

ABSTRACT

News curator is a Twitter user who follows, shares, re-shares, or gives feedback to the certain news article topic via tweet or retweet in Twitter. This user has the important role to help news article development process and as the alternative news source for public society. The instance of news curator in Indonesia is @kurawa, the Twitter account that help investigate the criminal case of Jakarta International School in 2016. Although this user is important, there is still lack of public trust due to various reason. Such as the tweet contain misleading information, hate speech, or hoax. Also, there are too many malicious users with fake/bot followers that can easily blow up that tweet, moreover verified user (has the blue tick mark) cannot guarantee to share the valuable tweet. So that, the challenge is how to assess and classify the credibility of news curator. Our contributions are providing the novel feature set related to the credibility class and the improvement performances of Naive Bayes Classifier to classify the credibility of Indonesian news curator in Twitter. In this study, the credibility is assessed into 3 different class: credible, seems credible and not-credible. And the class is annotated by public using user survey. Then, the classifier is built using Naive Bayes Classifier method with 159 features grouped by 7 kind of feature groups. Those are user-based, content-based retweet-based, sentiment learning-based, topic learning-based, spam lexicon-based, and sentiment lexicon-based. Its performance result is 66.5% F1-score. In the optimization, the classification is examined sequentially using different training and testing data ratios, different number of tweets, different selection feature methods, and different number of unlabelled news curators. And the results produce 77.2% F1-score. and it is obtained from using 90:10 ratio of training and testing data, 75.852 tweets, 65 features selected by Information Gain method, and 200 unlabelled news curators optimized by Expectation-Maximization method. Also, there are 3 feature groups that have the most relevant to credibility class. Those are retweet-based, sentiment learning-based and topic learning-based feature group.

Keywords: *news curator, twitter, credibility, classification, feature*