Logistic Regression and Naïve Bayes Comparison in Classifying Emotions on Indonesian X Social Media

Gerald Shabran Rasyad[1], Warih Maharani[2]
Fakultas Informatika, Universitas Telkom, Bandung
geraldsrasyad@student.telkomuniversity.ac.id,
wmaharani@telkomuniversity.ac.id,

Abstract
Emotions are integral to human interaction and decision-making, often expressed on social media platforms like X, which provides valuable data for sentiment analysis. However, analyzing texts from X poses challenges due to informal language, slang, and unique textual features. This study compares Logistic Regression and Naive Bayes in classifying emotions from Indonesian tweets, addressing gaps in prior research by exploring feature extraction methods, data split ratios, and hyperparameter tuning. Data were collected from 100 Telkom University students, resulting in 8,978 tweets labeled into four emotions: Happy, Sad, Angry, and Fear. After preprocessing, feature extraction methods TF-IDF and Bag of Words (BoW) were applied. Models were trained and tested on 10%, 20%, and 30% data splits, and performance was evaluated using accuracy, precision, recall, and F1-score. Hyperparameter tuning was conducted for Logistic Regression using GridSearch. Results showed Logistic Regression outperformed Naive Bayes, achieving 73.49% accuracy compared to 70.27%, with BoW yielding superior results over TF-IDF. The 20% data split provided the best balance for training and testing. This research demonstrates the effectiveness of Logistic Regression and highlights the importance of tailored feature extraction and parameter optimization for emotion classification in informal text datasets, particularly for Indonesian tweets.

Keywords: Logistic Regression, Naive Bayes, emotion classification, TF-IDF, BoW