BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai komparasi kinerja *DIETClassifier* dalam klasifikasi *intent* data sejarah Dewi Durga menggunakan *Framework Rasa*, maka kesimpulan yang diperoleh adalah sebagai berikut:

1. Perbandingan kinerja *DIETClassifier* dalam klasifikasi *intent* menunjukkan bahwa setiap konfigurasi *hyperparameter* memberikan pengaruh yang cukup signifikan terhadap performa model. Seluruh konfigurasi yang diuji menunjukkan kinerja yang baik dengan akurasi yang mendekati 100%, namun terdapat perbedaan dalam hal stabilitas generalisasi antar model. Beberapa konfigurasi memperlihatkan selisih yang cukup besar antara akurasi data pelatihan dan data pengujian (*gap accuracy*), yang mengindikasikan potensi *overfitting*. Sebaliknya, konfigurasi dengan nilai akurasi dan F1-score yang konsisten antara data latih dan uji menunjukkan kemampuan generalisasi yang lebih baik. Hal ini menandakan bahwa meskipun *DIETClassifier* merupakan model yang efisien dan fleksibel dalam menangani klasifikasi *intent* berbasis teks sejarah, pemilihan kombinasi *hyperparameter* yang tepat tetap menjadi faktor kunci untuk menghasilkan model yang optimal dan tidak bias terhadap data pelatihan.
2. Kombinasi model terbaik yang digunakan untuk mencapai akurasi maksimal dalam klasifikasi *intent* pada dataset sejarah Dewi Durga diperoleh melalui konfigurasi Config 1. Konfigurasi ini terdiri dari kombinasi *hyperparameter*: *epochs* = 100, *constrain_similarities* = *TRUE*, *fine_tune* = *TRUE*,

regularization_strength = 0.01, *batch_size* = 16, *learning_rate* = 0.001, *optimizer* = adamax, dan *dropout_rate* = 0.2. Berdasarkan hasil evaluasi, Config 1 berhasil mencapai akurasi tertinggi pada data pengujian sebesar 99,63%, dengan nilai F1-Score sebesar 99,57%, serta selisih akurasi (gap) yang sangat kecil antara data latih dan uji, yaitu -0,0007. Nilai ini mencerminkan kemampuan generalisasi model yang sangat baik. Selain itu, kurva akurasi dan *loss* menunjukkan proses pelatihan yang konvergen dan stabil, tanpa adanya fluktuasi tajam atau degradasi performa yang mengarah pada overfitting. Dengan demikian, Config 1 ditetapkan sebagai konfigurasi terbaik dalam penelitian ini karena mampu memberikan hasil klasifikasi yang paling optimal dan konsisten di antara sembilan skenario yang diuji.

5.2 Saran

Berdasarkan hasil dan keterbatasan yang ditemukan selama penelitian, beberapa saran yang dapat diajukan adalah:

1. Penelitian selanjutnya disarankan untuk mengeksplorasi kombinasi *pipeline* lain, termasuk penggunaan pre-trained embedding seperti BERT atau FastText, untuk melihat kemungkinan peningkatan performa klasifikasi *intent*.
2. Dataset yang digunakan dalam penelitian ini masih terbatas pada konteks kata "Durga" atau "Dewi Durga", sehingga dapat diperluas pada entitas sejarah atau budaya lain untuk membangun *chatbot* edukatif yang lebih luas dan inklusif. Model ini dapat diterapkan pada dataset sejarah lainnya, asalkan data tersebut memiliki format yang serupa, yaitu berupa pertanyaan dan jawaban, serta dilakukan pembersihan dan pemrosesan data yang sesuai

agar hasilnya tetap akurat dan optimal. Perlu dilakukan evaluasi lanjutan dari sisi pengguna (user experience) untuk mengukur kualitas interaksi *chatbot* secara langsung dalam lingkungan yang nyata, terutama dalam konteks edukasi sejarah dan budaya.

Halaman ini sengaja dikosongkan