

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi informasi dan komunikasi di era digital kini telah mengubah cara masyarakat dalam mengungkapkan pendapat, kritik, dan aspirasi, terutama melalui platform media sosial (Siahaan, 2024). X telah menjadi salah satu platform terpopuler untuk menyampaikan pendapat publik dengan cepat dan dalam waktu nyata yang saat ini disebut dengan sosial media X. Dengan banyaknya pengguna aktif di Indonesia, X sering kali berfungsi sebagai tempat untuk diskusi publik, terutama mengenai isu politik, pemerintahan, dan pelaksanaan pemilu. Data yang diperoleh dari X bersifat tidak teratur, berjumlah besar, dan memiliki tingkat perubahan yang tinggi, sehingga membutuhkan metode tertentu untuk bisa dianalisis dengan efektif (Alita et al., 2020).

Komisi Pemilihan Umum (KPU) selaku badan negara yang mengorganisir pemilihan umum memiliki fungsi vital dalam memastikan transparansi, kredibilitas, dan kepercayaan masyarakat. Namun, pada setiap pelaksanaan pemilu, KPU sering kali menjadi perhatian publik, baik melalui kritik, pujian, maupun penyebaran isu-isu tertentu. Pendapat masyarakat yang berkembang di media sosial bisa mencerminkan pandangan masyarakat tentang kinerja KPU, sehingga penting untuk dipetakan guna memahami isu-isu yang menjadi fokus utama. Oleh karena itu, analisis dialog di X tentang KPU dapat memberikan pandangan berharga bagi akademisi, pembuat kebijakan, serta KPU sendiri untuk memahami persepsi masyarakat dengan lebih mendalam (Fuad Amirullah et al., 2023).

Pemrosesan Bahasa Alami (*Natural Language Processing*/NLP) adalah cabang dari kecerdasan buatan yang bertujuan untuk mempelajari interaksi antara komputer dengan bahasa manusia. Salah satu teknik dalam NLP yang terkait dengan pemahaman percakapan publik adalah pemodelan topik. *Topic Modeling* merupakan metode *unsupervised learning* yang bisa menemukan serta mengelompokkan dokumen sesuai dengan topik-topik tersembunyi yang ada di dalamnya (Campbell et al., 2015). Metode ini sering digunakan dalam bidang Pemrosesan Bahasa Alami (NLP) dan Pencarian

Informasi. Metode ini memfasilitasi analisis data teks dalam jumlah besar secara otomatis dan menghasilkan kelompok topik yang biasanya direpresentasikan sebagai sekumpulan kata yang sering muncul bersama dalam tema tertentu(Wahyuni et al., 2025).

Latent Dirichlet Allocation (LDA) adalah pemodelan probabilistik yang membantu menemukan topik dalam kumpulan data. (Bagas Dwi Santosa et al., 2023). *LDA* adalah model hirarki Bayesian yang dapat digunakan untuk menampilkan model campuran dari berbagai topik dalam kumpulan data teks. Secara umum, *LDA* bekerja dengan memasukkan koleksi dokumen dengan parameter tertentu. Kemudian, metode *LDA* digunakan untuk menghasilkan model yang terdiri dari bobot yang dapat diubah menjadi probabilitas melalui normalisasi. Pemodelan topik *LDA* telah terbukti efektif dalam menemukan topik dominan dalam dokumen. Karena model *LDA* tradisional hanya mempertimbangkan kata-kata tunggal, informasi kontekstual penting dalam kalimat dapat hilang(Nurhaliza et al., 2023).

Selain *LDA*, metode lain yang lebih modern seperti *BERTopic* memanfaatkan *embedding* berbasis transformer sehingga mampu menghasilkan topik yang lebih koheren, khususnya pada teks pendek seperti komentar (Grootendorst, 2022). Dengan banyaknya opini dan diskusi masyarakat mengenai Pemilu Indonesia tahun 2024 di X, hal tersebut memberikan banyak topik yang dibicarakan oleh masyarakat. Analisis topik memungkinkan pemahaman lebih mendalam tentang opini dan sentimen publik terkait calon-calon, isu-isu, dan perkembangan Pemilu, yang dapat membantu kandidat dan partai politik merancang strategi kampanye yang lebih efektif. Selain itu, analisis topik dapat membantu dalam memahami preferensi dan keprihatinan pemilih muda, yang memiliki potensi pengaruh besar dalam Pemilu. Analisis topik dapat dilakukan menggunakan beberapa metode *Text Mining*, salah satunya adalah *LDA* untuk melihat topik apa saja yang banyak dibahas oleh masyarakat tentang Pemilu tahun 2024.

Studi tentang analisis percakapan publik dengan *Topic Modeling* sebelumnya telah dilakukan dalam berbagai konteks, seperti pemilu, kebijakan publik, dan masalah kesehatan. Namun, ada sedikit penelitian yang secara khusus memperhatikan persepsi publik terhadap KPU melalui data X. Oleh karena itu, penting untuk melakukan penelitian ini untuk mengidentifikasi masalah utama yang terkait dengan KPU. Dengan demikian,

penelitian ini dapat memberikan kontribusi ilmiah dalam bidang pengolahan bahasa dan manfaat praktis bagi lembaga yang terkait. Diharapkan hasil penelitian ini dapat menunjukkan topik-topik utama yang berkembang di masyarakat tentang KPU, memberikan gambaran umum tentang persepsi publik tentang KPU, serta dapat melihat perbandingan kinerja dari masing-masing metode, dan dapat melihat mana metode yang lebih baik cara kerjanya serta menjadi dasar untuk penelitian lebih lanjut dalam bidang seperti analisis sentimen, analisis jaringan, dan komunikasi politik.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, terdapat beberapa permasalahan utama yang perlu dianalisis dalam penelitian ini, yaitu:

1. Bagaimana implementasi model *LDA* dan *BERTopic* untuk identifikasi topik-topik percakapan publik mengenai KPU pada platform X?
2. Bagaimana kinerja *LDA* dan *BERTopic* berdasarkan hasil evaluasi koherensi dan interpretabilitas topik dalam menganalisis percakapan publik mengenai KPU?
3. Bagaimana metode *LDA* dan *BERTopic* mengidentifikasi topik-topik percakapan publik mengenai KPU pada platform X?

1.3 Tujuan

Penelitian ini bertujuan untuk:

1. Mengimplementasikan *LDA* dan *BERTopic* untuk melakukan analisis topik pada data X terkait KPU
2. Mengevaluasi kinerja *LDA* dan *BERTopic* berdasarkan nilai koherensi dan interpretasi topik dalam analisis percakapan publik mengenai KPU pada platform X.
3. Menganalisis dan membandingkan hasil identifikasi topik yang dihasilkan oleh *LDA* dan *BERTopic* pada data percakapan publik mengenai KPU.

1.4 Manfaat

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Penelitian ini memberikan kontribusi dalam pengembangan ilmu di bidang Pemrosesan Bahasa Alami (NLP), khususnya melalui penerapan dan perbandingan metode *LDA* dan *BERTopic*, sehingga dapat menjadi referensi bagi penelitian selanjutnya yang berfokus pada analisis percakapan publik di media sosial.
2. Hasil penelitian ini bermanfaat bagi KPU untuk memahami opini, kritik, dan isu utama yang berkembang di masyarakat, sehingga dapat digunakan sebagai bahan evaluasi dalam meningkatkan transparansi, kredibilitas, serta kepercayaan publik terhadap penyelenggaraan pemilu.
3. Penelitian ini dapat membantu pembuat kebijakan, kandidat, maupun partai politik dalam merumuskan strategi komunikasi politik yang lebih efektif dengan memahami isu-isu dominan yang menjadi perhatian masyarakat, serta dapat menjadi dasar pengembangan sistem analisis opini publik berbasis big data di masa depan.

1.5 Ruang Lingkup Masalah

Agar penelitian ini memiliki fokus yang jelas dan terarah, maka ruang lingkup penelitian dibatasi sebagai berikut:

1. Data yang digunakan dalam penelitian ini berasal dari percakapan publik di media sosial X yang memuat kata kunci terkait KPU dan Pemilu 2024.
2. Pengumpulan data dibatasi pada periode waktu tertentu yang relevan dengan pelaksanaan Pemilu 2024, sehingga data yang dianalisis benar-benar mencerminkan isu publik pada saat itu.
3. Metode pemodelan topik yang digunakan adalah *LDA* sebagai metode klasik berbasis probabilistik dan *BERTopic* sebagai metode modern berbasis transformer. Kedua metode ini dibandingkan untuk melihat kelebihan dan keterbatasannya dalam menganalisis teks pendek seperti komentar.

1.6 Kebaharuan

Kebaharuan penelitian ini terletak pada konteks, karakteristik data, dan pendekatan evaluasi yang digunakan. Berbeda dengan penelitian terdahulu yang umumnya menggunakan data berbahasa Inggris dan berfokus pada isu global seperti pariwisata, COVID-19, atau politik luar negeri, penelitian ini secara khusus menganalisis percakapan publik mengenai KPU di platform X dalam konteks sosial-politik Indonesia. Fokus kelembagaan nasional ini masih jarang dikaji dalam studi *Topic Modeling*. Dari sisi data, penelitian ini menggunakan teks pendek berbahasa Indonesia yang bersifat informal dan dinamis, sehingga menghadirkan tantangan linguistik yang berbeda dibandingkan data formal seperti judul berita atau artikel. Hal ini memberikan kontribusi empiris baru terkait penerapan *LDA* dan *BERTopic* pada karakteristik bahasa media sosial Indonesia.

Selain itu, kebaruan juga terletak pada pendekatan evaluasi, yaitu dengan mengombinasikan pengukuran koherensi topik secara kuantitatif dan analisis interpretatif untuk memahami keterkaitan topik dengan persepsi publik terhadap KPU. Dengan demikian, penelitian ini tidak hanya membandingkan performa teknis model, tetapi juga memberikan pemaknaan substantif terhadap wacana publik yang terbentuk. Secara keseluruhan, penelitian ini memberikan kontribusi baru dalam kajian *Topic Modeling* pada data media sosial Indonesia serta memperluas pemanfaatannya untuk memahami opini publik terhadap lembaga penyelenggara pemilu.