Comparative Analysis of Naive Bayes and SVM for Improved Emotion Classification on Social Media

Rio Ferdinand Putra Pratama [1], Warih Maharani[2]
Fakultas Informatika, Universitas Telkom, Bandung
ferdinandputra@student.telkomuniversity.ac.id,
wmaharani@telkomuniversity.ac.id,

Abstract

Emotion classification is important for understanding user interactions on social media, especially in identifying emotions such as happy, angry, sad, and fear. Howver, Indonesian text processing faces challenges due to language complexity and slang. This research compares Naïve Bayes and SVM models, focusing on evaluating the impact of preprocessing, feature extraction, and optimizing parameters to improve emotion classification. The dataset was collected from API X using crawling techniques and labeled manually by six annotators. The training process used full and half preprocessing datasets with TF-IDF, BoW, and Word2Vec feature extraction. Naïve Bayes and SVM models were evaluated using accuracy, precision, recall, and F1-score. Our findings show that full preprocessing improves accuracy, with TF-IDF + BoW achieving 78.01% using SVM and surpassing Naïve Bayes at 75.53%. The results categorize emotions into four classes: happy, sad, angry, and fear. This study demonstrates the value of preprocessing and feature selection to address slang and complexity in Indonesian texts. These findings provide insights for developing optimal models in emotion classification and offer applications in sentiment analysis, social media monitoring, and mental health detection.

Keywords: emotion classification; preprocessing; feature extraction; svm; naïve bayes.