

INTISARI

Seiring dengan perkembangan zaman, berkembang pula berbagai *platform* sosial media. Dari *platform* tersebut, pengguna dapat mencurahkan segala isi hatinya, salah satunya adalah opini terhadap suatu entitas produk, institusi, organisasi, maupun ketokohan. Opini tersebut memiliki berbagai aspek emosi yang jika dianalisis akan menjadi sebuah informasi yang sangat berharga, terutama kepada pengembang produk, pemangku kebijakan, hingga pengamat politik. Untuk mengklasifikasikan emosi dari tiap-tiap teks akan menjadi sebuah pekerjaan yang tidak efisien apabila dikerjakan secara manual, sehingga membutuhkan sebuah metode atau algoritma agar proses klasifikasi emosi tersebut dapat berjalan secara otomatis dan lebih akurat.

Penelitian ini mengeksplorasi dan membandingkan tingkat akurasi dari berbagai metode dan algoritma *natural language processing* mulai dari *machine learning* hingga *deep learning*. Secara spesifik, penelitian ini membandingkan metode *Logistic Regression*, *Naive Bayes*, *Support Vector Machine*, *LSTM*, dan IndoBERT berbasis *Transformers*. Selain itu, penelitian ini juga membandingkan metode pembobotan dan ekstraksi tiap fitur yang sering digunakan dalam *natural language processing*. Dari penelitian ini akan ditentukan metode yang terbaik untuk membentuk sistem klasifikasi emosi spesifik secara otomatis. Hasil penelitian ini menunjukkan bahwa model IndoBERT yang dilakukan *finetuning* menunjukkan tingkat akurasi paling tinggi dibandingkan dengan model yang lainnya. Model IndoBERT ini memiliki tingkat presisi, *recall*, dan *F1-Score* masing-masing sebesar 0.78.

Kata kunci : *natural language processing*, *deep learning*, *bert*, *indobert*, *emotion classification*

ABSTRACT

As the world evolves, there are many social media was developed. From these platforms, users can express their innermost thoughts, one of which is their opinion about a product entity, institution, organization, or public figure. These opinions embody various emotional aspects, and analyzing these emotions can yield valuable information, especially to product developers, policymakers, and political observers. However, manually classifying the emotions in each text can be an inefficient task, necessitating a method or algorithm so that the emotion classification process can be automated and faster.

This study explores and compares the accuracy levels of various natural language processing methods and algorithms, ranging from machine learning to deep learning. Specifically, this study compares the Logistic Regression, Naive Bayes, Support Vector Machine, LSTM, and IndoBERT methods based on Transformers. This study also compares the weighting methods and feature extraction that often used in natural language processing. From this study, the best method to create an automatic specific emotion classification system will be determined. The results of this study show that the fine-tuned IndoBERT model exhibits the highest accuracy level compared to the other models. This IndoBERT model has precision, recall, and F1-Score rates of 0.78, respectively.

Keywords : natural language processing, deep learning, bert, indobert, emotion classification