
Comparison of Random Forest and Decision Tree Methods for Emotion Classification based on Social Media Posts

Muhammad Abiyyu Tsaqif¹, Warih Maharani²

Fakultas Informatika, Universitas Telkom, Bandung

abiyyutsaqif@student.telkomuniversity.ac.id,

wmaharani@telkomuniversity.ac.id

Abstract

Social media platforms like X (formerly Twitter) have become essential for expressing emotions and opinions, making emotion classification a critical task with applications in mental health, public sentiment monitoring, and customer feedback analysis. This study compares Random Forest and Decision Tree algorithms for classifying emotions such as joy, sadness, anger, and fear which are from social media posts. Data collection involved crawling tweets and manual labeling. Preprocessing included tokenization, stemming, and stopword removal, with feature extraction using TF-IDF and Bag of Words. Experimental scenarios tested data split ratios, resampling for class balance, and parameter tuning. Decision Tree parameters included criterion (gini, entropy), max depth (none, fixed values), min samples split (2, 5), and min samples leaf (1, 2). Random Forest parameters tuned were n_estimators (100–400), max depth (none, fixed values), min samples split (2, 5, 10), and min samples leaf (1, 2). Results showed Random Forest achieving a maximum accuracy of 76.17%, outperforming Decision Tree's 72.62%. The combination of TF-IDF and Bag of Words delivered the highest accuracy for both models. This study underscores the importance of preprocessing, balanced datasets, and parameter optimization for effective emotion classification. The findings offer insights into advancing sentiment analysis and natural language processing, enabling practical applications in public sentiment tracking, customer experience enhancement, and crisis management.

Keywords: Bag of Words; Decision Tree; Emotion Classification; Random Forest; TF-IDF
