

DETEKSI UJARAN KEBENCIAN DALAM POSTINGAN TWITTER MENGGUNAKAN PENDEKATAN MACHINE LEARNING

Oleh

YUDA SEPTIAWAN

ABSTRAK

Pada era digital saat ini, konektivitas internet semakin meluas di Indonesia, memungkinkan akses informasi dan transaksi online dengan efektif dan efisien. Media sosial menjadi salah satu platform utama untuk berkomunikasi dan berbagi informasi, tetapi juga rentan digunakan untuk menyebarkan konten negatif seperti ujaran kebencian. Penelitian ini bertujuan untuk menguji efektivitas empat metode klasifikasi (Support Vector Machine, Logistic Regression, K-Nearest Neighbor, dan Decision Tree) dalam mendeteksi ujaran kebencian dalam postingan Twitter dengan hashtag #Indonesia. Data tweet dikumpulkan melalui API Twitter dan dilabeli menggunakan *sentistrength*. Metode klasifikasi diterapkan untuk mengidentifikasi sentiment tweet dan hasilnya dievaluasi berdasarkan akurasi dan F-measure pada dua rasio pembagian dataset, yaitu 90:10 dan 80:20 untuk data latih dan data uji. Hasil penelitian menunjukkan bahwa metode Support Vector Machine (SVM) memiliki kinerja paling baik, dengan akurasi sekitar 84% (90:10) dan 83% (80:20), serta F-measure sekitar 83% pada kedua rasio tersebut. Meskipun Logistic Regression menunjukkan akurasi stabil sekitar 82% pada kedua rasio dataset, F-measure cenderung lebih rendah dibandingkan SVM, terutama pada rasio 80:20. Kesimpulannya, SVM memiliki kemampuan yang konsisten dan baik dalam mengklasifikasikan data ujaran kebencian pada platform Twitter, menunjukkan tingkat akurasi tinggi dan keseimbangan antara presisi dan recall.

Kata kunci: *Sentiment analysis, Support Vector Machine (SVM),
Logistic Regression, K-Nearest Neighbor, Decision Tree*

DETEKSI UJARAN KEBENCIAN DALAM POSTINGAN TWITTER MENGGUNAKAN PENDEKATAN MACHINE LEARNING

Oleh

YUDA SEPTIAWAN

ABSTRACT

In today's digital era, internet connectivity is increasingly widespread in Indonesia, enabling effective and efficient access to information and online transactions. Social media has become one of the main platforms for communicating and sharing information, but is also vulnerable to being used to spread negative content such as hate speech. This study aims to test the effectiveness of four classification methods (Support Vector Machine, Logistic Regression, K-Nearest Neighbor, and Decision Tree) in detecting hate speech in Twitter posts with the hashtag #Indonesia. The tweet data was collected through the Twitter API and labelled using *sentistrength*. Classification methods were applied to identify sentiment tweets and the results were evaluated based on accuracy and F-measure at two dataset sharing ratios, namely 90:10 and 80:20 for training data and test data. The results show that the Support Vector Machine (SVM) method performs best, with accuracy of about 84% (90:10) and 83% (80:20), and F-measure of about 83% at both ratios. Although Logistic Regression showed a stable accuracy of about 82% in both dataset ratios, the F-measure tended to be lower than SVM, especially in the 80:20 ratio. In conclusion, SVM has a consistent and good ability to classify hate speech data on the Twitter platform, showing high accuracy and a balance between precision and recall.

Keyword: *Sentiment analysis, Support Vector Machine (SVM),
Logistic Regression, K-Nearest Neighbor, Decision Tree*

