

Identifikasi Target, Kategori, dan Level Ujaran Kebencian pada Data Twitter Berbahasa Daerah Menggunakan Classical Machine Learning dan Deep Learning = Identification of Target, Category, and Level of Hate Speech from Regional Language on Twitter Data Using Classical Machine Learning and Deep Learning

Ajeng Dwi Asti, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20528558&lokasi=lokal>

Abstrak

Ujaran kebencian dapat menyebabkan terjadinya konflik dan pembantaian di masyarakat sehingga harus segera ditangani. Indonesia memiliki lebih dari 700 bahasa daerah dengan karakteristik masing-masing. Ujaran kebencian yang ada di Indonesia juga pernah dilakukan menggunakan bahasa daerah. Media sosial Twitter paling sering digunakan dalam menyebarkan ujaran kebencian. Identifikasi target, kategori, serta level ujaran kebencian dapat membantu Polri dan Kemenkominfo dalam menentukan prioritas penanganan ujaran kebencian sehingga dapat meminimalisir dampaknya. Penelitian ini melakukan identifikasi ujaran kasar dan ujaran kebencian beserta target, kategori, dan level ujaran kebencian pada data Twitter berbahasa daerah menggunakan algoritma classical machine learning dan deep learning. Penelitian ini menggunakan data lima bahasa daerah di Indonesia dengan penutur terbanyak yaitu Jawa, Sunda, Madura, Minang, dan Musi. Pada data Bahasa Jawa performa terbaik diperoleh menggunakan algoritma Support Vector Machine (SVM) dengan transformasi data Classifier Chains (CC) serta kombinasi fitur word unigram, bigram, dan trigram dengan F1-score 70,43%. Algoritma SVM dengan transformasi data CC serta kombinasi fitur word unigram dan bigram memberikan performa terbaik pada data Bahasa Sunda dan Madura dengan masing-masing F1-score 68,79% dan 78,81%. Sementara itu, pada data Bahasa Minang dan Musi hasil terbaik diperoleh menggunakan algoritma SVM dengan transformasi data CC serta fitur word unigram dengan F1-score 83,57% dan 80,72%. Penelitian ini diharapkan dapat digunakan sebagai masukan bagi Polri dan Kemenkominfo dalam pembangunan sistem identifikasi ujaran kasar, ujaran kebencian serta target, kategori, dan level ujaran kebencian pada media sosial.

.....Hate speech can lead to conflict and massacres in society so it must be dealt immediately. Indonesia has more than 700 regional languages with their own characteristics. Hate speech in Indonesia has also been carried out using regional languages. Twitter is the most frequently used social media to spread hate speech. Identification of targets, categories, and levels of hate speech can help the National Police and the Ministry of Communication and Information to determine priorities for handling hate speech to minimize its impact. This study identifies abusive speech and hate speech along with the target, category, and level of hate speech on regional language Twitter data using classical machine learning and deep learning algorithms. This study uses data from five regional languages in Indonesia with the most speakers, namely Javanese, Sundanese, Madurese, Minang, and Musi. In Java language data, the best performance is obtained using the Support Vector Machine (SVM) algorithm with Classifier Chains (CC) data transformation and a combination of unigram, bigram, and trigram word features with an F1-score of 70.43%. The SVM algorithm with CC data transformation and the combination of unigram and bigram word features provides the best performance on Sundanese and Madurese data with F1-scores of 68.79% and 78.81%, respectively. Meanwhile, in Minang and Musi language data, the best results were obtained using the SVM algorithm with CC data

transformation and word unigram features with F1-scores of 83.57% and 80.72%, respectively. This research is expected to be used as input for the National Police and the Ministry of Communication and Information in developing a system for identifying harsh speech, hate speech and the target, category, and level of hate speech on social media.