

Klasifikasi Emosi Berdasarkan Pola Unggahan berbahasa Indonesia di Media Sosial X menggunakan metode RoBERTa

Zaidan Ibnuabil Iryanto¹, Warih Maharani²
Fakultas Informatika, Universitas Telkom, Bandung
ibnusabil@student.telkomuniversity.ac.id,
wmaharani@telkomuniversity.ac.id

Abstrak

Klasifikasi emosi merupakan aspek penting dalam *natural language processing*, terutama dalam memahami interaksi pengguna di media sosial. Media sosial X menjadi sumber data yang kaya di mana pengguna mengekspresikan berbagai emosi melalui unggahan mereka. Penelitian ini memperkenalkan perbandingan strategi preprocessing untuk menangani slang dalam posts media sosial berbahasa Indonesia, dengan menggunakan RoBERTa untuk mengklasifikasikan emosi dalam cuitan berbahasa Indonesia. Dataset yang digunakan terdiri dari 8.978 cuitan berlabel yang dikategorikan ke dalam empat kelas emosi, yaitu bahagia, marah, sedih, dan takut. Penelitian ini membandingkan metode *full preprocessing* dan *half preprocessing*. *Full preprocessing*, yang mencakup *normalization*, *stemming*, dan *stopword removal*, mencapai akurasi tertinggi sebesar 75,01% dan F1-score sebesar 0,750. Sebaliknya, *half preprocessing*, yang berfokus pada *case folding*, *cleansing*, dan *tokenization* dengan tetap mempertahankan slang dan bahasa informal, mencapai akurasi 70,68% dan F1-score sebesar 0,714. Mempertahankan slang dalam *half preprocessing* meningkatkan keberagaman bahasa tetapi juga menambah *noise*, sehingga menyulitkan RoBERTa dalam mengidentifikasi pola emosi yang konsisten. Namun, *full preprocessing* menghilangkan slang dan bahasa informal, mengurangi *noise*, serta memungkinkan RoBERTa untuk melakukan generalisasi yang lebih baik terhadap dataset. Temuan ini menyoroti dampak strategi *preprocessing* terhadap kemampuan RoBERTa dalam menangani dataset media sosial berbahasa Indonesia serta menunjukkan potensinya dalam meningkatkan tugas klasifikasi emosi untuk teks yang informal dan penuh *noise*.

Kata Kunci : emotion classification, RoBERTa, social media, Indonesian tweets, *preprocessing*.
