

BAB V

PENUTUP

5.1 Kesimpulan

Berikut ini kesimpulan yang dapat ditarik berdasarkan rangkaian penelitian skripsi yang telah dilaksanakan:

1. Analisis sentimen berbasis aspek dalam penelitian ini dilakukan dengan merancang 3 jenis model, yaitu model deteksi aspek, model deteksi sentimen, dan model gabungan untuk deteksi aspek dan sentimen. Ketiga model tersebut dibangun menggunakan algoritma *Bidirectional Long Short-Term Memory* (Bi-LSTM) dan dilatih dengan dataset dengan label berformat BIO.
2. Untuk menghasilkan nilai *accuracy* dan F1-score yang baik, setiap model melalui tahap *hyperparameter tuning*. *Hyperparameter* model seperti arsitektur model dan panjang *input* (*input length*) model memberikan dampak signifikan pada kinerja model. Arsitektur model yang optimal adalah arsitektur yang tidak terlalu sederhana dan tidak terlalu kompleks, sehingga model tidak mengalami *underfitting* ataupun *overfitting* terhadap dataset. Panjang *input* yang optimal adalah yang tidak terlalu pendek sehingga model tidak kehilangan informasi penting dalam data, namun tidak terlalu panjang untuk menghindari kompleksitas berlebih pada model. Konsep “optimal” dalam hal ini mengacu pada tercapainya keseimbangan yang tepat antara kompleksitas model dan kemampuan model untuk menangkap pola yang terdapat pada data. Sebuah nilai optimal untuk *input length* dan *arsitektur model* dapat tercapai ketika model mampu mencapai tingkat kinerja yang tinggi, diukur dengan perolehan *validation accuracy* model yang selalu meningkat dalam sejumlah siklus pelatihan (*epoch*) saat fase *training* serta perolehan skor *evaluation metrics* saat fase *testing*, seperti F1-score.
3. Setiap jenis model dalam penelitian ini diuji menggunakan data *testing* dan menghasilkan nilai kinerja yang diukur melalui *accuracy* dan F1-score. *Accuracy* dan F1-score terbaik secara berturut-turut sebesar 0.865 dan 0.739

untuk model deteksi aspek, 0.846 dan 0.681 untuk model deteksi sentimen, serta 0.836 dan 0.611 untuk model deteksi aspek dan sentimen. Model dengan kinerja terbaik tersebut diimplementasikan pada halaman web, sehingga menjadi suatu *web application* (*web app*).

4. Analisis model dilakukan dengan menganalisis perbedaan antara hasil prediksi model (*predicted*) dan hasil yang diharapkan (*actual*). Kekeliruan model dapat dirinci sebagai berikut:
 - a. Model memberikan prediksi label yang keliru pada token saat token tersebut memiliki label yang tidak konsisten (ambigu) pada data *training*, akibat ambiguitas pelabelan yang dihasilkan *annotator*.
 - b. Model memberikan prediksi label yang keliru dalam beberapa kondisi akibat kurangnya kemampuan model dalam melakukan *generalisasi* yang baik pada dataset.
 - c. Model memberikan prediksi label yang keliru pada token berlabel “O”, karena token berlabel “O” memiliki lebih dari satu label.
 - d. Model gabungan aspek dan sentimen menghasilkan prediksi label sentimen yang berbeda ketika seharusnya menghasilkan prediksi label sentimen yang sama pada suatu rangkaian kata dengan label aspek yang sama.

5.2 Saran

Berikut ini beberapa saran untuk peneliti berikutnya berdasarkan proses penelitian skripsi ini:

1. Pelabelan dataset yang dilakukan pada penelitian ini telah melalui serangkaian tahap yang komprehensif, melibatkan validasi secara manual oleh manusia serta otomatis dengan bantuan *software* dan kode program. Meskipun demikian, dalam kondisi yang ideal, direkomendasikan agar proses pelabelan suatu baris data dapat melibatkan pertimbangan lebih dari satu *annotator* agar dapat meminimalisir subjektivitas *annotator* dan ambiguitas pemahaman *annotator*. Dengan cara seperti ini, suara mayoritas dapat menjadi penentu label resmi untuk setiap token (kata) dalam baris data tersebut.

2. Untuk memperkuat justifikasi di balik label sentimen pada dataset, disarankan untuk melibatkan kolom nilai atau *rating* produk sebagai pertimbangan *annotator* untuk memberi label sentimen. Hal ini akan memberikan landasan lebih kokoh berdasarkan nilai evaluasi yang jelas langsung dari pengguna terhadap produk, sehingga dapat memperjelas makna pada kalimat ulasan yang bersifat ambigu.
3. Penelitian ini telah menghasilkan model yang dapat digunakan untuk analisis sentimen berbasis aspek. Meskipun demikian, penelitian selanjutnya diharapkan dapat berfokus untuk mengembangkan teknik NLP yang lebih canggih dan variatif, seperti melakukan praproses data, menggunakan *pre-trained word embedding*, menggunakan algoritma yang berbeda, atau menambahkan arsitektur jaringan yang lebih kompleks pada model. Dalam konteks pengembangan algoritma, disarankan agar peneliti berikutnya mengeksplorasi dan mempertimbangkan integrasi metode-metode yang telah terbukti efektif sebagaimana metode yang disajikan dalam penelitian terkait yang digunakan oleh penelitian skripsi ini. Sebagai contoh, dapat dipertimbangkan untuk mengadopsi algoritma Bi-LSTM dan CRF sebagaimana penelitian yang dilakukan oleh Fauzi & Romadhony (2021) untuk menemukan hasil pemodelan yang lebih baik dalam analisis sentimen berbasis aspek pada ulasan produk elektronik laptop. Bi-LSTM dapat memahami makna (konteks) suatu kata secara umum, sedangkan CRF dapat membantu mengatasi kelemahan Bi-LSTM dengan memastikan kesesuaian urutan makna suatu kata, termasuk hubungan antara kata-kata yang berada di sekitarnya. Dengan penggunaan kedua algoritma tersebut yang disusun dalam satu kesatuan arsitektur model, maka kinerja model diharapkan dapat semakin meningkat.
4. Untuk meningkatkan dampak dari model pada kasus analisis sentimen berbasis aspek dalam implementasi sesungguhnya, penulis merekomendasikan agar penelitian ini dapat dilanjutkan hingga tahap aplikasi praktis dalam industri, seperti mengimplementasikan model pada *dashboard* dengan tambahan visualisasi data terkait produk yang dapat membantu pemangku kepentingan dalam memonitor kondisi setiap produk

yang dijual secara *real-time*. Implementasi *dashboard* produk secara *real-time* memungkinkan pemangku kepentingan untuk memberikan respon yang cepat terhadap perubahan sentimen konsumen, sehingga analisis menjadi lebih dinamis dan responsif terhadap perubahan tren pasar, perubahan citra merek, atau peristiwa khusus yang dapat mempengaruhi opini publik. Dengan demikian, hasil dari penelitian ini dapat memberikan manfaat praktis dalam dunia nyata (*real-world application*), memungkinkan pengambilan keputusan yang tepat dalam mengelola reputasi produk dan strategi pemasaran produk.