

Detection of Indonesian Hate Speech in the Comments Column of Indonesian Artists' Instagram Using the RoBERTa Method

Adhe Akram Azhari¹⁾, Yuliant Sibaroni²⁾, Sri Suryani Prasetyowati³⁾

1, 2, 3) Informatics, School of Computing, Telkom University, Bandung, Indonesia

Article Info

Keywords: Instagram, RoBERTa, Full PreProcessing, Non Full-PreProcessing, Hate Speech

Article history:

Received 17 August 2018

Revised 15 February 2019

Accepted 4 April 2019

Available online 4 April 2019

DOI :

<https://doi.org/10.29100/jipi.v4i1.781>

* Corresponding author.

Corresponding Author

E-mail address:

adheakramazhari@students.telkomuniversity.ac.id

yuliant@telkomuniversity.ac.id

srisuryani@telkomuniversity.ac.id

ABSTRACT

This study detects hate speech comments from Instagram post comments where the method used is RoBERTa. Roberta's model was chosen based on the consideration that this model has a high level of accuracy in classifying text in English compared to other models, and possibly has good potential in detecting Indonesian as used in this research. There are two test scenarios namely full-preprocessing and non full-preprocessing where the experimental results show that non full-preprocessing has an average value of accuracy higher than full-preprocessing, and the average value of non full-preprocessing accuracy is 85.09%. Full-preprocessing includes several preprocessing stages, namely cleansing, case folding, normalization, tokenization, and stemming. While non full-preprocessing includes all processes in preprocessing except the stemming process. This shows that RoBERTa predicts comments well when not using full-preprocessing.