MODEL ABSA BERBASIS BERT DENGAN BIO TAGGING DAN CROSS-ATTENTION UNTUK ANALISIS MAKROEKONOMI DALAM BERITA FINANSIAL

Ezra Abednego, Hindriyanto Dwi Purnomo
Teknologi Informasi, Universitas Kristen Satya Wacana, Indonesia
672021118@student.uksw.edu

ABSTRAK

Perkembangan volume pemberitaan ekonomi global yang semakin masif menuntut sistem analisis teks finansial yang mampu mengekstraksi sentimen secara granular terhadap berbagai indikator makroekonomi. Pendekatan sentiment analysis konvensional yang hanya menghasilkan satu label sentimen per kalimat terbukti tidak memadai untuk teks berita finansial yang memuat banyak aspek dengan polaritas berbeda dalam satu kalimat. Selain itu, model Aspect-Based Sentiment Analysis (ABSA) yang ada belum mampu mengekstraksi hubungan eksplisit antara aspek, opini, dan sentimen secara simultan pada domain makroekonomi. Penelitian ini bertujuan membangun model ABSA berbasis Bidirectional Encoder Representations from Transformers (BERT) yang mengintegrasikan skema BIO Tagging dan mekanisme cross-attention untuk mengekstraksi triplet aspek–opini–sentimen secara end-to-end dari teks berita makroekonomi. Dataset terdiri atas 3.873 kalimat dari Investing.com, Bloomberg Economics, dan CNBC International yang dianotasi secara manual sebagai triplet aspek–opini–sentimen dengan skema BIO. Model dilatih secara end-to-end menggunakan pendekatan multi-task learning. Hasil eksperimen menunjukkan model mencapai kinerja tinggi pada seluruh komponen, dengan F1 pairing 0,94–0,95, akurasi klasifikasi sentimen 0,86–0,89 dengan Macro-F1 0,80–0,85, serta triplet F1 pada kisaran 0,87–0,88.

Kata kunci: *ABSA, BERT, BIO Tagging, Cross-Attention, Analisis Sentimen Finansial, Makroekonomi*

1. PENDAHULUAN

Perkembangan teknologi informasi dan digitalisasi media telah menghasilkan lonjakan volume pemberitaan ekonomi global yang sangat masif. Ribuan artikel berita finansial diproduksi setiap harinya oleh berbagai sumber seperti Bloomberg, Reuters, CNBC, dan Financial Times, mencakup isu-isu makroekonomi mulai dari kebijakan suku bunga bank sentral, dinamika inflasi, pertumbuhan GDP, hingga kondisi pasar tenaga kerja. Volume informasi ini jauh melampaui kapasitas analis manusia untuk memprosesnya secara manual, sementara keputusan investasi dan kebijakan fiskal sering kali bergantung pada pemahaman yang cepat dan akurat terhadap sentimen dalam berita tersebut [1]. Dalam konteks ini, analisis sentimen berbasis kecerdasan buatan menjadi solusi yang semakin relevan [2]. Namun, pendekatan sentiment analysis konvensional yang menghasilkan satu label sentimen per kalimat terbukti tidak memadai untuk teks finansial — satu kalimat berita ekonomi dapat memuat beberapa indikator dengan arah sentimen berbeda secara bersamaan, misalnya kalimat "The Fed raised rates aggressively as inflation surged, while GDP growth showed unexpected resilience" mengandung sentimen negatif terhadap inflasi, positif terhadap pertumbuhan GDP, dan netral terhadap kebijakan moneter sekaligus. Pendekatan satu-label-per-kalimat akan kehilangan granularitas informasi yang justru paling dibutuhkan oleh pelaku pasar dan pembuat kebijakan [3].

Aspect-Based Sentiment Analysis (ABSA) hadir sebagai jawaban atas keterbatasan tersebut, dengan kemampuannya mengekstraksi sentimen pada level aspek secara terpisah untuk setiap entitas yang disebutkan dalam satu kalimat [2]. Namun, meskipun ABSA menawarkan pendekatan yang lebih granular, penerapannya pada domain makroekonomi masih menghadapi keterbatasan yang mendasar. Pertama, teks berita finansial memiliki karakteristik linguistik yang jauh lebih kompleks dibanding domain ulasan produk yang selama ini menjadi tolok ukur pengembangan model ABSA — opini ekonomi jarang diekspresikan secara eksplisit, melainkan tersembunyi dalam perubahan indikator kuantitatif seperti "yields climbed" atau "output contracted" yang maknanya sangat bergantung pada konteks ekonomi makro secara keseluruhan [4]. Kedua, satu kalimat berita ekonomi kerap memuat beberapa aspek sekaligus dengan polaritas yang berbeda atau bahkan berlawanan, sehingga model yang tidak mampu membedakan relasi antar-aspek berisiko menghasilkan interpretasi sentimen yang menyesatkan [3]. Ketiga dan paling krusial, model ABSA yang ada hingga saat ini umumnya memformulasikan ekstraksi aspek, opini, dan sentimen sebagai tugas-tugas yang terpisah atau bergantung pada daftar aspek yang dipre-definisikan — pendekatan yang secara inheren tidak mampu menangkap triplet aspek–opini–sentimen secara simultan dan end-to-end dalam teks berita makroekonomi yang dinamis [5].

Penelitian ini bertujuan untuk merancang dan mengevaluasi model ABSA yang mampu mengekstraksi triplet aspek–opini–sentimen secara simultan dari teks berita makroekonomi berbahasa Inggris, dengan memanfaatkan representasi kontekstual berbasis BERT yang diperkuat dengan mekanisme BIO Tagging dan cross-attention sebagai fondasi arsitekturalnya.

Untuk menjawab permasalahan tersebut, penelitian ini mengusulkan arsitektur tiga komponen yang bekerja secara terpadu. Encoder BERT digunakan sebagai backbone representasi kontekstual yang mampu menangkap dependensi semantik jangka panjang antar token dalam kalimat berita finansial [6]. Di atasnya, skema BIO Tagging diterapkan untuk mengidentifikasi batas span aspek dan opini pada level token secara presisi. Selanjutnya, modul cross-attention dua arah digunakan untuk memodelkan hubungan relasional antara span aspek dan opini secara eksplisit, menggantikan pendekatan pipeline konvensional yang rawan propagasi kesalahan [5]. Keseluruhan arsitektur dilatih secara end-to-end menggunakan strategi multi-task learning yang mengoptimalkan tagging loss, pairing loss, dan sentiment classification loss secara simultan — memastikan setiap komponen tidak hanya belajar dari tugasnya masing-masing, tetapi juga saling memperkuat dalam membentuk triplet yang koheren dan akurat.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Maia et al. memperkenalkan dataset FiQA melalui kompetisi WWW'18, yang menjadi benchmark awal analisis sentimen berbasis aspek pada teks finansial. Penelitian ini menegaskan bahwa teks finansial memerlukan pendekatan khusus karena karakteristik linguistiknya yang berbeda dari domain umum, meskipun fokusnya masih terbatas pada aspek tingkat perusahaan dan belum menyentuh indikator makroekonomi [4]. Costa dan Silva melanjutkan dengan membangun sistem ABSA untuk dataset FiQA menggunakan fitur linguistik dan representasi semantik, dan mengonfirmasi bahwa sentimen dalam teks finansial bersifat sangat implisit dan domain-dependent — namun seperti pendahulunya, masih bergantung pada daftar aspek yang dipre-definisikan [7].

Pada ranah ekstraksi triplet, Peng et al. memperkenalkan paradigma Aspect Sentiment Triplet Extraction (ASTE) yang mengekstraksi aspek, opini, dan polaritas sentimen secara simultan, dan menunjukkan bahwa pendekatan end-to-end menghasilkan performa lebih baik dibandingkan pipeline bertahap [5]. Aziz et al memperluas pendekatan ini dengan mengintegrasikan BERT dan multi-layered Graph Convolutional Network untuk menangani seluruh sub-tugas ABSA secara terpadu, meskipun evaluasinya masih terbatas pada dataset ulasan konsumen umum [8].

Wang mengusulkan model ABSA berbasis *weak supervision* pada dokumen teks makroekonomi, khususnya risalah rapat FOMC yang mencakup diskusi kebijakan moneter, inflasi, dan ketenagakerjaan [9]. Penelitian ini menjadi salah satu upaya awal yang secara eksplisit menerapkan ABSA pada teks makroekonomi dan menunjukkan korelasi signifikan antara sentimen aspek dengan pergerakan indikator ekonomi riil, namun masih menggunakan pendekatan pipeline dan belum mengintegrasikan ekstraksi triplet secara end-to-end.

Secara keseluruhan, gap yang konsisten dari penelitian-penelitian tersebut adalah belum adanya model yang mengintegrasikan ekstraksi triplet aspek–opini–sentimen secara simultan pada domain teks berita makroekonomi — gap yang menjadi motivasi utama penelitian ini.

2.2. Sentiment Analysis dan ABSA

Sentiment analysis telah banyak dimanfaatkan dalam berbagai konteks, termasuk analisis opini publik dan business intelligence, karena kemampuannya merangkum pandangan dan sentimen secara otomatis dalam skala besar [2]. Pendekatan konvensional umumnya mengasumsikan satu sentimen terhadap satu topik per kalimat, sehingga kurang mampu merepresentasikan kompleksitas opini yang melibatkan lebih dari satu aspek [10]. Keterbatasan ini mendorong berkembangnya ABSA, yang mengidentifikasi aspek-aspek spesifik dan menetapkan nilai sentimen secara terpisah pada masing-masing aspek. Penerapan ABSA pada domain finansial menghadapi kompleksitas lebih tinggi karena penggunaan bahasa bernuansa, terminologi teknis, serta ketergantungan pada konteks ekonomi dan pengetahuan domain [4], [11].

2.3. Aspect Sentiment Triplet Extraction (ASTE)

ASTE merupakan perluasan dari ABSA yang mengekstraksi tiga elemen secara simultan — aspek, opini, dan polaritas sentimen — dalam satu kesatuan triplet relasional [5]. Berbeda dengan pendekatan ABSA konvensional yang memformulasikan setiap elemen sebagai tugas terpisah, ASTE dirancang untuk menangkap hubungan eksplisit antara ketiganya sekaligus, sehingga mengurangi propagasi kesalahan yang umum terjadi pada pendekatan pipeline bertahap.

Perkembangan arsitektur ASTE bergerak dari pipeline menuju model end-to-end yang lebih terpadu. Pendekatan berbasis sequence labeling dan span prediction terbukti meningkatkan akurasi ekstraksi pada kalimat multi-token dan multi-triplet [12], [13], sementara model generatif menunjukkan keunggulan dalam menangkap batas span yang lebih presisi [14], [15]. Integrasi informasi sintaktik melalui dependency parsing juga terbukti memperkuat struktur relasional antar entitas [16]. Meski demikian, seluruh perkembangan ini masih didominasi oleh evaluasi pada domain ulasan produk, sehingga penerapannya pada teks berita makroekonomi yang memiliki struktur linguistik lebih kompleks masih menjadi tantangan terbuka.

2.4. BERT dan Mekanisme Cross-Attention

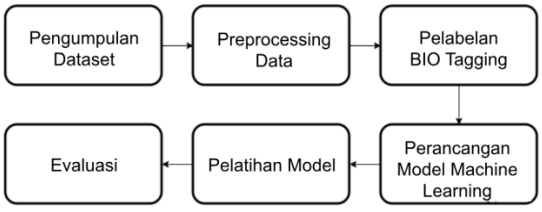
Devlin et al. [6] memperkenalkan BERT yang menghasilkan representasi kontekstual bidirectional pada level token melalui mekanisme masked language modeling. Vaswani et al. [17] menunjukkan bahwa

mekanisme self-attention pada arsitektur Transformer mampu memodelkan dependensi jarak jauh antar token. Meskipun BERT efektif dalam menangkap konteks global, arsitektur ini tidak secara eksplisit memodelkan hubungan relasional terstruktur antar token seperti relasi pasangan aspek–opini. Dozat dan Manning [18] menunjukkan efektivitas pendekatan biaffine attention dalam pemodelan relasi antar token, yang menjadi inspirasi bagi pengembangan mekanisme cross-attention pada penelitian ini.

3. METODE PENELITIAN

Penelitian ini bertujuan membangun model Aspect-Based Sentiment Analysis (ABSA) yang mampu mengekstraksi triplet aspek–opini–sentimen secara simultan dari teks berita makroekonomi, sebagai jawaban atas keterbatasan pendekatan ABSA yang ada dalam menangani kompleksitas linguistik domain finansial. Secara spesifik, model dirancang untuk menyelesaikan dua permasalahan utama: pertama, mengidentifikasi span aspek makroekonomi dan span opini secara akurat pada tingkat token dalam kalimat berita finansial; dan kedua, memetakan hubungan relasional antara kedua entitas tersebut sekaligus menentukan polaritas sentimennya dalam satu kerangka pembelajaran terpadu. Untuk mencapai tujuan tersebut, model dibangun dengan mengintegrasikan tiga komponen utama, yaitu encoder berbasis BERT sebagai backbone representasi kontekstual, skema BIO Tagging untuk ekstraksi entitas pada level token, dan mekanisme cross-attention untuk pemodelan relasi aspek–opini secara eksplisit. Seluruh komponen dilatih secara end-to-end menggunakan pendekatan multi-task learning, sehingga proses ekstraksi entitas, pemetaan relasi, dan klasifikasi sentimen dapat saling memperkuat dalam satu proses optimasi yang terintegrasi.

Pengembangan dan evaluasi model tersebut dilakukan menggunakan pendekatan kuantitatif komputasional, dengan fokus pada pengolahan data teks menggunakan algoritma machine learning dan pengukuran performa model melalui serangkaian metrik numerik. Pendekatan ini memungkinkan proses evaluasi yang objektif, terukur, dan dapat direproduksi.



Gambar 1. Metode Penelitian

3.1. Dataset

Dataset dalam penelitian ini dikumpulkan secara manual dengan mengakses portal berita ekonomi dan keuangan melalui website resmi masing-masing sumber seperti Investing.com, Bloomberg Economics, dan CNBC International. Pengumpulan data dilakukan menggunakan kata kunci yang berkaitan dengan indikator makroekonomi, seperti inflation, interest rate, GDP growth, economic outlook, monetary policy, bond yields, labor market.

Proses ekstraksi dilakukan pada level kalimat dengan memisahkan paragraf menjadi kalimat-kalimat individual. Selanjutnya, setiap kalimat hasil ekstraksi diseleksi secara manual. Kalimat dipilih apabila memenuhi dua kriteria, yaitu: (1) memuat indikator makroekonomi secara eksplisit, dan (2) mengandung pernyataan evaluatif yang menunjukkan penilaian atau perubahan terhadap indikator tersebut (mis. “rising”, “weakened”, “strong”, “slowed”). Kalimat yang hanya bersifat deskriptif tanpa evaluasi tidak dimasukkan.

Total kalimat yang berhasil dikumpulkan dalam penelitian ini sebanyak 3873 kalimat. Dataset kemudian dibagi ke dalam data pelatihan dan validasi dengan rasio 80:20. Dataset disimpan dalam format terstruktur yang terdiri atas teks kalimat dan daftar pasangan aspek–opini–sentimen. Apabila satu kalimat memiliki lebih dari satu aspek/opini, seluruh pasangan dicatat dalam entri kalimat yang sama.

Secara konseptual, setiap pasangan aspek–opini–sentimen direpresentasikan sebagai suatu triplet p = (a, o, s) dengan:

a. Aspect (a) : Entitas makroekonomi yang menjadi fokus informasi
b. Opinion (o) : Kata atau frasa yang menggambarkan kondisi, perubahan, atau evaluasi terhadap aspek tersebut
c. Sentiment (s) : Polaritas yang melekat pada opini (positive, negative, atau neutral)

3.2. Preprocessing Data

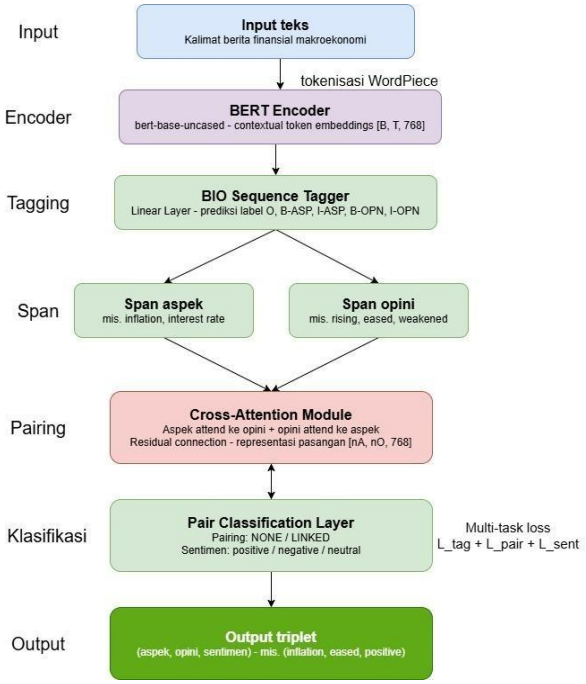
Data teks yang diperoleh melalui serangkaian tahap preprocessing meliputi case folding, cleaning, stopword removal, dan tokenisasi berbasis BERT untuk memastikan konsistensi representasi input dengan encoder transformer [6]. Selanjutnya, pelabelan token-level diterapkan menggunakan skema BIO yang terdiri atas tiga label: B (Begin) untuk token awal entitas, I (Inside) untuk token lanjutan, dan O (Outside) untuk token di luar entitas. Skema ini digunakan untuk menandai batas span aspek makroekonomi dan opini secara eksplisit pada level token [19]. Ketika satu token tersegmentasi menjadi beberapa subword oleh WordPiece tokenizer, label BIO dipetakan dengan menetapkan label awal pada subword pertama dan label lanjutan pada subword berikutnya, sementara token khusus seperti [CLS] dan [SEP] diabaikan dalam proses pelatihan.

3.3. Perancangan Model Machine Learning

Perancangan model ABSA dalam penelitian ini terdiri atas tiga komponen utama yang saling terintegrasi untuk mengekstraksi dan menganalisis hubungan aspek–opini dalam teks berita finansial.

Komponen pertama adalah encoder berbasis BERT (bert-base-uncased) yang berfungsi sebagai

backbone utama dalam menghasilkan representasi token kontekstual dari teks masukan. Representasi ini mampu menangkap dependensi semantik jangka panjang antar token dalam satu kalimat [6], [17]. Representasi token tersebut selanjutnya diproses oleh BIO sequence tagger — lapisan linear yang mengidentifikasi batas awal dan lanjutan entitas aspek maupun opini dalam kalimat, yang kemudian diekstraksi menjadi span melalui operasi mean pooling.



Gambar 2. Arsitektur Model

Komponen kedua adalah modul CrossAttentionPairing yang memodelkan hubungan dua arah antara span aspek dan opini secara eksplisit. Modul ini mengimplementasikan dua jalur atensi secara paralel — aspek-ke-opini dan opini-ke-aspek — sehingga model dapat menangkap keterkaitan semantik dua arah dan memilih pasangan yang paling relevan secara kontekstual [17], [18], [19]. Pendekatan cross-attention dipilih karena terbukti efektif dalam memodelkan ketergantungan antar entitas pada tugas ekstraksi relasi dibandingkan metode pipeline konvensional.

Komponen ketiga adalah SentimentHead, lapisan klasifikasi yang menerima representasi gabungan pasangan aspek–opini hasil cross-attention dan menghasilkan prediksi polaritas sentimen — positif, negatif, atau netral. Melalui komponen ini, model tidak hanya mengekstraksi aspek dan opini secara terpisah, tetapi juga menghasilkan penilaian sentimen yang spesifik terhadap setiap relasi aspek–opini yang terbentuk.

Model dilatih secara end-to-end menggunakan pendekatan multi-task learning untuk mempelajari ekstraksi aspek, pemetaan pasangan aspek–opini, dan klasifikasi sentimen secara simultan. Pendekatan ini memungkinkan seluruh komponen model berbagi representasi yang sama selama proses pembelajaran, sehingga informasi dari masing-masing sub-tugas dapat saling memperkuat. Pendekatan multi-task learning diketahui mampu meningkatkan kemampuan generalisasi model serta mengurangi propagasi kesalahan antar sub-tugas, sebagaimana dilaporkan dalam studi survei dan penerapan multi-task learning pada tugas pemrosesan bahasa alami sebelumnya [20], [21].

3.4. Pelatihan Model

Model dilatih secara end-to-end menggunakan strategi optimasi yang dirancang untuk meningkatkan stabilitas pembelajaran sekaligus meminimalkan risiko overfitting. Dataset dibagi menggunakan stratified split dengan rasio 70:15:15 untuk pelatihan, validasi, dan pengujian, memastikan distribusi polaritas sentimen tetap seimbang pada setiap subset.
Proses pelatihan menggunakan algoritma optimasi AdamW dengan skema differentiated learning rate, di mana laju pembelajaran backbone BERT dibuat lebih kecil dibandingkan komponen task-specific heads untuk menjaga stabilitas pembaruan parameter [22]. Learning rate scheduler berbasis cosine annealing dengan warm-up digunakan untuk membantu konvergensi model secara lebih stabil [23], sementara early stopping dengan patience tujuh epoch diterapkan untuk mencegah overfitting [24].
Seluruh proses pelatihan menggunakan formulasi multi-objective loss yang terdiri dari tiga komponen utama.

a. *Token-level tagging loss (*L_{tag}*)* mengukur prediksi label BIO pada ekstraksi aspek dan opini:
$$L_{tag} = -\sum_{t=1}^{T'} \log P(y_t \mid x_t) \quad (1)$$
b. *Pairing loss (*L_{pair}*)* mengukur prediksi hubungan antar-span aspek dan opini melalui mekanisme *cross-attention*:
$$L_{pair} = -\frac{1}{N}\sum_{i=1}^{N} \log P(r_i \mid s_i) \quad (2)$$
c. Sentiment classification loss (L_{sent}) mengukur prediksi polaritas sentimen pada setiap pasangan aspek–opini yang valid:
$$L_{sent} = -\frac{1}{M}\sum_{j=1}^{M} \log P(c_i \mid p_i) \quad (3)$$

Ketiga komponen tersebut digabungkan secara terintegrasi menggunakan formulasi (4):
$$L=\lambda_1 L_{tag}+\lambda_2 L_{pair}+\lambda_3 L_{sent} \quad (4)$$
dengan λ_1=1.0, λ_2=1.0, λ_3=0.1 sehingga kontribusi utama pembelajaran difokuskan pada komponen tagging dan pairing, sementara sentiment berperan sebagai sinyal supervisi tambahan.

3.5. Evaluasi

Tahap evaluasi dilakukan untuk menilai kinerja model ABSA yang dikembangkan dalam penelitian ini, yang terdiri atas tiga komponen utama, yaitu BIO *Tagging* untuk mengekstraksi aspek dan opini, *Cross-Attention Pairing* untuk menentukan hubungan aspek dan opini, serta *Sentiment Classification* untuk

menentukan polaritas sentimen setiap pasangan. Evaluasi disusun secara bertahap, dengan tujuan memastikan setiap sub-komponen berfungsi secara optimal sebelum dilakukan evaluasi sebagai satu sistem terpadu.

Tahap pertama merupakan evaluasi terhadap kemampuan model dalam mengekstraksi aspek dan opini. Kinerja pada tahap ini diukur melalui *token-level metrics* yang mencakup *precision*, *recall*, dan *F1-score* terhadap skema BIO (B-ASP, I-ASP, B-OPN, I-OPN, O). Pengujian ini bertujuan menilai ketepatan model dalam mengidentifikasi token terkait entitas makroekonomi seperti inflasi, suku bunga, pertumbuhan ekonomi, kebijakan moneter, maupun dinamika pasar. Selain itu, dilakukan pengujian pada tingkat *span-level*, di mana aspek dan opini harus mampu dikenali sebagai satu rentang yang utuh. Metrik yang digunakan pada tahap evaluasi ini adalah *span-level* F1, untuk memastikan bahwa entitas makroekonomi dalam teks berita finansial yang sering muncul dalam frasa panjang, seperti "moderating inflation pressure" atau "Federal Reserve rate decision", dapat dikenali secara utuh dan akurat.

Pada tahap kedua, performa model *Cross-Attention Pairing* dievaluasi menggunakan *pair-level* F1. Metrik tersebut mengukur kemampuan model dalam memetakan pasangan aspek-opini secara akurat, terutama pada kalimat yang mengandung beberapa isu ekonomi yang saling tumpang tindih. Mengingat struktur kalimat dalam berita finansial cenderung kompleks dan memuat relasi semantik yang tersebar di seluruh kalimat, evaluasi pasangan menjadi krusial untuk memastikan kemampuan model dalam memetakan hubungan antar entitas secara tepat.

Tahap ketiga adalah evaluasi klasifikasi polaritas sentimen. Pada tahap ini, digunakan *accuracy* dan *macro*-F1, dimana *macro*-F1 dipilih untuk mengatasi ketidakseimbangan distribusi polaritas yang umum terjadi pada berita ekonomi, khususnya dominasi sentimen netral. Evaluasi ini mengukur kemampuan model dalam membedakan polaritas positif, negatif, dan netral terhadap isu makro seperti inflasi, suku bunga, dan *outlook* ekonomi.

Pada tahap terakhir, dilakukan pengujian komprehensif melalui *Triplet Extraction Metrics*, yaitu *triplet F1* untuk struktur lengkap *(aspect, opinion, sentiment)*. Metrik tersebut memberikan gambaran menyeluruh mengenai kemampuan model dalam mengekstraksi informasi makroekonomi secara terintegrasi, mulai dari identifikasi entitas hingga interpretasi sentimen dalam satu kesatuan analisis. Secara keseluruhan, rangkaian evaluasi ini memberikan landasan sistematis dan menyeluruh untuk menilai efektivitas model ABSA dalam konteks analisis berita finansial, sekaligus memastikan kemampuan generalisasi model terhadap variasi struktur bahasa ekonomi modern.

4. HASIL DAN PEMBAHASAN

Mengacu pada tahapan penelitian yang telah diuraikan pada bab sebelumnya, bagian ini memaparkan secara sistematis hasil implementasi dan evaluasi model, serta pembahasan komprehensif terhadap temuan pada setiap tahap penelitian.

4.1. Deskripsi Dataset dan Hasil Preprocessing

Dataset yang digunakan dalam penelitian ini terdiri atas 3.098 berita finansial internasional yang diperoleh melalui *Investing.com*, *Bloomberg Economics*, dan *CNBC International.* Seluruh dokumen kemudian diproses melalui tahap *sentence splitting* untuk mengekstraksi setiap berita menjadi unit kalimat tunggal, dengan melalui seleksi berdasarkan relevansi terhadap isu makroekonomi seperti inflasi, kebijakan moneter, pertumbuhan ekonomi, dan dinamika pasar keuangan.

Proses *preprocessing* dilakukan melalui beberapa tahapan, yaitu *case folding*, pembersihan teks, *stopword removal*, tokenisasi berbasis BERT, serta pelabelan menggunakan skema BIO Tagging untuk menandai posisi aspek dan opini pada tingkat token. Tabel 1 menampilkan contoh hasil *preprocessing* dan konversi kalimat ke format BIO.

Tabel 1. Contoh Hasil *Preprocessing* dan BIO Tagging

Tahap	Hasil
Kalimat Asli	"The Federal Reserve is expected to maintain a restrictive policy as inflation moderates."
Kalimat Bersih	"Federal reserve expected maintain restrictive policy inflation moderates"
BIO Tagging	B-ASP – I-ASP – O – O – B-OPN – O – B-ASP – B-OPN

Berdasarkan contoh pada Tabel 1, "Federal Reserve" dikenali sebagai aspek makroekonomi pada kategori *monetary authority*, "restrictive" ditandai sebagai opini, dan "inflation" sebagai aspek kedua.

Distribusi sentimen terdiri atas 39% positif, 14.5% netral dan 46.5% negatif. Distribusi ini relatif seimbang dan mencerminkan karakteristik pemberitaan ekonomi global yang umumnya memuat sentimen campuran.

4.2. Hasil Pelatihan BIO Tagging

Tahap pertama evaluasi difokuskan pada pengukuran kinerja model dalam melakukan pelabelan token menggunakan skema BIO. Hasil pelatihan menunjukkan bahwa model berbasis BERT mampu mempelajari struktur linguistik berita finansial dengan baik, ditandai oleh peningkatan performa yang stabil dari awal hingga epoch akhir. Rentang performa pada epoch 41–45 menunjukkan bahwa model telah mencapai kondisi pembelajaran yang matang, di mana nilai F1 untuk keempat label BIO berada pada kisaran yang relatif tinggi dan konsisten. Tabel 2 menyajikan ringkasan performa token-level.

Tabel 2. Token-Level Performance

Label	Precision	Recall	F1 Score
B-ASP	0.766 – 0.773	0.812 – 0.823	0.78 – 0.79
I-ASP	0.839 – 0.846	0.860 – 0.868	0.85
B-OPN	0.816 – 0.833	0.833 – 0.849	0.82 – 0.83
I-OPN	0.79 – 0.81	0.74 – 0.76	0.78

Label B-ASP pada model mencapai nilai precision sebesar 0.766–0.773 dan recall 0.812–0.823, menghasilkan skor F1 pada kisaran 0.78–0.79. Pola asimetri antara precision dan recall pada label ini mengindikasikan bahwa model cenderung lebih mudah mendeteksi keberadaan token awal aspek daripada memastikan bahwa token yang terdeteksi memang merupakan aspek yang valid secara kontekstual. Kondisi ini dapat dipahami mengingat domain makroekonomi memiliki banyak frasa yang secara permukaan menyerupai aspek namun tidak bersifat evaluatif, seperti "the economy" atau "market participants", sehingga model sesekali melakukan over-detection pada frasa-frasa tersebut.

Pada label I-ASP, performa model menunjukkan hasil yang lebih kuat, dengan precision 0.839–0.846, recall 0.860–0.868, dan skor F1 yang konsisten di sekitar 0.85. Keseimbangan antara precision dan recall pada label ini menunjukkan bahwa model memiliki kemampuan yang stabil dalam mempertahankan span aspek secara konsisten, terutama ketika aspek terdiri atas lebih dari satu token. Hal ini penting dalam konteks berita makroekonomi, di mana aspek ekonomi umumnya berbentuk frasa multi-token seperti "interest rate policy" atau "monetary tightening cycle".

Label B-OPN memperoleh precision pada kisaran 0.816–0.833 dan recall 0.833–0.849, dengan skor F1 berada pada rentang 0.82–0.83. Hasil ini menunjukkan bahwa model mampu mengidentifikasi token awal opini secara akurat, meskipun variasi ekspresi opini yang bersifat adjektival dalam teks ekonomi masih menimbulkan sejumlah kesalahan deteksi. Dalam konteks praktis, kemampuan mendeteksi token awal opini secara tepat sangat krusial karena menentukan apakah ekspresi evaluatif seperti "eased", "tightened", atau "accelerated" berhasil ditangkap sebagai sinyal sentimen yang relevan.

Adapun label I-OPN merupakan kategori yang paling menantang, dengan precision sebesar 0.79–0.81 dan recall yang lebih rendah yaitu 0.74–0.76, sehingga menghasilkan skor F1 sekitar 0.78. Rendahnya nilai recall pada label ini mengindikasikan bahwa model masih mengalami kesulitan dalam mempertahankan konsistensi opinion span ketika opini memiliki bentuk multi-token yang kompleks, seperti "slowing economic pressure" atau "moderating inflation trend". Dari sudut pandang aplikasi, kelemahan ini berarti bahwa ekspresi opini panjang berpotensi terpotong hanya pada token awalnya — sehingga nuansa makna yang terkandung dalam frasa opini yang lebih panjang berisiko tidak tertangkap sepenuhnya oleh sistem.

Secara keseluruhan, konsistensi skor F1 pada rentang menengah hingga tinggi menunjukkan bahwa model berhasil mempelajari struktur aspek dan opini dalam teks finansial. Performa terbaik dicapai pada label I-ASP, sedangkan performa terendah terdapat pada label I-OPN. Pola ini konsisten dengan karakteristik linguistik berita finansial yang kaya akan frasa adjektival dan konstruksi klausa kompleks, di mana opini multi-token jauh lebih umum dibandingkan domain ulasan produk yang menjadi basis pengembangan sebagian besar model ABSA sebelumnya.

4.3. Evaluasi Pairing Aspect–Opinion

Tahap kedua dari model ABSA adalah pemetaan pasangan aspek–opini melalui modul cross-attention. Performa pada tahap pairing menunjukkan peningkatan yang konsisten sepanjang proses pelatihan. Pada epoch awal, nilai F1 berada di kisaran 0.82 dan meningkat secara stabil hingga mencapai 0.94–0.95 pada epoch 31–45. Tabel 3 menunjukkan evaluasi komponen pairing yang dilakukan menggunakan tiga metrik utama, yaitu precision, recall, dan F1-Score.

Tabel 3. Pair-Level Performance

Metrik	Nilai
Precision	0.91 – 0.92
Recall	0.97 – 0.98
F1-Score	0.94 – 0.95

Rentang nilai precision dalam 0.91–0.92 dan recall yang sangat tinggi pada 0.97–0.98 menghasilkan pola yang informatif secara interpretatif. Recall yang mendekati sempurna mengindikasikan bahwa model hampir tidak pernah melewatkan pasangan aspek–opini yang seharusnya terhubung — dari seluruh relasi valid yang ada dalam kalimat, hampir semuanya berhasil diidentifikasi. Di sisi lain, gap antara recall dan precision mengindikasikan adanya sejumlah kecil false positive, yaitu pasangan yang diprediksi terhubung padahal secara ground-truth tidak memiliki relasi. Dalam konteks aplikasi nyata seperti pemantauan sentimen kebijakan moneter, karakteristik ini lebih dapat diterima dibandingkan sebaliknya — sistem yang sesekali melaporkan relasi yang tidak valid jauh lebih berguna daripada sistem yang secara konsisten melewatkan relasi yang penting.

Berdasarkan pair-level performance yang dilakukan, diperoleh kinerja kuat yang mencerminkan efektivitas mekanisme cross-attention dalam menangkap hubungan semantik antara aspek ekonomi dan opini yang terkait. Struktur kalimat dalam berita finansial umumnya memiliki pola logis dan kausal yang eksplisit, misalnya "Fed rate cuts lifted market optimism" atau "slowing inflation eased pressure on policymakers", sehingga memberikan sinyal kontekstual yang membantu model dalam menentukan pasangan aspek–opini secara akurat. Capaian F1=0.94–0.95 ini juga melampaui kisaran performa

umum model ABSA pada domain non-finansial yang umumnya berada pada kisaran 0.80–0.88, yang mengonfirmasi bahwa integrasi mekanisme cross-attention dua arah efektif meningkatkan akurasi pemetaan relasional secara signifikan dibandingkan pendekatan sekuensial konvensional.

Meskipun demikian, sejumlah kesalahan pairing masih ditemukan, terutama pada kalimat yang memuat lebih dari satu aspek, seperti kombinasi inflation, employment data, dan interest rate projection dalam satu kalimat. Pada kondisi tersebut, model sesekali menghubungkan opini dengan aspek yang secara posisi berdekatan namun tidak relevan secara semantik. Kesalahan jenis ini berdampak langsung pada kualitas triplet yang dihasilkan — sebuah pasangan yang salah terhubung akan menghasilkan triplet yang keliru meskipun aspek dan opininya masing-masing terdeteksi dengan benar. Oleh karena itu, peningkatan kemampuan disambiguasi pada kalimat dengan banyak aspek menjadi arah pengembangan yang relevan untuk versi model selanjutnya.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa modul cross-attention mampu mengurangi ambiguitas hubungan antar entitas dan memberikan performa yang lebih baik dibandingkan pendekatan rule-based maupun model sekuensial tradisional. Dengan demikian, komponen ini berkontribusi signifikan terhadap peningkatan kualitas ekstraksi struktur ABSA secara end-to-end.

4.4. Evaluasi Klasifikasi Sentimen

Tahap selanjutnya dalam model ABSA adalah klasifikasi sentimen, yaitu proses menentukan polaritas sentimen positif, netral, atau negatif yang terkait dengan setiap pasangan aspek–opini yang telah teridentifikasi. Evaluasi pada tahap ini dilakukan dengan menilai akurasi dan Macro-F1 untuk setiap pasangan yang telah berhasil dipetakan oleh model. Performa pelatihan menunjukkan peningkatan paling stabil dibandingkan dua komponen evaluasi sebelumnya.

Pada epoch awal, akurasi performa model masih rendah berkisar di angka 0.46 dan Macro-F1 sebesar 0.25, yang menandakan bahwa model belum mampu mengenali pola sentimen secara memadai. Namun, terdapat peningkatan secara signifikan pada performa model ketika memasuki rentang epoch 31–45. Tabel 4 menampilkan peningkatan performa klasifikasi sentimen.

Tabel 4. Sentiment Classification Performance

Metrik	Nilai
Accuracy	0.86 – 0.89
Macro-F1	0.80 – 0.85

Accuracy sebesar 0.86–0.89 menunjukkan bahwa dari seluruh pasangan aspek–opini yang diklasifikasikan, sekitar 86–89 dari setiap 100 pasangan berhasil ditentukan polaritas sentimennya dengan benar. Namun, interpretasi accuracy saja tidak cukup untuk menggambarkan kualitas model secara menyeluruh. Gap antara accuracy (0.86–0.89) dan Macro-F1 (0.80–0.85) mengungkap adanya ketidakseimbangan performa antar kelas sentimen — accuracy yang tinggi sebagian disumbang oleh dominasi kelas negatif (46.5% dari dataset) dan positif (39%), sehingga model yang kuat pada dua kelas mayoritas ini secara otomatis menghasilkan accuracy keseluruhan yang tinggi meskipun performanya pada kelas netral lebih lemah.

Kelemahan pada kelas netral (14.5% dari dataset) dapat dipahami dari karakteristik linguistik teks ekonomi. Sentimen netral dalam berita finansial tidak disampaikan melalui ekspresi emosional yang eksplisit, melainkan melalui pernyataan faktual seperti "the central bank held rates steady" atau "GDP growth remained at 2.1 percent" — kalimat yang secara permukaan sulit dibedakan dari sentimen negatif ringan oleh model. Frasa-frasa ambigu seperti "cautious optimism", "moderating inflation", atau "mixed economic data" memperparah kesulitan ini karena tidak secara tegas menunjukkan kecenderungan ke salah satu polaritas.

Dari perspektif praktis, nilai Macro-F1 sebesar 0.80–0.85 mengindikasikan bahwa model ini cukup andal untuk digunakan sebagai lapisan pertama dalam sistem pemantauan sentimen otomatis. Bagi analis risiko di lembaga keuangan, performa ini berarti bahwa dari setiap 100 pasangan aspek–opini yang diklasifikasikan, sekitar 80–85 di antaranya akan menghasilkan label polaritas yang dapat dipercaya tanpa perlu verifikasi manual. Kelemahannya yang terletak pada kelas netral dapat diatasi secara operasional dengan menetapkan threshold kepercayaan — prediksi dengan skor keyakinan rendah pada kelas netral dapat secara otomatis diarahkan ke analis manusia untuk dikonfirmasi, sementara prediksi positif dan negatif dengan skor tinggi dapat langsung diproses oleh sistem.

Tren peningkatan performa pada seluruh tahap pelatihan menunjukkan bahwa model mampu menggeneralisasi pola sentimen makroekonomi secara efektif, terutama setelah memasuki fase pembelajaran menengah hingga akhir. Temuan ini mengonfirmasi bahwa representasi kontekstual yang dihasilkan oleh arsitektur model mampu menangkap dinamika sentimen yang kompleks dalam teks berita finansial, meskipun masih terdapat ruang pengembangan pada pemodelan ekspresi sentimen yang bersifat implisit dan ambigu.

4.5. Evaluasi Triplet (Aspect - Opinion - Sentiment)

Tahap akhir evaluasi dilakukan dengan mengukur kemampuan model dalam mengekstraksi triplet lengkap yang mencakup aspek, opini, dan polaritas sentimen. Triplet tersebut merupakan representasi informasi paling komprehensif dalam ABSA, karena adanya penggabungan antara tiga

komponen utama secara konsisten yaitu deteksi aspek, identifikasi opini, penentuan keterhubungan, dan klasifikasi sentimen. Evaluasi triplet yang dilakukan menggunakan metrik F1 Triplet hanya memberikan skor positif apabila ketiga elemen yaitu span aspek, span opini, dan kelas sentimen, berhasil diprediksi secara tepat dalam satu unit struktur. Dengan demikian, ketidaktepatan pada salah satu tahap — seperti aspek dan opini benar namun polaritas sentimen salah — akan tetap menghasilkan kegagalan pada level triplet.

Hasil evaluasi pada tingkat triplet menunjukkan bahwa performa awal model masih rendah dengan nilai F1 sekitar 0.48, namun mengalami peningkatan signifikan pada epoch 41–45 hingga mencapai nilai puncak 0.87–0.88. Peningkatan drastis dari 0.48 ke 0.87–0.88 ini mencerminkan bahwa model membutuhkan waktu pelatihan yang cukup untuk mempelajari konsistensi antar tiga komponen secara simultan — karakteristik yang umum ditemukan pada arsitektur multi-task dengan propagasi kesalahan berlapis, di mana setiap komponen harus terlebih dahulu stabil sebelum ketiganya dapat bekerja secara koheren dalam satu struktur triplet.

Untuk memberikan gambaran konkret, nilai triplet F1 sebesar 0.87–0.88 dapat diinterpretasikan sebagai berikut: dari 100 triplet aspek–opini–sentimen yang seharusnya diekstraksi dari kumpulan kalimat berita finansial, model berhasil mengekstraksi sekitar 87–88 triplet dengan ketiga elemennya benar secara simultan. Capaian ini tergolong tinggi mengingat triplet F1 adalah metrik paling ketat dalam kerangka ABSA — satu kesalahan kecil pada batas span atau satu kesalahan polaritas saja sudah cukup untuk menggugurkan seluruh triplet, meskipun dua elemen lainnya benar.

Meskipun demikian, capaian pada level triplet tetap lebih rendah dibandingkan metrik pada tingkat token, span, maupun pair. Pola ini umum ditemukan pada model ABSA multikomponen dan disebabkan oleh dua sumber utama propagasi kesalahan: pertama, kesalahan sentimen pada pasangan aspek–opini yang sebenarnya sudah benar; dan kedua, ketidaktepatan kecil pada batas span yang mengganggu konsistensi struktur triplet secara keseluruhan. Kesalahan paling sering muncul ketika opini disampaikan secara implisit atau ketika satu aspek memiliki beberapa opini dengan polaritas berbeda. Sebagai contoh, pada kalimat "Inflation eased but remains above the Fed's target", model seringkali hanya mengekstraksi satu opini atau salah menentukan polaritas dominan — karena kalimat tersebut secara bersamaan mengandung sinyal positif (eased) dan negatif (remains above target) terhadap aspek yang sama.

Dalam konteks implementasi nyata, triplet F1 sebesar 0.87–0.88 menunjukkan bahwa model ini sudah siap digunakan sebagai alat bantu primer, bukan sekadar alat eksperimental. Bagi bank sentral, sistem berbasis model ini dapat dimanfaatkan untuk memantau pergeseran sentimen pasar terhadap kebijakan moneter secara real-time dari ribuan artikel berita harian — sebuah tugas yang secara manual membutuhkan puluhan analis. Bagi manajer investasi, triplet yang dihasilkan seperti (inflation, eased, positive) atau (labor market, weakened, negative) dapat langsung diintegrasikan ke dalam dashboard risiko untuk memberikan sinyal awal sebelum dampak kebijakan tercermin dalam harga aset. Dengan tingkat keberhasilan 87–88%, model ini cukup untuk menjadi filter pertama yang menyisihkan volume pemberitaan yang tidak relevan, sehingga tenaga analis manusia dapat difokuskan hanya pada subset kalimat yang mengandung sinyal makroekonomi bermakna.

Secara keseluruhan, hasil ini menunjukkan bahwa keberhasilan struktur triplet sangat bergantung pada akurasi klasifikasi sentimen dan konsistensi batas span. Meski demikian, model terbukti mampu mengekstraksi struktur semantik makroekonomi yang beragam, mulai dari sinyal kebijakan seperti (monetary policy, tightened, negative) hingga indikator pertumbuhan seperti (GDP growth, accelerated, positive), sehingga memiliki potensi kuat untuk mendukung analisis ekonomi terstruktur dan pemantauan dinamika kebijakan secara otomatis dalam skala besar.

4.6. Ablation Study

Ablation study dilakukan untuk menganalisis kontribusi mekanisme pairing dan supervisi relasional pada model yang diusulkan. Evaluasi difokuskan pada tiga konfigurasi utama, yaitu model penuh (A0), model tanpa cross-attention yang digantikan dengan concatenation pooling (A1), dan model tanpa pairing loss $\square_{pair}$ (A2). Seluruh varian dilatih dan dievaluasi menggunakan konfigurasi, loss function, serta metrik yang identik, sehingga perbedaan performa yang diamati secara langsung merefleksikan dampak penghilangan komponen tertentu.

Tabel 5 menyajikan hasil ablation pada metrik yang paling relevan dengan tujuan ekstraksi relasi dan triplet. Penggantian cross-attention dengan concatenation pooling (A1) menyebabkan penurunan pada Pair F1 dan Triplet performance, yang menunjukkan bahwa mekanisme cross-attention berperan penting dalam memodelkan interaksi dua arah antara aspek dan opini. Sementara itu, penghilangan pairing loss (A2) menunjukkan penurunan yang sangat signifikan pada metrik Pair F1 dan Triplet performance, sementara performa klasifikasi sentimen relatif tetap tinggi. Fenomena ini mengindikasikan bahwa tanpa supervisi eksplisit pada tahap pairing, model masih mampu mempelajari pola polaritas sentimen secara lokal berdasarkan representasi aspek dan opini, namun gagal membangun asosiasi yang benar antara keduanya. Dengan kata lain, classifier sentimen dapat menghasilkan prediksi polaritas yang akurat secara individual, tetapi prediksi tersebut tidak lagi terikat secara konsisten pada pasangan aspek–opini yang tepat. Kondisi ini menyebabkan degradasi tajam pada

kualitas ekstraksi triplet secara end-to-end. Temuan ini menegaskan bahwa pairing loss memiliki peran krusial dalam memaksa model untuk mempelajari relasi aspek–opini yang koheren, serta mencegah model mengandalkan sinyal sentimen parsial yang tidak terstruktur dalam konteks ekstraksi triplet.

Tabel 5. Performance Sentiment Classification

Model	Pair F1	Sentiment F1	Sentiment Accuracy	Triplet F1
(A_0) Full Model	0.9480	0.8354	0.8851	0.8789
(A_1) Concatenation Pooling	0.9176	0.8412	0.8876	0.8725
(A_2) Tidak Menggunakan Pairing Loss	0.3680	0.8539	0.8988	0.3404

4.7. Pembahasan Umum

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model ABSA yang dikembangkan mampu mencapai performa yang stabil dan kompetitif pada seluruh komponen pemrosesan, mulai dari BIO Tagging hingga Triplet Extraction. Peningkatan performa paling signifikan terjadi setelah memasuki epoch 30, ketika seluruh metrik utama menunjukkan pola konvergensi yang konsisten. Stabilitas ini mengindikasikan bahwa arsitektur berbasis BERT yang dipadukan dengan mekanisme *cross-attention* mampu mempelajari struktur linguistik dalam berita finansial secara bertahap dan efektif.

Komponen *pairing* aspek-opini menunjukkan performa paling kuat, dengan nilai F1 yang mencapai 0,95 pada epoch optimal. Model terbukti memiliki kemampuan yang sangat baik dalam memetakan hubungan semantik antar aspek ekonomi dan opini terkait. Capaian ini melampaui kisaran performa umum pada penelitian ABSA klasik yang berada pada kisaran 0.80 – 0.88, sehingga menegaskan bahwa integrasi *cross-attention* mampu meningkatkan akurasi pemetaan relasional secara signifikan.

Tantangan terbesar ditemukan pada label I-OPN dan pada sentimen netral, yang secara konsisten menunjukkan hasil lebih rendah dibandingkan komponen lainnya. Hal ini sejalan dengan karakteristik teks ekonomi yang cenderung padat informasi, minim ekspresi emosional eksplisit, dan sering menyampaikan opini secara implisit. Dominasi frasa deskriptif seperti moderating *inflation*, *mixed economic data*, atau *persistent headwinds* menyulitkan model dalam menentukan batas opini dan polaritas sentimen secara presisi.

Performa pada Triplet Extraction yang melibatkan integrasi aspek, opini, dan sentimen secara simultan, tergolong sangat tinggi untuk konteks berita finansial. Nilai F1 pada rentang 0.87–0.88 menunjukkan bahwa meskipun terdapat propagasi kesalahan pada komponen sebelumnya, model mampu menangkap struktur semantik kompleks dalam berita makroekonomi secara konsisten. Capaian ini mengindikasikan bahwa model tidak sekadar memahami token dan hubungan antar frasa, tetapi juga mampu menginterpretasikan arah evaluatif dari konteks ekonomi secara menyeluruh.

Keseluruhan hasil memperlihatkan bahwa model yang dikembangkan sangat sesuai untuk domain makroekonomi, terutama karena kemampuannya dalam mengenali dan memproses pola frasa khas berita ekonomi seperti *tightening cycle*, *easing inflation*, atau *robust demand*. Kemampuan tersebut penting karena sentimen dalam teks ekonomi lebih bergantung pada interpretasi perubahan indikator fundamental daripada penggunaan kata sifat emosional secara langsung.

Secara umum, hasil penelitian menegaskan bahwa kombinasi BERT, BIO Tagging, dan mekanisme *cross-attention* pada tahap pairing memberikan fondasi yang kuat untuk ekstraksi aspek–opini–sentimen dalam teks finansial. Kinerja yang dicapai menunjukkan potensi besar penerapan model ini dalam analisis sentimen makroekonomi, pemantauan perkembangan kebijakan, dan sistem peringatan dini berbasis teks yang memanfaatkan dinamika pemberitaan ekonomi global.

5. KESIMPULAN DAN SARAN

Secara keseluruhan, penelitian ini menyimpulkan bahwa kombinasi encoder BERT, BIO Tagging, dan cross-attention pairing dalam kerangka multi-tugas merupakan pendekatan yang efektif untuk analisis sentimen makroekonomi pada berita finansial. Model yang dikembangkan mampu mengekstraksi aspek, opini, dan sentimen secara terstruktur dengan tingkat ketelitian yang tinggi, sehingga berpotensi dimanfaatkan untuk berbagai aplikasi lanjutan, seperti pemantauan sentimen kebijakan moneter, analisis risiko makroekonomi, maupun sistem peringatan dini berbasis teks. Ke depan, pengembangan lebih lanjut dapat diarahkan pada peningkatan pemodelan opini implisit, penanganan multi-sentimen dalam satu aspek, integrasi pengetahuan domain makroekonomi yang lebih eksplisit, serta perluasan dataset ke sumber berita lain agar kemampuan generalisasi model terhadap variasi gaya bahasa dan rezim ekonomi dapat semakin diperkuat.

DAFTAR PUSTAKA

[1] A. Case, Justin and Clements, “The Impact of Sentiment in the News Media on Daily and Monthly Stock Market Returns,” pp. 180–195, 2021.

[2] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, “A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges,” *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 11, pp. 11019–11038, 2023, doi: 10.1109/TKDE.2022.3230975.

[3] S. Consoli, L. Barbaglia, and S. Manzan, “Fine-grained, aspect-based semantic sentiment analysis within the economic and financial

domains," *Proc. - 2020 IEEE 2nd Int. Conf. Cogn. Mach. Intell. CogMI 2020*, pp. 52–61, 2020, doi: 10.1109/CogMI50398.2020.00017.

[4] M. Maia, S. Handschuh, B. Davis, R. Mcdermott, and A. Balahur, "WWW ' 18 Open Challenge : Financial Opinion Mining and Question Answering," vol. 1.

[5] H. Peng, L. Xu, L. Bing, F. Huang, W. Lu, and L. Si, "Knowing what, how and why: A near complete solution for aspect-based sentiment analysis," *AAAI 2020 - 34th AAAI Conf. Artif. Intell.*, pp. 8600–8607, 2020, doi: 10.1609/aaai.v34i05.6383.

[6] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. Mlm, pp. 4171–4186, 2019.

[7] D. de França Costa and N. F. F. da Silva, "INF-UFG at FiQA 2018 Task 1: Predicting Sentiments and Aspects on Financial Tweets and News Headlines," *Web Conf. 2018 - Companion World Wide Web Conf. WWW 2018*, vol. 2018-January, pp. 1967–1971, 2018, doi: 10.1145/3184558.3191828.

[8] K. Aziz, D. Ji, P. Chakrabarti, T. Chakrabarti, M. S. Iqbal, and R. Abbasi, "Unifying aspect - based sentiment analysis BERT and multi - layered graph convolutional networks for comprehensive sentiment dissection," *Sci. Rep.*, pp. 1–22, 2024, doi: 10.1038/s41598-024-61886-7.

[9] Y. Wang, "Aspect-based Sentiment Analysis in Document - FOMC Meeting Minutes on Economic Projection".

[10] E. Cambria, "Affective Computing and Sentiment Analysis," *IEEE Intell. Syst.*, vol. 31, no. 2, pp. 102–107, 2016, doi: 10.1109/MIS.2016.31.

[11] S. Yang, J. Rosenfeld, and J. Makutonin, "Financial Aspect-Based Sentiment Analysis using Deep Representations," 2018, [Online]. Available: http://arxiv.org/abs/1808.07931

[12] I. Naglik and M. Lango, "ASTE-Transformer: Modelling Dependencies in Aspect-Sentiment Triplet Extraction," *EMNLP 2024 - 2024 Conf. Empir. Methods Nat. Lang. Process. Find. EMNLP 2024*, pp. 2324–2339, 2024, doi: 10.18653/v1/2024.findings-emnlp.129.

[13] F. Yang, M. Zhang, G. Hu, and X. Zhou, "A Pairing Enhancement Approach for Aspect Sentiment Triplet Extraction," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 14119 LNAI, pp. 399–410, 2023, doi: 10.1007/978-3-031-40289-0_32.

[14] S. Zhou and T. Qian, "On the Strength of Sequence Labeling and Generative Models for Aspect Sentiment Triplet Extraction," pp. 12038–12050, 2023.

[15] Y. Chen and K. Chen, "A Span-level Bidirectional Network for Aspect Sentiment Triplet Extraction," pp. 4300–4309, 2022.

[16] J. Shang, Y. Zhang, L. Zhong, and R. Li, "Syntactic-Enhanced Multi-Task Learning Model for Aspect Sentiment Triplet Extraction," *Data Sci. Eng.*, vol. 10, no. 3, pp. 515–531, 2025, doi: 10.1007/s41019-025-00289-8.

[17] A. Vaswani *et al.*, "Attention Is All You Need," *Int. Conf. Inf. Knowl. Manag. Proc.*, no. Nips, pp. 4752–4758, 2023, doi: 10.1145/3583780.3615497.

[18] T. Dozat and C. D. Manning, "Deep biaffine attention for neural dependency parsing," *5th Int. Conf. Learn. Represent. ICLR 2017 - Conf. Track Proc.*, no. 2014, pp. 1–8, 2017.

[19] M. P. Marcus, "Text Chunking using Transformation-Based Learning 1 Introduction," *third Work. Very Large Corpora*, pp. 82–94, 1995.

[20] S. Ruder, "An Overview of Multi-Task Learning in Deep Neural Networks," no. May, 2017, [Online]. Available: http://arxiv.org/abs/1706.05098

[21] X. Liu, P. He, W. Chen, and J. Gao, "Multi-task deep neural networks for natural language understanding," *ACL 2019 - 57th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf.*, pp. 4487–4496, 2020, doi: 10.18653/v1/p19-1441.

[22] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *7th Int. Conf. Learn. Represent. ICLR 2019*, 2019.

[23] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," *5th Int. Conf. Learn. Represent. ICLR 2017 - Conf. Track Proc.*, pp. 1–16, 2017.

[24] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7700 LECTURE NO, pp. 437–478, 2012, doi: 10.1007/978-3-642-35289-8_26.