

Klasifikasi Emosi Berbasis Teks Bahasa Indonesia dengan Perbandingan CNN, LSTM dan BERT = Indonesian Text Based Emotion Classification with Comparison of CNN, LSTM and BERT

Maulana Nurhendronoto, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920543951&lokasi=lokal>

Abstrak

Emosi adalah perasaan yang muncul dalam diri seseorang sebagai respon dari situasi tertentu. Perasaan ini dapat memengaruhi pikiran, perilaku, dan persepsi seseorang terhadap suatu peristiwa. Klasifikasi emosi adalah bagian dari analisis sentimen yang bertujuan untuk menganalisis dan memperoleh emosi dari suatu data. Penelitian klasifikasi emosi berbasis teks perlu dilakukan karena dapat diimplementasikan pada berbagai bidang, seperti kesehatan dan pendidikan. Bahasa Indonesia menduduki peringkat 11 bahasa dengan penutur terbanyak di dunia dengan 200 juta penutur. Namun, penelitian klasifikasi emosi berbasis teks bahasa Indonesia masih sedikit dilakukan. Algoritma machine learning dapat digunakan untuk mengatasi berbagai tantangan dalam penelitian klasifikasi emosi seperti memahami emosi dan menganalisis emosi dari data yang tidak terstruktur. Penelitian ini berfokus pada pengembangan model machine learning dengan teknik convolutional neural network (CNN), long short-term memory (LSTM), dan bidirectional encoder representation from transformer (BERT). Berdasarkan pengujian yang dilakukan, metode convolutional neural network (CNN) mendapatkan F1 score sebesar 84,2%, metode long short term memory mendapatkan F1 score sebesar 82%, metode BERT en uncased mendapatkan F1 score sebesar 22%, dan metode BERT multi cased mendapatkan F1 score sebesar 32%. Hasil pengujian ini menandakan metode CNN merupakan metode dengan hasil pengujian terbaik dan BERT en uncased merupakan metode dengan hasil pengujian terburuk dibanding ketiga metode lainnya.

.....Emotions are feelings that arise within a person in response to a particular situation. These feelings can affect a person's thoughts, behavior, and perception of an event. Emotion classification is a part of sentiment analysis that aims to analyze and derive emotions from data. Text-based emotion classification research needs to be done because it can be implemented in various fields, such as health and education. Indonesian is ranked the 11th most spoken language in the world with 200 million speakers. However, there is still little research on Indonesian text-based emotion classification. Machine learning algorithms can be used to overcome various challenges in emotion classification research such as understanding emotions and analyzing emotions from unstructured data. This research focuses on developing machine learning models with convolutional neural network (CNN), long short-term memory (LSTM), and bidirectional encoder representation from transformer (BERT) techniques. Based on the tests conducted, the convolutional neural network (CNN) method gets an F1 score of 84,2%, the long short term memroy method gets an F1 score of 82%, the BERT en uncased method gets an F1 score of 22%, and the BERT multi cased method gets an F1 score of 32%. These results indicate that the CNN is the bets method while the BERT en uncased is the worst method compared to the three other methods.