

BAB II

TINJAUAN LITERATUR

2.1 Analisis Sentimen

Analisis sentimen merupakan teknik dalam text mining yang bertujuan mengidentifikasi kecenderungan opini dalam teks menjadi kategori tertentu seperti positif, negatif, atau netral. Metode ini banyak diterapkan pada media sosial untuk memahami persepsi publik terhadap isu sosial maupun politik.

Menurut Manullang et al., analisis sentimen dapat dilakukan menggunakan pendekatan lexicon based maupun machine learning. Sementara itu, perkembangan terbaru menunjukkan bahwa metode deep learning seperti LSTM dan BERT memberikan hasil yang lebih baik dalam memahami konteks bahasa alami dibanding metode konvensional.

Penelitian terbaru menunjukkan bahwa analisis sentimen sangat krusial untuk mengungkap opini publik dari komentar YouTube, terutama pada topik kontroversial seperti film dokumenter "Dirty Vote". Sebuah studi pada tahun 2024 secara spesifik membuktikan efektivitas model LSTM dalam kasus ini, memperkuat argumen bahwa pendekatan deep learning ini esensial untuk menangani teks informal dan konteks lokal Indonesia.

Studi berjudul "Analisis Sentimen Masyarakat Terhadap Film Dirty Vote" (2024) dari repository Mercubuana menganalisis komentar pada video "Dirty Vote" di kanal YouTube PSHK Indonesia. Peneliti menggunakan InSet Lexicon

untuk pelabelan otomatis, SVM, dan LSTM sebagai algoritma klasifikasi, serta SMOTE untuk mengatasi overfitting. LSTM mencapai akurasi tertinggi 96,3% dengan loss function 0,098 pada data stemming dan batch size 16, menunjukkan superioritasnya dalam menangkap pola sentimen negatif dominan (79,74%) dibandingkan metode lain.

Temuan ini menegaskan pentingnya analisis sentimen karena mendistribusikan opini publik secara akurat positif 20,26% vs negatif 79,74% yang tidak mungkin didapat tanpa model deep learning seperti LSTM untuk teks YouTube yang noisy. Dibandingkan SVM (97,58% akurasi tapi kurang stabil pada overfitting), LSTM lebih unggul dalam mempertahankan konteks urutan kata pada komentar jangka panjang dan informal, sedangkan penelitian ini hanya memakai satu algoritma yaitu LSTM yang menjadi pembeda dari penelitian ini ialah di data youtube dari versi konten yang berbeda dan algoritma yang digunakan, penelitian sebelumnya menggunakan dua algoritma serta penelitian yang ditulis oleh penulis sekarang menggunakan satu algoritma saja yaitu LSTM.

Untuk kasus spesifik "Dirty Vote II 03" menggunakan LSTM, jurnal ini menjadi bukti kuat (kurang dari 2 tahun lalu) bahwa analisis sentimen esensial guna mengukur reaksi publik terhadap konten politik sensitif di YouTube. Penelitian serupa pada komentar Ruangguru (menggunakan LSTM dengan akurasi 90%) juga mendukung, menyoroti efektivitas deep learning untuk data lokal Indonesia. Pendekatan ini krusial karena komentar YouTube sering kali menjadi indikator penerimaan publik yang cepat dan masif, terutama pada isu pemilu.

Beberapa penelitian terdahulu juga menunjukkan bahwa analisis sentimen efektif digunakan untuk memetakan opini publik pada isu politik, termasuk pemilu dan kebijakan pemerintah (Bagaskara et al., 2024; Putra Prasetyo et al., 2024).

2.2 Media Sosial sebagai Sumber Data

YouTube merupakan salah satu platform media sosial dengan jumlah pengguna terbesar yang menyediakan data teks dalam bentuk komentar. Data komentar tersebut mencerminkan opini publik yang spontan dan aktual terhadap suatu isu.

Penelitian oleh Hendrawan & Sela (2024) serta Rahmawati et al. (2025) menunjukkan bahwa komentar YouTube dapat dimanfaatkan sebagai sumber data yang relevan dalam analisis sentimen berbasis deep learning. Dinamika sentimen dapat berubah sesuai dengan topik yang dibahas, terutama pada isu politik dan ekonomi.

Dalam konteks film dokumenter *Dirty Vote*, respons publik yang muncul di media sosial menunjukkan adanya polarisasi opini sebagaimana juga dibahas oleh Bagaskara et al. (2024) pada platform Twitter.

2.3 Pra-Pemrosesan Data Teks

Pra-pemrosesan (preprocessing) merupakan tahapan penting sebelum proses klasifikasi dilakukan. Tahapan ini bertujuan untuk membersihkan dan menyiapkan data agar lebih terstruktur dan siap diproses oleh model.

Menurut Teknokrat (2022), tahapan pra-pemrosesan umumnya meliputi:

1. Case folding
2. Tokenisasi
3. Stopword removal
4. Normalisasi
5. Stemming

Proses ini membantu mengurangi noise dalam teks, terutama pada bahasa media sosial yang sering mengandung singkatan, emotikon, serta kata tidak baku. Pratiwi & Kamayani (2024) juga menegaskan bahwa kualitas pelabelan dan preprocessing sangat memengaruhi performa model klasifikasi sentimen.

2.4 Pelabelan Data

Pelabelan data merupakan proses pemberian kategori sentimen pada setiap komentar. Pendekatan pelabelan dapat dilakukan secara manual maupun otomatis. Pratiwi & Kamayani (2024) menjelaskan bahwa perbedaan metode pelabelan dapat memengaruhi distribusi kelas dan performa model. Oleh karena itu, konsistensi dalam pemberian label sangat penting agar model dapat belajar secara optimal.

2.5 Long Short-Term Memory (LSTM)

LSTM merupakan pengembangan dari Recurrent Neural Network (RNN) yang dirancang untuk mengatasi permasalahan vanishing gradient. Model ini mampu menangkap ketergantungan jangka panjang dalam data sekuensial seperti teks.

Menurut Aljabar & Abd Karim (2022), LSTM efektif digunakan dalam analisis sentimen media sosial karena mampu memahami konteks kalimat secara lebih mendalam dibanding metode klasik seperti Naïve Bayes.

Beberapa penelitian yang membuktikan efektivitas LSTM antara lain:

1. Dimas Anugrah Firdlous (2023) pada analisis sentimen Pemilu 2024
2. Septian Ade Putra (2024) pada komentar YouTube
3. Siti Mutmainah et al. (2024) pada aplikasi PeduliLindungi
4. Wicaksono & Nastiti (2024) pada YouTube Indonesia Lawyers Club

Hasil penelitian tersebut menunjukkan bahwa LSTM mampu menghasilkan akurasi yang tinggi dalam klasifikasi teks berbahasa Indonesia.

2.6 Perbandingan Model

Beberapa penelitian membandingkan LSTM dengan model lain seperti SVM, Naïve Bayes, dan BERT.

Ardiansyah (2024) membandingkan SVM dan LSTM pada analisis sentimen film *Dirty Vote* dan menemukan bahwa model deep learning cenderung lebih unggul dalam menangkap konteks kalimat kompleks.

Bagaskara et al. (2024) menggunakan Naïve Bayes dengan optimasi Grid Search, sementara Informasi (2025) serta Khadapi et al. membandingkan LSTM dengan BERT dan menunjukkan bahwa model transformer seperti BERT memiliki performa tinggi, namun memerlukan sumber daya komputasi lebih besar.

Dengan demikian, LSTM menjadi pilihan yang seimbang antara performa dan kompleksitas komputasi.

2.7 Evaluasi Model

Evaluasi model klasifikasi sentimen umumnya menggunakan confusion matrix dan metrik seperti:

1. Accuracy
2. Precision
3. Recall
4. F1-score

Menurut Aljabar & Abd Karim (2022) serta Rahmawati et al. (2025), penggunaan beberapa metrik evaluasi diperlukan untuk memberikan gambaran performa model secara menyeluruh, terutama ketika distribusi data tidak seimbang.

2.8 Penelitian Terkait

Beberapa penelitian yang relevan dengan topik ini antara lain:

- Ardiansyah (2024): Analisis sentimen film *Dirty Vote* menggunakan SVM dan LSTM.
- Bagaskara et al. (2024): Analisis sentimen *Dirty Vote* di Twitter menggunakan Naïve Bayes.
- Rahmawati et al. (2025): Analisis sentimen komentar YouTube berbasis LSTM.

- Putra Prasetyo et al. (2024): LSTM pada isu penundaan Pemilu 2024.

Berdasarkan penelitian-penelitian tersebut, belum banyak kajian yang secara spesifik menganalisis komentar YouTube pada video “*Dirty Vote II 03*” menggunakan pendekatan LSTM, sehingga penelitian ini memiliki kontribusi kebaruan pada objek dan waktu kajian.