

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pemilu adalah fondasi utama dari demokrasi, namun tidak cukup hanya dengan adanya pemilu untuk membuktikan bahwa suatu negara menjalankan demokrasi tanpa pemilu yang dilakukan secara bebas, adil, dan kompetitif. Untuk mencapai demokrasi yang sejati, dibutuhkan institusi yang kokoh, perlindungan terhadap hak asasi manusia, serta budaya politik yang sesuai dengan prinsip demokrasi. Dalam konteks Indonesia, Pemilihan kepala daerah adalah salah satu wujud praktik demokrasi di tingkat daerah. Pilkada serentak yang dijadwalkan pada 27 November 2024 memberikan kepada masyarakat kesempatan untuk memilih pemimpin daerah mereka yang akan menjabat selama lima tahun mendatang. Meskipun pilkada dapat mendukung kemajuan demokrasi, acara ini juga membawa tantangan berupa konflik yang terkait dengan aturan teknis, independensi penyelenggaraan, dan potensi kekerasan massa yang muncul akibat perselisihan hasil pemilihan [1][2].

Twitter merupakan salah satu platform informasi yang sangat digemari di Indonesia, berada di posisi kelima sebagai negara dengan jumlah pengguna terbanyak. Situs ini memberikan kesempatan bagi penggunanya untuk mengungkapkan pendapat atau memberikan respon terhadap berbagai masalah, termasuk pemilihan kepala daerah. Para calon pejabat daerah dapat memanfaatkan umpan balik serta kritik yang muncul di Twitter [4]. Dalam konteks ini, analisis sentimen sebagai elemen dari text mining menjadi penting untuk memahami pandangan masyarakat mengenai pemilihan kepala daerah [3]. Dengan analisis sentimen, tweet dapat dikelompokkan menjadi kategori positif, netral, atau negatif [5][6]. Deep learning, seperti Recurrent Neural Network (RNN) dan Bidirectional Long Short-Term Memory (BiLSTM), menunjukkan performa yang lebih baik dalam mengolah data besar untuk menghasilkan prediksi yang lebih tepat dibandingkan dengan algoritma pembelajaran mesin tradisional[8].

Salah satu langkah krusial dalam analisis sentimen adalah penggambaran kata menggunakan metode Word Embedding. Metode ini mengatur kata-kata dalam

ruang vektor yang membantu mesin dalam memahami bahasa manusia. Word Embedding mampu mengambil informasi sintaktik dan semantik dari kata-kata yang ada dalam korpus. Angka kesamaan yang hampir sama dengan satu menunjukkan bahwa dua kata.

memiliki arti yang mirip, sedangkan nilai yang lebih rendah menunjukkan perbedaan makna [7]. Dengan menggabungkan teknologi deep learning dan representasi vektor kata, analisis sentimen dapat dilakukan lebih efektif dalam menangkap opini publik terkait isu-isu demokrasi di media sosial.

Dalam analisis sentimen yang menggunakan media sosial, data yang dikumpulkan sering kali menunjukkan distribusi kelas yang tidak seimbang, di mana salah satu kelas sentimen (misalnya positif, negatif, atau netral) dapat mendominasi jumlah data. Masalah ini muncul karena model pembelajaran mesin biasanya cenderung bias terhadap kelas mayoritas dan tidak cukup akurat dalam mengklasifikasikan kelas minoritas, yang dapat mengurangi performa evaluasi model secara keseluruhan[19]. Untuk mengatasi ketidakseimbangan data, penelitian ini menggunakan metode augmentasi dengan WordNet untuk mengevaluasi kinerja Bi LSTM. WordNet menggunakan sinonim, penambahan acak, dan penghapusan acak. Augmentasi sinonim meningkatkan variasi teks dengan mengganti kata dengan sinonimnya. Tujuannya adalah untuk memperbanyak data teks sambil mempertahankan makna keseluruhan teks. Teknik sinonim yang mengelompokkan kata berdasarkan sinonimnya di WordNet memungkinkan penggantian kata yang sesuai dengan konteks kalimat saat ini [9]. Penghapusan acak adalah penghapusan acak dari setiap kata dalam kalimat [10]. Ketidakseimbangan distribusi kelas sentimen dalam data media sosial dapat membuat model pembelajaran mesin menjadi bias terhadap kelas mayoritas dan kurang tepat dalam mengklasifikasikan kelas minoritas[19]. Oleh karena itu penerapan teknik penyeimbangan data seperti augmentasi berbasis WordNet sangat penting untuk menghasilkan distribusi data yang lebih seimbang, meningkatkan generalisasi model, dan memastikan keakuratan prediksi yang merata di seluruh kategori sentimen [16].

Penelitian sebelumnya menunjukkan bahwa pemilihan metode word embedding dan algoritma klasifikasi memainkan peran penting dalam meningkatkan kinerja model dalam analisis sentimen teks. Saragih et al. [11] menemukan bahwa Bi-LSTM dengan FastText mencapai akurasi tertinggi 99% dalam deteksi emosi, sementara Agustiningih et al. [12] menunjukkan bahwa self-trained GloVe dan FastText unggul dalam analisis sentimen terkait vaksin COVID-19 dengan akurasi masing-masing 92.5% dan 92.3%. Firdaus dan Rochmawati [13] membandingkan CNN dan Bi-LSTM menggunakan Word2Vec, di mana Bi-LSTM unggul dengan akurasi analisis emosi 84.36% dan sentimen 69.48%. Selain itu, Muzakir dan Suriani [14] menunjukkan bahwa Bi-LSTM dengan GloVe mencapai akurasi 97.5% dalam deteksi berita palsu, dengan AUC sebesar 99.4%. Selanjutnya, penelitian Aditya et al. [15] menerapkan metode LSTM dengan Word2Vec dan teknik oversampling SMOTE untuk analisis sentimen terhadap tiga kandidat presiden Indonesia tahun 2024 menggunakan data Twitter. Hasil penelitian menunjukkan bahwa integrasi LSTM dan SMOTE mampu meningkatkan performa model secara signifikan, menghasilkan akurasi rata-rata 89.42% dengan peningkatan akurasi sebesar 9.69% dibandingkan LSTM tanpa SMOTE. Penelitian ini menegaskan bahwa penanganan ketidakseimbangan kelas serta pemilihan representasi kata berbasis Word2Vec berkontribusi terhadap peningkatan performa klasifikasi sentimen pada data media sosial yang bersifat tidak terstruktur.

Meskipun penelitian terdahulu telah membandingkan ketiga teknik embedding tersebut dan menunjukkan kinerja yang baik[11]. perbandingan ulang dalam penelitian ini tetap diperlukan karena perbedaan domain, karakteristik data, dan konteks linguistik yang spesifik. Data Pilkada 2024 di platform media sosial X memiliki tingkat noise, variasi bahasa informal, dan dinamika kata yang unik, termasuk penggunaan singkatan, kata tidak baku, dan istilah politik lokal yang mungkin tidak tercakup dalam korpus umum[17]. Oleh karena itu, perlu diuji *embedding* mana yang paling *robust* dalam menangani kompleksitas tersebut, terutama mengingat FastText dikenal mampu menangani kata *out-of-vocabulary* (OOV) melalui informasi subword, relevansi pada media sosial sangat tinggi karena bahasa media sosial Indonesia memiliki OOV variasi ejaan tidak baku[18], sementara GloVe dan Word2Vec lebih bergantung pada statistik korpus

yang ada. Selain itu, penelitian ini menggunakan teknik augmentasi berbasis WordNet yang berbeda dari penelitian sebelumnya[15]. sehingga dapat memengaruhi performa embedding secara berbeda dan memberikan wawasan baru dalam penanganan ketidakseimbangan kelas untuk data teks berbahasa Indonesia [20].

Penelitian ini dilakukan dengan tujuan untuk mengevaluasi efektivitas teknik Word Embedding FastText, GloVe, dan Word2Vec dengan model Bi-LSTM dalam analisis sentimen komentar masyarakat terkait Pilkada 2024 di media sosial X. Penelitian ini berdasarkan temuan penelitian sebelumnya yang menunjukkan bahwa algoritma deep learning seperti LSTM lebih baik daripada metode berbasis probabilistik dalam Mengidentifikasi opini positif, negatif, atau netral. Data dikumpulkan menggunakan kata kunci "Pilkada 2024" dan dikumpulkan melalui crawling Twitter dari tanggal 18 hingga 22 November 2024 dan 29 November hingga 15 Desember 2024.

1.2 Rumusan Masalah

Rumusan masalah yakni :

- a. Bagaimana performa masing-masing word embedding GloVe, Word2Vec, dan FastText pada model Bi-LSTM pada analisis sentimen terkait tweet Pilkada 2024?
- b. Bagaimana performa algoritma Bi LSTM dalam mengklasifikasikan sentimen berlabel positif, negatif, dan netral pada dataset tweet terkait Pilkada 2024?

1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk mengevaluasi dan membandingkan efektivitas tiga teknik word embedding GloVe, Word2Vec, dan FastText dalam meningkatkan kinerja model BiLSTM untuk analisis sentimen tweet terkait Pilkada 2024 serta mengukur performa algoritma BiLSTM dalam mengklasifikasikan sentimen berlabel positif, negatif, dan netral berdasarkan metrik accuracy, precision, recall, dan f1score, sehingga dapat mengidentifikasi teknik word embedding yang paling

optimal serta model deep learning yang efektif dalam memahami opini dari publik terkait isu politic di sosial media.

1.4 Batasan Masalah

Batasan masalah yakni:

- a. Penelitian ini hanya fokus pada komentar masyarakat di media sosial X yang berkaitan dengan Pilkada 2024 yaitu masa sebelum Pilkada 2024 dari tanggal 18 November 2024 hingga 22 November 2024 dan masa setelah Pilkada dari tanggal 29 November hingga 15 Desember 2024.
- b. Data yang dianalisis terbatas pada 16.929 tweet yang diambil menggunakan kata kunci 'pilkada 2024'.
- c. Model yang digunakan dalam penelitian ini adalah Bi LSTM dengan teknik balancing data WordNet.
- d. Klasifikasi sentimen dibatasi pada tiga kategori: positif, negatif, dan netral.

