

Deteksi Ujaran Kebencian dan Ujaran Kasar Terkait Covid-19 Berbahasa Indonesia di Twitter = Hate Speech and Abusive Language Detection Related to COVID-19 in Indonesian Language on Twitter

Mohammad Rizky Adrian, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20529167&lokasi=lokal>

Abstrak

Salah satu upaya pengendalian konten negatif media sosial seperti ujaran kebencian dan ujaran kasar adalah dengan mengotomasi proses filter dari konten media sosial. Dalam konteks COVID19, proses otomasi ini dapat dimanfaatkan oleh KOMINFO, virtual police, satuan tugas COVID19, ataupun para akademisi. Data dikumpulkan dari Twitter selama bulan Mei sampai Juni 2021. Penelitian memanfaatkan korpus dari penelitian terdahulu untuk mengetahui apakah pengetahuan dari penelitian terdahulu dapat digunakan pada domain COVID19. Dataset dievaluasi menggunakan algoritma Support Vector Machine (SVM), Naïve Bayes, Random Forest Decision Tree (RFDT), Logistic Regression, dan ADABOOST, dengan variasi SMOTE dan undersampling. Unigram-bigram kata digunakan sebagai fitur dikombinasikan dengan fitur lexicon dan orthogonal, serta diekstraksi menggunakan Term Frequency-Inverse Document Frequency dan Count Vectorizer. Hasil anotasi menunjukkan perbandingan data imbalance sebesar 1:73 untuk ujaran kebencian dan 1:24 untuk ujaran kasar. Evaluasi dari hasil penelitian menunjukkan bahwa pemanfaatan model klasifikasi dari penelitian terdahulu (2019) dikombinasikan dengan dataset COVID19 memiliki nilai recall dan F1 klasifikasi ujaran kebencian (nilai recall 69.23%) dan ujaran kasar (nilai recall 71.3%) yang lebih baik. Algoritma pembangun model terbaik didominasi oleh algoritma SVM dan ADABOOST. Hasil dari penelitian perlu ditindaklanjuti agar dapat dirasakan manfaatnya secara langsung, misalnya dengan membungkus model klasifikasi pada API (application programmable interface).

.....One of the efforts to control negative aspect of social media like hate speech and abusive language is by automating the filtering process of content on social media. In the context of COVID19, KOMINFO, the virtual police, the COVID19 task force, or academics can benefit from this solution. Data was collected from Twitter in the period of May to June 2021. The study utilizes the corpus from previous studies to find out whether previous research knowledge can be used in the COVID19 domain. The COVID19 dataset uses the Support Vector Machine (SVM), Naïve Bayes, Random Forest Decision Tree (RFDT), Logistic Regression, and ADABOOST algorithms, with variations of data imbalances handling (SMOTE and undersampling). Unigram-bigram words, lexicon, and orthogonal are used as features extracted by TF-IDF and Count Vectorizer. The annotation results show a comparison of the imbalanced data of 1:73 for hate speech and 1:24 for abusive language in COVID19 dataset. Results of the study shows that the use of the classification model from previous studies (2019) combined with the COVID19 dataset has a better recall value and F1 classification of hate speech (with recall score of 69.23%) and abusive language (with recall score of 71.3%). The best classifier models mostly built using SVM and ADABOOST. The results of this research need to be followed up so that they can be used directly, for example by wrapping the best classifier model on API (application programmable interface).