

ABSTRAKSI

Pengguna media sosial saat ini cenderung menggunakan bahasa campur kode untuk mengungkapkan pendapatnya. Pengguna media sosial telah meningkat secara eksponensial di beberapa negara seperti Indonesia, telah memunculkan volume besar data campur kode, di mana pengguna menggunakan lebih dari satu bahasa dalam satu teks. Data dengan campur kode seringkali berisik dan yang paling penting, model monolingual biasanya tidak bekerja dengan baik. Ini menjadi tantangan bagi Natural Language Processing (NLP) untuk memproses dan menganalisis data. Dalam karya ini, kami melakukan eksperimen analisis sentimen pada data campuran kode Inggris-Indonesia. Pendekatannya yaitu dengan memanfaatkan model multilingual pre-trained mBERT. Dengan menganalisis prediksi model analisis sentimen, kita dapat menilai seberapa efektif model dapat menyesuaikan diri dengan kebisingan implisit yang melekat pada data campuran kode. Performa model klasifikasi diuji menggunakan parameter batch size dan epochs untuk menemukan dan mendapatkan akurasi tertinggi. Hasil percobaan menunjukkan bahwa akurasi tertinggi yang kami peroleh dari model mBERT yang dilatih dengan dataset kami diperoleh adalah 76%, dengan ukuran batch 16 dan epoch yang digunakan 7.

Index Terms: Campur kode, Analisis sentimen, mBERT, Natural Language Processing