

In this study, will be conducted a paraphrase identification on a user generated sentences where all the datas taken from a social media Twitter that is tweet which is limited to 140 characters using an algortihm that focusing on finding a score between two strings so namely Levensthein Distance, and also will be conducted an accuracy test on the algorithm it self using a confusion matrix and also a testing scheme using a K-Fold Cross Validation. From this study we can conclude that we have an accuracy around 0,7695, f1 score aroun 0.6906, precission score around 0,8090 and lastly a recall score which is 0,6024.