

BAB 1 PENDAHULUAN

1.1. Latar Belakang

Khususnya di Jakarta, Pemilihan Kepala Daerah (Pilkada) merupakan titik penting dalam sistem politik Indonesia. Situasi sosial, politik, dan ekonomi masyarakat tercermin dalam Pilkada, yang lebih dari sekadar proses pemilihan pemimpin. Mengingat bahwa X dan platform media sosial serupa merupakan sarana utama bagi masyarakat untuk mengekspresikan ide dan keyakinan mereka, maka sangat penting untuk memahami perasaan masyarakat pada Pilkada Jakarta 2024 [3]. Dengan basis pengguna yang terus berkembang, X adalah tambang emas informasi tentang pemikiran dan perasaan orang yang tidak terekspresikan tentang berbagai topik, termasuk Pilkada Jakarta 2024 [28].

Pentingnya memahami proses pembentukan opini publik dalam kerangka demokrasi lokal, khususnya di Jakarta, membuat penelitian ini menjadi hal yang mendesak. Kebijakan publik, pertumbuhan kota, dan kesejahteraan masyarakat akan sangat terpengaruh oleh Pilkada Jakarta 2024. Reaksi publik terhadap kandidat, rencana, dan kebijakan kampanye Pilkada dapat diperoleh dari pemeriksaan opini publik. Penilaian pemilih selama pemilihan dan dalam menilai kinerja pejabat terpilih dipengaruhi oleh sikap masyarakat, oleh karena itu hal ini menjadi signifikan. Selain itu, para pemangku kepentingan dapat menggunakan analisis sentimen untuk mengukur prioritas masyarakat dan menggunakan informasi tersebut untuk menginformasikan upaya-upaya di masa depan untuk meningkatkan kebijakan dan layanan publik.

Beberapa penelitian sebelumnya sudah mencoba untuk menganalisis sentimen di media sosial, terutama menggunakan *deep learning*. Misalnya, Mollah (2022) mengembangkan model LSTM untuk analisis sentimen di *Twitter*, yang mampu memahami emosional dalam teks dengan baik [1]. Murthy, et. al., (2020) juga menggunakan LSTM untuk menganalisis sentimen dalam teks, dengan akurasi yang tinggi dalam membedakan sentimen positif, dan negatif [2]. Studi lain, Gandhi et. al., (2021) menggunakan CNN dan LSTM juga untuk menganalisis sentimen di

Twitter, menunjukkan potensi besar dari metode ini untuk memahami sentimen masyarakat [4]. Selain itu, penelitian oleh Andriawan et. al., (2024) menggunakan algoritma *Gated Recurrent Unit* (GRU) dan Natural Language Processing untuk klasifikasi analisis sentimen partai politik di Twitter, yang menunjukkan efektivitas GRU dalam memahami isu-isu politik [29]. Pada saat yang sama, Geni et al., (2023) menggunakan model bahasa IndoBERT untuk menganalisis suasana hati tweet menjelang Pemilihan Umum Indonesia 2024. Hal ini membantu pemerintah mempersiapkan pemilu dengan mengungkapkan opini populer [30].

Analisis sentimen terhadap pemilu Jakarta 2024 masih sangat bermanfaat, bahkan jika pemilu tersebut diadakan pada bulan November tahun itu. Salah satu manfaat dari penggunaan analisis sentimen adalah dapat memberikan gambaran lengkap mengenai reaksi pemilih terhadap berbagai bagian dari pemilu, seperti program, kandidat, dan kebijakan. Manfaat kedua adalah perencanaan strategi politik di masa depan dapat diinformasikan oleh penelitian sentimen. Kampanye yang dijalankan oleh kandidat dan partai politik dapat disesuaikan dengan permintaan konstituen mereka dengan lebih baik jika mereka memiliki pemahaman yang kuat tentang tren opini publik. Kegunaan ketiga dari riset sentimen adalah untuk menemukan kekhawatiran yang tidak akan hilang setelah pemilu, seperti legislasi yang tidak populer atau inisiatif yang tidak efektif. Perbaikan kebijakan dan program di masa depan dapat dibangun berdasarkan hal ini. Terakhir, menganalisis opini publik tentang peristiwa-peristiwa sebelumnya dapat mengajarkan kita tentang dinamika sosial dan politik masyarakat Jakarta, yang dapat menjadi bahan penelitian di masa depan atau menjadi sumber informasi bagi mereka yang bekerja di bidang komunikasi dan politik.

Dalam penelitian ini, empat model Deep Learning: IndoBERT, LSTM, CNN, dan GRU akan digunakan untuk menilai bagaimana perasaan masyarakat Jakarta terhadap Pilkada 2024. Mengingat pentingnya menangkap asosiasi jangka panjang dalam data teks untuk analisis sentimen, keempat model tersebut dipilih untuk tugas ini. Untuk membuat model yang lebih akurat, penelitian ini juga akan mengidentifikasi karakteristik yang relevan dan melakukan persiapan data yang lebih baik.

Dengan menggunakan empat model Deep Learning yang berbeda, penelitian ini bertujuan untuk membangun sistem kategorisasi sentimen opini publik Platform X untuk Pilkada Jakarta 2024. Penelitian ini juga bertujuan untuk membandingkan dan membedakan kinerja keempat model dalam menentukan apakah dokumen teks (seperti komentar, tweet, atau diskusi) memiliki sentimen positif, negatif, atau netral, serta mengevaluasi akurasi, presisi, recall, dan f1-score dari keempat model tersebut. Selain itu, penelitian ini bermaksud untuk menilai bagaimana manajemen ketidakseimbangan kelas dan metode prapemrosesan data (seperti penghapusan kata dan normalisasi bahasa informal) berdampak pada kinerja model dalam kerangka analisis sentimen politik lokal.

1.2. Rumusan Masalah

Penelitian ini dirumuskan untuk menjawab beberapa permasalahan utama yang terkait dengan analisis sentimen masyarakat terhadap Pilkada Jakarta 2024 pada platform X menggunakan metode *Long Short-Term Memory* (LSTM). Permasalahan tersebut dapat diuraikan sebagai berikut:

1. Kebutuhan Model Analisis Sentimen yang Efektif untuk Data Teks Media Sosial:
 - a. Dibutuhkan model *Deep Learning* yang dapat menangkap hubungan temporal dan pola kompleks dalam teks, seperti IndoBERT, LSTM, CNN, dan GRU yang unggul dalam memahami data teks sekuensial dibandingkan metode lain seperti *Naive Bayes* atau SVM.
 - b. Perlu adanya metode *preprocessing* yang tepat, seperti normalisasi teks, dan penghapusan elemen yang tidak relevan.
2. Kekhususan Dataset untuk Konteks Lokal Pilkada Jakarta 2024:
 - a. Sebagian besar penelitian terdahulu menggunakan dataset generik, seperti ulasan film atau produk, yang tidak spesifik terhadap isu lokal seperti Pilkada Jakarta 2024.
 - b. Masalah yang dihadapi adalah Pilkada Jakarta 2024, dan yang dibutuhkan adalah dataset yang unik dan relevan, yang mencerminkan pandangan masyarakat Jakarta secara real-time. Kata kunci terkait Pilkada Jakarta 2024 harus digunakan untuk mendapatkan informasi ini secara langsung dari platform X.

- c. Tantangan lain adalah memastikan dataset yang dikumpulkan mencakup berbagai perspektif masyarakat, sehingga analisis sentimen yang dihasilkan tidak bias dan dapat mewakili opini publik secara keseluruhan.
3. Batasan dan Asumsi Penelitian:
- a. Data yang diambil dari platform X mencerminkan opini masyarakat Jakarta secara umum, meskipun terdapat kemungkinan bias tertentu, seperti dominasi suara dari kelompok tertentu.
 - b. Positif, negatif, dan netral adalah tiga cara dasar untuk mengklasifikasikan sentimen. Kami melakukan ini untuk membuat analisis lebih mudah dan temuan lebih mudah dipahami.
 - c. Lingkup penelitian dibatasi pada data yang dikumpulkan selama periode tertentu untuk menjaga relevansi analisis dengan isu Pilkada Jakarta 2024..
4. Tantangan Evaluasi dan Validasi Model:
- a. Pelatihan, pengujian, dan penyempurnaan parameter dalam model Deep Learning membutuhkan sumber daya komputer yang besar. Jika mempertimbangkan waktu dan biaya yang dibutuhkan untuk komputasi, hal ini menjadi kendala yang signifikan.
 - b. Untuk memastikan model dapat diandalkan untuk klasifikasi sentimen, model harus dievaluasi menggunakan metrik seperti F1-score, recall, akurasi, dan presisi.

1.3. Tujuan dan Manfaat

Berdasarkan rumusan masalah dan pendekatan penelitian yang telah dibahas, maka tujuan dari Tugas Akhir ini dapat dinyatakan sebagai berikut:

1. Mengembangkan model *Deep Learning* untuk analisis sentimen) yang mampu mengklasifikasikan sentimen masyarakat Jakarta terkait Pilkada 2024 pada platform X dengan akurasi minimal 80%.
2. Merancang metode *preprocessing* data yang efektif untuk menangani karakteristik data tidak terstruktur dari platform X, meliputi:
 - a. Normalisasi teks bahasa Indonesia informal
 - b. Penanganan singkatan dan *emoticon*

c. Pembersihan data dari informasi yang tidak relevan

3. Melakukan evaluasi komprehensif model IndoBERT, LSTM, CNN, dan GRU dengan menggunakan metrik: Akurasi, Presisi, *Recall*, *F1-Score*.

Memberikan kontribusi akademis dalam pengembangan metode analisis sentimen berbasis kecerdasan buatan untuk mengidentifikasi dinamika opini publik pada peristiwa sosial-politik lokal.

1.4. Batasan Masalah

Kelayakan dan kepraktisan dari penelitian ini telah menginformasikan pemilihan batasan-batasan tertentu. Pertama, pengumpulan data dibatasi hanya pada konten publik di Platform X yang menggunakan kata kunci spesifik terkait Pilkada Jakarta 2024. Pembatasan ini dilakukan karena keterbatasan waktu untuk melakukan *crawling* data serta potensi bias akibat dominasi topik tertentu di luar isu Pilkada. Kedua, pelabelan sentimen (positif, negatif, netral) dilakukan secara otomatis dengan bantuan *IndoBERT pretrained*. Hal ini disebabkan belum adanya korpus berlabel khusus untuk konteks politik lokal Jakarta yang siap pakai, sehingga memerlukan penyesuaian umum dan validasi subjektif

1.5. Metode Penelitian

Prosedur berikut merupakan bagian dari studi kuantitatif ini: (1) Studi literatur untuk mengidentifikasi metode analisis sentimen berbasis LSTM dan teknik preprocessing yang relevan; (2) Pengumpulan data dari Platform X menggunakan *twitter auth token* selama periode Juli–November 2024; (3) *Preprocessing data* teks meliputi penghapusan kata (URL, emoji), normalisasi kata informal, tokenisasi, *stop word removal*; (4) Perancangan model LSTM dengan lapisan embedding dan dropout untuk mengurangi overfitting; (5) Untuk melatih model, kami menggunakan dataset berlabel dengan 80% data latih dan 20% data uji. Untuk mengevaluasi performa model, kami menggunakan metrik seperti recall, akurasi, presisi, dan F1-score. Terakhir, untuk melihat model deep learning mana yang lebih baik, kami membandingkannya dengan empat skenario yang berbeda. Semua ini dilakukan dengan menggunakan bahasa pemrograman Python dan kerangka kerja TensorFlow, dengan bantuan tinjauan literatur untuk mendukung teori.

1.6. Jadwal Pelaksanaan

Jadwal pelaksanaan pada penelitian ini disajikan dalam tabel 1.1 berikut:

Tabel 1. 1 Jadwal Pelaksanaan Tugas Akhir.

No.	Deskripsi Tahapan	Bulan 1	Bulan 2	Bulan 3	Bulan 4	Bulan 5	Bulan 6
1	Studi Literatur						
2	Pengumpulan Data						
3	Preprocessing Data						
4	Perancangan dan Pembuatan Model						
5	Pengujian Model						
6	Analisis Hasil						
7	Penyusunan Laporan/Buku TA						