

BAB V

SIMPULAN DAN SARAN

5.1 Simpulan

Model BILSTM yang dikombinasikan dengan penyesuaian bobot kelas dan tuning hyperparameter menunjukkan performa terbaik dengan akurasi 94% dan recall kelas positif 86%, menandakan keseimbangan optimal antara sensitivitas dan presisi. Model ini melampaui BILSTM dengan penyesuaian bobot kelas saja yang mencapai 93% akurasi dan recall positif 79%, serta BILSTM standar dengan akurasi 92% dan recall positif 77%. Model LSTM memiliki performa bervariasi antara 92–94%, namun tuning hyperparameter justru menurunkan akurasi menjadi 92%. Sementara itu, model CNN mempertahankan kestabilan performa pada kisaran 90–91%, meskipun dengan presisi dan recall kelas positif yang lebih rendah dibanding model LSTM dan BILSTM.

Secara keseluruhan, model BILSTM yang dikombinasikan dengan penyesuaian bobot kelas dan tuning hyperparameter menunjukkan performa terbaik dengan akurasi makro tertinggi mencapai 94%, mengungguli LSTM dan CNN yang masing-masing memiliki akurasi sekitar 93% dan 90%. Model LSTM dengan pengaturan class weights bahkan mencapai akurasi tertinggi 94%, serta performa metrik precision, recall, dan F1-score di atas 92%, menunjukkan keseimbangan optimal antara ketepatan dan sensitivitas. Setelah tuning class weights dan hyperparameter, LSTM mempertahankan performa stabil dengan akurasi 94%, sementara BiLSTM juga menampilkan hasil kompetitif terutama pada precision dan F1-score di atas 90%. CNN konsisten berada di posisi paling bawah dengan akurasi dan metrik lainnya mendekati 90%. Hasil uji statistik Friedman dengan p-value sekitar 0,022–0,023 menegaskan adanya perbedaan signifikan dalam performa ketiga model tersebut.

Kontribusi utama dari penelitian ini terletak pada pengembangan sebuah aplikasi analisis sentimen yang terintegrasi secara menyeluruh, mulai dari tahap pra-pemrosesan teks, pemilihan model deep learning, proses klasifikasi sentimen,

hingga visualisasi hasil yang interaktif dan informatif. Aplikasi ini dilengkapi dengan fitur seperti word cloud dan visualisasi distribusi sentimen yang memungkinkan pengguna menggali informasi lebih mendalam dari data, baik berdasarkan kata kunci maupun kategori sentimen tertentu.

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan. Dataset yang digunakan hanya mencakup ulasan dari satu platform aplikasi (Tinder), sehingga penerapan model pada domain lain masih perlu diteliti lebih lanjut. Selain itu, meskipun model deep learning memberikan performa tinggi, proses pelatihannya membutuhkan sumber daya komputasi yang besar dan waktu yang relatif lama, sehingga perlu mempertimbangkan efisiensi dalam penerapan praktisnya.

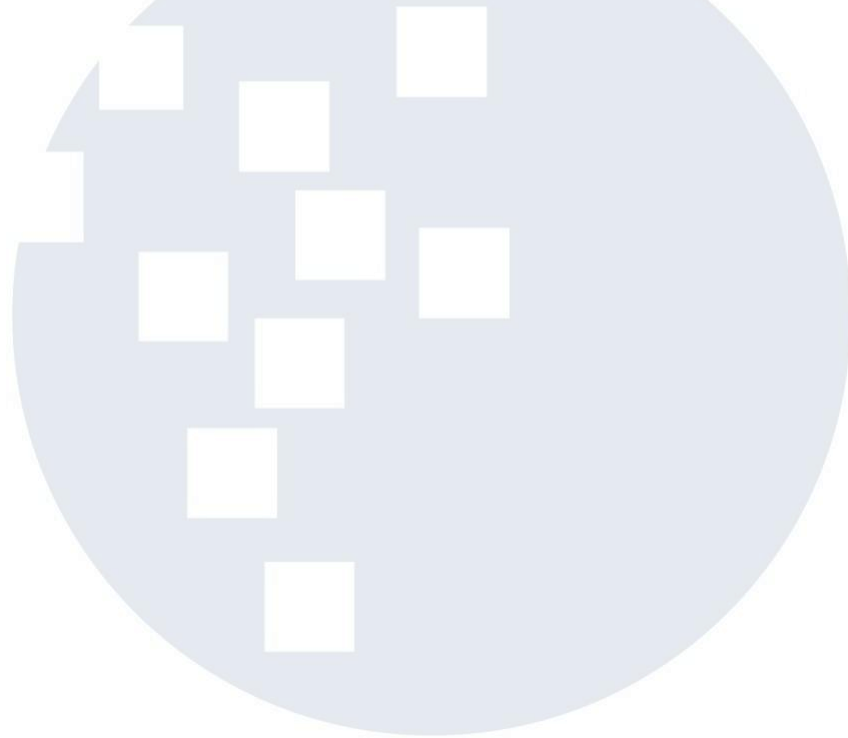
Sebagai tindak lanjut, penelitian ini dapat dikembangkan lebih lanjut dengan menerapkan pendekatan transfer learning atau memanfaatkan model bahasa terlatih sebelumnya (seperti BERT) untuk mengevaluasi kemungkinan peningkatan performa. Selain itu, integrasi dengan data secara real-time dan penerapan analisis aspek (aspect-based sentiment analysis) juga menjadi arah pengembangan yang potensial di masa mendatang.

5.2 Saran

Berdasarkan hasil penelitian ini, beberapa saran yang dapat dijadikan acuan untuk pengembangan studi selanjutnya antara lain:

1. Disarankan menggunakan data dari berbagai platform seperti Twitter, Instagram, atau marketplace digital untuk menguji kemampuan generalisasi model dengan ragam gaya bahasa dan konteks sosial yang beragam.
2. Perlu dilakukan evaluasi model deep learning seperti GRU, Transformer, serta model pretrained seperti BERT dan RoBERTa. Bandingkan juga dengan model klasik seperti Random Forest, XGBoost, dan LightGBM untuk melihat keunggulan tiap metode.
3. DANN dapat membantu mengatasi perbedaan distribusi antara data latih dan uji dari sumber berbeda, sehingga meningkatkan generalisasi model dalam skenario domain adaptasi.

4. Analisis aspek memungkinkan pengenalan opini terhadap fitur spesifik dalam ulasan panjang dan kompleks, sehingga menghasilkan informasi sentimen yang lebih detail dan bernilai.



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA