BAB V
PENUTUP

5.1 Kesimpulan

Berdasarkan rangkaian tahapan yang dilakukan, penelitian ini menunjukkan bahwa komentar TikTok sebagai data teks bebas tidak dapat langsung digunakan untuk pemodelan karena penulisannya sangat beragam dan masih mengandung banyak elemen yang tidak relevan, seperti emoji, simbol, tautan, pengulangan karakter, serta variasi ejaan dan singkatan. Oleh karena itu, tahap preprocessing menjadi langkah penting agar data lebih siap diolah. Hasil preprocessing memperlihatkan bahwa proses pembersihan teks, penyamaan huruf, normalisasi kata, penghapusan kata-kata umum yang kurang informatif, serta penghapusan data duplikat dan kosong berhasil membuat komentar lebih seragam dan lebih fokus pada inti kalimat. Dengan data yang lebih rapi, proses pelabelan dan pelatihan model dapat dilakukan secara lebih terkontrol.

Pelabelan otomatis berbasis leksikon InSet menghasilkan tiga kelas sentimen, yaitu negatif, netral, dan positif. Distribusi sentimen menunjukkan bahwa data didominasi oleh sentimen negatif sebesar 40% dan netral sebesar 39,3%, sedangkan sentimen positif memiliki porsi paling kecil yaitu 20,7%. Temuan ini mengindikasikan bahwa pembahasan isu mafia BBM di TikTok banyak memunculkan respons kritis dari warganet, namun juga cukup banyak komentar yang bersifat informatif atau tidak menunjukkan sikap secara tegas. Dengan demikian, pemetaan sentimen relevan untuk menggambarkan arah opini publik secara lebih sistematis.

Pada tahap pemodelan, dataset dibagi menjadi data latih, validasi, dan uji menggunakan stratified split agar proporsi kelas tetap terjaga. Pembagian data menghasilkan 5.796 data latih, 724 data validasi, dan 725 data uji. Selanjutnya dilakukan pengujian beberapa konfigurasi pelatihan melalui empat metode pencarian hyperparameter, yaitu manual tuning, grid search, random search, dan genetic algorithm. Berdasarkan evaluasi pada data validasi, metode random search

memberikan konfigurasi terbaik dengan optimizer AdamW, learning rate 5e-05, class weight balanced, oversampling aktif, warmup ratio 0,05, dan epoch 2. Proses pelatihan dipantau menggunakan validation loss dan menerapkan early stopping, sehingga model yang dipilih mempertimbangkan kestabilan generalisasi pada data validasi, bukan hanya performa pada data latih.

Evaluasi akhir pada data uji menunjukkan bahwa IndoBERT mampu melakukan klasifikasi sentimen dengan performa yang baik dan konsisten, dengan accuracy sebesar 0,87 dan F1 score macro sebesar 0,87 Jika dilihat per kelas, model menunjukkan kinerja terbaik pada sentimen negatif dengan F1 score sebesar 0,90, disusul sentimen positif dengan F1 score sebesar 0,88, dan sentimen netral dengan F1 score sebesar 0,86. Hasil evaluasi per kelas juga mengindikasikan bahwa kelas netral memiliki recall yang tinggi namun precision lebih rendah, yang menandakan masih ada sebagian komentar dari kelas lain yang tertarik menjadi netral. Hal ini wajar mengingat komentar TikTok cenderung singkat dan sering ambigu, sehingga perbedaan antara netral dan sentimen tertentu tidak selalu tegas.

5.2 Saran

Penelitian Penelitian ini membuktikan bahwa IndoBERT dapat digunakan untuk mengklasifikasikan sentimen komentar TikTok dengan hasil yang baik. Untuk memperkaya kontribusi dan memperluas manfaat penelitian, beberapa pengembangan berikut dapat dipertimbangkan pada penelitian selanjutnya.

1. **Pendalaman interpretasi hasil klasifikasi**

   Penelitian lanjutan dapat menambahkan analisis yang memperjelas makna di balik hasil klasifikasi, misalnya dengan menampilkan kata kunci atau frasa dominan pada tiap kelas sentimen, serta beberapa contoh komentar yang mewakili masing-masing kategori. Tambahan ini akan membantu pembaca memahami konteks kemunculan sentimen dan membuat hasil analisis opini publik lebih komunikatif.

2. **Penguatan karakteristik kelas netral**

   Kelas netral dapat dijadikan fokus pengembangan lebih lanjut melalui

analisis contoh kasus yang khas, seperti komentar yang bersifat ambigu, informatif, atau bercampur unsur emosi. Dengan memahami karakteristik tersebut, normalisasi bahasa tidak baku, singkatan, dan ragam slang TikTok dapat diperkaya sehingga pemisahan antara netral dan sentimen lainnya menjadi semakin jelas.

3. **Penyempurnaan pelabelan otomatis**

   Pelabelan berbasis leksikon InSet sudah efektif untuk menghasilkan label secara cepat dan konsisten. Pada penelitian selanjutnya, proses ini dapat diperkaya melalui pengecekan manual pada sebagian kecil sampel sebagai kontrol kualitas, serta penambahan kosakata domain yang relevan dengan isu yang dikaji. Langkah ini dapat memperkuat kesesuaian label dengan konteks komentar yang sering singkat dan bergantung situasi.

4. **Eksplorasi konfigurasi pelatihan yang lebih luas**

   Penelitian berikutnya dapat memperluas ruang eksplorasi hyperparameter, seperti variasi batch size, panjang maksimum token, warmup ratio, jumlah epoch, serta pengaturan lain yang relevan. Eksplorasi yang lebih luas dapat membantu menemukan konfigurasi pelatihan yang semakin optimal untuk karakteristik komentar TikTok yang cenderung pendek namun sangat bervariasi.

5. **Perbandingan model untuk memperkaya temuan**

   Untuk menambah wawasan, penelitian lanjutan dapat membandingkan IndoBERT dengan model transformer lain yang relevan seperti IndoRoBERTa, DistilBERT, atau model multilingual. Perbandingan ini dapat memberikan gambaran model mana yang paling efektif pada kondisi tertentu, sekaligus memperkuat posisi IndoBERT dalam konteks klasifikasi sentimen bahasa Indonesia.

6. **Pengujian pada variasi data yang lebih beragam**

   Agar hasil penelitian dapat menggambarkan performa model pada kondisi yang lebih luas, penelitian selanjutnya dapat menguji model pada data

TikTok dari periode waktu berbeda atau dari topik yang relevan namun bervariasi. Pengujian ini dapat menunjukkan konsistensi model dalam menghadapi dinamika tren, ragam bahasa, dan gaya penulisan pengguna TikTok yang terus berkembang.