

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pidato politik merupakan sarana penting bagi para pemimpin dan tokoh politik untuk menyampaikan visi, misi, serta kebijakan kepada publik. Melalui pidato, pesan-pesan strategis disampaikan dengan tujuan membangun citra, memengaruhi opini publik, dan memperoleh dukungan politik. Analisis terhadap pidato politik membantu mengungkap nilai dan cara komunikasi dari orator. Oleh karena itu, analisis terhadap pidato politik menjadi sangat penting untuk memahami strategi, maksud tersembunyi, dan dampak pidato terhadap masyarakat.

Pada perkembangan teknologi, analisis pidato politik dapat ditingkatkan menggunakan metode *Natural Language Processing* (NLP). Salah satu metode yang relevan untuk analisis teks adalah *Named Entity Recognition* (NER), atau pengenalan entitas bernama. NER merupakan teknik dalam NLP yang bertujuan mengidentifikasi dan mengklasifikasikan entitas bernama dalam teks, seperti nama orang, organisasi, lokasi, waktu, dan lainnya (Jurafsky & Martin, 2023). Melalui penerapan NER, analisis terhadap pidato politik dapat dilakukan secara lebih sistematis dan mendalam, memungkinkan identifikasi aktor, institusi, serta elemen penting lainnya yang disebutkan dalam pidato. Metode ini memberikan efisiensi dalam analisis teks politik yang sebelumnya dilakukan secara manual.

Penerapan NER dalam analisis pidato politik memiliki beberapa manfaat signifikan. NER memungkinkan identifikasi otomatis terhadap entitas penting yang sering disebut dalam pidato, sehingga mempermudah pemetaan jaringan sosial dan politik yang dibangun oleh orator (Collobert et al., 2011). Misalnya, penyebutan nama tokoh atau organisasi tertentu dalam pidato dapat mencerminkan hubungan politik, aliansi, atau bahkan konflik yang ingin disampaikan secara implisit. Pengenalan entitas utama memungkinkan analisis terhadap prioritas kebijakan atau pesan inti yang ingin disampaikan kepada publik. Selain itu, NER mempermudah identifikasi bias atau kecenderungan dalam pidato melalui analisis frekuensi penyebutan entitas tertentu, yang

berpotensi menunjukkan agenda politik tersembunyi.

Beberapa penelitian sebelumnya telah menunjukkan efektivitas NER dalam analisis teks berbahasa Indonesia. Permana & Purnamasari (2019) menerapkan metode *Bidirectional LSTM* untuk pengenalan entitas bernama dalam teks bahasa Indonesia, yang mencapai akurasi sebesar 89,5%. Namun, penerapan NER secara spesifik untuk analisis pidato politik masih terbatas, baik dalam jumlah studi maupun implementasinya pada konteks politik Indonesia.

Pemilihan metode NER dalam penelitian ini didasarkan pada statusnya sebagai salah satu teknik *state-of-the-art* (SOTA) dalam NLP. Dengan berkembangnya model berbasis *Transformer* seperti BERT (*Bidirectional Encoder Representations from Transformers*), kinerja NER telah meningkat secara signifikan dalam berbagai tugas pengenalan entitas, termasuk dalam bahasa yang kurang terwakili seperti bahasa Indonesia. Model berbasis *Transformer* mampu menangkap konteks semantik yang kompleks melalui representasi *bidirectional*. Hal ini meningkatkan akurasi identifikasi entitas dan memungkinkan analisis relasi antar entitas dalam teks politik. Hal ini membuat NER sangat cocok untuk diaplikasikan dalam analisis pidato politik yang sering kali sarat dengan elemen-elemen semantik dan pragmatik yang kompleks.

IndoBERT, adaptasi dari BERT untuk bahasa Indonesia. Menurut Wilie et al. (2020), IndoBERT mencapai *F1-score* sebesar 81,9% pada tugas NER untuk teks berbahasa Indonesia, mengungguli model-model sebelumnya seperti *Bidirectional LSTM*. Keunggulan ini didukung oleh kemampuan IndoBERT dalam menangkap konteks semantik yang lebih mendalam dan kompleks, terutama dalam teks yang mengandung banyak elemen budaya dan politik lokal.

Penelitian ini akan menggunakan dataset pidato politik dari sumber resmi, seperti arsip pidato presiden atau tokoh politik lainnya. Dataset dianalisis menggunakan NER berbasis *Transformer*, yakni IndoBERT. Model terbukti efektif dalam berbagai tugas NLP karena kemampuannya menangkap konteks semantik secara mendalam (Devlin et al., 2019). Selain itu, pendekatan berbasis *transfer learning* akan diterapkan untuk meningkatkan akurasi analisis pada teks berbahasa Indonesia. Kombinasi metode ini diharapkan dapat memberikan hasil yang lebih akurat dibandingkan penelitian sebelumnya yang menggunakan

metode tradisional. Pada penelitian ini akan menggunakan model *cahya/bert-base-indonesian-NER*.

Kontribusi penelitian ini juga terletak pada relevansinya terhadap konteks politik Indonesia. Analisis terhadap pidato politik di Indonesia dapat memberikan wawasan baru mengenai dinamika politik lokal, termasuk cara tokoh politik menyampaikan pesan, membangun citra, dan memengaruhi publik. Sebagai contoh, penelitian Ekayanta (2022) mengungkapkan bahwa pidato kenegaraan sering kali digunakan sebagai alat untuk membentuk opini publik dan memperkuat legitimasi politik. Dengan menerapkan NER, penelitian ini diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai lanskap politik Indonesia, termasuk identifikasi pola komunikasi yang digunakan oleh tokoh politik.

1.2 Rumusan Masalah

Berikut rumusan masalah:

1. Bagaimana kinerja tugas *Named Entity Recognition* (NER) berbasis IndoBERT dalam mengidentifikasi entitas penting (seperti nama tokoh, organisasi, lokasi, waktu, dan angka *cardinal*) pada transkrip pidato politik berbahasa Indonesia?
2. Bagaimana distribusi entitas penting (seperti nama tokoh, organisasi, lokasi, dan waktu) yang teridentifikasi dalam transkrip pidato politik?

1.3 Tujuan Penelitian

Tujuan penelitian ini meliputi:

1. Mengetahui hasil kinerja *Named Entity Recognition* (NER) berbasis IndoBERT untuk mengidentifikasi entitas bernama dalam pidato politik, seperti nama tokoh, organisasi, lokasi, dan waktu, sehingga dapat memberikan analisis yang sistematis.
2. Mengetahui distribusi entitas penting (seperti nama tokoh, organisasi, lokasi, dan waktu) yang teridentifikasi dalam transkrip pidato politik.

1.4 Manfaat Penelitian

Berikut beberapa manfaat dari penelitian ini yang terbagi menjadi manfaat teoritis dan manfaat praktis:

Manfaat Teoritis

1. Memberikan kontribusi pada pengembangan NLP, khususnya dalam penerapan NER berbasis *Transformer* pada teks berbahasa Indonesia.
2. Memperluas wawasan dalam bidang analisis komunikasi politik dengan pendekatan teknologi NLP modern, seperti analisis menggunakan tugas NER.

Manfaat Praktis

1. Memberikan alat analisis yang efisien dan akurat bagi peneliti, akademisi, atau praktisi komunikasi politik dalam memahami isi pidato politik.
2. Memberikan wawasan baru terkait pola komunikasi dan hubungan sosial-politik yang terkandung dalam pidato politik di Indonesia.

1.5 Batasan Penelitian

Pada penelitian ini terdapat batasan, yang meliputi:

1. Penelitian ini hanya menggunakan transkrip video pidato resmi Presiden Joko Widodo, berjumlah 186 video yang diterbitkan pada kanal resmi YouTube Sekretariat Kabinet Republik Indonesia secara publik pada periode 2014–2024 dengan kategori pidato pers. Data pidato yang tidak tersedia secara publik atau tidak terdokumentasi dengan baik, tidak dianalisis.
2. Penelitian hanya menggunakan *Named Entity Recognition* (NER) untuk analisis entitas penting. Analisis lebih lanjut seperti sentimen atau wacana tidak termasuk dalam ruang lingkup penelitian.
3. *Labeling* data dilakukan dengan menggunakan *library Hugging Face Transformers*, yang menyediakan model IndoBERT dan tokenizer untuk tugas NER.
4. Proses *preprocessing* tidak melibatkan *stemming* dan *lemmatization* karena dapat mengganggu konteks identifikasi entitas dalam tugas NER. Selain itu,

data tidak dibagi berdasarkan kalimat, melainkan berdasarkan satuan dokumen (per pidato).

5. Fokus penelitian adalah pada entitas yang umum diidentifikasi oleh NER yaitu: PER, ORG, LOC, DATE, CRD. Entitas abstrak dan ambigu tidak dianalisis.
6. Hasil interpretasi jaringan entitas dibatasi pada keterkaitan aktor dan isu yang teridentifikasi melalui data pidato, tanpa menyimpulkan niat atau maksud politis tertentu.
7. Penelitian ini mengandalkan kemampuan NER dalam bahasa Indonesia. Ke tidak sempurnaan alat NER dalam menangani variasi bahasa dan konteks dapat memengaruhi hasil analisis.

