

ABSTRAK

Fenomena *body shaming* di media sosial X (*Twitter*) semakin marak terjadi dan menimbulkan berbagai dampak psikologis bagi korbannya. Tingginya aktivitas pengguna media sosial menyebabkan penyebaran komentar negatif terkait bentuk tubuh semakin sulit dikendalikan. Oleh karena itu, diperlukan sistem yang mampu melakukan analisis sentimen secara otomatis untuk mengetahui opini publik terhadap fenomena tersebut. Penelitian ini bertujuan untuk menerapkan algoritma *Extreme Gradient Boosting* (XGBoost) dalam mengklasifikasikan sentimen publik terhadap fenomena *body shaming* di media sosial X serta mengetahui hasil analisis sentimen yang diperoleh. Data penelitian dikumpulkan menggunakan teknik web scraping melalui *Tweet Harvest* dan menghasilkan 1.383 *tweet* berbahasa Indonesia yang kemudian diklasifikasikan secara manual ke dalam tiga kelas sentimen, yaitu positif, negatif, dan netral. Tahap *text preprocessing* meliputi *case folding*, *cleansing*, *tokenizing*, normalisasi, dan *filtering* tanpa penerapan stemming karena terbukti menurunkan performa model pada data teks media sosial. Pembobotan fitur dilakukan menggunakan metode TF-IDF, sedangkan pembagian data menggunakan rasio 80:20 dengan penerapan *Random Over Sampling* (ROS) untuk mengatasi ketidakseimbangan kelas. Model XGBoost dibangun menggunakan parameter $n_estimators = 300$, $learning_rate = 0,05$, dan $max_depth = 5$. Hasil evaluasi menggunakan *confusion matrix* menunjukkan nilai akurasi sebesar 80,14%, dengan nilai *F1-score* pada kelas negatif sebesar 0.85, kelas netral sebesar 0.69, dan kelas positif sebesar 0.80. Hasil penelitian menunjukkan bahwa algoritma XGBoost mampu mengklasifikasikan sentimen publik terhadap fenomena *body shaming* dengan kinerja yang cukup baik. Selain itu, sistem analisis sentimen berbasis web berhasil diimplementasikan untuk memudahkan proses klasifikasi secara otomatis dan terstruktur.

Kata Kunci: *Body Shaming*, Analisis Sentimen, Media Sosial X, *Extreme Gradient Boosting*, *Term Frequency-Inverse Document Frequency*

ABSTRACT

The phenomenon of body shaming on social media platform X (Twitter) has become increasingly widespread and has caused various psychological impacts on its victims. The high level of social media activity has made the spread of negative comments related to body shape more difficult to control. Therefore, a system capable of automatically performing sentiment analysis is needed to identify public opinions regarding this phenomenon. This study aims to implement the Extreme Gradient Boosting (XGBoost) algorithm in classifying public sentiment toward the body shaming phenomenon on social media X and to determine the sentiment analysis results obtained. The research data were collected using a web scraping technique through Tweet Harvest, resulting in 1,383 Indonesian-language tweets which were manually classified into three sentiment classes: positive, negative, and neutral. The text preprocessing stage included case folding, cleansing, tokenizing, normalization, and filtering without applying stemming, as it was proven to reduce model performance on social media text data. Feature weighting was carried out using the TF-IDF method, while the data were divided using an 80:20 ratio with the implementation of Random Over Sampling (ROS) to address class imbalance. The XGBoost model was built using parameters of $n_estimators = 300$, $learning_rate = 0.05$, and $max_depth = 5$. The evaluation results using a confusion matrix showed an accuracy value of 80.14%, with F1-scores of 0.85 for the negative class, 0.69 for the neutral class, and 0.80 for the positive class. The results indicate that the XGBoost algorithm is capable of classifying public sentiment toward the body shaming phenomenon with fairly good performance. In addition, a web-based sentiment analysis system was successfully implemented to facilitate the automatic and structured sentiment classification process.

Keywords: Body Shaming, Sentiment Analysis, Social Media X, Extreme Gradient Boosting, Term Frequency-Inverse Document Frequency