ABSTRAK

Penelitian ini membandingkan algoritma Recurrent Neural Networks (RNN) dan Support Vector Machine (SVM) untuk klasifikasi ujaran kebencian (hate speech) dan bahasa kasar (abusive) pada teks berbahasa Indonesia. Dataset yang digunakan adalah ID Multi-Label Hate Speech yang terdiri dari 13.169 data dan terbagi dalam empat kelas kombinasi hate speech dan abusive. Proses preprocessing meliputi case folding, tokenizing, normalisasi dan stemming. Fitur representasi untuk SVM menggunakan TF-IDF, FastText, dan IndoBERTweet, sedangkan RNN menggunakan embedding dari token mentah, FastText, dan IndoBERTweet. Evaluasi dilakukan dengan 5-fold cross-validation menggunakan metrik akurasi dan macro F1-score. Hasil menunjukkan bahwa SVM dengan fitur TF-IDF menghasilkan akurasi tertinggi sebesar 80,51%, sementara RNN dengan embedding IndoBERTweet memperoleh akurasi 78,52%. Hasil ini menunjukkan bahwa pemilihan algoritma, fitur representasi, dan kualitas preprocessing berpengaruh besar terhadap kinerja klasifikasi teks hate speech.

Kata kunci: Hate Speech, Abusive, RNN, SVM

ABSTRACT

This study compares the Recurrent Neural Networks (RNN) and Support Vector Machine (SVM) algorithms for classifying hate speech and abusive language in Indonesian text. The dataset used is the ID Multi-Label Hate Speech, consisting of 13,169 entries divided into four classes based on combinations of hate speech and abusive content. The preprocessing steps include case folding, tokenization, normalization and stemming. For feature representation, SVM utilizes TF-IDF, FastText, and IndoBERTweet, while RNN uses embeddings from raw tokens, FastText, and IndoBERTweet. Evaluation is conducted using 5-fold cross validation with accuracy and macro F1-score as performance metrics. The results show that SVM with TF-IDF features achieved the highest accuracy of 80.51%, while RNN with IndoBERTweet embeddings achieved an accuracy of 78.52%. These findings indicate that the choice of algorithm, feature representation, and the quality of preprocessing significantly influence the performance of hate speech classification models.

Keywords: Hate Speech, Abusive, RNN, SVM