* (LDA) dengan memanfaatkan representasi semantik kontekstual dari *transformer models*. Prosedur c-TF-IDF yang diusulkan menghasilkan representasi topik yang lebih interpretatif dibandingkan dengan metode konvensional dengan memperlakukan semua dokumen dalam satu kluster sebagai satu dokumen besar dan membandingkannya dengan dokumen dari kluster lain. Metodologi ini sangat sesuai untuk teks media sosial yang pendek dan informal, karena tidak memerlukan asumsi distribusional yang ketat seperti LDA.

Keunggulan BERTopic untuk teks media sosial divalidasi secara empiris oleh Egger dan Yu (2022), yang membandingkan BERTopic dengan LDA, NMF, dan Top2Vec pada dataset 31.800 *tweet* terkait COVID-19 dan perjalanan. Evaluasi mereka mendemonstrasikan superioritas BERTopic untuk teks pendek dan tidak terstruktur, khususnya dalam hal koherensi topik dan interpretabilitas. Metrik evaluasi menunjukkan bahwa BERTopic menghasilkan topik yang lebih koheren dan mudah dipahami dibandingkan dengan metode tradisional, terutama untuk data yang memiliki karak-

teristik serupa dengan media sosial: panjang dokumen pendek, penggunaan bahasa informal, dan struktur gramatikal yang longgar. Temuan ini memberikan justifikasi kuat untuk pemilihan BERTopic dalam penelitian ini, mengingat pesan Telegram finansial memiliki karakteristik yang sangat mirip dengan *tweet*: pendek, informal, dan sering menggunakan singkatan atau jargon.

II.5.2 Perkembangan Terkini dalam *Topic Modeling Neural*

Angelov dan Inkpen (2024) mengembangkan metodologi *topic modeling* lebih lanjut dengan memperkenalkan *Contextual-Top2Vec*, yang menggunakan *contextual token embeddings* dari dokumen untuk menghasilkan topik hierarkis, *topic spans* dalam dokumen, dan label topik berbasis frasa. Penelitian mereka mengusulkan penggunaan *BERTScore* untuk mengevaluasi koherensi dan informativeness topik, mengatasi keterbatasan metrik evaluasi tradisional seperti koherensi berbasis PMI. Model mereka mengungguli pendekatan *state-of-the-art* pada metrik evaluasi komprehensif, mendemonstrasikan bahwa representasi kontekstual dapat meningkatkan kualitas topik yang dihasilkan. Meskipun penelitian ini menggunakan BERTopic sebagai metodologi dasar, wawasan dari *Contextual-Top2Vec* menginformasikan pertimbangan desain, terutama dalam hal evaluasi kualitas topik dan interpretasi hasil *clustering*.

II.6 Telegram sebagai Platform Komunitas

Telegram telah berkembang menjadi platform komunikasi yang dominan untuk komunitas online, terutama di Indonesia. Bagian ini mengulas karakteristik unik Telegram sebagai platform untuk komunitas berskala besar, pola komunikasi yang terjadi di dalamnya, dan implikasi untuk desain sistem analitik dan rangkuman otomatis.

II.6.1 Karakteristik dan Dinamika Platform Telegram

Nobari, Reshadatmand, dan Neshati (2017) menyediakan analisis pionir tentang struktur dan aspek topikal Telegram menggunakan data yang di-*crawl* dari kanal publik. Penelitian mereka mengeksplorasi pola jaringan, klasifikasi pesan (spam/ham), dan arsitektur Telegram sebagai platform *messaging* hibrida yang menggabungkan fitur komunikasi pribadi dan publik. Mereka mengidentifikasi bahwa Telegram memiliki karakteristik unik dibandingkan platform media sosial lain: kombinasi antara privasi komunikasi satu-satu dengan kemampuan penyebaran informasi massal melalui kanal dan grup besar. Struktur ini menciptakan dinamika komunikasi yang berbeda dari platform seperti Twitter atau Facebook, di mana informasi dapat menyebar

dengan cepat dalam komunitas tertutup yang besar tanpa eksposur publik yang luas. Pemahaman tentang arsitektur ini penting untuk desain sistem rangkuman yang harus mempertimbangkan privasi komunitas sambil memfasilitasi akses informasi yang efisien.

Hashemi dan Chahooki (2019) mengembangkan metodologi untuk mengukur kualitas grup Telegram melalui analisis perilaku pengguna, mengeksplorasi lebih dari 900.000 kanal dan 300.000 *supergroup* berbahasa Persia. Mereka mengusulkan fitur-fitur pengukuran kualitas yang membedakan grup berkualitas tinggi dari grup berkualitas rendah berdasarkan pola aktivitas pengguna, konsistensi partisipasi, dan struktur komunikasi. Penelitian mereka menunjukkan bahwa grup berkualitas tinggi cenderung memiliki partisipasi yang lebih seimbang, topik diskusi yang lebih fokus, dan tingkat spam yang lebih rendah. Temuan ini relevan dengan konteks komunitas Michael Yeoh yang memiliki struktur berjenjang, di mana grup "Advanced" diasumsikan memiliki kualitas diskusi yang lebih tinggi dibandingkan grup "Cuap Cuap" umum. Sistem rangkuman harus dirancang dengan mempertimbangkan perbedaan kualitas dan karakteristik diskusi antar grup.

II.6.2 Analisis Skala Besar Komunitas Telegram

Morgia dkk. (2024) menyediakan dataset Telegram terbesar hingga saat ini dengan 120.979 kanal dan lebih dari 400 juta pesan, bersama dengan infrastruktur untuk pengumpulan dan analisis data skala besar. Penelitian mereka menyediakan alat dan metodologi untuk penelitian komunitas Telegram yang mencakup analisis dinamika komunitas, propagasi konten, dan pola komunikasi tingkat platform. Kontribusi mereka mendemonstrasikan bahwa Telegram telah berkembang menjadi ekosistem informasi yang kompleks dengan dinamika komunitas yang beragam. Skala dataset ini memberikan konteks untuk memahami bahwa komunitas finansial dengan 22.000+ anggota seperti yang diteliti dalam penelitian ini merupakan bagian dari fenomena komunikasi online yang lebih luas, di mana platform seperti Telegram telah menjadi infrastruktur kritis untuk pertukaran informasi dalam komunitas tematik berskala besar.

Penelitian terbaru oleh Perlo dkk. (2025) menyediakan analisis per-topik pertama dari grup Telegram, mengeksplorasi 51 juta pesan dari 669 grup di berbagai topik termasuk Pendidikan, Politik, dan Cryptocurrency. Penelitian mereka menganalisis pola aktivitas pengguna, kehadiran bot, dan berbagi media, serta menyediakan alat *open-source* untuk pengumpulan pesan otomatis. Temuan kunci mereka adalah bahwa grup dengan topik berbeda menunjukkan pola komunikasi yang sangat berbeda

dalam hal volume pesan, distribusi waktu aktivitas, dan tipe konten yang dibagikan. Khususnya untuk kategori Cryptocurrency, yang memiliki kesamaan dengan komunitas finansial yang diteliti, mereka menemukan tingkat aktivitas yang sangat tinggi dengan puncak aktivitas yang berkorelasi dengan jam perdagangan pasar. Temuan ini memvalidasi desain sistem rangkuman per jam dalam penelitian ini, karena pola aktivitas temporal yang kuat mengindikasikan bahwa rangkuman periodik sesuai dengan ritme alami diskusi komunitas.

II.7 Arsitektur Pemrosesan Data *Real-Time*

Implementasi sistem rangkuman otomatis untuk komunitas Telegram yang aktif memerlukan arsitektur pemrosesan data yang dapat menangani aliran pesan secara *real-time* atau *near-real-time*. Bagian ini mengulas pendekatan arsitektural untuk pemrosesan *stream* teks dengan NLP, pola *event-driven architecture*, dan pertimbangan desain untuk sistem yang responsif dan *scalable*.

II.7.1 *Pipeline* NLP untuk Pemrosesan *Stream* Teks

Hamidi dkk. (2021) mengusulkan arsitektur *pipeline* NLP yang dinamis dan terdistribusi untuk pemrosesan *stream* teks *real-time*. Penelitian mereka menggunakan Apache Storm dan Apache Kafka untuk menerapkan tugas-tugas NLP pada aliran data tekstual, memungkinkan pengembang untuk menginjeksi modul NLP melalui berbagai bahasa pemrograman dan mendukung sumber data multipel. Arsitektur mereka mendemonstrasikan bagaimana komponen NLP yang berbeda (*tokenisasi*, *part-of-speech tagging*, *named entity recognition*, analisis sentimen) dapat dikomposisikan dalam *pipeline* yang dapat di-*scale* secara horizontal. Prinsip-prinsip desain ini relevan dengan penelitian ini, di mana sistem harus mengintegrasikan beberapa komponen NLP (analisis sentimen, NER, *topic modeling*, generasi rangkuman) dalam arsitektur yang koheren dan dapat diperluas.

Isah dkk. (2019) menyediakan survei komprehensif tentang *framework* pemrosesan *stream* data terdistribusi, termasuk Apache Storm, Spark Streaming, Apache Flink, dan Kafka Streams. Penelitian mereka membandingkan model pemrosesan, karakteristik operasional, dan pola arsitektural dari berbagai *framework*, memberikan taksonomi yang membantu dalam pemilihan teknologi yang sesuai untuk kasus penggunaan spesifik. Untuk konteks komunitas Telegram dengan volume pesan yang tinggi namun tidak ekstrem (ratusan hingga ribuan pesan per jam), mereka mengidentifikasi bahwa pendekatan *micro-batch processing* dengan Kafka atau arsitektur berbasis *event* yang lebih sederhana dapat memberikan keseimbangan optimal anta-

ra latensi, throughput, dan kompleksitas implementasi. Wawasan ini menginformasikan desain sistem dalam penelitian ini, yang mengadopsi arsitektur berbasis *event* dengan penjadwalan periodik daripada pemrosesan *stream* kontinu, sesuai dengan kebutuhan rangkuman per jam.

II.7.2 Arsitektur Berbasis *Event* dan Performa

Rubert dan Farias (2023) menyediakan bukti empiris tentang dampak *event-driven architecture* (EDA) terhadap performa melalui studi eksploratif yang membandingkan arsitektur berbasis *event* dengan arsitektur monolitik. Penelitian mereka mengukur penggunaan CPU, memori, waktu respons, *throughput*, dan paket jaringan, mendemonstrasikan bahwa EDA dapat memberikan keunggulan signifikan dalam *scalability* dan responsivitas, terutama untuk sistem yang menangani beban kerja yang bervariasi dan tidak dapat diprediksi. Temuan mereka menunjukkan bahwa EDA sangat cocok untuk skenario di mana komponen sistem perlu berkomunikasi secara asinkron dan dapat di-*scale* secara independen. Konteks komunitas Telegram finansial, di mana volume pesan bervariasi drastis sepanjang hari (tinggi selama jam perdagangan, rendah di malam hari), sangat sesuai dengan karakteristik ini. Sistem yang dikembangkan dalam penelitian ini mengadopsi prinsip EDA dengan pemisahan antara komponen pengumpulan pesan, pemrosesan NLP, dan generasi rangkuman, memungkinkan setiap komponen untuk beroperasi secara independen dan di-*scale* sesuai kebutuhan.

II.7.3 Pertimbangan *Micro-batching* dan Frekuensi Data

Meskipun pemrosesan *stream* murni memberikan latensi terendah, penelitian menunjukkan bahwa pendekatan *micro-batching* sering kali lebih praktis untuk banyak aplikasi NLP. Konsep rangkuman per jam dalam penelitian ini dapat dipandang sebagai bentuk *macro-batching*, di mana pesan dikumpulkan selama satu jam sebelum diproses sebagai satu *batch*. Pendekatan ini memberikan beberapa keuntungan: (1) memungkinkan agregasi konteks yang lebih kaya untuk tugas-tugas seperti *topic modeling* yang memerlukan pandangan holistik terhadap data, (2) mengurangi beban komputasi dengan menghindari pemrosesan setiap pesan individual secara *real-time*, dan (3) sesuai dengan ritme natural konsumsi informasi pengguna, yang jarang memerlukan pembaruan lebih sering dari setiap jam untuk analisis tren komunitas.

II.7.4 Tantangan Inferensi LLM dalam Sistem *Real-Time*

Bottleneck arsitektural yang paling signifikan dalam sistem rangkuman otomatis berbasis LLM adalah latensi inferensi model bahasa besar. Tidak seperti komponen NLP tradisional seperti tokenisasi atau *named entity recognition* yang dapat dieksekusi dalam milidetik, generasi teks dengan LLM skala besar (puluhan hingga ratusan miliar parameter) dapat memerlukan waktu detik hingga menit untuk satu panggilan inferensi, tergantung pada panjang konteks input dan panjang output yang dihasilkan. Tantangan ini menjadi kritis dalam konteks sistem rangkuman per jam, di mana seluruh *pipeline* pemrosesan harus selesai dalam jendela waktu yang terbatas (idealnya kurang dari 5-10 menit dari akhir periode pengumpulan data) untuk menjaga relevansi temporal rangkuman.

Penelitian terkini menunjukkan bahwa kecepatan inferensi LLM sangat dipengaruhi oleh tiga faktor utama: (1) ukuran model (jumlah parameter), (2) panjang konteks input, dan (3) infrastruktur komputasi yang digunakan (CPU, GPU, atau akselerator khusus seperti TPU). Untuk konteks rangkuman komunitas Telegram dengan potensi ratusan pesan per jam, panjang konteks input dapat dengan mudah mencapai puluhan ribu *token*, yang secara signifikan meningkatkan waktu inferensi pada infrastruktur standar. Lebih lanjut, model yang lebih besar seperti Llama 3.1 70B, meskipun memberikan kualitas output superior dibandingkan model yang lebih kecil, memerlukan sumber daya komputasi yang substansial dan dapat mengalami latensi yang tidak dapat diterima pada infrastruktur konvensional.

Pemilihan Groq API dengan model Llama 3.1 70B dalam penelitian ini merupakan solusi arsitektural yang dirancang khusus untuk mengatasi bottleneck inferensi LLM ini. Groq menggunakan *Language Processing Unit* (LPU), arsitektur akselerator khusus yang dirancang untuk inferensi LLM dengan kecepatan ekstrem. Platform Groq telah mendemonstrasikan kemampuan untuk menghasilkan ratusan hingga ribuan *tokens per second*, secara signifikan melampaui solusi berbasis GPU tradisional. Kecepatan inferensi yang tinggi ini adalah faktor pendukung (*enabler*) teknis yang krusial yang memungkinkan implementasi rangkuman per jam yang praktis; tanpa throughput inferensi yang tinggi, sistem akan mengalami akumulasi *backlog* dan keterlambatan yang tidak dapat diterima, merusak nilai temporal dari rangkuman yang dihasilkan.

Selain itu, penggunaan *API* cloud seperti Groq menghilangkan kompleksitas operasional dalam mengelola infrastruktur inferensi LLM lokal, termasuk manajemen GPU, optimasi *batch size*, dan penjadwalan sumber daya. Pendekatan ini memung-

kinkan sistem untuk fokus pada logika bisnis dan integrasi komponen NLP, sementara kebutuhan komputasi intensif untuk inferensi LLM di-*offload* ke platform yang dioptimalkan. Arsitektur ini mencerminkan prinsip *separation of concerns* dalam desain sistem terdistribusi, di mana setiap komponen (pengumpulan data, pemrosesan NLP, inferensi LLM) diimplementasikan menggunakan teknologi yang paling sesuai untuk karakteristik komputasinya.

II.8 Peringkasan untuk Media Sosial dan Percakapan Online

Peringkasan konten media sosial memiliki tantangan unik dibandingkan dengan peringkasan dokumen formal atau artikel berita. Bagian ini mengulas penelitian tentang peringkasan media sosial, khususnya Reddit dan forum diskusi online, yang memiliki karakteristik serupa dengan komunitas Telegram.

II.8.1 Pemanfaatan Sinyal Komunitas dalam Peringkasan

Kano dkk. (2018) mengusulkan pendekatan inovatif untuk peringkasan ekstraktif percakapan online dengan memanfaatkan ukuran popularitas dalam media sosial (seperti *votes*, *shares*, *bookmarks*) sebagai *distant labels*. Penelitian mereka mendemonstrasikan bahwa metrik keterlibatan komunitas dapat berfungsi sebagai indikator kualitas dan relevansi konten, dengan memisahkan kontribusi konten dari faktor kontekstual. Mereka menemukan bahwa pesan dengan *upvotes* tinggi dalam forum diskusi cenderung mengandung informasi yang lebih substantif dan relevan dengan topik diskusi utama. Pendekatan ini sangat relevan dengan penelitian ini, di mana struktur hierarkis komunitas Telegram (pemimpin komunitas versus anggota biasa) dapat dimanfaatkan sebagai sinyal implisit tentang kredibilitas dan pentingnya informasi. Implementasi *weighted sentiment analysis* dalam sistem yang dikembangkan mengadopsi prinsip ini dengan memberikan bobot yang berbeda berdasarkan peran pengirim pesan.

Fabbri dkk. (2021) memperkenalkan *ConvoSumm*, benchmark peringkasan percakapan yang mencakup empat dataset dari berbagai bentuk percakapan online: komentar berita, forum diskusi, forum Q&A komunitas, dan *thread* email. Penelitian mereka menggunakan *framework issues-viewpoints-assertions* yang menginkorporasikan *argument mining* melalui konstruksi graf untuk memodelkan struktur percakapan. Pendekatan ini mengakui bahwa percakapan online sering kali melibatkan pertukaran pandangan yang kompleks dengan argumen pendukung, yang memerlukan pemahaman struktural lebih mendalam daripada sekadar ekstraksi kalimat penting. Kontribusi metodologis mereka relevan dengan penelitian ini kare-

na diskusi finansial di Telegram sering kali mengandung perdebatan antara strategi investasi berbeda, analisis yang bertentangan, atau perspektif pasar yang beragam. Sistem rangkuman yang efektif harus dapat mensintesis pandangan-pandangan yang berbeda ini daripada hanya mengekstrak opini dominan.

II.8.2 Peringkasan Percakapan *Multi-speaker*

Kompleksitas percakapan *multi-speaker* dibahas secara komprehensif oleh Chen dan Yang (2021), yang mengusulkan pendekatan yang sadar struktur (*structure-aware*) untuk peringkasan percakapan abstraktif. Penelitian mereka memodelkan relasi wacana dan *action triples* ("siapa-melakukan-apa") menggunakan graf terstruktur untuk menangkap karakteristik kompleks interaksi manusia-manusia di media sosial. Mereka mengimplementasikan mekanisme *decoding* multi-granularitas yang dapat menghasilkan rangkuman pada tingkat abstraksi berbeda. Pendekatan ini sangat relevan dengan konteks komunitas Telegram berjenjang, di mana percakapan melibatkan puluhan hingga ratusan partisipan dengan peran dan tingkat keahlian yang berbeda. Sistem harus dapat mengenali struktur interaksi (misalnya, pertanyaan-jawaban, klaim-dukkungan, analisis-counter-analisis) untuk menghasilkan rangkuman yang koheren dan informatif.

II.8.3 Peringkasan Ekstrim untuk Konten Media Sosial

Sotudeh dkk. (2021) memperkenalkan dataset TLDR9+ dengan lebih dari 9 juta *instance* pelatihan dari Reddit untuk peringkasan ekstrim (rangkuman satu kalimat). Penelitian mereka mengatasi agregasi konten media sosial skala besar dan mencakup subset berkualitas tinggi (TLDRHQ) melalui anotasi manusia. Mereka mendemonstrasikan bahwa peringkasan ekstrim, yang menghasilkan rangkuman sangat singkat dengan tingkat kompresi dan abstraksi tinggi, dapat efektif untuk konten media sosial yang cenderung verbose dan mengandung banyak informasi redundan. Konsep peringkasan ekstrim ini relevan dengan kebutuhan komunitas Telegram finansial, di mana anggota mungkin memerlukan "snapshot" cepat dari diskusi komunitas tanpa membaca rangkuman panjang. Sistem yang dikembangkan dalam penelitian ini dapat mengadaptasi prinsip ini dengan menyediakan "headline" atau "key takeaway" di awal setiap rangkuman per jam sebelum menyajikan detail yang lebih mendalam.

II.9 Kesenjangan Penelitian dan Kontribusi

Tinjauan literatur komprehensif di atas mengidentifikasi beberapa kesenjangan penelitian kritis yang ditangani oleh penelitian ini. Pertama, tidak ada penelitian yang

mengintegrasikan NLP bahasa Indonesia (khususnya model berbasis RoBERTa dan BERT) dengan analisis domain finansial untuk konteks media sosial. Sebagian besar penelitian analisis sentimen finansial fokus pada bahasa Inggris atau Tiongkok, sementara penelitian NLP Indonesia cenderung menggunakan domain umum atau ulasan produk. Penelitian ini menjembatani kesenjangan ini dengan mengembangkan sistem yang mengintegrasikan *indonesian-roberta-base-emotion-classifier* dan *bert-base-indonesian-NER* untuk analisis teks finansial Indonesia dalam konteks media sosial.

Kedua, meskipun terdapat penelitian ekstensif tentang komunitas Telegram dan peringkasan media sosial secara terpisah, sangat sedikit penelitian yang menangani sistem peringkasan otomatis khusus untuk aplikasi bot Telegram. Penelitian tentang bot Telegram cenderung fokus pada interaksi percakapan sederhana atau otomasi tugas administratif, bukan pada sintesis intelijen informasi yang kompleks. Penelitian ini memberikan kontribusi dengan mendemonstrasikan implementasi praktis sistem peringkasan otomatis yang diintegrasikan dengan Telegram Bot API untuk penyampaian rangkuman periodik.

Ketiga, penelitian tentang *pipeline* NLP *real-time* yang mengintegrasikan multiple model (analisis sentimen, NER, *topic modeling*, generasi rangkuman dengan LLM) dalam arsitektur *event-driven* untuk platform *messaging* masih terbatas. Sebagian besar penelitian mengevaluasi model-model ini secara terpisah atau dalam konteks dataset benchmark offline. Penelitian ini memberikan kontribusi metodologis dengan mendemonstrasikan integrasi empat komponen NLP berbeda dalam sistem yang koheren dan operasional.

Keempat, tidak ada penelitian yang menangani aliran informasi hierarkis dan sintesis lintas grup dalam komunitas berjenjang. Sebagian besar penelitian peringkasan media sosial mengasumsikan satu sumber data atau agregasi horizontal dari multiple sumber yang setara. Penelitian ini menangani skenario unik di mana informasi mengalir melalui struktur hierarkis (grup Cuap Cuap, Swing Plan, Advanced) dengan karakteristik diskusi yang berbeda, dan sistem harus mensintesis informasi lintas tingkat hierarki ini.

Kelima, pendekatan rangkuman per jam untuk diskusi finansial dengan *weighted sentiment* berbasis peran pengguna merupakan kontribusi novel. Sementara penelitian sebelumnya mengeksplorasi pemanfaatan sinyal popularitas atau metrik keterlibatan, penelitian ini mengimplementasikan *weighting* eksplisit berdasarkan struktur peran komunitas yang mencerminkan kredibilitas dan pengaruh relatif kontributor

dalam ekosistem informasi finansial.

Secara keseluruhan, penelitian ini mengintegrasikan teori *information overload* dan *cognitive load*, metodologi NLP terkini untuk bahasa Indonesia dan domain finansial, arsitektur pemrosesan *event-driven*, dan pemahaman tentang dinamika komunitas Telegram untuk mengembangkan solusi holistik yang mengatasi tantangan praktis dalam komunitas finansial Indonesia. Kontribusi utama penelitian ini terletak pada demonstrasi bahwa sistem otomatis berbasis *AI* dapat secara efektif mereduksi *information overload* dalam komunitas online berskala besar melalui sintesis intelligen multi-sumber, lintas-grup (*cross-group*), dan hierarkis yang disesuaikan dengan konteks spesifik ekosistem informasi komunitas.

BAB III

ANALISIS MASALAH

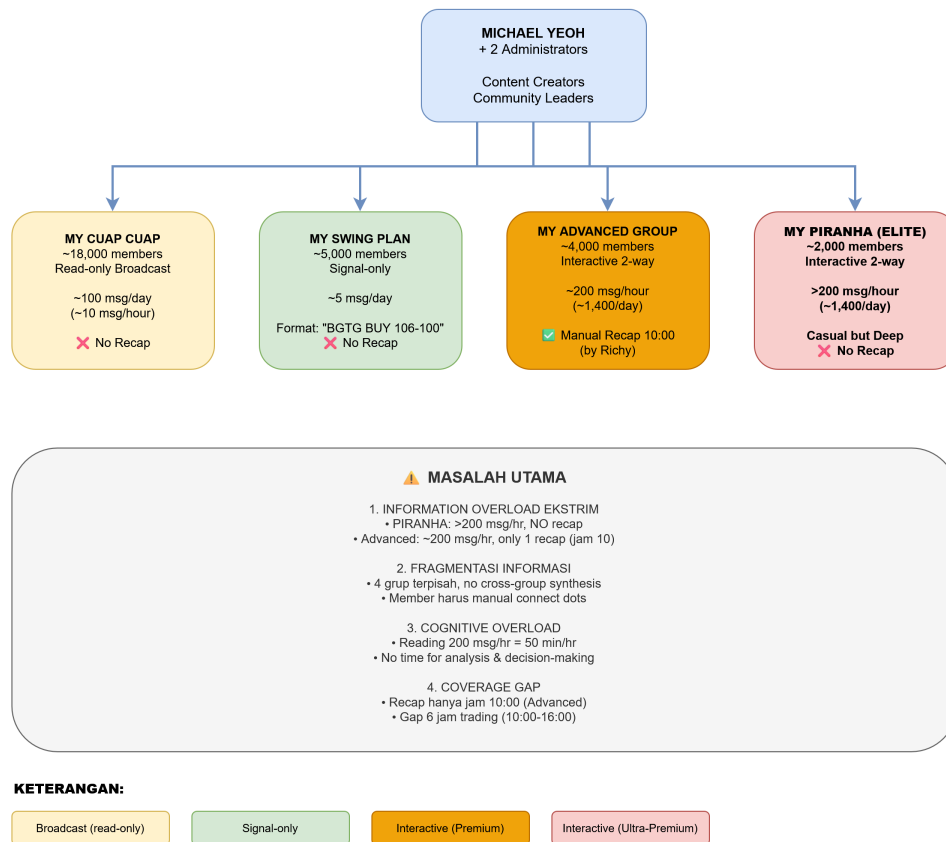
Bab ini menyajikan analisis mendalam terhadap kondisi sistem komunitas Telegram Michael Yeoh saat ini, identifikasi masalah spesifik yang dihadapi, analisis kebutuhan sistem rangkuman otomatis, serta evaluasi berbagai alternatif solusi dengan justifikasi pemilihan pendekatan yang diusulkan. Analisis ini menjadi fondasi untuk desain konsep solusi yang akan diuraikan pada Bab IV.

III.1 Analisis Kondisi Saat Ini

III.1.1 Model Konseptual Komunitas Telegram Eksisting

Komunitas investasi saham Indonesia Michael Yeoh di Telegram merupakan ekosistem informasi finansial berskala besar dengan struktur hierarkis empat tingkat yang dirancang untuk menyediakan nilai berbeda pada setiap level keanggotaan. Gambar III.1 mengilustrasikan arsitektur sistem komunikasi komunitas saat ini beserta aliran informasi antar komponen.

SISTEM KOMUNITAS TELEGRAM MICHAEL YEOH - KONDISI SAAT INI



Gambar III.1 Model konseptual sistem komunikasi komunitas Telegram Michael Yeoh saat ini

Struktur hierarkis komunitas terdiri dari empat grup dengan karakteristik berbeda yang dapat dilihat pada Tabel III.1. Setiap grup memiliki fungsi spesifik dalam ekosistem informasi komunitas, dengan kompleksitas komunikasi yang meningkat seiring level hierarki.

Tabel III.1 Karakteristik grup dalam komunitas Telegram Michael Yeoh

Aspek	Cuap Cuap	Swing Plan	Advanced	PIRANHA
Jumlah Member	18.000	5.000	4.000	2.000
Tipe Komunikasi	Broadcast (<i>read-only</i>)	Signal-only	Interaktif 2-arah	Interaktif 2-arah
Volume Pesan	100 msg/hari (10/jam)	5 msg/hari	200 msg/jam	>200 msg/jam
Konten Utama	Edukasi umum, berita pasar	Sinyal beli/ju- al terstruktur	Diskusi analisis teknikal	Debat strategi expert
Recap Manual	Tidak ada	Tidak ada	1× per hari (jam 10:00)	Tidak ada
Tingkat Masalah	Rendah	Minimal	Tinggi	Sangat Tinggi

Setiap grup dalam hierarki komunitas memiliki peran dan dinamika komunikasi yang berbeda. Grup MY Cuap Cuap berfungsi sebagai kanal informasi umum dengan mode *read-only*, di mana hanya administrator dan Michael Yeoh yang dapat mengirim pesan. Konten utama meliputi edukasi pasar saham, analisis makroekonomi, dan pengumuman penting. Dengan volume sekitar 100 pesan per hari atau rata-rata 10 pesan per jam, grup ini memiliki tingkat *information overload* yang relatif rendah.

Grup MY Swing Plan merupakan kanal khusus untuk distribusi sinyal *trading* dengan format terstruktur seperti "BGTG BUY 106-100 SL 98 porsi 5%". Volume pesan sangat rendah (5 pesan per hari) karena sifatnya yang spesifik dan terarah. Grup ini tidak mengalami masalah *information overload*, namun menjadi bagian penting dari ekosistem informasi yang perlu disintesis dengan diskusi di grup lain untuk memberikan konteks komprehensif.

Grup MY Advanced Group adalah grup premium pertama dengan komunikasi interaktif dua arah, di mana sekitar 4.000 anggota dapat berdiskusi secara langsung. Volume pesan mencapai sekitar 200 pesan per jam selama jam perdagangan aktif (09:00-16:00), menghasilkan total sekitar 1.400 pesan per hari. Grup ini memiliki satu rangkuman manual yang dibuat oleh asisten bernama Richy Matthew Ashari setiap hari pada jam 10:00 pagi. Rangkuman manual ini mencakup kondisi pasar dan pergerakan saham-saham prioritas (*PP stocks*) dengan format yang mencantumkan level support, resistance, dan analisis kekuatan saham. Namun, rangkuman ini

hanya mencakup satu jam pertama perdagangan (09:00-10:00), meninggalkan gap informasi selama enam jam perdagangan berikutnya (10:00-16:00).

Grup MY PIRANHA merupakan grup ultra-premium dengan 2.000 anggota elite dan mengalami volume pesan tertinggi, melebihi 200 pesan per jam, menghasilkan lebih dari 1.400 pesan per hari. Berbeda dengan Advanced Group, PIRANHA tidak memiliki rangkuman manual sama sekali. Diskusi di grup ini cenderung lebih kasual dalam gaya komunikasi namun sangat mendalam secara konten, dengan debat strategi investasi antar investor berpengalaman. Paradoks "kasual namun *high-volume*" ini menciptakan tantangan ekstraksi sinyal yang unik, karena informasi penting sering tersembunyi dalam percakapan informal yang memerlukan pemahaman konteks mendalam.

Aliran informasi dalam ekosistem komunitas bersifat hierarkis dan fragmentasi. Michael Yeoh dan dua administrator lainnya dapat memposting di semua grup, menciptakan aliran informasi vertikal dari puncak hierarki. Namun, anggota di level berbeda tidak dapat mengakses informasi di grup yang lebih tinggi kecuali mereka menjadi anggota grup tersebut. Hal ini menciptakan fragmentasi informasi di mana analisis makro di Cuap Cuap, sinyal konkret di Swing Plan, diskusi implementasi di Advanced, dan debat strategi di PIRANHA terpisah satu sama lain. Anggota yang merupakan bagian dari multiple grup harus secara manual melakukan sintesis informasi lintas grup, yang memerlukan waktu dan kemampuan analisis yang signifikan.

III.1.2 Identifikasi Masalah Sistem Saat Ini

Berdasarkan analisis terhadap model konseptual sistem saat ini, terdapat tiga kategori masalah utama yang menjadi fokus penelitian ini: *information overload* dengan intensitas ekstrim di grup premium, fragmentasi informasi lintas struktur hierarkis, dan keterbatasan kognitif anggota dalam memproses volume informasi tinggi.

III.1.2.1 *Information Overload* Ekstrim di Grup PIRANHA

Grup PIRANHA mengalami tingkat *information overload* paling parah dalam ekosistem komunitas. Dengan volume melebihi 200 pesan per jam selama jam perdagangan (09:00-16:00), grup ini menghasilkan lebih dari 1.400 pesan per hari. Tidak adanya mekanisme rangkuman otomatis atau manual membuat anggota dihadapkan pada dua pilihan yang sama-sama suboptimal: (1) memantau diskusi secara *real-time* yang memerlukan perhatian konstan selama tujuh jam perdagangan, atau (2) melakukan *catch-up* dengan membaca ratusan pesan secara retrospektif, yang beri-

siko kehilangan konteks temporal dan relevansi diskusi.

Karakteristik unik dari diskusi PIRANHA yang kasual namun mendalam memperburuk masalah ini. Tidak seperti diskusi terstruktur di grup lain, informasi kritis di PIRANHA sering tersembunyi dalam percakapan informal tanpa penanda eksplisit seperti format sinyal trading atau header topik. Ekstraksi sinyal dari diskusi kasual memerlukan pemahaman konteks yang dalam, yang sulit dilakukan ketika anggota harus memproses volume pesan yang sangat tinggi dalam waktu terbatas.

Dampak kuantitatif dari masalah ini dapat diestimasi menggunakan kerangka *Cognitive Load Theory*. Jika diasumsikan rata-rata waktu baca dan pemahaman per pesan adalah 15 detik, maka untuk membaca 200 pesan memerlukan 50 menit, mendekati satu jam penuh. Selama jam perdagangan yang berlangsung selama tujuh jam, anggota yang mencoba membaca semua pesan akan menghabiskan hampir seluruh waktu mereka hanya untuk konsumsi informasi, tanpa waktu untuk analisis dan pengambilan keputusan. Situasi ini jelas tidak sustainable dan bertentangan dengan tujuan komunitas untuk memfasilitasi pengambilan keputusan investasi yang efektif.

III.1.2.2 *Information Overload* Tinggi di Grup Advanced dengan *Coverage Gap*

Grup Advanced Group, meskipun memiliki satu rangkuman manual harian, masih mengalami masalah *information overload* yang signifikan dengan beberapa dimensi. Pertama, volume pesan sekitar 200 pesan per jam menghasilkan total sekitar 1.400 pesan per hari, angka yang hampir identik dengan PIRANHA. Kedua, rangkuman manual yang ada hanya mencakup satu jam pertama perdagangan (09:00-10:00), meninggalkan gap informasi selama enam jam berikutnya.

Gap *coverage* temporal ini menciptakan masalah asimetri informasi. Anggota yang dapat memantau diskusi pada pagi hari memiliki akses ke rangkuman manual jam 10:00 yang menyediakan konteks diskusi pagi. Namun, anggota yang bergabung setelah jam 10:00 atau yang tidak dapat memantau diskusi siang hingga sore hari (10:00-16:00) kehilangan konteks penting dari diskusi di periode tersebut. Mengingat bahwa kondisi pasar dan sentimen dapat berubah drastis dalam beberapa jam, gap ini dapat menyebabkan pengambilan keputusan berdasarkan informasi yang tidak lengkap atau *outdated*.

Keterbatasan rangkuman manual juga terletak pada skalabilitas dan konsistensi. Pem-

buatan rangkuman bergantung pada satu individu (Richy Matthew Ashari) dengan ketersediaan waktu dan perspektif subjektif tertentu. Model ini tidak dapat di-*scale* untuk memberikan *coverage* yang lebih komprehensif sepanjang hari tanpa menambah jumlah personel secara signifikan, yang tidak efisien dari segi biaya dan koordinasi. Lebih lanjut, tidak ada jaminan konsistensi format dan kedalaman analisis antar rangkuman karena sifatnya yang manual.

III.1.2.3 Fragmentasi Informasi Lintas Struktur Hierarkis

Struktur hierarkis empat tingkat komunitas, meskipun dirancang untuk menyediakan nilai berbeda di setiap level, menciptakan masalah fragmentasi informasi yang sistemik. Informasi yang relevan untuk pengambilan keputusan investasi tersebar di empat grup dengan sedikit mekanisme sintesis eksplisit. Sebagai contoh, skenario tipikal dapat melibatkan: (1) Michael Yeoh memposting analisis makroekonomi tentang kondisi likuiditas pasar di grup Cuap Cuap, (2) sinyal konkret untuk beberapa saham diberikan di Swing Plan berdasarkan analisis tersebut, (3) diskusi tentang implementasi strategi dan manajemen risiko berlangsung di Advanced Group, dan (4) debat mendalam tentang potensi *upside* versus *downside* terjadi di PIRANHA.

Untuk mendapatkan pemahaman komprehensif, seorang anggota yang merupakan bagian dari semua grup harus secara manual: (1) membaca dan memahami konten di keempat grup, (2) mengidentifikasi hubungan semantik antar informasi, (3) mensintesis informasi menjadi *actionable insight*, dan (4) melakukan ini secara berulang sepanjang hari karena informasi terus mengalir. Proses ini memerlukan *cognitive switching cost* yang tinggi karena anggota harus berpindah antar konteks komunikasi yang berbeda (broadcast vs interaktif, formal vs kasual, umum vs spesifik).

Masalah ini diperparah oleh fakta bahwa sebagian besar anggota tidak memiliki akses ke semua grup. Anggota di Cuap Cuap dan Swing Plan tidak dapat mengakses diskusi mendalam di Advanced atau PIRANHA, menciptakan asimetri informasi struktural. Sementara itu, anggota premium yang memiliki akses ke multiple grup menghadapi beban kognitif yang berlebihan untuk melakukan sintesis manual. Tidak ada mekanisme dalam sistem saat ini yang secara otomatis mengagregasi dan mensintesis informasi lintas grup untuk memberikan pandangan holistik.

III.1.2.4 Keterbatasan Kapasitas Kognitif dan Dampaknya terhadap Kualitas Keputusan

Berdasarkan kerangka *Cognitive Load Theory* yang telah dibahas dalam Bab II, kapasitas memori kerja manusia memiliki keterbatasan fundamental dalam memproses informasi secara simultan. Dalam konteks komunitas Telegram finansial dengan volume pesan yang sangat tinggi, anggota menghadapi beban kognitif yang melampaui kapasitas optimal pada beberapa dimensi.

Pertama, *intrinsic cognitive load* yang tinggi inheren dalam domain finansial itu sendiri. Analisis investasi saham memerlukan pemahaman tentang analisis teknis, fundamental, sentimen pasar, manajemen risiko, dan psikologi *trading*. Kompleksitas intrinsik ini tidak dapat dikurangi karena merupakan bagian esensial dari domain.

Kedua, *extraneous cognitive load* yang ditambahkan oleh desain sistem saat ini sangat signifikan. Antarmuka Telegram dengan notifikasi konstan, fragmentasi informasi antar grup, kurangnya struktur hierarkis dalam penyajian pesan, dan tidak adanya mekanisme filter atau prioritas menciptakan beban kognitif tambahan yang tidak berkontribusi langsung pada pembelajaran atau pengambilan keputusan. Desain sistem yang mengharuskan anggota untuk scroll ratusan pesan, membedakan informasi penting dari *noise*, dan melakukan sintesis manual meningkatkan *extraneous load* secara dramatis.

Ketiga, kapasitas untuk *germane cognitive load* yang berkontribusi pada pemahaman mendalam dan pembentukan skema mental investasi yang baik menjadi terbatas. Ketika sebagian besar sumber daya kognitif digunakan untuk mengatasi beban informasi (*intrinsic* dan *extraneous*), kapasitas untuk refleksi, analisis mendalam, dan integrasi pengetahuan baru menjadi terbatas.

Dampak dari situasi ini terhadap kualitas pengambilan keputusan dapat bermanifestasi dalam beberapa cara. Anggota mungkin mengalami *decision fatigue*, di mana kualitas keputusan menurun seiring waktu karena deplesi sumber daya kognitif. Alternatifnya, anggota mungkin mengembangkan strategi *satisficing*, di mana mereka membuat keputusan berdasarkan subset informasi yang mudah diakses daripada melakukan analisis komprehensif. Dalam kasus terburuk, anggota mungkin mengalami *analysis paralysis*, di mana volume informasi yang berlebihan menyebabkan ketidakmampuan untuk membuat keputusan sama sekali.

Penelitian empiris yang akan dilakukan dalam tugas akhir ini bertujuan untuk mengukur dampak dari masalah-masalah ini dan mengevaluasi sejauh mana sistem rangkuman otomatis dapat mereduksi beban kognitif dan meningkatkan efektivitas pengambilan keputusan anggota komunitas.

III.2 Analisis Kebutuhan

Berdasarkan identifikasi masalah pada sistem saat ini, bagian ini menganalisis kebutuhan sistem rangkuman otomatis dari perspektif *stakeholder* yang berbeda, serta mendefinisikan kebutuhan fungsional dan non-fungsional yang harus dipenuhi oleh solusi yang diusulkan.

III.2.1 Identifikasi *Stakeholder* dan *Pain Points*

Sistem rangkuman otomatis yang diusulkan melayani beberapa kelompok *stakeholder* dengan *pain points* dan kebutuhan yang berbeda. Pemahaman mendalam terhadap kebutuhan setiap *stakeholder* penting untuk desain sistem yang efektif dan *user-centric*.

III.2.1.1 Anggota Grup PIRANHA (Target Primer)

Anggota grup PIRANHA, yang berjumlah sekitar 2.000 investor ultra-premium, merupakan *stakeholder* primer dengan *pain points* paling akut. *Pain point* utama adalah *information overload* ekstrim dengan volume lebih dari 200 pesan per jam yang tidak memiliki mekanisme rangkuman sama sekali. Saat ini, mereka mengandalkan *workaround* manual seperti scrolling periodik atau pencarian pengguna spesifik yang dipercaya, namun pendekatan ini memiliki keterbatasan signifikan dalam hal efisiensi waktu dan risiko kehilangan informasi penting.

Kebutuhan spesifik kelompok ini meliputi: (1) rangkuman periodik yang dapat memberikan snapshot diskusi tanpa harus membaca ratusan pesan, (2) identifikasi topik dan saham yang paling banyak dibahas untuk prioritas perhatian, (3) ekstraksi sinyal dari percakapan kasual yang informal namun substantif, dan (4) diferensiasi kontribusi dari investor berpengalaman versus diskusi umum. Mengingat karakteristik anggota PIRANHA sebagai investor berpengalaman dengan *time value* yang tinggi, efisiensi waktu akses informasi menjadi nilai proposition utama.

III.2.1.2 Anggota Grup Advanced (Target Sekunder)

Anggota grup Advanced, dengan sekitar 4.000 anggota premium, menghadapi *pain points* yang serupa namun dengan nuansa berbeda. Meskipun mereka memiliki akses ke rangkuman manual jam 10:00, terdapat gap *coverage* signifikan untuk perdagangan siang hingga sore (10:00-16:00). *Workaround* saat ini adalah memeriksa grup secara periodik, namun ini mengganggu produktivitas dan fokus pada aktivitas lain.

Kebutuhan kelompok ini meliputi: (1) rangkuman tambahan sepanjang hari untuk melengkapi rangkuman manual eksisting, (2) informasi tentang perubahan sentimen atau topik baru yang muncul setelah jam 10:00, dan (3) sintesis lintas grup yang menghubungkan informasi dari Cuap Cuap dan Swing Plan dengan diskusi di Advanced. Kelompok ini juga akan mendapat manfaat dari standarisasi format rangkuman yang konsisten, mengingat variabilitas dalam rangkuman manual.

III.2.1.3 Anggota Cuap Cuap dan Swing Plan

Anggota grup Cuap Cuap (18.000) dan Swing Plan (5.000) menghadapi *pain points* yang berbeda, yaitu keterbatasan akses ke diskusi mendalam yang terjadi di grup premium. Meskipun mereka tidak mengalami *information overload* di grup mereka sendiri, mereka kehilangan konteks penting tentang bagaimana analisis umum atau sinyal trading diimplementasikan dan didiskusikan oleh komunitas premium.

Kebutuhan kelompok ini meliputi: (1) akses (meskipun mungkin terbatas) ke insight dari diskusi premium tanpa harus upgrade keanggotaan, (2) pemahaman tentang sentimen komunitas secara keseluruhan terhadap saham atau kondisi pasar tertentu, dan (3) validasi bahwa sinyal yang mereka terima memiliki dukungan diskusi substantif di tingkat yang lebih tinggi. Sistem rangkuman dapat menyediakan nilai ini sambil tetap menghormati struktur premium komunitas.

III.2.1.4 Michael Yeoh dan Tim Administrator

Michael Yeoh dan dua administrator lainnya memiliki perspektif yang berbeda sebagai *content creator* dan pengelola komunitas. *Pain point* utama mereka adalah kesulitan memastikan bahwa informasi penting yang mereka distribusikan benar-benar sampai dan dipahami oleh semua anggota yang relevan. Keterbatasan rangkuman manual adalah bahwa hanya satu orang (Richy) yang dapat mengalokasikan waktu untuk ini, dan tidak *scalable* untuk memberikan *coverage* yang lebih komprehensif.

Kebutuhan kelompok ini meliputi: (1) alat untuk broadcast insight teragregasi yang dapat menjangkau lebih banyak anggota secara efektif, (2) pemahaman tentang topik dan sentimen yang dominan dalam komunitas untuk informed *content strategy*, (3) reduksi beban manual dalam pembuatan rangkuman, dan (4) mekanisme untuk memastikan konsistensi kualitas rangkuman terlepas dari ketersediaan personel.

III.2.2 Kebutuhan Fungsional

Berdasarkan analisis *stakeholder* dan masalah sistem saat ini, Tabel III.2 merangkum kebutuhan fungsional sistem rangkuman otomatis yang diusulkan. Kebutuhan ini dikategorikan berdasarkan prioritas menggunakan metode MoSCoW (*Must have*, *Should have*, *Could have*, *Won't have*).

Tabel III.2 Kebutuhan fungsional sistem rangkuman otomatis

ID	Kebutuhan Fungsional	Prioritas	Keterangan
FR-01	Sistem harus dapat mengumpulkan pesan dari empat grup Telegram (MY Cuap Cuap, MY Swing Plan, MY Advanced Group, MY PIRANHA) secara <i>real-time</i>	MUST	Fondasi sistem untuk akuisisi data
FR-02	Sistem harus dapat mengekstrak entitas finansial (nama perusahaan, kode ticker saham) menggunakan model <i>cahya/bert-base-indonesian-NER</i> dan <i>whitelist</i> ticker dari <i>Indonesia Stock Exchange</i>	MUST	Identifikasi aset yang dibahas
FR-03	Sistem harus dapat menganalisis emosi pesan menggunakan <i>indonesian-roberta-base-emotion-classifier</i> untuk mendeteksi spektrum emosi (ketakutan, kegembiraan, kemarahan)	MUST	Analisis sentimen psikologis pasar
FR-04	Sistem harus dapat mengidentifikasi topik diskusi dominan dalam setiap periode waktu menggunakan BERTopic	MUST	Pemodelan topik untuk kategorisasi
FR-05	Sistem harus menerapkan <i>weighted influence system</i> dengan bobot: Michael Yeoh (4×), Administrator (3×), VIP members (2×), Regular members (1×)	MUST	Filter intelijen berbasis hierarki

Lanjut ke halaman berikutnya

Tabel III.2 – lanjutan dari halaman sebelumnya

ID	Kebutuhan Fungsional	Prioritas	Keterangan
FR-06	Sistem harus menghasilkan rangkuman secara otomatis pada interval waktu tetap (11:00, 12:00, 14:00, 15:00, 16:00)	MUST	Otomasi periodik
FR-07	Rangkuman harus merupakan sintesis lintas-grup (<i>cross-group merged</i>) yang mengintegrasikan informasi dari keempat grup dalam satu dokumen koheren	MUST	Sintesis holistik
FR-08	Format rangkuman harus <i>mixed</i> yang menggabungkan bullet points untuk informasi terstruktur dan paragraf naratif untuk konteks	MUST	Readability dan comprehensive-ness
FR-09	Sistem TIDAK boleh memproses pesan yang telah di- <i>unsend</i> oleh pengirim untuk menghormati privasi	MUST	Perlindungan privasi eksplisit
FR-10	Sistem harus mengirim rangkuman yang dihasilkan ke kanal Telegram output yang dedicated melalui Telegram Bot API	MUST	Distribusi otomatis
FR-11	Sistem harus menyimpan rangkuman historis beserta metadata (timestamp, jumlah pesan diproses, topik terdeteksi) dalam basis data	SHOULD	Persistensi data dan audit trail
FR-12	Sistem harus mendukung query untuk mengakses rangkuman periode tertentu (harian, mingguan) melalui perintah bot	COULD	Akses retroaktif

III.2.3 Kebutuhan Non-Fungsional

Kebutuhan non-fungsional mendefinisikan atribut kualitas sistem yang harus dipenuhi untuk memastikan sistem dapat beroperasi secara efektif dalam konteks *real-world* komunitas Telegram finansial. Tabel III.3 merangkum kebutuhan non-fungsional beserta target metrik yang dapat diukur.

Tabel III.3 Kebutuhan non-fungsional sistem rangkuman otomatis

ID	Aspek	Target Metrik	Keterangan
NFR-01	<i>Response Time</i>	Rangkuman harus selesai dihasilkan dalam waktu kurang dari 5 menit setelah akhir periode pengumpulan	Relevansi temporal untuk grup PIRANHA yang sangat aktif
NFR-02	Akurasi NER	<i>Precision</i> dan <i>Recall</i> deteksi ticker saham Indonesia harus mencapai minimal 80%	Kualitas ekstraksi entitas finansial
NFR-03	Kualitas <i>Topic Modeling</i>	<i>Coherence score</i> BERTopic harus mencapai minimal 0.5	Topik yang dihasilkan harus interpretatif dan bermakna
NFR-04	<i>System Uptime</i>	Ketersediaan sistem harus mencapai minimal 95% selama jam perdagangan (09:00-16:00)	Reliabilitas operasional
NFR-05	<i>Scalability</i>	Sistem harus mampu menangani lebih dari 250 pesan per jam per grup tanpa degradasi performa	<i>Future-proof</i> untuk pertumbuhan komunitas
NFR-06	Privasi	Zero kebocoran data; pesan yang di- <i>unsend</i> harus dikecualikan dari pemrosesan	Kepatuhan terhadap prinsip privasi

Lanjut ke halaman berikutnya

Tabel III.3 – lanjutan dari halaman sebelumnya

ID	Aspek	Target Metrik	Keterangan
NFR-07	<i>Usability</i>	Rangkuman harus dapat dibaca dan dipahami dalam waktu kurang dari 3 menit oleh anggota rata-rata	Target <i>user experience</i>
NFR-08	Lokalisasi	Semua output rangkuman harus dalam Bahasa Indonesia dengan terminologi finansial yang sesuai	Aksesibilitas untuk komunitas lokal
NFR-09	Efisiensi Biaya	Biaya operasional Groq API untuk inferensi LLM harus kurang dari \$50 per bulan untuk 18 rangkuman per hari	Keberlanjutan finansial

Target metrik pada kebutuhan non-fungsional ini akan menjadi dasar untuk evaluasi sistem pada fase *Evaluation* metodologi DSRM. Beberapa metrik seperti akurasi NER dan kualitas topic modeling dapat diukur secara objektif melalui evaluasi teknis. Metrik lain seperti *usability* dan efisiensi waktu akses akan diukur melalui survei *feedback* pengguna yang akan dilakukan selama fase pengujian sistem.

III.3 Analisis Alternatif Solusi

Untuk mengatasi masalah *information overload* dan fragmentasi informasi yang telah diidentifikasi, terdapat beberapa pendekatan alternatif yang dapat dipertimbangkan. Bagian ini menganalisis empat alternatif utama beserta kelebihan dan kekurangan masing-masing, diikuti dengan analisis pemilihan solusi menggunakan *decision matrix*.

III.3.1 Alternatif Pendekatan untuk Mereduksi *Information Overload*

III.3.1.1 Alternatif 1: Ekspansi *Manual Curation* (Status Quo yang Ditingkatkan)

Alternatif pertama adalah mempertahankan pendekatan rangkuman manual namun dengan ekspansi signifikan dalam frekuensi dan *coverage*. Pendekatan ini melibatkan perekrutan dua hingga tiga asisten tambahan untuk membuat rangkuman multiple kali per hari di semua grup, termasuk PIRANHA yang saat ini tidak memiliki rangkuman manual.

Kelebihan:

- Kualitas penilaian manusia yang superior dalam memahami nuansa konteks dan relevansi informasi
- Tidak ada risiko teknis atau ketergantungan pada sistem *machine learning* yang mungkin menghasilkan output yang tidak akurat
- Fleksibilitas dalam menyesuaikan gaya dan format rangkuman berdasarkan feedback komunitas secara *ad-hoc*

Kekurangan:

- Biaya operasional tinggi: dengan asumsi kompensasi Rp 5 juta per bulan per asisten, total biaya untuk tiga asisten adalah Rp 15 juta per bulan, atau Rp 180 juta per tahun
- Masalah skalabilitas fundamental: untuk memberikan *coverage* enam kali per hari di empat grup memerlukan 24 instansi rangkuman manual per hari, yang sulit dikoordinasikan dan memastikan konsistensi
- Bias subjektif dan variabilitas antar-*curator*: setiap asisten mungkin memiliki perspektif dan prioritas berbeda dalam memilih informasi yang dianggap penting
- *Human fatigue* dan potensi kesalahan: membuat rangkuman berkualitas tinggi secara konsisten memerlukan fokus dan energi kognitif yang tinggi, yang sulit dipertahankan dalam jangka panjang
- Ketidakmampuan untuk menerapkan *weighted influence* secara algoritmik dan konsisten

III.3.1.2 Alternatif 2: Bot Filter Berbasis Aturan (*Rule-Based Keyword Bot*)

Alternatif kedua adalah mengembangkan bot Telegram sederhana yang memfilter pesan berdasarkan aturan *keyword* dan pola teks yang telah ditentukan sebelumnya. Bot akan mendeteksi kata kunci seperti nama saham (PTRO, BBKA, GOTO), kata-

kata sinyal (*buy, sell, resistance, support*), dan mengirimkan pesan yang cocok ke kanal output.

Kelebihan:

- Implementasi cepat dan mudah: sistem berbasis aturan dapat dibangun dalam 1-2 minggu
- Biaya operasional sangat rendah: hosting bot sederhana dapat dilakukan dengan biaya kurang dari \$5 per bulan
- Transparan dan dapat diprediksi: aturan filter jelas dan dapat disesuaikan oleh administrator tanpa keahlian *machine learning*

Kekurangan:

- Tidak ada pemahaman konteks: sistem *keyword matching* tidak dapat membedakan antara diskusi substantif tentang saham versus sekadar menyebut nama saham secara kasual
- *False positive* tinggi: banyak pesan yang mengandung keyword tetapi tidak informatif akan difilter
- *False negative* tinggi: informasi penting yang tidak menggunakan keyword eksplisit akan terlewat
- Tidak dapat melakukan sintesis atau agregasi: bot hanya meneruskan pesan individual tanpa menciptakan rangkuman yang koheren
- *Brittle* dan memerlukan pemeliharaan konstan: setiap perubahan dalam cara komunitas berkomunikasi memerlukan update manual aturan
- Tidak dapat menerapkan *weighted influence* atau analisis sentimen yang kompleks

III.3.1.3 Alternatif 3: Bot Rangkuman Berbasis AI dengan NLP Khusus Domain (DIUSULKAN)

Alternatif ketiga, yang merupakan solusi yang diusulkan dalam penelitian ini, adalah mengembangkan sistem rangkuman otomatis berbasis *artificial intelligence* dengan *pipeline* NLP yang dioptimalkan untuk domain finansial Indonesia. Sistem ini mengintegrasikan multiple komponen: ekstraksi entitas menggunakan *bert-based-indonesian-NER*, analisis emosi dengan *indonesian-roberta-base-emotion-classifier*, pemodelan topik dengan BERTopic, *weighted influence algorithm*, dan generasi rangkuman dengan *large language model* (Llama 3.1 70B via Groq API).

Kelebihan:

- Pemahaman konteks berbasis semantik: model NLP dapat memahami makna

pesan di luar sekadar *keyword matching*

- Skalabilitas tinggi: setelah dibangun, sistem dapat memproses empat grup dengan enam rangkuman per hari (total 24 rangkuman) tanpa penambahan biaya proporsional
- *Weighted influence system* algoritmik: dapat secara konsisten menerapkan bobot berbeda berdasarkan peran pengirim (Michael Yeoh 4×, admin 3×, VIP 2×, regular 1×) yang sulit dilakukan secara manual
- Optimisasi untuk domain finansial Indonesia: penggunaan model yang di-*fine-tune* untuk bahasa Indonesia dan entitas finansial lokal (kode ticker BEI, nama perusahaan Indonesia)
- Sintesis lintas-grup otomatis: kemampuan untuk mengintegrasikan informasi dari empat grup dalam satu rangkuman koheren
- Perlindungan privasi built-in: mekanisme untuk mengecualikan pesan yang di-*unsubscribe*
- Biaya operasional terjangkau: estimasi \$30-50 per bulan untuk Groq API, jauh lebih rendah dibanding Rp 15 juta per bulan untuk solusi manual
- Konsistensi output: format dan kualitas rangkuman stabil karena dihasilkan oleh sistem yang sama

Kekurangan:

- Kompleksitas teknis tinggi: memerlukan keahlian dalam NLP, *machine learning*, dan integrasi sistem
- Ketergantungan pada kualitas model: akurasi sistem terbatas pada performa model NER (target 80%), topic modeling, dan LLM
- *Vendor lock-in risk*: ketergantungan pada Groq API untuk inferensi LLM menimbulkan risiko jika layanan mengalami *downtime* atau perubahan kebijakan
- Waktu pengembangan lebih lama: estimasi 2-3 bulan untuk implementasi dan *fine-tuning* sistem lengkap
- Potensi kesalahan model: meskipun target akurasi 80%, tetap ada 20% margin error yang dapat menghasilkan informasi yang tidak akurat atau tidak relevan

III.3.1.4 Alternatif 4: Solusi SaaS *Generic* dengan LLM *Third-Party*

Alternatif keempat adalah menggunakan layanan SaaS *third-party* seperti API ChatGPT (OpenAI) atau Claude (Anthropic) secara langsung untuk rangkuman tanpa *pipeline* NLP kustom. Pendekatan ini melibatkan pengiriman pesan Telegram ke API LLM generik dengan *prompt engineering* untuk menghasilkan rangkuman.

Kelebihan:

- Deployment cepat: dapat diimplementasikan dalam 1-2 minggu dengan *prompt engineering*
- Kualitas rangkuman umum yang baik: LLM seperti GPT-4 atau Claude memiliki kemampuan rangkuman yang superior untuk teks umum
- Tidak memerlukan maintenance model NLP: vendor menangani update dan pemeliharaan model

Kekurangan:

- Tidak ada ekstraksi entitas finansial Indonesia: model generik tidak dioptimalkan untuk mengenali kode ticker BEI atau nama perusahaan Indonesia
- Tidak ada *weighted influence logic*: sistem tidak dapat membedakan bobot kontribusi berdasarkan peran pengirim
- Risiko privasi signifikan: semua data pesan dikirim ke server pihak ketiga (OpenAI/Anthropic) yang berada di luar Indonesia
- Biaya operasional tinggi: untuk volume 1.400+ pesan per hari dengan context window besar, biaya GPT-4 atau Claude dapat mencapai \$100-200 per bulan
- Kurang customizable: sulit untuk mengintegrasikan logika bisnis spesifik seperti *cross-group synthesis* atau perlindungan privasi untuk pesan yang di-*unsend*

III.3.2 Analisis Pemilihan Solusi

Untuk memilih alternatif solusi yang optimal, dilakukan analisis menggunakan *decision matrix* dengan enam kriteria yang diberi bobot berdasarkan prioritas untuk konteks komunitas Telegram finansial ini. Tabel III.4 menyajikan *decision matrix* dengan skor untuk setiap alternatif pada setiap kriteria.

Tabel III.4 *Decision matrix* untuk pemilihan alternatif solusi

Kriteria	Bobot	Alt 1: Manu- al	Alt 2: Rule	Alt 3: AI NLP	Alt 4: SaaS
Skalabilitas (24 re- cap/hari)	25%	1/5	5/5	5/5	4/5
Akurasi (konteks fi- nansial)	20%	5/5	2/5	4/5	3/5
Efisiensi Biaya	15%	1/5	5/5	4/5	3/5
<i>Customizability</i> (we- ighted)	20%	3/5	2/5	5/5	2/5
Privasi & Kepatuhan	10%	5/5	5/5	5/5	2/5
Kecepatan Deplo- yment	10%	4/5	5/5	3/5	4/5
Skor Tertimbang	100%	2.75	3.60	4.35	3.00

Berdasarkan *decision matrix*, Alternatif 3 (Bot Rangkuman Berbasis AI dengan NLP Khusus Domain) memperoleh skor tertimbang tertinggi sebesar 4.35 dari skala 5.0. Pemilihan alternatif ini dijustifikasi oleh beberapa alasan strategis dan teknis berikut.

Pertama, dari perspektif *return on investment* (ROI), solusi AI NLP menawarkan efisiensi biaya yang dramatis. Dengan biaya operasional sekitar \$40 per bulan (sekitar Rp 600.000 dengan kurs Rp 15.000 per dolar), solusi ini 96% lebih murah dibandingkan solusi manual yang memerlukan Rp 15 juta per bulan untuk tiga asisten, sambil memberikan *coverage* yang sama atau bahkan lebih baik (24 rangkuman per hari).

Kedua, dari perspektif superioritas teknis, solusi AI NLP adalah satu-satunya alternatif yang dapat mengimplementasikan *weighted influence system* secara algoritmik dan konsisten. Kemampuan untuk memberikan bobot 4× untuk Michael Yeoh, 3× untuk administrator, 2× untuk VIP members, dan 1× untuk regular members secara otomatis dalam setiap rangkuman adalah keunggulan yang tidak dapat ditiru oleh solusi lain. Selain itu, optimisasi untuk domain finansial Indonesia melalui model NER khusus dan *emotion classifier* Bahasa Indonesia memberikan akurasi kontekstual yang tidak dapat dicapai oleh solusi generik.

Ketiga, skalabilitas adalah kriteria kritis mengingat volume pesan ekstrim di grup

PIRANHA (>200 msg/jam) dan Advanced (200 msg/jam). Solusi manual tidak dapat di-*scale* untuk menangani volume ini secara sustainable, sementara solusi *rule-based* menghasilkan terlalu banyak *false positive/negative*. Solusi AI NLP dapat memproses ratusan pesan per jam dengan performa konsisten tanpa degradasi kualitas.

Keempat, kemampuan sintesis lintas-grup adalah diferensiator utama. Tidak ada alternatif lain yang dapat secara otomatis dan konsisten mengintegrasikan informasi dari empat grup dengan karakteristik berbeda (broadcast, signal, diskusi formal, diskusi kasual) menjadi satu rangkuman koheren yang menyediakan pandangan holistik. Kemampuan ini sangat penting mengingat fragmentasi informasi adalah salah satu masalah inti yang diidentifikasi dalam sistem saat ini.

Kelima, dari perspektif kontribusi riset, Alternatif 3 menawarkan novelitas akademis tertinggi. Implementasi *weighted community intelligence* untuk komunitas finansial Indonesia, optimisasi NLP untuk domain lokal yang *underrepresented* dalam literatur, dan validasi *real-world* dengan komunitas 22.000+ anggota memberikan kontribusi signifikan terhadap badan pengetahuan dalam bidang NLP finansial dan sistem informasi komunitas.

Meskipun Alternatif 3 memiliki kompleksitas teknis tertinggi dan waktu pengembangan terpanjang, trade-off ini dapat dijustifikasi mengingat ini adalah proyek tugas akhir yang bertujuan untuk mengembangkan solusi inovatif dengan kontribusi riset yang substantif. Lebih lanjut, periode pengembangan 2-3 bulan masih feasible dalam konteks timeline tugas akhir.

Aspek perlindungan privasi melalui eksklusi pesan yang di-*unsend* juga merupakan keunggulan yang membedakan Alternatif 3 dari solusi SaaS *third-party* yang memiliki risiko privasi lebih tinggi karena data dikirim ke server eksternal. Implementasi lokal dengan Groq API (yang tidak menyimpan data training dari input) memberikan balance antara kemampuan LLM *state-of-the-art* dan perlindungan data komunitas.

Berdasarkan analisis komprehensif ini, Alternatif 3 dipilih sebagai solusi yang akan diimplementasikan dalam penelitian tugas akhir ini. Desain konsep solusi yang detail akan diuraikan pada Bab IV.

BAB IV

DESAIN KONSEP SOLUSI

Ilustrasikan desain konsep solusi dalam bentuk model konseptual dan penjelasan secara ringkas, beserta perbedaannya dengan sistem saat ini. Ilustrasi harus dapat dibandingkan (*before and after*). Karena masih berupa proposal, bab ini hanya berisi gambar desain konsep solusi tersebut dan penjelasan perbandingannya dengan gambar sistem yang ada saat ini (yang tergambar di awal Bab III).

BAB V

RENCANA SELANJUTNYA

Jelaskan secara detail langkah-langkah rencana selanjutnya, hal-hal yang diperlukan atau akan disiapkan, dan risiko dan mitigasinya, yang meliputi:

1. Rencana implementasi, termasuk alat dan bahan yang diperlukan, lingkungan, konfigurasi, biaya, dan sebagainya.
2. Desain pengujian dan evaluasi, misalnya metode verifikasi dan validasi.
3. Analisis risiko dan mitigasi, misalnya tindakan selanjutnya jika ada yang tidak berjalan sesuai rencana.

DAFTAR PUSTAKA

- Angelov, Dimo, dan Diana Inkpen. 2024. "Topic Modeling: Contextual Token Embeddings Are All You Need". Dalam *Findings of EMNLP 2024*. <https://doi.org/10.18653/v1/2024.findings-emnlp.790>.
- AppMagic. 2024. *Number of Telegram App Downloads Worldwide in 2023, by Country*. Statista. Indonesia: approximately 27.21 million downloads. <https://www.statista.com/statistics/1336855/telegram-downloads-by-country/>.
- Araci, Doğu. 2019. "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models". *arXiv preprint arXiv:1908.10063*, <https://arxiv.org/abs/1908.10063>.
- Arnold, Miriam, Mascha Goldschmitt, dan Thomas Rigotti. June 2023. "Dealing with Information Overload: A Comprehensive Review". *Frontiers in Psychology* 14 (): 1122200. <https://doi.org/10.3389/fpsyg.2023.1122200>.
- Beebom. 2023. *20 Best Telegram Bots to Enhance Your Messaging Experience*. Beebom. Diakses 21 Oktober 2025. <https://beebom.com/telegram-bots/>.
- Bollen, Johan, Huina Mao, dan Xiaojun Zeng. 2011. "Twitter Mood Predicts the Stock Market". *Journal of Computational Science* 2 (1): 1–8. <https://doi.org/10.1016/j.jocs.2010.12.007>.
- Bursa Efek Indonesia. September 2024. *Jumlah Investor Saham di Indonesia Lampau 6 Juta SID*. Siaran Pers 2224. Total investor: 6.001.573 SID per 25 September 2024. Pertumbuhan 744.000 investor baru di tahun 2024. Bursa Efek Indonesia. <https://www.idx.co.id/>.
- Chapman, Pete, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, dan Rüdiger Wirth. 2000. *CRISP-DM 1.0: Step-by-Step Data Mining Guide*. SPSS Inc. CRISP-DM Consortium. <https://www.kde.cs.uni-kassel.de/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf>.

- Chen, Jiaao, dan Diyi Yang. 2021. "Structure-Aware Abstractive Conversation Summarization via Discourse and Action Graphs". Dalam *Proceedings of NAACL 2021*. <https://doi.org/10.18653/v1/2021.naacl-main.109>.
- Cheng, Ing-Haw, Shan Ge, dan Yan Li. 2023. "Retail Investor Attention and the Idiosyncratic Volatility Puzzle". Analisis perilaku investor ritel dan peran media sosial dalam keputusan investasi, *Journal of Financial Economics* 147 (2): 540–564. <https://doi.org/10.1016/j.jfineco.2022.12.003>.
- DeYoung, Jay, Iz Beltagy, Madeleine van Zuylen, Bailey Kuehl, dan Lucy Lu Wang. 2024. "Do Multi-Document Summarization Models Synthesize?" *Transactions of the Association for Computational Linguistics* 12. https://doi.org/10.1162/tacl_a_00687. <https://arxiv.org/abs/2301.13844>.
- Edmunds, Angela, dan Anne Morris. 2000. "The Problem of Information Overload in Business Organisations: A Review of the Literature". *International Journal of Information Management* 20 (1): 17–28. [https://doi.org/10.1016/S0268-4012\(99\)00051-1](https://doi.org/10.1016/S0268-4012(99)00051-1).
- Egger, Roman, dan Joanne Yu. 2022. "A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts". *Frontiers in Sociology* 7. <https://doi.org/10.3389/fsoc.2022.886498>.
- Eppler, Martin J., dan Jeanne Mengis. 2004. "The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines". *The Information Society* 20 (5): 325–344. <https://doi.org/10.1080/01972240490507974>.
- Fabbri, Alexander, Faiaz Rahman, Imad Rizvi, Borui Wang, Haoran Li, Yashar Mehdad, dan Dragomir Radev. 2021. "ConvoSumm: Conversation Summarization Benchmark and Improved Abstractive Summarization with Argument Mining". Dalam *Proceedings of ACL-IJCNLP 2021*. <https://doi.org/10.18653/v1/2021.acl-long.535>.
- Gliwa, Bogdan, Iwona Mochol, Maciej Biesek, dan Aleksander Wawer. 2019. "SAM-Sum Corpus: A Human-annotated Dialogue Dataset for Abstractive Summarization". Dalam *Proceedings of the 2nd Workshop on New Frontiers in Summarization*. <https://doi.org/10.18653/v1/D19-5409>.

- Grootendorst, Maarten. 2022. "BERTopic: Neural topic modeling with a class-based TF-IDF procedure". *arXiv preprint arXiv:2203.05794*, <https://arxiv.org/abs/2203.05794>.
- Hamidi, Hojat, Anouar Ben Thameur, Ali S. Alhejaili, Shafaq Anwar, dan Omar Al-fandi. 2021. "Real-time Text Stream Processing: A Dynamic and Distributed NLP Pipeline". Dalam *2021 International Symposium on Electrical, Electronics and Information Engineering*. <https://doi.org/10.1145/3459104.3459198>.
- Hashemi, Ali, dan Mohammad Ali Zare Chahooki. 2019. "Telegram group quality measurement by user behavior analysis". *Social Network Analysis and Mining* 9 (33). <https://doi.org/10.1007/s13278-019-0575-9>.
- Heitmayer, Maxi. 2025. "The Second Wave of Attention Economics: Attention as a Universal Symbolic Currency on Social Media and beyond". *Interacting with Computers* 37 (1): 18–29. <https://doi.org/10.1093/iwc/iwae035>.
- International Data Corporation. March 2021. *Worldwide Global DataSphere Forecast, 2021–2025: The World Keeps Creating More Data — Now, What Do We Do with It All?* Document #US46410421. Forecast projects 181 zettabytes by 2025. IDC Corporate USA. <https://www.idc.com/getdoc.jsp?containerId=US46410421>.
- Isah, Hazem, Tariq Abughofa, Sazia Mahfuz, Dharmitha Ajerla, Farhana Zulker-nine, dan Shahzad Khan. 2019. "A Survey of Distributed Data Stream Processing Frameworks". *IEEE Access* 7. <https://doi.org/10.1109/ACCESS.2019.2946884>.
- Junaidi, Junaidi, dan Nurhidayah Nurhidayah. 2023. "Social Media Impact on Trading Behavior: An Examination Among Indonesian Young Adult Investors with Capital Market Literacy as a Mediator". 314 responden Indonesia, *JEMA: Jurnal Ilmiah Bidang Akuntansi dan Manajemen* 20 (1): 136–155. <https://doi.org/10.31106/jema.v20i1.19687>.
- Kano, Ryuji, Yasuhide Miura, Motoki Taniguchi, Yan-Ying Chen, Francine Chen, dan Tomoko Ohkuma. 2018. "Harnessing Popularity in Social Media for Extractive Summarization of Online Conversations". Dalam *Proceedings of EMNLP 2018*. <https://doi.org/10.18653/v1/D18-1144>.

- Khairunnisa, Siti Oryza, Aizhan Imankulova, dan Mamoru Komachi. 2020. “Towards a Standardized Dataset on Indonesian Named Entity Recognition”. Dalam *Proceedings of ACL-IJCNLP 2020: Student Research Workshop*. <https://doi.org/10.18653/v1/2020.aacl-srw.10>.
- Koto, Fajri, Afshin Rahimi, Jey Han Lau, dan Timothy Baldwin. 2020. “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP”. Dalam *Proceedings of COLING 2020*. <https://doi.org/10.18653/v1/2020.coling-main.66>.
- Liu, Yang, dan Mirella Lapata. 2019. “Text Summarization with Pretrained Encoders”. Dalam *Proceedings of EMNLP-IJCNLP 2019*. <https://arxiv.org/abs/1908.08345>.
- Morgia, Massimo La, Alessandro Mei, Alberto Maria Mongardini, dan Jie Wu. 2024. “TGDataset: Collecting and Exploring the Largest Telegram Channels Dataset”. Dalam *Proceedings of ACM SIGKDD 2024*. <https://doi.org/10.1145/3690624.3709397>.
- Nematzadeh, Azadeh, Giovanni Luca Ciampaglia, Yong-Yeol Ahn, dan Alessandro Flammini. 2019. “Information Overload in Group Communication: From Conversation to Cacophony in the Twitch Chat”. *Royal Society Open Science* 6 (10): 191412. <https://doi.org/10.1098/rsos.191412>.
- Nobari, Arash Dargahi, Negar Reshadatmand, dan Mahmood Neshati. 2017. “Analysis of Telegram, An Instant Messaging Service”. Dalam *Proceedings of CIKM 2017*. <https://doi.org/10.1145/3132847.3133132>.
- Nugroho, Kemal Surya, Ary Yusuf Sukmadewa, Hafizh Wuswilahaken DW, Fajar Ade Bachtiar, dan Novanto Yudistira. 2021. “BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews”. Dalam *Proceedings of SIET 2021*. <https://doi.org/10.1145/3479645.3479679>. <https://arxiv.org/abs/2107.06802>.
- Otoritas Jasa Keuangan. 2022. *Laporan Tahunan OJK 2022*. Total investor pasar modal: 10.3 juta (akhir 2022), pertumbuhan 1.000% sejak 2017. Jakarta: Otoritas Jasa Keuangan. <https://www.ojk.go.id/>.

- Pawar, Devika, Pratiksha Kokate, Kanchan Warke, Shubham Kanse, dan Satish R. Devane. 2022. “Stock Market Price Prediction Using LSTM RNN and Sentiment Analysis”. Menggunakan LSTM dan analisis sentimen Twitter untuk prediksi harga saham, *ECS Transactions* 107 (1): 14087–14097. <https://doi.org/10.1149/10701.14087ecst>.
- Perlo, Alessandro, Giordano Paoletti, Nikhil Jha, Luca Vassio, Jussara Almeida, dan Marco Mellia. 2025. “Topic-wise Exploration of the Telegram Group-verse”. Dalam *Companion Proceedings of the ACM on Web Conference 2025*, 1792–1801. Sydney, NSW, Australia. <https://doi.org/10.1145/3701716.3717506>.
- Pilault, Jonathan, Raymond Li, Sandeep Subramanian, dan Chris Pal. 2020. “On Extractive and Abstractive Neural Document Summarization with Transformer Language Models”. Dalam *Proceedings of EMNLP 2020*. <https://doi.org/10.18653/v1/2020.emnlp-main.748>.
- Ravaut, Mathieu, Aixin Sun, Nancy Chen, dan Shafiq Joty. 2024. “On Context Utilization in Summarization with Large Language Models”. Dalam *Proceedings of ACL 2024 - Volume 1: Long Papers*. <https://doi.org/10.18653/v1/2024.acl-long.153>.
- Rubert, Marcos, dan Kleinner Farias. 2023. “On the impact of event-driven architecture on performance: An exploratory study”. *Future Generation Computer Systems* 153. <https://doi.org/10.1016/j.future.2023.10.021>.
- Shah, Agam, Abhinav Gullapalli, Ruchit Vithani, Michael Galarnyk, dan Sudheer Chava. 2023. “FiNER-ORD: Financial Named Entity Recognition Open Research Dataset”. *arXiv preprint arXiv:2302.11157*, <https://arxiv.org/abs/2302.11157>.
- Shahrzadi, Leila, Khadijeh Esfandyari, Maryam Nematollahi, Elham Monaghesh, dan Mehdi Kazemi. 2024. “Causes, consequences, and strategies to deal with information overload: A scoping review”. *International Journal of Information Management Data Insights* 4. <https://doi.org/10.1016/j.jjime.2024.100261>.
- Shen, Chenhui, Liying Cheng, Xuan-Phi Nguyen, Yang You, dan Lidong Bing. 2023. “Large Language Models are Not Yet Human-Level Evaluators for Abstractive Summarization”. Dalam *Findings of EMNLP 2023*. <https://doi.org/10.18653/v1/2023.findings-emnlp.278>.

- Si, Jianfeng, Arjun Mukherjee, Bing Liu, Sinno Jialin Pan, Qing Li, dan Huayi Li. 2014. “Exploiting Social Relations and Sentiment for Stock Prediction”. Dalam *Proceedings of EMNLP 2014*. <https://doi.org/10.3115/v1/D14-1120>.
- Sidnam-Mauch, Emily, dan Peter Monge. 2024. “Individual Differences and Technology Affordances Combine to Predict Mobile Social Media Distraction Behaviors and Consequences”. Dalam *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3613904.3641950>.
- Skulmowski, Alexander, dan Günter Daniel Rey. 2022. “Understanding Cognitive Load in Digital and Online Learning: a New Perspective on Extraneous Cognitive Load”. *Educational Psychology Review* 34:171–196. <https://doi.org/10.1007/s10648-021-09624-7>.
- Sotudeh, Sajad, Hanieh Deilamsalehy, Franck Dernoncourt, dan Nazli Goharian. 2021. “TLDR9+: A Large Scale Resource for Extreme Summarization of Social Media Posts”. Dalam *Proceedings of the Third Workshop on New Frontiers in Summarization*. <https://doi.org/10.18653/v1/2021.newsum-1.15>.
- Sweller, John. 1988. “Cognitive Load During Problem Solving: Effects on Learning”. *Cognitive Science* 12 (2): 257–285. https://doi.org/10.1207/s15516709cog1202_4.
- Tian, Yuanhe, Fei Xia, dan Yan Song. 2024. “Dialogue Summarization with Mixture of Experts based on Large Language Models”. Dalam *Proceedings of ACL 2024*. <https://doi.org/10.18653/v1/2024.acl-long.385>.
- Wang, Qingyue, Liang Ding, Yanan Cao, Zhiliang Tian, Shi Wang, Dacheng Tao, dan Li Guo. 2023. “Recursively Summarizing Enables Long-Term Dialogue Memory in Large Language Models”. *arXiv preprint arXiv:2308.15022*, <https://arxiv.org/abs/2308.15022>.
- Wilie, Bryan, Karissa Vincentio, Genta Indra Winata, Samuel Cahyawijaya, Xiaohong Li, Zhi Yuan Lim, Sidik Soleman, dkk. 2020. “IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding”. Dalam *Proceedings of AACL-IJCNLP 2020*. <https://doi.org/10.18653/v1/2020.aacl-main.85>.

Xu, Yichong, Xiaodong Liu, Chunyuan Li, Hoifung Poon, dan Jianfeng Gao. 2021. “DialogSum: A Real-Life Scenario Dialogue Summarization Dataset”. Dataset dan metode untuk merangkum percakapan dialog menggunakan NLP, *arXiv preprint arXiv:2105.06762*, <https://arxiv.org/abs/2105.06762>.

Zhang, Tianyi, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, dan Yoav Artzi. 2020. “BERTScore: Evaluating Text Generation with BERT”. Dalam *International Conference on Learning Representations*. <https://arxiv.org/abs/1904.09675>.

Zhang, Yuzhe, dan Hong Zhang. 2022. “FinBERT-MRC: Financial Named Entity Recognition Using BERT Under the Machine Reading Comprehension Paradigm”. *arXiv preprint arXiv:2205.15485*, <https://arxiv.org/abs/2205.15485>.