

ABSTRAK

Kesehatan mental remaja merupakan isu global yang semakin mendesak di era digital, di mana penggunaan media sosial yang intensif berkorelasi dengan peningkatan gejala depresi, kecemasan, dan kondisi psikologis negatif lainnya. Penelitian ini bertujuan untuk membangun model analisis sentimen berbasis *machine learning* guna mendeteksi dan mengklasifikasikan kondisi kesehatan mental remaja berdasarkan teks dari media sosial menggunakan algoritma *Support Vector Machine* (SVM) dengan kernel *RBF*. Dataset yang digunakan adalah "*Sentiment Analysis for Mental Health*" dari *Kaggle* yang terdiri dari 52.681 data teks berlabel setelah proses pembersihan data, mencakup tujuh kategori kondisi kesehatan mental, yaitu *Normal*, *Depression*, *Suicidal*, *Anxiety*, *Bipolar*, *Stress*, dan *Personality Disorder*. Proses penelitian meliputi tahap *preprocessing* teks berupa *lowercasing*, penghapusan karakter non-huruf, penghapusan *stop words*, dan *lemmatization*, dilanjutkan dengan ekstraksi fitur menggunakan *TF-IDF* (*Term Frequency-Inverse Document Frequency*) dengan konfigurasi *max_features* = 10.000 dan *bigram* (*ngram_range*=(1,2)). Pembagian data dilakukan secara *stratified* dengan proporsi 70:30 untuk data latih dan data uji. *Hyperparameter tuning* menggunakan *GridSearchCV* dengan *5-fold cross validation* menghasilkan parameter terbaik $C = 10$ dan $\gamma = 0,1$ dengan nilai F1-Score cross validation sebesar 73,30%. Hasil evaluasi model pada data uji (15.805 sampel) menunjukkan *accuracy* sebesar 75,60%, *precision* sebesar 75,54%, *recall* sebesar 75,60%, dan F1-score sebesar 75,42%. Nilai *Macro Average AUC* sebesar 0,9520 mengonfirmasi kemampuan diskriminasi model yang sangat kuat pada seluruh kelas. Kelas *Normal* memperoleh performa terbaik (F1-score = 0,90) sementara kelas *Stress* memperoleh performa terendah (F1-score = 0,57). Tantangan utama yang teridentifikasi adalah tingginya misklasifikasi antara kelas *Depression* dan *Suicidal* akibat kemiripan pola bahasa secara linguistik. Penelitian ini membuktikan bahwa pendekatan SVM berbasis TF-IDF efektif digunakan sebagai sistem deteksi dini kondisi kesehatan mental remaja melalui analisis teks media sosial.

Kata Kunci: Analisis Sentimen, Kesehatan Mental Remaja, Media Sosial, *Support Vector Machine*, *TF-IDF*, Klasifikasi Teks, *Machine Learning*

ABSTRACT

Adolescent mental health is an increasingly urgent global issue in the digital era, where intensive social media use correlates with increased symptoms of depression, anxiety, and other negative psychological conditions. This study aims to develop a machine learning-based sentiment analysis model to detect and classify adolescent mental health conditions based on social media text using the Support Vector Machine (SVM) algorithm with an RBF kernel. The dataset used is the "Sentiment Analysis for Mental Health" from Kaggle, consisting of 52,681 labeled text records after data cleaning, encompassing seven mental health condition categories: Normal, Depression, Suicidal, Anxiety, Bipolar, Stress, and Personality Disorder. The research process includes text preprocessing stages such as lowercasing, removal of non-letter characters, stop word removal, and lemmatization, followed by feature extraction using TF-IDF (Term Frequency-Inverse Document Frequency) with a configuration of `max_features = 10,000` and `bigram (ngram_range=(1,2))`. Data splitting was performed in a stratified manner with a 70:30 ratio for training and testing sets. Hyperparameter tuning using GridSearchCV with 5-fold cross validation yielded the best parameters of $C = 10$ and $\gamma = 0.1$, with a cross-validation F1-Score of 73.30%. Model evaluation on the test data (15,805 samples) yielded an accuracy of 75.60%, precision of 75.54%, recall of 75.60%, and F1-score of 75.42%. A Macro Average AUC of 0.9520 confirms the model's strong discriminative capability across all classes. The Normal class achieved the best performance (F1-score = 0.90), while the Stress class had the lowest performance (F1-score = 0.57). The primary challenge identified was the high misclassification between the Depression and Suicidal classes due to their linguistic similarity. This study demonstrates that TF-IDF-based SVM is an effective approach for early detection of adolescent mental health conditions through social media text analysis.

Keywords: *Sentiment Analysis, Adolescent Mental Health, Social Media, Support Vector Machine, TF-IDF, Text Classification, Machine Learning*