

ABSTRAK

Perkembangan media sosial, khususnya *platform X*, telah mendorong meningkatnya penggunaan meme sebagai medium komunikasi multimodal yang memadukan teks dan gambar untuk mengekspresikan emosi. Penelitian ini bertujuan mengembangkan sistem analisis emosi berbasis multimodal dengan menggabungkan model IndoBERT untuk pemrosesan teks berbahasa Indonesia dan YOLOv8 untuk analisis visual, guna mendeteksi emosi pada konten meme di *platform X*. Data dikumpulkan melalui teknik scraping, sebanyak 2810 meme berbahasa Indonesia yang diklasifikasikan ke enam kategori emosi berdasarkan teori Basic Emotion Paul Ekman, yaitu Bahagia, Jijik, Marah, Sedih, Takut, dan Terkejut. Dataset dibagi dengan rasio 80:10:10 untuk pelatihan, validasi, dan pengujian. Penggabungan kedua model dilakukan menggunakan metode *weighted late fusion* dengan pencarian nilai alpha optimal. Hasil evaluasi menunjukkan model gabungan dengan nilai alpha terbaik 0,6 menghasilkan akurasi 72,55%, *Precision macro* 73,03%, *Recall macro* 72,76%, dan *F1-score macro* 72,66%, menunjukkan bahwa pendekatan multimodal mampu menghasilkan klasifikasi emosi yang lebih representatif dibandingkan model unimodal.

Kata kunci: Analisis emosi, *Meme*, Multimodal, IndoBERT, YOLOv8, *Platform X*, *Weighted late fusion*

ABSTRACT

The development of social media, especially platform X, has encouraged the increasing use of memes as a multimodal communication medium that combines text and images to express emotions. This study aims to develop a multimodal-based emotion analysis system by combining the IndoBERT model for Indonesian language text processing and YOLOv8 for visual analysis, to detect emotions in meme content on platform X. Data were collected through scraping techniques, a total of 2810 Indonesian language memes classified into six emotion categories based on Paul Ekman's Basic Emotion theory, namely Happiness, Disgust, Anger, Sadness, Fear, and Surprise. The dataset was divided with a ratio of 80:10:10 for training, validation, and testing. The combination of the two models was carried out using the weighted late fusion method with the search for the optimal alpha value. The evaluation results show that the combined model with the best alpha value of 0.6 produces an Accuracy of 72.55%, a macro Precision of 73.03%, a macro Recall of 72.76%, and a macro F1-score of 72.66%, indicating that the multimodal approach is able to produce a more representative emotion classification than the unimodal model.

Keywords: Emotion analysis, Memes, Multimodal, IndoBERT, YOLOv8, Platform X, Weighted late fusion