

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi informasi dan komunikasi saat ini telah mengubah cara masyarakat berkomunikasi dan memperoleh informasi, terutama di bidang politik. Platform YouTube menjadi salah satu media utama untuk menyampaikan informasi politik, termasuk pidato dari tokoh-tokoh politik. Di era digital seperti sekarang ini, pidato politik tidak hanya dapat dilihat dan didengar secara langsung oleh publik tetapi juga sering menjadi objek diskusi di kolom komentar. Komentar-komentar ini mencerminkan pandangan, opini, dan sentimen publik terhadap isi pidato, sehingga analisis terhadap komentar ini menjadi penting untuk memahami bagaimana pesan politik diterima dan dimaknai oleh masyarakat (Liu, 2020).

Analisis sentimen adalah metode yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini atau emosi yang terkandung dalam teks, seperti positif, negatif, atau netral. Dalam konteks politik, analisis sentimen dapat memberikan wawasan penting bagi para politisi, tim kampanye, dan peneliti untuk memahami respons masyarakat terhadap pidato atau kebijakan tertentu. Sentimen ini dapat berfungsi sebagai indikator kepuasan atau ketidakpuasan publik terhadap tokoh atau kebijakan politik tertentu (Pang & Lee, 2008). Namun, tantangan utama dalam analisis sentimen adalah menangkap kompleksitas bahasa, termasuk konteks, ironi, dan slang yang sering muncul dalam komentar di media sosial (Davidov, 2010).

Komentar-komentar yang ada pada video pidato politik di YouTube sering kali memiliki arti yang sangat kompleks, yang melibatkan ironi, sindiran, atau bahkan sindiran sosial yang sulit dipahami tanpa mengetahui konteksnya. Sebagai contoh, komentar seperti "Wah, pidato ini bener-bener 'memukau'... kalau bisa tidur, pasti tidur!" mungkin secara literal terdengar positif, namun sebenarnya menunjukkan ketidakpuasan terhadap pidato yang disampaikan. Dalam kasus seperti ini, analisis sentimen tradisional yang hanya mengandalkan kata-kata secara terpisah akan kesulitan untuk mengidentifikasi makna sesungguhnya, yang sebenarnya adalah negatif. Oleh karena itu, penting untuk mengembangkan teknik yang dapat menangkap maksud dari konteks ini, yang bisa berupa ironi, humor, atau sindiran.

Salah satu pendekatan yang populer dalam analisis sentimen adalah penggunaan *word embedding*, yaitu metode representasi kata dalam bentuk vektor numerik yang mempertimbangkan hubungan semantik antar kata. Teknik seperti Word2Vec (Mikolov et al., 2013), GloVe (Speer & Chin, 2016), dan FastText (Bojanowski et al., 2017b) telah menunjukkan hasil yang baik dalam berbagai aplikasi pemrosesan bahasa alami (Natural Language Processing, NLP). Dengan menggunakan *word embedding*, model dapat memahami hubungan antar kata dalam kalimat, serta menangkap makna yang lebih mendalam dari kata-kata yang digunakan dalam konteks tertentu, seperti dalam komentar-komentar di YouTube.

Pendekatan *word embedding* memungkinkan analisis sentimen untuk lebih fleksibel dalam menangani berbagai variasi bahasa, termasuk slang, kode campuran (code-mixing), dan penggunaan kata-kata yang tidak baku. Dengan mengubah kata-kata menjadi vektor, teknik ini memungkinkan model untuk menangkap hubungan semantik yang lebih kompleks, serta memahami bagaimana kata-kata berinteraksi dalam kalimat. Sebagai contoh, kata "hebat" dalam konteks pidato politik dapat diartikan secara positif, tetapi ketika digunakan dalam kalimat yang lebih kompleks, seperti "pidato ini hebat kalau dilihat dari segi tidur," dapat diartikan negatif, sesuatu yang lebih mudah ditangkap menggunakan teknik *word embedding* dibandingkan dengan metode konvensional yang lebih sederhana.

Penelitian ini bertujuan untuk menganalisis sentimen komentar publik terhadap pidato politik berbahasa Indonesia di YouTube dengan menggunakan metode *word embedding*. Pemilihan YouTube sebagai sumber data didasarkan pada popularitas platform ini sebagai media diskusi publik yang sangat berpengaruh dalam membentuk opini masyarakat (Bärtl, 2018). Data komentar pada video pidato politik berisi beragam pandangan masyarakat, sehingga cocok untuk analisis.

Selain itu, penelitian ini juga mencoba menjawab tantangan yang muncul dalam pengolahan bahasa Indonesia, seperti slang, campuran bahasa (code-mixing), dan penggunaan kata-kata tidak baku. Dengan memanfaatkan teknik *word embedding* yaitu FastText, penelitian ini diharapkan dapat memberikan kontribusi dalam memahami respons publik terhadap isu-isu politik serta meningkatkan akurasi dan efisiensi dalam analisis sentimen berbahasa Indonesia.

Beberapa penelitian sebelumnya telah membahas analisis sentimen terhadap isu-isu politik menggunakan pendekatan berbasis *machine learning* dan teknik *word embedding* seperti Word2Vec, GloVe, dan FastText. Penelitian oleh (Putri et al., 2024) meneliti klasifikasi sentimen terhadap isu pemindahan ibu kota Indonesia menggunakan model CNN dengan representasi teks berbasis GloVe dan FastText. Sementara itu, (Alfariqi et al., 2020) melakukan klasifikasi sentimen pada data Twitter menggunakan pendekatan CNN dan FastText, dan menunjukkan bahwa FastText memiliki performa yang efektif dalam menangani teks berbahasa Indonesia.

Namun, kedua penelitian tersebut belum secara khusus mengkaji sentimen dalam konteks komentar publik terhadap video pidato politik di YouTube, yang umumnya mengandung kompleksitas bahasa seperti *slang*, ironi, dan *code-mixing*, serta memerlukan pemahaman konteks yang lebih mendalam. Oleh karena itu, penelitian ini bertujuan untuk mengisi celah tersebut dengan menerapkan FastText dalam analisis sentimen komentar publik terhadap pidato politik berbahasa Indonesia di YouTube.

Hasil dari penelitian ini diharapkan tidak hanya bermanfaat dalam konteks akademik, tetapi juga memiliki manfaat langsung. Misalnya, para pengambil kebijakan dan tim komunikasi politik dapat menggunakan wawasan dari analisis sentimen ini untuk menyusun strategi komunikasi yang lebih efektif dan responsif terhadap opini publik. Selain itu, pendekatan metodologi yang digunakan dalam penelitian ini dapat menjadi acuan bagi penelitian selanjutnya yang berfokus pada analisis sentimen dalam bahasa Indonesia atau konteks politik lainnya.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah ditulis, maka dapat dituliskan rumusan masalah sebagai berikut :

1. Bagaimana distribusi sentimen (positif, negatif, netral) yang terdapat dalam komentar pengguna YouTube terhadap video pidato politik berbahasa Indonesia?
2. Bagaimana kinerja FastText dalam analisis sentimen komentar berbahasa Indonesia pada platform YouTube?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini selain menerapkan dan mengetahui hasil implementasi menggunakan FastText, juga bertujuan untuk :

1. Mengidentifikasi dan mengklasifikasikan sentimen komentar pengguna terhadap video pidato politik, baik yang bersifat positif, negatif, maupun netral.
2. Mengevaluasi sejauh mana FastText dapat meningkatkan kinerja dalam mengidentifikasi sentimen komentar yang lebih tepat dibandingkan dengan metode analisis sentimen tradisional.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut :

1. Dengan menggunakan FastText, penelitian ini dapat meningkatkan akurasi dalam analisis sentimen dibandingkan dengan metode tradisional, memberikan hasil yang lebih dapat diandalkan.
2. Hasil dari penelitian ini dapat digunakan untuk mengevaluasi sejauh mana FastText dapat meningkatkan akurasi dalam mengidentifikasi sentimen komentar yang lebih tepat dibandingkan dengan metode analisis sentimen tradisional.
3. Hasil penelitian ini dapat menjadi referensi bagi penelitian-penelitian selanjutnya yang berkaitan dengan analisis sentimen, baik di bidang politik maupun bidang lainnya, serta mendorong penelitian lebih lanjut dalam konteks media sosial.

1.5 Batasan Penelitian

Untuk menghindari suatu permasalahan dalam penelitian dalam skala yang lebih luas maka perlu adanya batasan masalah. Batasan masalah dalam penelitian ini adalah :

1. Penelitian ini hanya menggunakan data komentar dari video pidato politik tentang keterangan pers Presiden Joko Widodo yang diunggah di kanal YouTube Sekretariat Presiden dan Sekretariat Kabinet RI selama 10 tahun yaitu dari tahun 2014-2024 sebanyak 12.102 komentar sebagai sumber data.

2. Data yang digunakan terbatas pada komentar berbahasa Indonesia, sehingga hasil penelitian tidak dapat digeneralisasi untuk komentar dalam bahasa lain.
3. Dalam penelitian ini, metode pelabelan otomatis menggunakan model MDHugol yang diakses melalui pustaka *transformers* dari Hugging Face.
4. Analisis sentimen dilakukan dengan klasifikasi sentimen dasar (positif, negatif, atau netral)
5. Model yang dibangun dalam penelitian ini dievaluasi berdasarkan metrik standar seperti akurasi, presisi, *recall*, dan F1-score.

