

Analisis dan Deteksi Hate Speech pada Media Sosial Twitter Berbahasa Indonesia = Hate Speech Analysis and Detection in Indonesian Language in Twitter Social Media

Tansa Trisna Astono Putri, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20468381&lokasi=lokal>

Abstrak

ABSTRAK

Kebebasan berpendapat melalui media sosial untuk mengungkapkan pikiran, pendapat dan tanggapan terhadap suatu topik tertentu menimbulkan dampak negatif berupa konten yang menebarkan kebencian. Penelitian ini bertujuan untuk melakukan deteksi sebuah informasi yang merupakan ujaran kebencian di media sosial Twitter. Data yang digunakan berjumlah 4.002 data sentimen terkait topik politik, agama, suku dan ras di Indonesia. Pada pembangunan model, penelitian ini menggunakan metode klasifikasi sentimen dengan algoritma machine learning seperti Na ve Bayes, Multi Level Perceptron, AdaBoost Classifier, Random Forest Decision Tree dan Support Vector Machine SVM . Di samping itu, penelitian ini juga melakukan perbandingan performa model dengan menggunakan unigram, bigram dan unigram-bigram dalam proses fitur ekstraksi dan penggunaan SMOTE untuk mengatasi imbalanced data. Evaluasi dari percobaan yang dilakukan menunjukkan bahwa algoritma AdaBoost menghasilkan model terbaik dengan nilai recall tertinggi yaitu 99.5 yang memiliki nilai akurasi sebesar 70.0 dan nilai F1-score sebesar 82.2 untuk klasifikasi ujaran kebencian apabila menggunakan bigram.

<hr>

ABSTRACT

Freedom of expression through social media to express idea, opinion and view about current topic causes negative impact as the rise of hateful content. This study aims to detect a hate speech information through Twitter. Dataset of this study consists of 4.002 sentiment data related to politic, race, religion and clan topic. The model development of this study conducted by sentiment classification method with machine learning algorithm such as Na ve Bayes, Multi Level Perceptron, AdaBoost Classifier, Random Forest Decision Tree and Support Vector Machine SVM . We also conduct a comparison of model performance that used unigram, bigram, unigram bigram feature and SMOTE to handle imbalanced data. Evaluation of this study showed that AdaBoost algorithm resulted the best classification model with the highest recall model which was 99.5 , accuracy score as much as 70.0 and F1 score 82.2 to classify hate speech when using bigram features.