

Comparison of *Support Vector Machine* and *Naïve Bayes* in Classification of SARA speech on Twitter

Amirul Mahdi, Yuliant Sibaroni²

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹primumtertium@students.telkomuniversity.ac.id, ²yuliant@telkomuniversity.ac.id

Abstract

SARA (Tribe, Religion, Race, and Between groups) SARA is a thought of sentiment about personal identity which related to, heredity, religion, nationality or ethnic and group. At recent days most people that abuse the social media as a tool to spread the hatred with speech of SARA, and media where is quit common for people to use is *Twitter*. SARA is not just spread the hatred to the target but also can cause some people that are related to the target to feel offended about the speech of SARA. This research is to apply the classification of sentiment analysis. to find out about the habits of Indonesian people that still using *Twitter* as media to share hatred through speech of SARA. What is being done in this journal is to compare between 2 Classification method *Support Vector Machine (SVM)* with *Naïve Bayes* to find the accuracy level of those 2 and which is more higher from the classification speech of SARA and will also be doing weighting of words with *TF-IDF* method. As for other use of this research is to find out the performance of *Support Vector Machine* from doing the classification speech of SARA *Tweet*. After doing test for 2 times, it shows that implementation with Multinomial Naïve Bayes from the case “Classification of SARA speech on twitter” Naïve Bayes got higher accuracy, recall, F1-score and precision on the second test attempt with 82,97%, 82,97%, 81,82%, 83,68% and in the first attempt SVM got higher result with 56,95%, 56,95%, 56,84%, 56,83%. And also we doing comparison with two different kernel that is RBF and Polynomial and we get accuracy 59.57%, recall 59.57%, F1-score 55.14%, dan precision 72.90% and for RBF we got 29.78%, recall 29.78%, F1-score 16.22%, dan precision 46.11%. and for the third experiment SVM got the best result at C = 2 and gamma = 2 also when C = 1 and gamma = 3 where the accuracy 80.85%, recall 80.85%, F1-score 80.84% and precision 81.81%.

Keywords: SARA, Suport Vector Machine, Naïve Bayes, Sentiment analysis
