

INTISARI

PENINGKATAN KLASIFIKASI *INTENT* DAN EKSTRAKSI *ENTITY* MELALUI EKSTRAKSI FITUR BERBASIS SEMANTIK DAN *DIET CLASSIFIER* PADA CHATBOT (STUDI KASUS DI PUSAT PELATIHAN KERJA DAERAH JAKARTA UTARA)

Oleh

Agus Wahyu Priyanto
23/530419/PPA/06733

Permasalahan utama dalam pelayanan informasi di Pusat Pelatihan Kerja Daerah (PPKD) Jakarta Utara adalah tingginya volume pertanyaan masyarakat yang selama ini hanya dilayani melalui media sosial Instagram dan WhatsApp admin. Hal ini menyebabkan keterlambatan respons, ketergantungan pada sumber daya manusia, serta layanan yang tidak tersedia selama 24 jam. Selain itu, chatbot yang ada umumnya berbasis aturan (*rule-based*) atau fitur sintaksis sehingga kurang mampu memahami variasi bahasa alami secara kompleks, yang berdampak pada rendahnya akurasi klasifikasi *intent* dan ekstraksi *entity*.

Penelitian ini mengusulkan pengembangan chatbot berbasis kerangka kerja Rasa dengan mengintegrasikan ekstraksi fitur berbasis semantik menggunakan *IndoBERT Language Model Featurizer* serta *Dual Intent and Entity Transformer (DIET) Classifier*. Pendekatan ini memungkinkan chatbot untuk memahami makna kata dan konteks percakapan secara lebih baik, sehingga meningkatkan kemampuan klasifikasi *intent* dan ekstraksi *entity* dibandingkan pendekatan sintaksis.

Hasil pengujian menunjukkan bahwa penggunaan fitur semantik meningkatkan performa model secara konsisten pada tiga metode evaluasi, yaitu *held-out*, *cross-validation* dan data uji aktual. Pada *held-out*, peningkatan *F1-score* untuk klasifikasi *intent* mencapai 1,83% - 5,66%, dan ekstraksi *entity* meningkat 1,44% - 2,67%. Pada *cross-validation*, peningkatan klasifikasi *intent* sebesar 0,18% - 1,5%, dan ekstraksi *entity* sebesar 0,01% - 0,82%. Sedangkan pada data uji aktual, klasifikasi *intent* meningkat 0,82% - 0,88%, dan ekstraksi *entity* sebesar 0,09% - 0,42%. Meski terjadi sedikit penurunan pada metrik *macro average*, fitur semantik tetap menunjukkan kinerja lebih baik secara keseluruhan. Dengan demikian, pendekatan berbasis semantik terbukti meningkatkan kemampuan chatbot dalam memahami *intent* dan konteks bahasa alami.

Kata Kunci : *Chatbot*, Pemrosesan Bahasa Alami, *Dual Intent and Entity Transformer (DIET)*, *Bidirectional Encoder Representation from Transformers (BERT)*, *Intent*, Entitas.

ABSTRACT

IMPROVING INTENT CLASSIFICATION AND ENTITY EXTRACTION THROUGH SEMANTIC-BASED FEATURE EXTRACTION AND DIET CLASSIFIER IN A CHATBOT (A CASE STUDY AT THE NORTH JAKARTA REGIONAL JOB TRAINING CENTER)

By

Agus Wahyu Priyanto
23/530419/PPA/06733

The main problem in information services at the North Jakarta Regional Job Training Center (Pusat Pelatihan Kerja Daerah Jakarta Utara) lies in the high volume of public inquiries, which have so far been handled only through social media platforms such as Instagram and WhatsApp by administrators. This condition leads to delayed responses, dependency on human resources, and limited service availability that does not operate 24 hours. Moreover, existing chatbots are generally rule-based or rely on syntactic features, making them less capable of understanding complex variations in natural language, which results in low accuracy in intent classification and entity extraction.

This research proposes the development of a chatbot based on the Rasa framework by integrating semantic-based feature extraction using IndoBERT Language Model Featurizer and the Dual Intent and Entity Transformer (DIET) Classifier. This approach enables the chatbot to better understand the meaning of words and conversational context, thereby improving the performance of intent classification and entity extraction compared to syntactic-based approaches.

The experimental results show that applying semantic features consistently improves model performance across three evaluation methods: held-out, cross-validation, and actual test data. In the held-out method, the F1-score for intent classification increased by 1.83%–5.66%, while entity extraction improved by 1.44%–2.67%. In the cross-validation method, intent classification improved by 0.18%–1.5% and entity extraction by 0.01%–0.82%. Meanwhile, in the actual test data, intent classification improved by 0.82%–0.88%, and entity extraction by 0.09%–0.42%. Although a slight decrease occurred in the macro average metric, semantic features still demonstrated overall better performance. Thus, the semantic-based approach effectively enhances the chatbot's capability to comprehend intent and the natural language context.

Keywords : Chatbot, Natural Language Processing, Dual Intent and Entity Transformer (DIET), Bidirectional Encoder Representation from Transformers (BERT), Intent, Entity.