ABSTRACT

Hoax Detection on Twitter

Titi Widaretna

Supervisor   : Jimmy Tirtawangsa, Ph.D

Co-Supervisor   : Dr. Ade Romadhony

Hoax is misleading information that can be very dangerous. This research aims to detect tweets as hoax or non-hoax based on the context of the tweet using semantic approach and classification techniques. To find the context of a tweet, Doc2Vec is used to generate a semantic vector representation that captures the context. The semantic vector generated by Doc2Vec will be used as input in the classification process. Several classifier algorithms are used to find the best algorithm to detect hoax accurately.

Hoax detection on Twitter focuses on measuring the value of accuracy and precision, where the level of truth in detecting hoax and non-hoax must be high (True Positive and True Negative), while the error rate for detecting hoax as non-hoax must be low (False Positive). The experimental results of the proposed method, the best results are obtained using the Doc2Vec for the tweet representation model and SVM as a classifier with an accuracy of 93.02% and a precision of 93.02% respectively.

Keywords: Hoax Detection, Doc2Vec, Natural Language Preprocessing