

BAB I

PENDAHULUAN

1.1 Latar Belakang

Transformasi sistem keuangan global dalam dua dekade terakhir ditandai oleh integrasi teknologi digital yang semakin masif, mencakup sistem pembayaran elektronik, *mobile banking*, *financial technology*, hingga *digital asset*. Digitalisasi ini mendorong efisiensi transaksi, peningkatan inklusi keuangan, serta lahirnya instrumen investasi berbasis teknologi [1]. Salah satu inovasi paling signifikan dalam lanskap keuangan digital adalah *cryptocurrency*, yaitu aset kripto berbasis teknologi *blockchain* yang memungkinkan pencatatan transaksi secara terdesentralisasi, transparan, dan aman tanpa bergantung pada otoritas pusat [2]. Sejak diperkenalkannya Bitcoin pada 2008, ekosistem *cryptocurrency* berkembang pesat dan mencakup ribuan aset digital dengan berbagai fungsi, mulai dari *medium of exchange* hingga instrumen investasi spekulatif [3]. Secara global, kapitalisasi pasar *cryptocurrency* telah mencapai lebih dari USD 2,5 triliun pada 2023 dan menunjukkan pertumbuhan signifikan sepanjang 2024, dengan peningkatan kapitalisasi kuartalan lebih dari 60% pada periode awal tahun akibat peningkatan partisipasi investor institusional dan sentimen pasar yang positif [4]. Pertumbuhan ini juga ditopang oleh peningkatan adopsi produk investasi berbasis kripto di pasar keuangan formal [5]. Skala transaksi yang besar, basis investor yang luas, serta volatilitas harga yang tinggi menjadikan *cryptocurrency* sebagai fenomena finansial yang sangat dipengaruhi oleh dinamika informasi dan persepsi publik [6]. Dalam konteks pasar yang berbasis ekspektasi dan spekulasi tersebut, pembentukan opini publik dan sentimen kolektif menjadi faktor yang berperan penting dalam memengaruhi perilaku investasi dan fluktuasi harga aset kripto.

Seiring dengan karakteristik pasar *cryptocurrency* yang sangat sensitif terhadap ekspektasi dan persepsi investor, media sosial berkembang menjadi ruang utama pembentukan opini dan penyebaran informasi terkait aset kripto [7]. Platform seperti *Twitter (X)*, *Reddit*, *Telegram*, dan berbagai forum daring memungkinkan pertukaran informasi berlangsung secara *real-time* tanpa mekanisme kurasi formal

sebagaimana media konvensional [8]. Tingginya aktivitas diskusi mengenai cryptocurrency di platform Twitter tercermin dari penelitian yang berhasil mengumpulkan dan menganalisis 53.580.759 tweet terkait cryptocurrency selama periode 2023–2024 sebagai sumber data dalam pengukuran efisiensi pasar aset kripto [9]. Intensitas percakapan yang tinggi mengenai pergerakan harga, kebijakan regulasi, peluncuran proyek baru, maupun rumor pasar menciptakan arus sentimen kolektif yang dapat dengan cepat memengaruhi keputusan jual-beli [10]. Lonjakan sentimen negatif sering dikaitkan dengan peningkatan tekanan jual dan volatilitas yang lebih tajam, sementara sentimen positif dan euforia komunitas kerap mendorong kenaikan harga dalam jangka pendek [11]. Selain itu, fenomena seperti *fear of missing out* (FOMO) dan *fear, uncertainty, and doubt* (FUD) terbentuk melalui narasi yang viral dan diperkuat oleh efek jejaring (*network effect*), sehingga mempercepat transmisi sentimen ke dalam mekanisme pasar [12]. Besarnya volume data teks yang dihasilkan setiap hari, penggunaan bahasa informal, istilah teknis kripto, serta keberadaan sarkasme dan hiperbola menjadikan diskursus *cryptocurrency* di platform Twitter bersifat kompleks dan tidak terstruktur [13]. Kondisi ini dapat menyebabkan pemantauan sentimen secara manual menjadi tidak efisien dan sulit dilakukan secara konsisten, sehingga diperlukan pendekatan analisis sentimen berbasis komputasi yang mampu mengidentifikasi dan mengklasifikasikan opini publik secara otomatis untuk memahami dinamika pengaruhnya terhadap pergerakan harga *cryptocurrency*.

Dalam konteks ini, tugas klasifikasi teks menjadi komponen penting, di mana pendekatan berbasis *machine learning* klasik sering digunakan sebagai *baseline* karena relatif sederhana dan efisien secara komputasi [14]. Metode seperti *Support Vector Machine* (SVM) dan *Naïve Bayes* masih umum dipakai dalam analisis sentimen untuk memetakan opini menjadi kelas positif, netral, atau negatif [15]. Dalam sebuah eksperimen klasifikasi sentimen teks media sosial, algoritma SVM mencapai akurasi sebesar 88,85%, sedangkan *Naïve Bayes* hanya sekitar 72,64% ketika dievaluasi pada dataset unggahan media sosial berlabel tiga kelas, menunjukkan perbedaan performa yang cukup besar antar metode klasik tersebut [16]. Pada penelitian lain, analisis sentimen Twitter terkait vaksin *COVID-19* di Indonesia menunjukkan bahwa algoritma SVM memiliki performa lebih tinggi

dibandingkan *Naïve Bayes* dan *LSTM* dalam klasifikasi sentimen publik [17]. Hasil lain menunjukkan bahwa kombinasi *Word2Vec* + SVM pada dataset ulasan *Amazon* mencapai akurasi sebesar 83,45%, dan kombinasi TF-IDF + SVM yang mencapai akurasi tertinggi sebesar 89,67% pada pengujian yang sama [18]. Meskipun pendekatan berbasis TF-IDF maupun *word embedding* yang dikombinasikan dengan algoritma klasik mampu memberikan performa yang cukup baik pada dataset tertentu, model-model tersebut cenderung bergantung pada kualitas representasi fitur serta proses prapemrosesan yang dilakukan.

Perkembangan model berbasis *transformer* memperbaiki keterbatasan pendekatan klasik karena mampu membangun representasi kontekstual dua arah melalui mekanisme *self-attention*, sehingga lebih efektif menangkap relasi antarkata, dependensi jarak jauh, serta nuansa makna dibanding *bag-of-words*, TF-IDF, maupun RNN/LSTM [19]. Keunggulan ini terlihat pada perbandingan langsung di tugas sentimen media sosial pada dataset *Twitter*, dimana metode klasik seperti *Naïve Bayes*, SVM, dan LSTM tertinggal dari model berbasis *transformer*, misalnya RoBERTa-LSTM, RoBERTa-GRU, hingga varian hibrida berbasis RoBERTa yang mencapai akurasi tertinggi dengan F1 naik secara signifikan [20]. Pada *benchmark* sentimen SST-2, model *transformer* juga menunjukkan performa tinggi dengan BERT dan TextCNN+RoBERTa, sedangkan varian *transformer* lain pada studi yang sama membuktikan capaian akurasi yang tertinggi diantara model lain [21]. Pada dataset ulasan *Amazon book reviews*, RoBERTa dibuktikan mengungguli model tradisional dan LSTM dengan 96,30% akurasi serta *F1-score* 98,11%, memperlihatkan konsistensi keunggulan *transformer* pada teks yang kaya konteks [22]. Meskipun BERT memiliki kemampuan yang baik dalam memahami konteks bahasa, keterbatasannya dalam mengenali istilah, jargon, dan karakteristik khusus pada domain *cryptocurrency* menjadi alasan diperlukannya model BERT yang telah dilatih secara spesifik pada data *cryptocurrency* untuk menghasilkan klasifikasi sentimen yang lebih akurat.

Dalam domain spesifik, pendekatan *domain adaptation* menunjukkan peningkatan performa yang lebih jelas dibandingkan model bahasa umum [23]. Beberapa penelitian terbaru menunjukkan bahwa model yang dipra-latih secara khusus pada korpus *cryptocurrency*, seperti CryptoBERT, mampu menghasilkan

representasi semantik yang lebih sesuai dengan konteks pasar kripto [24], [25], [26]. Salah satu studi menunjukkan bahwa integrasi sentimen berbasis CryptoBERT menghasilkan performa prediksi harga dengan koefisien determinasi R^2 sebesar 0,9874, yang mengindikasikan kontribusi signifikan terhadap peningkatan akurasi dibandingkan FinBERT dan model *Transformer* lainnya [25]. Studi lain menunjukkan bahwa model berbasis *Transformer* mengungguli pendekatan leksikon tradisional, dengan CryptoBERT menunjukkan kinerja paling optimal dibandingkan BERT, FinBERT, dan BERTweet, sehingga mengindikasikan bahwa model yang dilatih khusus pada korpus kripto mampu menghasilkan klasifikasi sentimen yang lebih relevan dan konsisten di bidang cryptocurrency dibandingkan model *domain* umum maupun finansial [24]. Selain itu, pada sistem perdagangan otomatis berbasis CBITS, pemanfaatan CryptoBERT dalam analisis sentimen dengan akurasi mencapai 81,42% menunjukkan peningkatan performa strategi perdagangan dibandingkan penggunaan model BERT umum [27]. Penelitian lain juga menunjukkan bahwa hasil analisis sentimen yang dihasilkan oleh model FinBERT dan CryptoBERT memiliki hubungan yang kuat dengan pergerakan harga *cryptocurrency* aktual, dimana nilai koefisien korelasi yang diperoleh mencapai 0,88 [28]. Meskipun demikian, model berbasis *Transformer* saja cenderung masih memiliki keterbatasan karena seringkali belum optimal dalam menangkap pola lokal dan variasi ekspresi singkat pada teks media sosial.

Sejumlah penelitian menunjukkan bahwa integrasi CNN setelah *Transformer* mampu mengatasi keterbatasan model *Transformer* dalam menangkap pola lokal, dependensi frasa pendek, serta fitur spasial yang bersifat kontekstual sempit, sehingga memperkuat representasi teks dan meningkatkan akurasi klasifikasi [29], [30], [31], [32]. Pada konteks analisis sentimen di Twitter yang memiliki karakteristik teks singkat, ambigu, dan beragam secara emosional, pengujian menunjukkan bahwa penggunaan BERT saja menghasilkan akurasi sebesar 85,61%, sementara integrasi dengan RNN meningkatkan akurasi menjadi 92,21%, dan kombinasi dengan CNN memberikan peningkatan tertinggi hingga 93,37%, sekaligus menunjukkan performa *precision*, *recall*, dan *F1-score* yang lebih baik dibandingkan varian tanpa arsitektur tambahan [32]. Dalam konteks analisis

sentimen *bullet screen* pada platform video yang memiliki karakteristik panjang teks bervariasi, banyak kata baru, dan ambiguitas emosional, pengujian menunjukkan bahwa akurasi RoBERTa sebesar 76,35% meningkat menjadi 84,19% setelah diintegrasikan dengan CNN, sekaligus memberikan peningkatan pada *precision*, *recall*, dan *F1-score* dalam mengklasifikasikan sentimen terkait isu Rusia–Ukraina [33]. Pada arsitektur *hybrid* serupa, masih ditemukan keterbatasan dalam memprioritaskan fitur yang paling relevan secara kontekstual, terutama ketika menghadapi teks yang mengandung emoji, ironi, atau sarkasme.

Representasi yang dihasilkan dari kombinasi beberapa lapisan masih berpotensi mengandung fitur yang kurang relevan atau redundan, sehingga diperlukan mekanisme tambahan yang mampu memprioritaskan informasi yang paling berkontribusi terhadap proses klasifikasi. Dalam konteks tersebut, penambahan *attention mechanism* pada arsitektur *hybrid Transformer–CNN* terbukti semakin memperkuat kemampuan model dalam memfokuskan bobot pada fitur yang paling relevan secara kontekstual [34], [35], [36]. Integrasi mekanisme *attention* pada arsitektur BERT–CNN–BiLSTM yang mengakomodasi pola penggunaan *emoji* membuktikan peningkatan akurasi dari 95,47% menjadi 97,12% serta *F1-score* dari 0,86 menjadi 0,94 pada dataset *SuicidEmoji*, menunjukkan bahwa *attention* membantu model menyoroti indikator linguistik dan fitur emoji yang paling relevan dalam deteksi depresi berbasis media sosial [37]. Hasil tersebut menunjukkan bahwa mekanisme *attention* berfungsi sebagai strategi penyempurnaan dalam arsitektur hibrida, karena mampu mengarahkan fokus model pada informasi yang paling signifikan, sehingga meningkatkan ketepatan, konsistensi, dan sensitivitas model terhadap variasi konteks dalam proses klasifikasi teks.

Berdasarkan keseluruhan kajian tersebut, penelitian ini diarahkan pada pengembangan model analisis sentimen *cryptocurrency* yang lebih adaptif dengan memanfaatkan arsitektur hibrida berbasis *Transformer–CNN* yang diperkuat dengan mekanisme *attention*. Pada tahap awal, penelitian ini mengevaluasi performa model *Transformer domain-spesifik* dibandingkan dengan varian hibrida untuk mengidentifikasi kontribusi ekstraksi fitur lokal terhadap peningkatan akurasi. Selanjutnya, mekanisme *attention* diintegrasikan untuk mengoptimalkan proses pembobotan fitur sehingga model mampu lebih responsif terhadap

karakteristik teks kripto yang umumnya singkat, mengandung *emoji*, *slang*, serta indikasi sarkasme. Hasil klasifikasi sentimen yang diperoleh kemudian dianalisis lebih lanjut menggunakan uji korelasi Spearman untuk mengidentifikasi hubungan antara sentimen publik dengan pergerakan harga *cryptocurrency* aktual. Melalui pendekatan tersebut, penelitian ini diharapkan tidak hanya menghasilkan model klasifikasi sentimen yang lebih akurat, tetapi juga memberikan pemahaman mendalam mengenai keterkaitan antara dinamika sentimen publik di media sosial dengan fluktuasi harga pada pasar *cryptocurrency* yang bersifat cepat dan volatil.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah tersebut, rumusan masalah dalam penelitian ini dapat diidentifikasi sebagai berikut:

1. Bagaimana hasil performa model CryptoBERT sebagai *domain-specific language model* dalam melakukan klasifikasi sentimen pada teks diskursus *cryptocurrency* dibandingkan dengan model BERT generik?
2. Bagaimana hasil pengaruh integrasi *Convolutional Neural Network* (CNN) dengan CryptoBERT dalam meningkatkan kemampuan ekstraksi fitur lokal pada analisis sentimen diskursus *cryptocurrency*?
3. Bagaimana hasil penerapan *attention mechanism* pada arsitektur *hybrid* CryptoBERT–CNN memengaruhi kemampuan model dalam memfokuskan fitur-fitur penting yang berkontribusi terhadap klasifikasi sentimen?
4. Bagaimana hasil perbandingan kinerja model CryptoBERT, CryptoBERT–CNN, dan CryptoBERT–CNN berbasis *Attention* ditinjau dari metrik evaluasi *Accuracy*, *F1-Score*, *Loss*, serta *Running Time* pada analisis sentimen diskursus *cryptocurrency* di platform Twitter?
5. Bagaimana hasil korelasi antara sentimen publik yang dihasilkan dari model analisis sentimen dengan pergerakan harga *cryptocurrency* aktual berdasarkan pengujian menggunakan metode korelasi *Spearman*?

1.3 Batasan Masalah

Batasan masalah yang diidentifikasi pada penelitian ini diantaranya:

1. Dataset yang digunakan dalam penelitian ini terbatas pada data teks berupa *tweets* yang diperoleh melalui platform *Kaggle*, di mana data tersebut merupakan hasil *scraping* dari *Twitter* (X). Oleh karena itu, hasil penelitian ini tidak ditujukan untuk digeneralisasikan pada platform media sosial lain seperti *Reddit*, *Telegram*, atau *YouTube*.
2. Dataset yang digunakan dalam penelitian ini masih berupa data berbahasa Inggris dan belum mencakup data berbahasa Indonesia maupun bahasa lainnya, sehingga hasil penelitian lebih merepresentasikan sentimen *cryptocurrency* pada konteks bahasa Inggris.
3. *Tweets* yang dianalisis merupakan data berbahasa Inggris, sehingga model yang dikembangkan tidak secara langsung merepresentasikan diskursus *cryptocurrency* dalam bahasa lain.
4. Data yang digunakan dibatasi pada *tweets* yang dipublikasikan selama periode tahun 2023, sehingga dinamika sentimen di luar rentang waktu tersebut tidak termasuk dalam cakupan penelitian.
5. Klasifikasi sentimen dibatasi pada tiga kategori, yaitu positif, negatif, dan netral, tanpa mempertimbangkan intensitas sentimen (misalnya sangat positif atau sangat negatif) maupun analisis aspek tertentu (*aspect-based sentiment analysis*).
6. Evaluasi performa model dibatasi pada metrik *Accuracy*, *F1-Score*, *Loss*, dan *Running Time*, tanpa melibatkan metrik tambahan seperti *AUC-ROC*, *Matthews Correlation Coefficient* (MCC), maupun analisis interpretabilitas lanjutan.
7. Implementasi model dilakukan menggunakan *framework deep learning* tertentu (misalnya *PyTorch* atau *TensorFlow*) pada lingkungan komputasi terbatas, sehingga hasil *running time* dan performa dapat dipengaruhi oleh spesifikasi perangkat keras yang digunakan.
8. Penelitian tidak mencakup tahap *deployment* model ke dalam sistem atau aplikasi nyata, sehingga hasil yang diperoleh hanya terbatas pada tahap pengembangan dan evaluasi model dalam kerangka kerja KDD, tanpa implementasi lebih lanjut pada lingkungan produksi.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Adapun tujuan dalam penelitian ini diantaranya:

1. Mengevaluasi performa CryptoBERT sebagai *domain-specific language model* dalam melakukan klasifikasi sentimen pada teks diskursus *cryptocurrency* dibandingkan dengan model BERT generik.
2. Menganalisis pengaruh integrasi *Convolutional Neural Network* (CNN) dengan CryptoBERT dalam meningkatkan kemampuan ekstraksi fitur lokal pada analisis sentimen diskursus *cryptocurrency*.
3. Mengkaji peran dan pengaruh penerapan *attention mechanism* pada arsitektur *hybrid* CryptoBERT–CNN dalam memfokuskan fitur-fitur penting yang berkontribusi terhadap proses klasifikasi sentimen.
4. Membandingkan kinerja model CryptoBERT, CryptoBERT–CNN, dan CryptoBERT–CNN berbasis *Attention* berdasarkan metrik evaluasi *Accuracy*, *F1-Score*, *Loss*, serta *Running Time* pada analisis sentimen diskursus *cryptocurrency* di platform Twitter.
5. Menganalisis korelasi antara sentimen publik yang dihasilkan dari model analisis sentimen dengan pergerakan harga *cryptocurrency* aktual menggunakan metode korelasi *Spearman*.

1.4.2 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan sumbangan yang bermakna terhadap pengembangan ilmu pengetahuan, baik dari sisi konseptual maupun implementatif. Kontribusi yang dihasilkan dalam penelitian ini adalah sebagai berikut:

1. Memperkaya literatur akademik mengenai penerapan *domain-specific language model* (CryptoBERT) dalam analisis sentimen pada sektor keuangan digital, serta memberikan pembuktian empiris mengenai efektivitas model berbasis domain dibandingkan model bahasa generik.
2. Memberikan kontribusi konseptual terkait pengembangan arsitektur *hybrid* yang mengintegrasikan *transformer* (CryptoBERT) dan *Convolutional*

Neural Network (CNN) untuk menggabungkan pemahaman konteks global dan ekstraksi fitur lokal dalam satu kerangka pemodelan terpadu.

3. Mengkaji peran *attention mechanism* tambahan dalam meningkatkan selektivitas fitur serta interpretabilitas model, sehingga memperluas pemahaman mengenai optimalisasi arsitektur *deep learning* pada teks media sosial yang dinamis dan tidak terstruktur.
4. Menyediakan hasil evaluasi berbasis metrik *Accuracy*, *F1-Score*, *Loss*, dan *Running Time* sebagai referensi empiris dalam membandingkan model tunggal dan model *hybrid* pada analisis sentimen *cryptocurrency*.
5. Mendorong penerapan teknologi NLP dan *deep learning* dalam ekosistem keuangan digital, khususnya dalam mengolah data teks media sosial yang berskala besar dan terus berkembang.
6. Menjadi referensi praktis bagi pengembang sistem *financial analytics* dalam membangun model yang lebih adaptif, efisien, dan memiliki keseimbangan antara akurasi prediksi dan efisiensi waktu komputasi.

1.5 Sistematika Penulisan

Penulisan skripsi ini disusun sesuai dengan pedoman dan format yang telah ditetapkan oleh program studi. Sistematika penulisan dirancang untuk memberikan alur pembahasan yang terstruktur dan sistematis, sehingga memudahkan pembaca dalam memahami keseluruhan isi penelitian. Skripsi ini terdiri atas lima bab utama dengan rincian sebagai berikut:

BAB I: PENDAHULUAN

Bab pertama berisi penjelasan mengenai latar belakang penelitian yang menguraikan pertumbuhan aset *cryptocurrency* secara global, peran media sosial dalam membentuk opini publik, serta urgensi analisis sentimen dalam memahami dinamika pasar kripto yang volatil. Selain itu, bab ini memuat rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, serta sistematika penulisan. Bab ini memberikan gambaran umum mengenai arah penelitian dan kontribusi yang diharapkan, serta menjadi landasan konseptual untuk pembahasan pada bab-bab selanjutnya.

BAB II: LANDASAN TEORI

Bab kedua membahas teori dan konsep yang mendukung penelitian, meliputi *Sentiment Analysis*, *Natural Language Processing* (NLP), serta perkembangan arsitektur *deep learning* berbasis *Transformer*. Pada bab ini dijelaskan konsep BERT dan mekanisme *self-attention*, serta pengembangan *domain-specific language* model seperti CryptoBERT dalam konteks teks finansial dan *cryptocurrency*. Selain itu, dipaparkan pula konsep *Convolutional Neural Network* (CNN) untuk ekstraksi fitur lokal pada teks serta *attention mechanism* untuk meningkatkan selektivitas fitur dalam proses klasifikasi. Bab ini juga menguraikan penelitian-penelitian terdahulu yang relevan sebagai dasar pembandingan dan penguatan metodologi.

BAB III: METODOLOGI PENELITIAN

Bab ketiga menjelaskan tahapan penelitian secara sistematis, dimulai dari pemilihan dan pemahaman dataset *tweets cryptocurrency* tahun 2023 yang diperoleh dari platform *Kaggle*. Tahap selanjutnya meliputi preprocessing data, seperti pembersihan teks, normalisasi, tokenisasi, dan penyesuaian label sentimen. Setelah itu dijelaskan perancangan dan implementasi arsitektur model yang terdiri dari CryptoBERT, CryptoBERT-CNN, serta CryptoBERT-CNN berbasis *Attention*. Bab ini juga memaparkan proses pelatihan (*training*), validasi, serta evaluasi model menggunakan metrik *Accuracy*, *F1-Score*, *Loss*, dan *Running Time*. Selain itu, dijelaskan spesifikasi perangkat keras dan perangkat lunak yang digunakan dalam eksperimen.

BAB IV: ANALISIS DAN HASIL PENELITIAN

Bab keempat menyajikan hasil eksperimen dari masing-masing model yang diuji. Analisis dilakukan dengan membandingkan performa CryptoBERT, CryptoBERT-CNN, dan CryptoBERT-CNN berbasis *Attention* berdasarkan metrik evaluasi yang telah ditentukan. Selain itu, dibahas interpretasi hasil, perbandingan dengan penelitian terdahulu, serta analisis kelebihan dan keterbatasan arsitektur yang diusulkan.

BAB V: SIMPULAN DAN SARAN

Bab kelima berisi kesimpulan penelitian yang merangkum temuan utama dan menjawab rumusan masalah. Selain itu, disampaikan saran untuk penelitian selanjutnya, seperti perluasan dataset, eksplorasi arsitektur transformer lain, atau pengembangan sistem analisis sentimen secara real-time dalam konteks cryptocurrency.



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA