

Mendeteksi Cyberhate pada Twitter menggunakan Text Classification dan Crowdsourced Labelling

Hadi Kurniawan Sidiq¹, Dana Sulistyio Kusumo, S.T., M.T., PhD², Indra Lukmana Sardi, S.T., M.T.³

^{1,2,3}S1 Informatika, Universitas Telkom, Bandung
Indonesia

¹hadikaes@student.telkomuniversity.ac.id, ²danakusumo@telkomuniversity.ac.id,

³indraluk@telkomuniversity.ac.id

Abstrak

Pada masa kampanye pemilihan umum presiden 2019 di Indonesia, banyak dukungan yang dilakukan masyarakat dengan berbagai macam bentuk dukungan seperti dukungan berupa bentuk sebaran dan juga tulisan pada sosial media. Sosial media seperti Twitter misalnya, pada masa pemilihan presiden terdapat banyak tagar-tagar dukungan seperti #2019gantipresiden, #2019tetapjokowi dan tagar lainnya yang terkait pemilihan presiden Indonesia. Namun, tidak jarang terdapat banyak ujaran kebencian atau *hate speech* yang terkandung pada *tweet* dengan tagar tersebut. *Hate speech* yang disebarluaskan di Internet (*cyberhate*) dapat menyebabkan munculnya perselisihan antar individu maupun kelompok. Pada masa kampanye pemilihan presiden ini, *cyberhate* dapat menyebabkan perselisihan antar kelompok pendukung kedua pasangan calon presiden dan dapat berakibat konflik seperti kerusuhan dan aksi lainnya yang merugikan Negara. Penelitian menggunakan algoritma SVM untuk mendeteksi *cyberhate* yang menghasilkan keakuratan terbaiknya sebesar 97%. Selain itu, penelitian ini juga menerapkan *crowdsourced labelling* dalam pelabelan dataset yang digunakan yang menghasilkan kevalidan data sebesar 98% data valid dengan 7 data yang salah.

Kata Kunci: *crowdsourced labelling, cyberhate tweets, hate speech detection, text classification.*

Abstract

During the 2019 presidential election campaign in Indonesia, a lot of support was made by the community with various forms of support such as poster distribution or even content on social media. Social media such as Twitter for example, during the presidential election there were many support tags such as #2019gantipresiden, #2019tetapjokowi and other hashtags related to the Indonesian presidential election. However, many hate speeches contained in tweets with the related hashtag. Hate speech on the Internet (cyberhate) could cause disputes between support groups of the two presidential candidate which cause conflicts such as riots and other actions that harm the State. The study uses the SVM algorithm to detect cyberhates that produce the best accuracy of 97%. In addition, this study also applies crowdsourced labeling in dataset labeling which result 98% valid data.

Kata Kunci: *crowdsourced labelling, cyberhate tweets, hate speech detection, text classification.*

1. Pendahuluan

1.1 Latar Belakang

Pada masa kampanye pemilihan umum presiden 2019 di Indonesia, banyak dukungan yang dilakukan masyarakat dengan berbagai macam bentuk dukungan seperti dukungan dalam bentuk sebaran maupun tulisan di sosial media. Sosial media seperti Twitter misalnya, pada masa pemilihan presiden terdapat banyak tagar-tagar dukungan seperti #2019gantipresiden, #2019tetapjokowi dan tagar lainnya yang terkait pemilihan presiden Indonesia. Namun, tidak jarang terdapat banyak ujaran kebencian atau *hate speech* yang terkandung pada *tweet* dengan tagar tersebut. *Hate speech* yang disebarluaskan di Internet (*cyberhate*) dapat menyebabkan munculnya perselisihan antar individu maupun kelompok.

Pada masa kampanye pemilihan presiden ini, *cyberhate* dapat menyebabkan perselisihan antar kelompok pendukung kedua pasangan calon presiden dan dapat berakibat konflik seperti kerusuhan dan aksi lainnya yang merugikan Negara. Pihak berwenang telah melakukan tindakan penanganan *cyberhate* ini berupa diaturnya hukuman bagi pelaku *cyberhate* dalam surat edaran kepolisian republik Indonesia tentang ujaran kebencian¹. Namun, tetap saja masih terdapat banyak pelaku *cyberhate* yang menyebarkan kebencian di Internet yang susah