BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan komunikasi telah mengubah cara masyarakat memperoleh informasi, berinteraksi, serta menyampaikan pendapat terhadap berbagai isu publik. Kehadiran media sosial memungkinkan masyarakat tidak hanya menjadi penerima informasi, tetapi juga berperan aktif dalam memberikan kritik, dukungan, maupun evaluasi terhadap kebijakan pemerintah secara terbuka dan *real-time* (Zordi et al., 2026). Di antara berbagai platform media sosial, X (sebelumnya *Twitter*) menjadi salah satu media yang paling aktif digunakan untuk mendiskusikan isu sosial, ekonomi, maupun politik karena mampu menghasilkan data opini publik dalam jumlah besar dalam waktu yang singkat (Asokere et al., 2025). Karakteristik tersebut menjadikan media sosial sebagai salah satu sumber informasi yang potensial untuk memahami persepsi masyarakat terhadap implementasi suatu kebijakan pemerintah melalui pendekatan *Natural Language Processing* (NLP) (Ridwan & Lubis, 2026).

Salah satu kebijakan pemerintah yang memperoleh perhatian luas dari masyarakat adalah Instruksi Presiden Republik Indonesia Nomor 1 Tahun 2025 tentang Efisiensi Belanja dalam Pelaksanaan Anggaran Pendapatan dan Belanja Negara (APBN) dan Anggaran Pendapatan dan Belanja Daerah (APBD) (Kemenkeu, 2025). Kebijakan ini diterbitkan sebagai upaya meningkatkan efektivitas pengelolaan keuangan negara melalui optimalisasi penggunaan anggaran dan realokasi belanja menuju program-program prioritas (AntaraNews, 2025). Implementasi kebijakan tersebut berdampak pada berbagai sektor strategis, seperti pendidikan, kesehatan, pembangunan infrastruktur, serta program sosial, sehingga memunculkan beragam tanggapan dari masyarakat (Sapay et al., 2025). Sebagian masyarakat memandang kebijakan tersebut sebagai langkah yang diperlukan untuk meningkatkan efisiensi dan akuntabilitas pengelolaan anggaran negara, sedangkan sebagian lainnya mengkhawatirkan dampaknya terhadap

kualitas pelayanan publik maupun keberlanjutan program pemerintah (Walizi, 2025). Kondisi tersebut menunjukkan bahwa opini masyarakat terhadap kebijakan efisiensi anggaran tidak hanya berupa penilaian positif atau negatif, tetapi juga mencerminkan berbagai aspek evaluasi kebijakan yang berbeda.

Besarnya volume opini yang dihasilkan media sosial mendorong pemanfaatan *Natural Language Processing* (NLP), khususnya *Sentiment Analysis*, untuk menganalisis respons masyarakat secara otomatis (Ritter et al., 2023). Berbagai penelitian mengenai kebijakan efisiensi anggaran di Indonesia telah menerapkan pendekatan tersebut menggunakan algoritma seperti *Support Vector Machine* (SVM) (Mumtaz et al., 2025) dan *Naïve Bayes* (Alghifari & Azis, 2025) untuk mengklasifikasikan opini masyarakat ke dalam kategori positif, negatif, maupun netral. Hasil penelitian tersebut mampu memberikan gambaran umum mengenai kecenderungan sentimen masyarakat terhadap kebijakan pemerintah, sehingga dapat menjadi salah satu indikator dalam memahami respons publik. Namun demikian hanya menghasilkan satu label sentimen untuk keseluruhan opini sehingga belum mampu mengidentifikasi bagian atau aspek kebijakan yang menjadi sumber munculnya sentiment (Mutmainah et al., 2025). Padahal, dalam konteks evaluasi kebijakan publik, pembuat kebijakan tidak hanya memerlukan informasi mengenai kecenderungan sentimen masyarakat, tetapi juga membutuhkan informasi mengenai aspek kebijakan apa yang memperoleh apresiasi maupun kritik sehingga dapat menjadi dasar dalam melakukan evaluasi dan penyempurnaan implementasi kebijakan (Tjilen & Maturbongs, 2025).

Informasi opini publik berbasis aspek menjadi penting karena dapat memberikan masukan yang lebih spesifik mengenai bagian-bagian kebijakan yang memperoleh apresiasi maupun kritik dari masyarakat. Informasi tersebut dapat dimanfaatkan oleh pemerintah sebagai *evidence-based* input dalam proses monitoring, evaluasi, dan penyempurnaan implementasi kebijakan. Dengan mengetahui aspek kebijakan yang paling banyak menjadi perhatian publik, pemerintah dapat menetapkan prioritas perbaikan secara lebih terarah dibandingkan

apabila hanya mengetahui kecenderungan sentimen masyarakat secara umum (Hua et al., 2024).

Perkembangan *Aspect-Based Sentiment Analysis* (ABSA) memberikan solusi terhadap keterbatasan *Sentiment Analysis* konvensional dengan mengidentifikasi aspek yang menjadi objek evaluasi sekaligus menentukan polaritas sentimen pada setiap aspek tersebut (Pramono et al., 2024). Berbeda dengan analisis sentimen yang hanya menghasilkan satu label untuk keseluruhan opini, ABSA mampu menghasilkan informasi yang lebih rinci mengenai bagian atau aspek tertentu yang memperoleh apresiasi maupun kritik dari masyarakat (Jimmy et al., 2026). Kemampuan tersebut menjadikan ABSA lebih sesuai untuk mendukung evaluasi kebijakan publik karena pembuat kebijakan tidak hanya memerlukan informasi mengenai kecenderungan sentimen masyarakat, tetapi juga membutuhkan informasi mengenai aspek kebijakan yang menjadi penyebab munculnya sentimen tersebut. Berbagai penelitian menunjukkan bahwa ABSA mampu menghasilkan informasi yang lebih kaya dibandingkan analisis sentimen konvensional, namun implementasinya pada domain kebijakan publik masih relatif terbatas dibandingkan domain ulasan produk maupun layanan komersial (Hua et al., 2024).

Kajian sistematis mengenai penerapan analisis sentimen pada kebijakan publik menunjukkan bahwa sebagian besar penelitian masih berfokus pada klasifikasi sentimen secara umum dan belum banyak mengembangkan pendekatan ekstraksi aspek secara spesifik (Harwenda et al., 2025). Selain itu, penelitian pada domain kebijakan publik masih menunjukkan variasi yang besar baik dari sisi metode, objek kebijakan, maupun dataset yang digunakan sehingga belum tersedia pendekatan yang baku untuk membangun maupun mengevaluasi sistem ABSA pada domain tersebut (Hua et al., 2024). Kondisi ini menunjukkan bahwa pengembangan ABSA pada domain kebijakan publik masih memiliki peluang penelitian yang luas, khususnya pada bahasa Indonesia yang masih tergolong sebagai bahasa dengan sumber daya terbatas (*low-resource language*).

Tantangan lain yang masih dihadapi dalam pengembangan ABSA adalah proses identifikasi aspek yang tidak selalu dinyatakan secara eksplisit dalam suatu opini.

Pada media sosial, masyarakat sering kali menyampaikan pendapat menggunakan kalimat yang singkat, tidak baku, mengandung singkatan, ironi, maupun konteks yang implisit sehingga aspek yang menjadi objek evaluasi harus disimpulkan berdasarkan keseluruhan makna kalimat (Vishwakarma & Wao, 2023). Berbagai penelitian terbaru menunjukkan bahwa ekstraksi implisit aspek masih menjadi salah satu tantangan utama dalam ABSA karena memerlukan pemahaman semantik dan konteks yang lebih mendalam dibandingkan ekstraksi aspek eksplisit (Alawwad et al., 2026). Kondisi tersebut menyebabkan pendekatan klasifikasi pada tingkat dokumen atau kalimat menjadi kurang memadai untuk menggambarkan hubungan antara aspek dan sentimen pada opini masyarakat di media sosial.

Untuk mengatasi permasalahan tersebut, penelitian ini menggunakan pendekatan *Named Entity Recognition* (NER) berbasis *sequence labeling*. Berbeda dengan pendekatan klasifikasi, NER melakukan pelabelan pada tingkat token sehingga model mampu mengenali bagian teks (*entity span*) yang secara langsung merepresentasikan suatu aspek beserta polaritas sentimennya (W.P et al., 2022). Pendekatan ini memungkinkan satu *Tweet* mengandung lebih dari satu aspek dengan polaritas sentimen yang berbeda sehingga lebih sesuai dengan karakteristik opini masyarakat mengenai kebijakan publk (Oswal et al., 2022). Meskipun demikian, penerapan NER pada data media sosial masih menghadapi tantangan yang cukup besar karena tingginya variasi bahasa, kesalahan penulisan, penggunaan singkatan, serta keterbatasan konteks yang menyebabkan proses penentuan batas awal dan batas akhir *entity span* menjadi lebih kompleks dibandingkan pada teks formal (Ritter et al., 2023).

Selain kompleksitas metode, berbagai penelitian juga menunjukkan bahwa kualitas dataset merupakan faktor yang sangat menentukan keberhasilan NER (Shvarts et al., 2026). Kajian mengenai NLP pada bahasa dengan sumber daya terbatas menegaskan bahwa keterbatasan korpus beranotasi, belum adanya standar anotasi, serta rendahnya reproduktibilitas penelitian masih menjadi tantangan utama dalam pengembangan model NLP (Chen, 2025). Pada penelitian ABSA, sebagian besar dataset dibangun menggunakan anotasi otomatis atau *silver*

annotation (Tolentino et al., 2021), sedangkan penelitian yang menyusun *gold-standard* dataset melalui proses anotasi manual, penyusunan *Annotation Guideline*, validasi oleh pakar, serta pengukuran reliabilitas anotasi masih relatif terbatas. Padahal, kualitas anotasi menjadi fondasi utama dalam membangun dataset yang konsisten dan dapat digunakan sebagai dasar pengembangan maupun evaluasi model ekstraksi aspek (Palla, 2024).

Dari sisi pemodelan, berbagai arsitektur *deep learning* telah diterapkan untuk menyelesaikan tugas *sequence labeling*, seperti SimpleRNN, BiGRU, BiLSTM, maupun Transformer Encoder. Masing-masing arsitektur memiliki karakteristik yang berbeda dalam mempelajari representasi konteks sehingga menghasilkan performa yang beragam pada setiap domain (Zulkarnain et al., 2026). Namun, sebagian besar penelitian sebelumnya hanya mengevaluasi satu atau dua model pada dataset yang berbeda-beda sehingga sulit menentukan apakah perbedaan performa disebabkan oleh arsitektur model atau oleh kualitas dataset yang digunakan (Raymond et al., 2025). Selain itu, penelitian yang mengkaji pengaruh penambahan ukuran dataset terhadap performa model pada tugas ekstraksi aspek berbasis NER juga masih sangat terbatas (Meng et al., 2025). Oleh karena itu, evaluasi berbagai arsitektur *deep learning* pada dataset yang dibangun melalui prosedur anotasi yang sama menjadi penting untuk memperoleh perbandingan yang lebih objektif sekaligus memahami pengaruh ukuran dataset terhadap kemampuan generalisasi model (Seow et al., 2025).

Berdasarkan uraian tersebut, dapat diidentifikasi beberapa kesenjangan penelitian. Pertama, penelitian mengenai opini publik terhadap kebijakan efisiensi anggaran masih didominasi oleh pendekatan *Sentiment Analysis*, sehingga belum mampu mengidentifikasi aspek kebijakan yang menjadi sumber munculnya sentimen. Kedua, implementasi ABSA berbasis NER pada domain kebijakan publik berbahasa Indonesia masih relatif terbatas, terutama untuk aspek-aspek yang bersifat implisit. Ketiga, penelitian sebelumnya masih minim dalam membangun dataset beranotasi yang tervalidasi melalui penyusunan *Annotation Guideline* dan pengukuran reliabilitas anotasi. Keempat, masih terbatas penelitian yang

membandingkan berbagai arsitektur *deep learning* menggunakan dataset yang sama serta mengevaluasi performanya melalui *Exact Span Matching* dan *Partial Span Matching*. Kelima, pengaruh penambahan ukuran dataset terhadap kemampuan model dalam melakukan ekstraksi aspek pada domain kebijakan publik juga belum banyak dilaporkan.

Berdasarkan kesenjangan tersebut, penelitian ini berfokus pada pembangunan dataset ABSA berbasis NER yang beranotasi dan tervalidasi untuk opini publik mengenai kebijakan efisiensi anggaran pada media sosial X. Dataset dikembangkan melalui penyusunan *Annotation Guideline* berdasarkan teori evaluasi kebijakan William N. Dunn, divalidasi oleh dua orang pakar menggunakan *Fleiss' Kappa*, serta dianotasi secara manual untuk menghasilkan data pelatihan yang konsisten. Selanjutnya, dataset tersebut digunakan untuk mengevaluasi dan membandingkan empat arsitektur *deep learning*, yaitu SimpleRNN-CRF, BiGRU-CRF, BiLSTM-CRF, dan Transformer Encoder-CRF, menggunakan metrik *Exact Span Matching* dan *Partial Span Matching*. Penelitian ini juga melakukan eksperimen penambahan ukuran dataset untuk menganalisis pengaruh jumlah data terhadap performa model. Dengan demikian, kontribusi utama penelitian ini tidak hanya terletak pada perbandingan model *deep learning*, tetapi juga pada pembangunan gold-standard dataset ABSA berbasis NER yang valid, reliabel, dan tervalidasi melalui annotation guideline, anotasi manual, validasi pakar, serta pengukuran Inter-Rater Reliability sehingga dapat dimanfaatkan sebagai sumber daya linguistik pada penelitian ABSA dan NER di domain kebijakan publik berbahasa Indonesia, yang diharapkan dapat menjadi referensi bagi penelitian selanjutnya.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Belum tersedianya gold-standard dataset ABSA berbasis NER pada domain kebijakan publik berbahasa Indonesia yang dibangun melalui *annotation guideline*, anotasi manual, validasi pakar, dan pengukuran reliabilitas

anotasi menyebabkan keterbatasan sumber daya data yang valid dan reliabel untuk mendukung pengembangan serta evaluasi model ekstraksi aspek.

2. Implementasi pendekatan ABSA berbasis NER pada opini masyarakat terhadap kebijakan efisiensi anggaran masih belum banyak dikaji, khususnya untuk menangani aspek-aspek yang diterapkan pada opini kebijakan yang banyak disampaikan secara kontekstual dan tidak langsung.
3. Evaluasi komprehensif terhadap berbagai arsitektur *Deep Learning* menggunakan dataset yang dibangun melalui prosedur anotasi yang sama masih terbatas, sehingga karakteristik performa masing-masing model pada domain kebijakan publik belum diketahui secara objektif.
4. Pengaruh penambahan ukuran dataset terhadap kemampuan generalisasi model dalam melakukan ekstraksi aspek pada domain kebijakan publik masih belum banyak dilaporkan dalam penelitian sebelumnya.

1.3 Pertanyaan Penelitian

Dari latar belakang dan rumusan masalah yang telah dijelaskan sebelumnya, dapat dirumuskan beberapa pertanyaan penelitian sebagai berikut:

1. Bagaimana membangun dan memvalidasi gold-standard dataset ABSA berbasis NER yang valid dan reliabel pada opini masyarakat terhadap kebijakan efisiensi anggaran di media sosial X sehingga layak digunakan sebagai sumber daya untuk pengembangan dan evaluasi model ekstraksi aspek?
2. Bagaimana mengimplementasikan pendekatan NER menggunakan arsitektur SimpleRNN-CRF, BiGRU-CRF, BiLSTM-CRF, dan Transformer Encoder-CRF untuk melakukan ekstraksi aspek dan sentimen pada opini masyarakat mengenai kebijakan efisiensi anggaran?
3. Bagaimana performa arsitektur SimpleRNN-CRF, BiGRU-CRF, BiLSTM-CRF, dan Transformer Encoder-CRF dalam melakukan ekstraksi aspek berdasarkan evaluasi *Exact Span Matching* dan *Partial Span Matching*?

4. Bagaimana pengaruh penambahan ukuran dataset terhadap kemampuan generalisasi model dalam melakukan ekstraksi aspek pada domain kebijakan efisiensi anggaran?

1.4 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah yang ada, tujuan dari penelitian ini adalah sebagai berikut:

1. Membangun dan memvalidasi gold-standard dataset ABSA berbasis NER yang valid dan reliabel pada opini masyarakat terhadap kebijakan efisiensi anggaran di media sosial X sebagai sumber daya untuk pengembangan dan evaluasi model ekstraksi aspek.
2. Mengimplementasikan pendekatan NER menggunakan arsitektur SimpleRNN-CRF, BiGRU-CRF, BiLSTM-CRF, dan Transformer Encoder-CRF untuk melakukan ekstraksi aspek dan sentimen pada opini masyarakat mengenai kebijakan efisiensi anggaran.
3. Mengevaluasi dan membandingkan performa arsitektur SimpleRNN-CRF, BiGRU-CRF, BiLSTM-CRF, dan Transformer Encoder-CRF berdasarkan metrik *Exact Span Matching* dan *Partial Span Matching*.
4. Menganalisis pengaruh penambahan ukuran dataset terhadap kemampuan generalisasi model dalam melakukan ekstraksi aspek pada domain kebijakan efisiensi anggaran.

1.5 Batasan Masalah

Agar penelitian ini lebih terarah dan tetap pada ruang lingkup yang ditentukan, maka diberikan batasan-batasan masalah sebagai berikut:

1. Objek penelitian dibatasi pada opini masyarakat mengenai kebijakan Efisiensi Anggaran dalam Pelaksanaan APBN dan APBD berdasarkan Instruksi Presiden Republik Indonesia Nomor 1 Tahun 2025 yang dipublikasikan melalui media sosial X (sebelumnya Twitter).

2. Data yang digunakan dalam penelitian ini berupa teks berbahasa Indonesia dari media sosial X dan tidak mencakup data dari platform media sosial lainnya terkait opini publik tentang kebijakan efisiensi anggaran.
3. Pengambilan data dibatasi pada periode setelah diterbitkannya Instruksi Presiden (Inpres) No. 1 Tahun 2025 (Januari 2025 hingga Oktober 2025) guna menjaga relevansi data dengan konteks kebijakan saat ini.
4. Penelitian ini membatasi ruang lingkup analisis pada isi utama (*body text*) tweet. Informasi yang terkandung dalam tagar (*hashtag*) tidak digunakan sebagai sumber informasi dalam proses ekstraksi aspek dan sentimen sehingga analisis dilakukan secara konsisten berdasarkan konteks linguistik pada badan tweet.
5. Aspek yang dianalisis dibatasi pada tiga aspek evaluasi kebijakan yang diadaptasi dari teori evaluasi kebijakan William N. Dunn, yaitu Efektivitas Kebijakan, Pengelolaan Sumber Daya, dan Keberlanjutan (*Sustainability*).
6. Pendekatan ABSA yang digunakan menerapkan metode NER berbasis *sequence labeling* menggunakan skema pelabelan BILUO.
7. Proses anotasi dataset dilakukan secara manual menggunakan perangkat lunak Prodigy berdasarkan *Annotation Guideline* yang telah disusun dan divalidasi oleh dua orang pakar melalui pengukuran *Inter-Rater Reliability* menggunakan Fleiss' Kappa.
8. Model yang dievaluasi pada penelitian ini dibatasi pada empat arsitektur deep learning, yaitu SimpleRNN-CRF, BiGRU-CRF, BiLSTM-CRF, dan Transformer Encoder-CRF. Penelitian ini tidak mengusulkan maupun mengembangkan arsitektur model baru.
9. Evaluasi performa model dilakukan menggunakan metrik *Exact Span Matching* dan *Partial Span Matching* yang diukur melalui *Precision, Recall,* dan *F1-Score.* Analisis kesalahan dilakukan berdasarkan *Confusion Matrix* dan *Error Analysis* untuk mengidentifikasi karakteristik prediksi model.
10. Eksperimen penambahan ukuran dataset dilakukan untuk menganalisis pengaruh jumlah data terhadap performa model. Penelitian ini tidak

membahas teknik augmentasi data sintetis maupun optimasi hyperparameter secara ekstensif sebagai fokus utama penelitian.

11. Seluruh proses pelatihan dan evaluasi model dilaksanakan menggunakan fasilitas *High Performance Computing* (HPC), sedangkan pembahasan penelitian difokuskan pada hasil ekstraksi aspek dan tidak membahas optimasi performa komputasi maupun arsitektur HPC.

1.6 Keaslian Penelitian

Penelitian mengenai ABSA dan NER telah banyak dilakukan pada berbagai domain, seperti ulasan produk, restoran, layanan, maupun media sosial. Selain itu, beberapa penelitian di Indonesia juga telah menerapkan analisis sentimen maupun ABSA pada isu-isu kebijakan publik. Meskipun demikian, berdasarkan studi literatur yang dilakukan, belum ditemukan penelitian yang memiliki ruang lingkup, metodologi, serta tujuan penelitian yang sama dengan penelitian ini.

Perbedaan penelitian ini dengan penelitian-penelitian sebelumnya tidak hanya terletak pada objek penelitian, tetapi juga pada pendekatan pembangunan dataset dan proses evaluasi model. Sebagian besar penelitian terdahulu berfokus pada pengembangan model klasifikasi sentimen atau penerapan satu hingga dua arsitektur *deep learning* pada dataset yang telah tersedia. Sementara itu, penelitian yang membangun dataset ABSA berbasis NER melalui penyusunan *Annotation Guideline*, validasi pakar menggunakan *Inter-Rater Reliability,* anotasi manual secara sistematis, serta evaluasi berbagai arsitektur *deep learning* pada dataset yang sama masih relatif terbatas, khususnya pada domain kebijakan publik berbahasa Indonesia.

Keaslian penelitian ini juga ditunjukkan melalui integrasi beberapa tahapan penelitian dalam satu kerangka kerja yang utuh, yaitu pembangunan dataset, validasi anotasi, implementasi empat arsitektur SimpleRNN-CRF, BiGRU-CRF, BiLSTM-CRF, dan Transformer Encoder-CRF, evaluasi menggunakan *Exact Span Matching* dan *Partial Span Matching,* serta eksperimen pengaruh penambahan ukuran dataset terhadap kemampuan generalisasi model. Kombinasi tahapan

tersebut belum ditemukan secara bersamaan pada penelitian-penelitian terdahulu yang menjadi rujukan dalam penelitian ini.

1.7 Manfaat Penelitian

Manfaat dari penelitian ini sesuai dengan batasan dan masalah yang ditemukan, antara lain:

a. Manfaat Praktis
    1. Bagi Peneliti

       Penelitian ini dapat menjadi referensi dalam pembangunan dataset ABSA berbasis NER, mulai dari proses penyusunan *Annotation Guideline*, validasi anotasi, hingga evaluasi model pada domain kebijakan publik maupun domain lainnya.
    2. Bagi Akademisi

       Dataset, metodologi, dan hasil evaluasi yang dihasilkan dapat dimanfaatkan sebagai referensi dalam pengembangan penelitian lanjutan pada bidang NLP, khususnya terkait ABSA, NER, *sequence labeling*, maupun evaluasi model *Deep Learning* pada bahasa Indonesia.
    3. Bagi Instansi Pemerintah

       Pendekatan yang diusulkan dapat menjadi salah satu alternatif dalam menganalisis opini masyarakat terhadap implementasi kebijakan publik secara lebih rinci, sehingga tidak hanya mengetahui kecenderungan sentimen masyarakat, tetapi juga mengidentifikasi aspek-aspek kebijakan yang memperoleh apresiasi maupun kritik sebagai bahan evaluasi dan perumusan kebijakan.
    4. Bagi Pengembang Sistem NLP

       Dataset beranotasi dan hasil evaluasi model yang dihasilkan dapat dimanfaatkan sebagai salah satu referensi dalam pengembangan sistem ekstraksi informasi maupun sistem analisis opini berbasis NER pada teks media sosial berbahasa Indonesia.
b. Manfaat Teoritis

Penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan ilmu pengetahuan, khususnya pada bidang NLP, ABSA dan NER. Kontribusi teoritis penelitian ini meliputi:

1. Menambah referensi ilmiah mengenai penerapan ABSA berbasis NER pada domain kebijakan publik berbahasa Indonesia, yang hingga saat ini masih relatif terbatas dibandingkan domain komersial seperti ulasan produk dan layanan.
2. Menyediakan dataset beranotasi yang dibangun melalui proses penyusunan *Annotation Guideline*, validasi pakar, pengukuran *Inter-Rater Reliability,* dan anotasi manual secara sistematis sehingga dapat menjadi salah satu sumber daya linguistik (*linguistic resources*) bagi penelitian NLP pada domain kebijakan publik.
3. Memberikan bukti empiris mengenai karakteristik performa arsitektur SimpleRNN-CRF, BiGRU-CRF, BiLSTM-CRF, dan Transformer Encoder-CRF pada tugas ekstraksi aspek berbasis NER menggunakan dataset yang dibangun melalui prosedur anotasi yang sama.
4. Pemahaman mengenai pengaruh ukuran dataset terhadap kemampuan generalisasi model pada tugas ekstraksi aspek berbasis NER sehingga dapat menjadi referensi dalam pengembangan penelitian selanjutnya.

1.8 Sistematika Penelitian

Penelitian tentang *Aspect-Based Sentiment Analysis* Terhadap Kebijakan Efisiensi Anggaran Pada Media Sosial X Menggunakan Pendekatan *Deep Learning* disusun dengan sistematika sebagai berikut:

BAB I: Pendahuluan

Bab ni menjelaskan latar belakang penelitian, rumusan masalah, pertanyaan penelitian, tujuan penelitian, batasan penelitian, keaslian penelitian, manfaat penelitian, serta sistematika penulisan tesis. Bab ini memberikan gambaran mengenai permasalahan penelitian, urgensi penelitian, *research gap*, dan kontribusi yang diharapkan dari penelitian.

BAB II: Tinjauan Pustaka

Bab ini menyajikan kajian pustaka yang menjadi landasan penelitian, meliputi teori-teori yang berkaitan dengan NLP, Opinion Mining, ABSA, NER) *sequence labeling*, *Conditional Random Field* (CRF), arsitektur SimpleRNN, BiGRU, BiLSTM, dan Transformer Encoder, serta penelitian-penelitian terdahulu yang relevan. Pada bab ini juga disajikan analisis *research gap* sebagai dasar penyusunan kerangka penelitian.

BAB III: Landasan Teori

Bab ini menjelaskan konsep-konsep teoritis yang digunakan dalam penelitian, meliputi teori evaluasi kebijakan publik William N. Dunn sebagai dasar penentuan aspek penelitian, konsep anotasi data, *Annotation Guideline, Inter-Rater Reliability, Fleiss' Kappa*, evaluasi model *menggunakan Exact Span Matching* dan *Partial Span Matching,* serta metrik evaluasi *Precision, Recall, dan F1-Score* sebagai dasar analisis performa model.

BAB IV: Metodologi Penelitian

Bab ini menjelaskan metode penelitian yang digunakan secara sistematis, mulai dari identifikasi permasalahan dan studi literatur, pengumpulan data dari media sosial X, penyusunan *Annotation Guideline*, proses anotasi menggunakan perangkat lunak Prodigy, validasi anotasi oleh pakar menggunakan Fleiss' Kappa, pembangunan dataset ABSA berbasis NER, persiapan data, implementasi model SimpleRNN-CRF, BiGRU-CRF, BiLSTM-CRF, dan Transformer Encoder-CRF, proses pelatihan model, evaluasi menggunakan *Exact Span Matching* dan *Partial Span Matching,* serta eksperimen penambahan ukuran dataset untuk menganalisis pengaruhnya terhadap performa model.

BAB V: Hasil Penelitian dan Pembahasan

Bab ini menyajikan hasil penelitian beserta pembahasannya, meliputi hasil pengumpulan data, pembangunan dataset, hasil anotasi dan validasi anotasi, analisis karakteristik dataset, implementasi dan pelatihan model, evaluasi performa

menggunakan *Exact Span Matching* dan *Partial Span Matching*, analisis *confusion matrix* dan *error analysis*, perbandingan performa keempat arsitektur *deep learning*, serta analisis pengaruh penambahan ukuran dataset terhadap kemampuan generalisasi model.

BAB VI: Penutup

Bab ini berisi kesimpulan yang diperoleh berdasarkan hasil penelitian serta saran yang dapat digunakan sebagai dasar pengembangan penelitian selanjutnya, baik terkait pengembangan dataset, metode ekstraksi aspek berbasis NER, maupun pengembangan model *deep learning* pada domain kebijakan publik.