

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pemilihan umum presiden dan wakil presiden (Pilpres) menjadi momen krusial yang mencerminkan dinamika politik serta aspirasi masyarakat terhadap proses demokrasi (Prakoso et al., 2023). Perkembangan pesat media sosial memberi kesempatan bagi pengguna untuk bebas mengutarakan opini. Analisis opini publik terhadap masing-masing calon pasangan menjadi kunci penting untuk memahami bagaimana pandangan masyarakat serta dinamika politik yang terjadi pada masa Pilpres (Indrawan et al., 2020).

Analisis opini publik dapat dilakukan di berbagai platform media sosial. Platform media sosial X menjadi salah satu platform yang sering digunakan untuk analisis ini. Hal ini karena platform media sosial X memiliki kemudahan dalam mendapatkan data cuitan atau *tweet* pengguna serta memiliki total pengguna di Indonesia yang mencapai 79 juta pada Januari tahun 2024 (Databooks, 2024). Volume data yang besar dan tidak terstruktur di X, serta kompleksitas sentimen yang terkandung di dalamnya, menghadirkan tantangan tersendiri dalam analisis opini publik. Pemanfaatan pendekatan seperti *network analysis*, *machine learning*, hingga *deep learning* menjadi sangat krusial untuk memproses dan memahami data ini dengan efisien dan efektif (Bello-Orgaz et al., 2016).

Berbagai macam metode *text mining* telah dikembangkan untuk melakukan analisis opini publik seperti analisis sentimen dan pemodelan topik. Analisis sentimen bertujuan untuk mengelompokkan *tweet* menjadi beberapa sentimen seperti positif, negatif atau netral (Yavari et al., 2022). Sedangkan pemodelan topik bertujuan untuk memahami isu-isu yang mendominasi percakapan, mengidentifikasi agenda kampanye yang diusung oleh masing-masing kubu, serta juga dapat memonitor pergeseran fokus diskusi seiring berjalannya waktu.

Latent Dirichlet Allocation (LDA) merupakan sebuah model probabilistik yang digunakan untuk mengidentifikasi topik-topik yang tersembunyi dalam sebuah kumpulan dokumen teks. LDA berasumsi bahwa setiap dokumen terdiri dari beberapa topik, dan setiap topik dibentuk oleh distribusi kata-kata tertentu (Blei, 2003) . Melalui analisis pola kemunculan kata-kata dalam dokumen, LDA dapat digunakan untuk mengidentifikasi topik-topik yang mendasari dokumen-dokumen tersebut.

Robustly Optimized BERT Pretraining Approach (RoBERTa) merupakan sebuah model *deep learning* berbasis *Transformer* yang telah menunjukkan performa *state-of-the-art* dalam berbagai tugas pemrosesan bahasa alami, termasuk analisis sentimen (Liu et al., 2019). Model ini dilatih dengan menggunakan pendekatan *pre-training* yang lebih *robust* dan dataset yang lebih besar dibandingkan dengan model *Bidirectional Encoder Representation from Transformer* (BERT) sebelumnya (Liu et al., 2019). RoBERTa menggunakan konsep *self-attention* dalam memahami struktur dan makna dari data sehingga

model ini mampu menangkap konteks dan nuansa bahasa secara lebih baik dan menghasilkan prediksi sentimen yang lebih akurat (Liu et al., 2019).

Penelitian terdahulu oleh Angraini dan Lhaksana (2024) menggunakan metode *Long Short Term Memory* (LSTM) dan *Convolutional Neural Network* (CNN) untuk memprediksi sentimen pemilihan legislatif pada platform X. Penelitian ini menunjukkan bahwa CNN merupakan model terbaik dengan akurasi sebesar 93.27%. Penelitian mengenai analisis sentimen juga dilakukan oleh Mantika et al. (2024) menggunakan metode *Naive Bayes* dan *Logistic Regression* dalam melakukan analisis sentimen Pilpres 2024. Penelitian ini mendapatkan model *Naive Bayes* adalah model terbaik dengan akurasi sebesar 63%. Penelitian oleh Wisnu et al. (2020) menggunakan metode *Support Vector Machine* (SVM) dan *Latent Dirichlet Allocation* (LDA) dalam menganalisis sentimen dan pemodelan topik pada data pemilihan gubernur Jawa Tengah. Penelitian ini memperoleh akurasi pada model SVM sebesar 92.9% dan mendapatkan sepuluh topik pada masing-masing kandidat menggunakan metode LDA.

Penelitian ini berfokus dalam mengkombinasikan pemodelan topik dan analisis sentimen pada data *tweet* Pilpres tahun 2024. Berdasarkan penelitian sebelumnya, kombinasi pemodelan topik dan analisis sentimen mampu memberikan hasil yang lebih akurat dan mendalam dibandingkan dengan menggunakan satu metode. Penelitian ini menggunakan metode *Latent Dirichlet Allocation* (LDA) yang merupakan sebuah metode yang mampu mengidentifikasi topik-topik dalam kumpulan dokumen teks untuk pemodelan topik. Analisis sentimen pada penelitian ini menggunakan *Robustly Bidirectional Encoder for Transformers approach* (RoBERTa) yang merupakan model *deep learning* berbasis

Transformer yang telah menunjukkan performa *state-of-the-art* dalam berbagai tugas pemrosesan bahasa alami, termasuk analisis sentimen. RoBERTa digunakan karena memiliki performa yang unggul dalam pemrosesan data teks (Liu et al., 2019). Selain itu, model ini juga dapat memiliki *pre-training* bahasa Indonesia sehingga proses *training* dan jumlah data yang dibutuhkan menjadi lebih efisien (Widarmanti et al., 2022).

Melalui penggabungan LDA dan RoBERTa, penelitian ini diharapkan dapat memberikan kontribusi metodologis dalam analisis media sosial dan politik, serta memberikan wawasan berharga bagi praktisi politik, pengamat, dan pembuat kebijakan. Lebih jauh lagi, penelitian ini diharapkan dapat meningkatkan transparansi dan partisipasi masyarakat dalam proses demokrasi dengan memberikan pemahaman yang lebih baik tentang opini publik terkait Pilpres 2024.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka terdapat beberapa permasalahan yang akan dibahas dalam penelitian ini :

1. Bagaimana distribusi *tweet* pada masing-masing pasangan calon presiden dan wakil presiden?
2. Bagaimana penerapan metode LDA untuk mengidentifikasi topik-topik dominan yang dibicarakan terkait Pilpres 2024 pada masing-masing pasangan calon di media sosial X?
3. Bagaimana hasil analisis menggunakan metode RoBERTa untuk distribusi sentimen (positif, negatif, netral) terhadap masing-masing topik dan pasangan calon dalam Pilpres 2024?

4. Bagaimana pergeseran topik atau perubahan sentimen terhadap calon pasangan tertentu seiring berjalannya waktu kampanye Pilpres 2024 berdasarkan analisis LDA dan RoBERTa?

1.3 Batasan Masalah

Penelitian ini terbatas pada data berupa cuitan di media sosial X dengan topik yang relevan dengan pemilihan presiden 2024, seperti program kerja, debat, kampanye, atau isu-isu terkait. Data dikumpulkan dari tanggal 31 Desember 2023 hingga 31 Januari 2024 melalui API X, dengan fokus analisis pada konten teks dari *tweet* tersebut. Penelitian ini hanya berfokus pada penerapan metode LDA untuk mengidentifikasi topik-topik utama yang dibicarakan, dan RoBERTa untuk menganalisis sentimen dari *tweet* terkait topik-topik tersebut. Data yang digunakan adalah data berbahasa Indonesia, dengan toleransi terhadap beberapa tambahan kalimat berbahasa Inggris di dalamnya.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai pada penelitian ini adalah sebagai berikut:

1. Menganalisis distribusi *tweet* pada masing-masing pasangan calon presiden dan wakil presiden;
2. Mengidentifikasi topik-topik dominan yang dibicarakan terkait Pilpres 2024 pada masing-masing pasangan calon di media sosial X menggunakan metode LDA;
3. Mengidentifikasi distribusi sentimen (positif, negatif, netral) terhadap masing masing topik dan pasangan calon dalam Pilpres 2024 yang dianalisis menggunakan RoBERTa;

4. Mengevaluasi pergeseran topik atau perubahan sentimen terhadap calon pasangan tertentu seiring berjalannya waktu kampanye Pilpres 2024 berdasarkan analisis LDA dan RoBERTa.