Comparison of Naive Bayes and SVM Methods for Identifying Anxiety Based on Social Media

Endri Rizki Nugraha[1], Warih Maharani[2]
Fakultas Informatika, Universitas Telkom, Bandung
exdxix@student.telkomuniversity.ac.id,
wmaharani@telkomuniversity.ac.id,

Abstract

This research aims to detect anxiety patterns from social media posts using Naive Bayes (NB) and Support Vector Machine (SVM) algorithms. Tweets are extracted using Data Crawling techniques, then continued their way into labeling using Depression Anxiety Stress Scale (DASS-42) questionnaire along with Random Oversampler to balance out the unbalanced dataset and NB and SVM were chosen for their effectiveness in text sentiment classification. This study integrates textual features obtained from the Term Frequency-Inverse Document Frequency (TF-IDF) and Bag of Words (BoW) methods. The study compares the performance of these algorithms in detecting anxiety using datasets from the X platform. The comparison aims to identify the advantages and limitations of each method in handling textual sentiment data. This research aims to analyze sentiment data by calculating accuracy, recall, and F1-score to determine the most optimal performance outcome. The results indicate that the SVM with TF-IDF feature extraction achieved the highest accuracy of 72% and an average F1-Score of 61%, while the NB with BoW achieved 56% accuracy and an average F1-Score of 49%. These findings highlight the effectiveness of combining SVM and TF-IDF features which improve model effectiveness with SVM producing the best overall result in identifying anxiety from social media data.

Keywords: Anxiety, Naive Bayes, SVM, DASS-42, Sentiment Analysis, Social Media.